



Mémoire présenté devant le jury de l'EURIA en vue de l'obtention du
Diplôme d'Actuaire EURIA
et de l'admission à l'Institut des Actuaire

le 08 Septembre 2022

Par : Robin PHILIPPE

Titre : Les modèles de Machine Learning au service de l'égalité Femme-Homme : comment les actuaires viennent challenger la réglementation ?

Confidentialité : Non

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membre présent du jury de l'Institut

des Actuaires :

Anaëlle LE BERRE

Signature :

Entreprise :

Galea

Signature :

GALEA & Associés
25 rue de Choiseul
75002 PARIS
Tél. 01 43 22 11 11
R.C.S. Paris - 492 379 839

Membres présents du jury de l'EURIA : **Directeur de mémoire en entreprise :**

Marine Habart

Franck Vermet

Nicolas LESNISSE

Signature :

**Autorisation de publication et de mise en ligne sur un site de diffusion
de documents actuariels**

(après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise :

Signature du candidat :

Résumé

Malgré un écart de taux d'activité qui s'est constamment réduit en France depuis les années 1970, l'égalité homme-femme au sein des entreprises reste un sujet particulièrement sensible et de plus en plus regardé. La loi « Rixain » du 24 décembre 2021 et la loi « Avenir Professionnel » adoptée en septembre 2018 comportent ainsi plusieurs mesures visant à une plus grande égalité entre les femmes et les hommes dans les entreprises.

Le suivi des indicateurs obligatoires à l'échelle de grosses structures peut nécessiter la manipulation d'un volume de données conséquent où l'intelligence artificielle saura apporter son efficacité. Le champ d'application de l'actuariat trouve ici une extension pour le moment assez peu exploitée.

Ce mémoire s'appuie sur un portefeuille d'entreprise de plusieurs milliers de salariés. L'étude va s'intéresser dans un premier temps aux indicateurs réglementaires de l'égalité Homme/Femme qui se veulent applicables à toute entreprise, tout en démontrant leurs limites. Dans un second temps, la mise en place d'indicateurs personnalisés qui permettront une meilleure visibilité des inégalités existantes sera étudiée en exploitant les possibilités offertes par la Data Science, dans une optique pour les Ressources Humaines de pouvoir disposer d'informations supplémentaires dans le processus de décisions, à adapter aux situations particulières de l'entreprise.

En effectuant un suivi fin des indicateurs, l'actuaire aura contribué au respect des exigences réglementaires, à la prévention et le pilotage de risques sociaux en évitant d'éventuelles pénalités. Cela permettra aussi à l'entreprise de mettre en avant sa politique égalitaire pour se démarquer dans un marché de l'emploi en tension permanente depuis quelques années. Par son travail, l'actuaire peut ainsi permettre à l'entreprise de continuer à assurer son attractivité auprès des talents féminins.

Mots clés : égalité Homme-Femme, prévention, risque sociaux, apprentissage automatique, interprétabilité, équité, biais, engagements sociaux, retraite, rémunération, promotion, augmentation.

Summary

Despite a gap in activity rates that has been steadily narrowing in France since the 1970s, gender equality within companies remains a particularly sensitive and increasingly watched subject. The « Rixain » law of 24 December 2021 and the « Avenir Professionnel » law adopted in September 2018 thus include several measures aimed at greater equality between women and men in companies.

The monitoring of mandatory indicators on the scale of large structures may require the manipulation of a significant volume of data where artificial intelligence will be able to provide its efficiency. The field of application of actuarial science finds here an extension that is currently not very well exploited.

This thesis is based on a portfolio of companies with several thousand employees. The study will first look at the regulatory indicators of gender equality that are intended to be applicable to all companies, while demonstrating their limitations. Secondly, the implementation of personalised indicators that will allow a better visibility of existing inequalities will be studied by exploiting the possibilities offered by Data Science, with a view to providing Human Resources with additional information in the decision-making process, to be adapted to the particular situation of the company.

By carrying out a fine monitoring of the indicators, the actuary will have contributed to the respect of the regulatory requirements, to the prevention and the management of social risks by avoiding possible penalties. This will also enable the company to highlight its equal opportunities policy in order to stand out in a job market that has been under constant pressure for several years. Through her work, the actuary can thus enable the company to continue to ensure its attractiveness to female talent.

Keywords : gender equality, prevention, social risks, machine learning, interpretability, fairness, bias, social commitments, pension, pay, promotion, pay rise

Note de Synthèse

La croissance des inégalités et tensions sociales continue d'occuper chaque année le milieu de classement des cartographies des risques émergents. L'intensification des attentes des populations sur les questions des inégalités sociétales, environnementales et économiques a mené à la mise en place de divers politiques et critères au sein des entreprises. Un index en particulier a été mis en place à partir de la loi Avenir Professionnel (Septembre 2018) afin de lutter contre les discriminations liées au sexe dans les entreprises. Cet indicateur a été créé de façon à être calculable pour toutes les entreprises et présente donc certaines limites, ce mémoire a pour objectif de proposer une méthodologie plus poussée afin que les entreprises puissent identifier plus finement les potentiels inégalités présentes en leur sein et ainsi prévenir au mieux les risques.

Cet index est composé de 4 à 5 sous-indicateurs selon la taille de l'entreprise. Chaque sous-indicateur note différents écarts entre les femmes et les hommes afin d'établir une note sur 100 points.

L'écart de rémunération femmes-hommes est noté sur 40 points.
L'écart de répartition des augmentations individuelles est noté sur 20 points.
L'écart de répartition des promotions est noté sur 15 points.
Le nombre de salariées augmentées après un congé maternité est noté sur 15 points.
La parité parmi les 10 plus hautes rémunérations est notée sur 10 points.

En cas d'index inférieur à 85 points, les entreprises doivent fixer et publier des objectifs de progression sur chacun des indicateurs. En cas d'index inférieur à 75 points, les entreprises doivent publier leurs mesures de correction et de rattrapage.

Le but de ce mémoire est de pouvoir proposer un outil de pilotage dans le cadre des inégalités entre les hommes et les femmes aux Ressources Humaines des entreprises. L'idée est de proposer un outil d'aide à la décision comprenant une liste d'individus potentiellement discriminés (selon chaque sous-indicateur) ainsi qu'une budgétisation de rattrapage salarial et différents indicateurs à travers des modèles d'apprentissage automatique afin de les appuyer dans la prévention des risques sociaux. L'actuaire est qualifié pour quantifier, analyser et prévenir les risques en mettant à profit différents outils actuariels et indicateurs.

Le sujet de l'égalité Homme-Femme apparaît ainsi comme un domaine où l'actuaire peut jouer un rôle, cependant les conclusions des modèles Machine Learning se limitent à la réalité terrain dont seuls les RH ont conscience : ainsi les conclusions sur les individus potentiellement discriminés ne peuvent se substituer à une analyse qualitative post modèle.

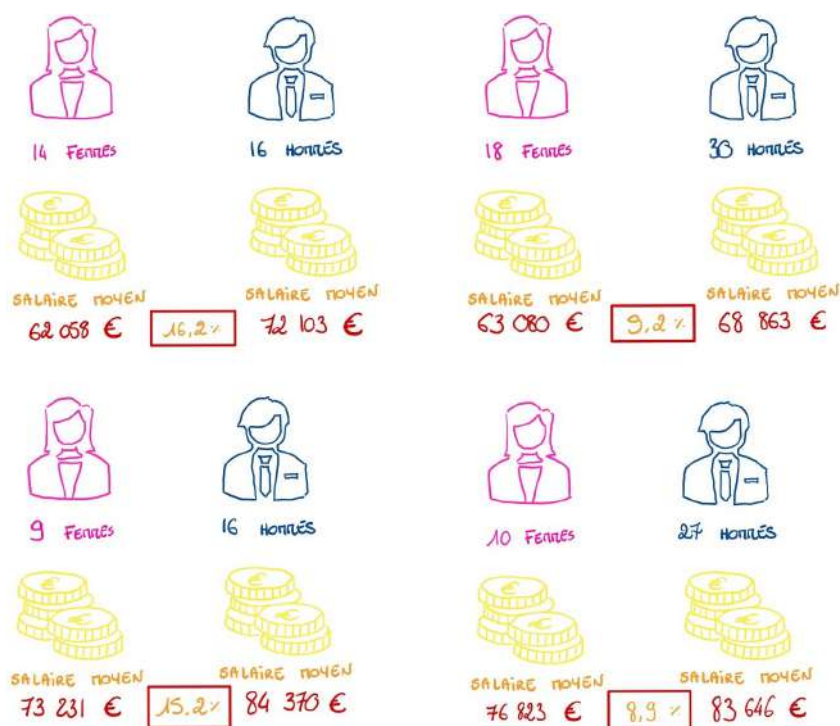
Pour chaque indicateur un modèle d'apprentissage a été implémenté afin d'analyser au mieux les potentielles discriminations présentes dans une entreprise.

Ecart de rémunération

L'entreprise considérée dans ce mémoire a obtenue 89 points au titre de l'index de l'égalité professionnelle, dont 39 points sur 40 aux écarts de rémunérations entre les sexes (par coefficient hiérarchique et âge).

Un modèle d'arbre de régression (CART) sur le salaire a été utilisé afin d'identifier le plus finement possible des groupes d'individus au sein desquelles sont présents des salariés potentiellement discriminés en termes de rémunération. Cela consistait à étudier les segmentations au niveau du sexe faites par l'arbre pour expliquer le salaire. Le modèle fournit en sortie une liste d'individus potentiellement discriminés qu'il convient de faire valider ensuite par les RH.

Le modèle identifie 4 groupes au sein desquelles pourrait exister des discriminations de salaire entre les hommes et les femmes toutes choses égales par ailleurs.



Cette liste d'individus permet d'établir un premier budget global de rattrapage salarial qui s'élève à **413 205€**.

Par la suite les écarts de Promotion et d'augmentation sont étudiés.

Ecart de promotion

L'idée de cette partie est de pouvoir identifier des individus potentiellement discriminés à l'obtention d'une promotion et de soumettre aux RH une liste pour une étude plus approfondie.

Un modèle de classification XGBoost (Extreme Gradient Boosting) sur la variable « Promotion » est alors établi en supprimant la variable « Sexe » pour que le modèle ne soit pas biaisé par le genre de l'individu.

Le but est de s'intéresser à tous les cas de faux positifs avec une forte probabilité de prédiction de promotion c'est-à-dire les individus prédits comme promus par le modèle mais qui ne l'ont pas été dans la réalité.

Puis il convient d'utiliser des outils d'interprétabilité des modèles comme SHAP pour être à même d'expliquer pourquoi le modèle a décidé de promouvoir ces individus.

Dans un premier temps, l'étude des métriques permettent de juger de la pertinence d'un modèle mis en place et ainsi de le valider :

Précision	Recall
89,7%	88,3%

FIGURE 1 – Métriques du modèle de classification

Le modèle possède un rappel et une précision proche de 1, le modèle a dans l'absolu bien prédit les promotions.

Le modèle étant globalement pertinent, il convient d'étudier plus précisément ses erreurs afin de vérifier qu'elles ne sont pas dues à une potentielle discrimination.

Il est observé un taux de faux positif de 5%, nous allons par la suite étudier les plus hautes prédictions du modèle parmi les faux positifs.

Il est nécessaire de comprendre pourquoi le modèle donne ces hautes probabilités de promotion à ces individus.

L'algorithme de SHAP, issu de la théorie des jeux, est un moyen efficace pour interpréter localement un modèle.

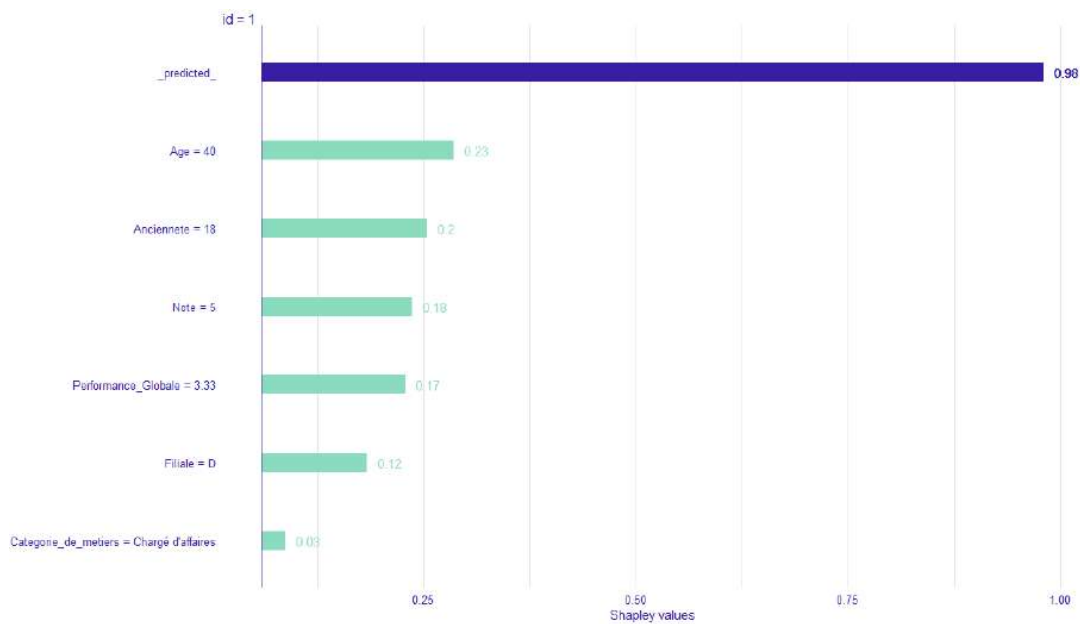


FIGURE 2 – Valeur de Shapley du premier individu

Ici, le modèle a attribué une promotion à l'individu par rapport son ancienneté importante et son coefficient hiérarchique. Les performances de l'individu jouent également en

faveur de l'obtention d'une promotion. L'âge est fortement responsable d'une promotion selon le modèle.

Cette méthodologie basée sur l'association des erreurs du modèle à des potentielles discriminations permet donc d'obtenir une liste d'individus potentiellement discriminés à soumettre aux RH pour une étude plus approfondie.

De la même manière, les écarts d'obtention et de montants d'augmentation ont été étudiés afin de venir identifier des individus potentiellement discriminés. Par la suite, les écarts dans les hautes rémunérations ont été étudiés.

Ecart dans les hautes rémunérations

Dans le cadre de l'étude des écarts dans les hautes rémunérations, les très hauts salaires ont été approchés à travers la théorie des valeurs extrêmes (*Peak over threshold*). L'idée est d'assimiler une haute rémunération comme une valeur extrême et de prédire ces valeurs extrêmes à l'aide d'arbres de Pareto, dans le but de déterminer des individus potentiellement discriminés dans les hautes rémunérations.

Le principe de ces arbres de régression est basé sur une modification de l'algorithme CART, la fonction de perte devient une log vraisemblance, celle d'une distribution de Pareto généralisée.

Il est donc possible d'obtenir une liste d'individus potentiellement discriminés par rapport au sexe dans l'obtention d'une haute rémunération.

Les approches réalisées (Extrem Gradient Boosting, Arbre de Pareto généralisé) dans le cadre de l'étude des hautes rémunérations posent problème puisqu'elles sont construites autour d'un biais. En effet, puisque très peu de femmes sont présentes dans les hauts salaires, les algorithmes vont avoir tendance à discriminer les individus en fonction du sexe pour l'attribution d'un haut salaire.

L'idéal serait d'avoir des outils pour étudier les biais dans les bases de données afin que les modèles ne les incluent pas dans leur construction et engendrent des erreurs dans les résultats.

Un pan de l'apprentissage automatique, appelé "Algorithm Fairness", autrement dit l'équité dans les algorithmes, tend à essayer d'étudier et de corriger les biais dans les processus de décisions automatisés des modèles de Machine Learning.

Avant même de pouvoir étudier les différences de montant de hautes rémunérations entre les hommes et les femmes, il serait intéressant de savoir si le fait d'être dans un haut salaire n'est pas sujet à discrimination, autrement dit si l'attribution des hauts salaires n'est pas biaisé en faveur d'un sexe. Dans cette logique, le simple fait d'être une femme suffirait pour les algorithmes de régression à baisser le montant de prédiction des hauts

salaires toutes choses égales par ailleurs.

On se ramène ainsi à un problème de classification. Un algorithme de préprocessing appelé "Reweight" a été utilisé pour étudier et corriger le biais induit par le sexe dans l'obtention d'un haut salaire.

Il est observé que le groupe des hommes obtient 14,6% de résultats positifs en plus vis à vis de l'obtention d'une haute rémunération par rapport aux femmes.

Par la suite, cet algorithme attribue un poids pour corriger ce biais dans les futurs modèles utilisés. Il suffit alors de modifier la fonction de perte des modèles pour que ceux-ci prennent en compte ces poids et atténuent le biais lié aux hautes rémunérations.

Après avoir budgétiser une campagne de revalorisation salarial, il peut être intéressant de se questionner sur l'impact de ces rattrapages salariaux sur les passifs sociaux des entreprises.

Engagement Sociaux

Ce mémoire s'intéresse plus particulièrement aux indemnités de fin de carrière. L'index a été créé pour être généralisable à tout type d'entreprise c'est pourquoi il est intéressant de pouvoir discuter de cette campagne de revalorisation selon les différentes méthodes comptables qui concernent les entreprises.

Deux méthodes sont distinguées, en accord avec les pratiques de reconnaissance prévues par la norme IAS 19 :

- la première, classique, consiste en un étalement linéaire de l'acquisition des droits entre la date d'entrée dans le régime et la date de départ en retraite.
- la seconde qui correspond à la recommandation de 2021 du comité d'interprétation des normes IFRS (IFRIC) consiste à linéariser l'acquisition des droits sur une période réduite. Cette méthode est recommandée pour l'établissement des comptes consolidés.

Le montant des engagements de l'entreprise sur l'année en cours envers une liste de salariées potentiellement discriminées avant et après revalorisation salariale est établie suivant ces deux méthodes comptables. Le graphique suivant résume l'impact de ces mesures.

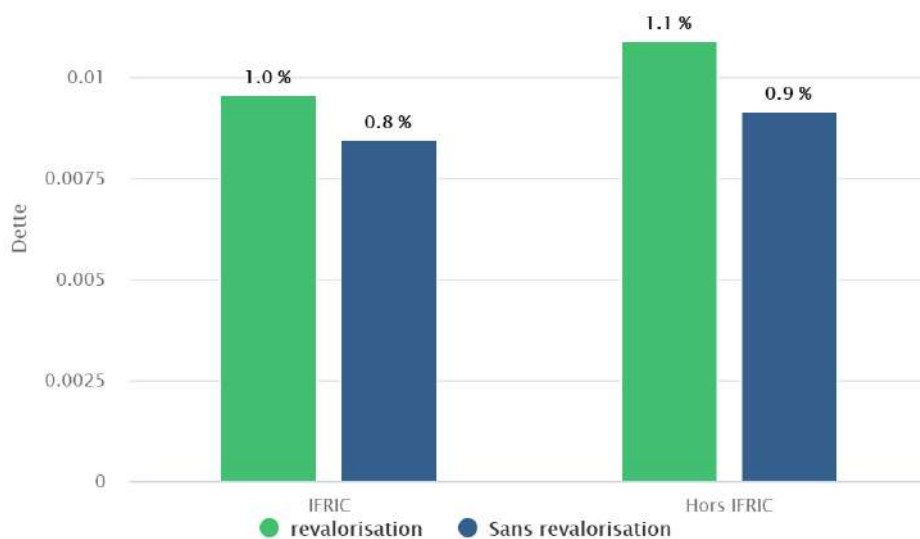


FIGURE 3 – Pourcentage que la dette vis-à-vis de la dette totale

La méthode comptable a un impact assez conséquent sur le montant que l'entreprise doit provisionner au titre des avantages du personnel.

La part de la dette actuarielle de l'entreprise envers ses salariés au titre de l'année en cours pour les individus discriminés varie de 10 points de base entre les deux méthodes. L'impact de la revalorisation sur la dette totale sur l'année en cours est de 20 points de base. Ceci signifie que le montant de la dette des individus discriminés par rapport à la dette totale a augmenté de 0.2%.

On observe une différence de 30 points de base entre aucune revalorisation en méthode comptable IFRIC par rapport à une revalorisation comptabilisé dans l'autre méthode.

L'impact sur les passifs sociaux pourrait être bien plus important qu'un régime IFC dans le cas de régimes à prestations définies octroyant des droits plus élevés.

Conclusion

Ce mémoire introduit une méthodologie pour permettre aux actuaires de se positionner sur un nouveau risque émergent. Le but est de proposer divers outils pour venir en aide aux RH dans la prévention des risques sociaux.

Ces outils d'aide à la décision peuvent être un *listing* d'individus à étudier comme potentiellement discriminés ; un budget théorique de revalorisation salariale à travers l'étude de salaires théoriques ainsi qu'un focus sur l'impact de ces reprises de salaires sur les passifs sociaux des entreprises.

Executive summary

The growth of social inequalities and tensions continues to occupy the middle of the emerging risk maps each year. The intensification of public expectations on issues of societal, environmental and economic inequalities has led to the implementation of various policies and criteria within companies. One index in particular has been put in place from the Avenir Professionnel law (September 2018) in order to fight against gender discrimination in companies. This indicator was created in such a way as to be calculable for all companies and therefore has certain limitations. The aim of this report is to propose a more advanced methodology so that companies can identify potential inequalities within their companies more accurately and thus better prevent the risks.

This index is made up of 4 to 5 sub-indicators depending on the size of the company. Each sub-indicator scores different gaps between women and men in order to establish a score out of 100 points.

The gender pay gap is scored out of 40 points.

The gender pay gap is scored out of 20 points.

The distribution gap for promotions is rated out of 15 points.

The number of female employees receiving raises after maternity leave is out of 15 points.

Parity among the top 10 earners is out of 10 points.

If the index is below 85 points, companies must set and publish targets for progress on each of the indicators. If the index is below 75 points, companies must publish their corrective and remedial measures.

The aim of this thesis is to propose a steering tool in the context of gender inequality to the Human Resources departments of companies.

The idea is to propose a decision-making tool comprising a list of potentially discriminated individuals (according to each sub-indicator) as well as a budgeting of salary catch-up and various indicators (such as theoretical salaries) in order to support them in preventing social risks.

The actuary is qualified to quantify, analyse and prevent risks by using different actuarial tools and indicators.

The subject of gender equality thus appears to be an area where the actuary can play a role, but the conclusions of the models are limited to the reality on the ground of which

only HR is aware : thus the conclusions on potentially discriminated individuals cannot replace a qualitative post-model analysis.

For each indicator, a learning model has been implemented in order to analyse the potential discrimination present in a company.

Pay Gap

The company considered in this report obtained 89 points for the professional equality index, including 39 points out of 40 for the gender pay gap (by hierarchical coefficient and age).

A regression tree model (CART) on pay was used to identify as finely as possible groups of individuals within which potentially discriminated employees are present in terms of pay. This consisted of studying the gender segmentations made by the tree to explain pay.

The model provides an output list of potentially discriminated individuals that needs to be validated by HR afterwards. The model identifies four groups within which there could be wage discrimination between men and women, all other things being equal.



This list of individuals makes it possible to establish an initial global budget for salary catching up which amounts to €413,205.

Subsequently, the differences in promotion and increase are studied.

Gap in promotion

The idea of this part is to be able to identify individuals who are potentially discriminated against in obtaining a promotion and to submit this list of individuals to HR for a more in-depth study.

An XGBoost (Extreme Gradient Boosting) classification model on the variable Promotion is then established by removing the gender variable so that the model is not biased by the gender of the individual.

The aim is to look at all the cases of false positives with a high probability of prediction of promotion, i.e. individuals predicted to be promoted (as almost certain) by the model but who were not in reality.

Then it is necessary to use model interpretability tools like SHAP to be able to explain why the model decided to promote these individuals.

In the first instance, the study of the metrics makes it possible to judge the relevance of a model set up and thus to validate it :

Précision	Recall
89,7%	88,3%

FIGURE 4 – Metrics of the classification model

The model has a recall and an accuracy close to 1, the model has in absolute terms predicted the promotions well. The model being globally relevant, it is advisable to study its errors more precisely in order to verify that they are not due to potential discrimination.

A false positive rate of 5% is observed, we will then study the highest predictions of the model among the false positives.

It is necessary to understand why the model gives these high probabilities of promotion to these individuals.

The SHAP algorithm, derived from game theory, is an efficient way to interpret a model locally.

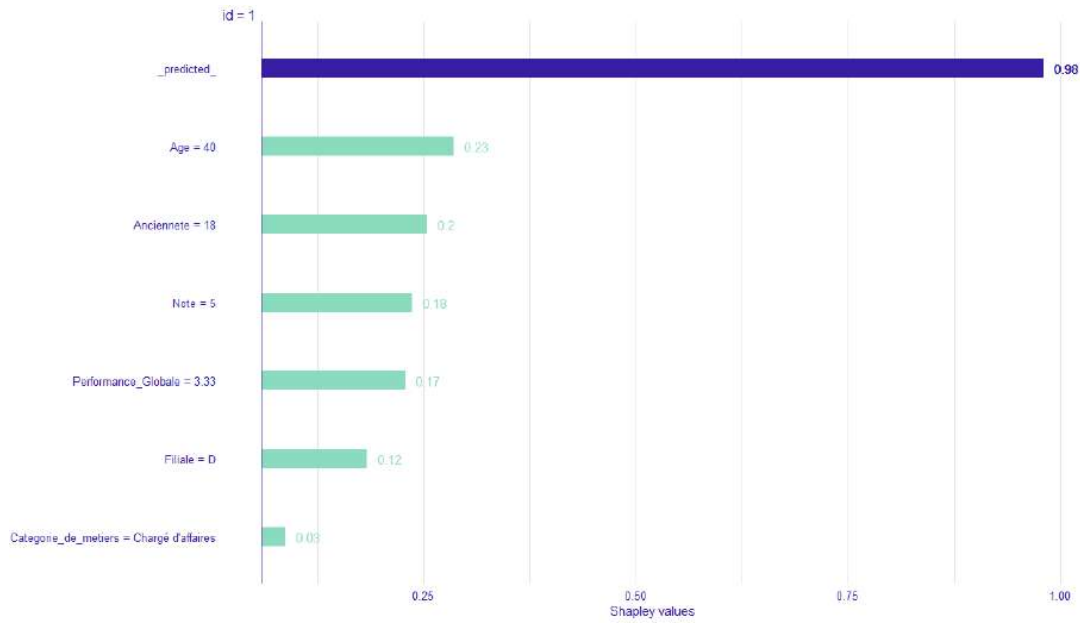


FIGURE 5 – Shapley value of the first individual

Here, the model attributed a promotion to the individual in relation to his high seniority and his hierarchical coefficient. The performance of the individual also plays a role in obtaining a promotion. The age is strongly responsible for the attribution of a promotion to this individual.

This methodology, based on the association of model errors with potential discrimination, makes it possible to obtain a list of potentially discriminated individuals to submit to HR for further study.

In the same way, the differences in the amount of pay rises were studied in order to identify potentially discriminated individuals. Subsequently, the gaps in high pay were studied.

Gap in high salaries

Within the framework of the study of the gaps in high salaries, very high salaries have been approached through the theory of extreme values (*Peak over threshold*). The idea is to assimilate a high salary as an extreme value and to predict these extreme values with the help of Pareto trees, in order to determine potentially discriminated individuals in high salaries.

The principle of these regression trees is based on a modification of the CART algorithm, the loss function becomes a log likelihood of a generalised Pareto distribution.

It is thus possible to obtain a list of individuals potentially discriminated against by sex in obtaining a high salary.

The approaches used (Extrem Gradient Boosting, Generalised Pareto Trees) in the context of the study of high salaries pose problems because they are built around a bias. Indeed, since very few women are present in high salaries, the algorithms will tend to discriminate against individuals on the basis of gender when awarding a high salary. The ideal would be to have tools to study the biases in the databases so that the models do not include them in their construction and generate errors in the results.

One branch of machine learning, called "Algorithm Fairness", tends to try to study and correct for biases in the automated decision processes of Machine Learning models.

Before we can even study the differences in the amount of high salaries between men and women, it would be interesting to know whether the fact of being in a high salary, in other words, whether the attribution of high salaries is not biased. According to this logic, the simple fact of being a woman would be sufficient for the regression algorithms to lower the prediction amount of high salaries, all other things being equal.

This brings us back to a classification problem. A pre-processing algorithm called "Reweight" was used to study and correct for the gender bias in obtaining a high salary.

It was found that the male group had 14.6% more positive outcomes for high pay than the female group.

Subsequently, this algorithm assigns a weight to correct for this bias in future models used. It is then sufficient to modify the loss function of the models so that they take into account these weights and attenuate the bias linked to high salaries.

After budgeting for a pay rise campaign, it may be interesting to consider the impact of these pay rises on the social liabilities of companies.

Social Benefits

This memorandum is more particularly interested in the end-of-career indemnities. The index was created to be generalisable to all types of companies, which is why it is interesting to be able to discuss this revaluation campaign according to the different accounting standards and methods that concern companies.

Two methods are distinguished, in accordance with the recognition practices provided for by IAS 19 : - the first, classic, consists of spreading the acquisition of rights on a straight-line basis between the date of entry into the scheme and the date of retirement.

- the second corresponds to the 2021 recommendation of the IFRS Interpretation Committee (IFRIC) and consists of linearising the acquisition of rights over a reduced period. This method is recommended for the preparation of consolidated accounts.

The amount of the company's commitments over the current year to a list of potentially discriminated employees before and after salary revaluation is established using these two accounting methods. The following graph summarises the impact of these measures.

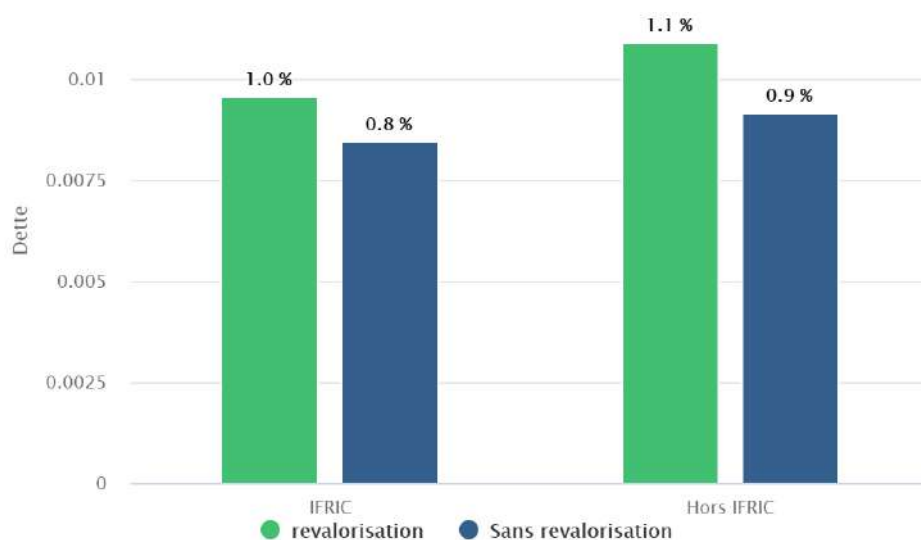


FIGURE 6 – Percentage of debt to total debt

The accounting method has a fairly significant impact on the amount that the company must provision for employee benefits.

The company's share of the actuarial liability to its employees in the current year for discriminated individuals varies by 10 basis points between the two methods. The impact of the revaluation on the total liability in the current year is 20 basis points. This means that the amount of debt of discriminated individuals in relation to the total debt has increased by 0.2%.

There is a difference of 30 basis points between no revaluation under the IFRIC accounting method and a revaluation under the other method.

It should be noted that the illustration made on an IFC scheme could be much more important in the case of defined benefit schemes with higher entitlements.

Conclusion

This thesis introduces a methodology to enable actuaries to position themselves on a new emerging risk. The aim is to propose various tools to assist HR in the prevention of social risks.

These decision-making tools can be a list of individuals to be studied as potentially discriminated against, a theoretical budget for salary increases through the study of theoretical salaries, and a focus on the impact of these salary increases on the social liabilities of companies.

Remerciements

Je souhaite exprimer tous mes remerciements à l'ensemble de Galea & Associés sans qui ce mémoire n'aurait jamais pris forme.

Je tiens bien sûr à exprimer toute ma reconnaissance à Norbert Gautron, président de Galea & Associés, pour sa proximité, son partage d'expérience et d'expertise sur le monde de l'actuariat.

Je remercie mes responsables de mémoire, Léonard Fontaine pour m'avoir accompagné tout au long de ce mémoire, pour sa bienveillance, son aide et la qualité de ses remarques et Nicolas Lesnisse, pour sa disponibilité, son professionnalisme, son soutien, ses conseils et son entrain.

Je remercie infiniment, Alexandre Eby, pour ses précieux conseils et corrections en matière de Machine Learning. Merci Alexandre pour ta disponibilité jusqu'à la dernière minute!

Je remercie Margaux Vaillant, pour m'avoir accueilli dans son équipe, et m'avoir accompagné durant toute cette année d'alternance.

Je remercie Carole Splichal pour la qualité de ses suggestions.

Je remercie Etienne Raynal et ses idées novatrices.

Je remercie également Mélissa Ragneau pour le partage de toute son expertise en matière de protection sociale.

Je souhaite également remercier chaleureusement toute l'équipe enseignante de l'Euria pour l'enseignement de qualité.

J'ai bien sûr une pensée toute particulière pour Franck Vermet, directeur de l'EURIA.

Enfin, j'exprime toute ma gratitude à ma famille et mes amis pour leur écoute, leur patience, leur soutien et leur curiosité.

Table des matières

Introduction	1
I Les risques liés aux écarts de rémunération Femme-Homme	2
1 Les risques liés aux inégalités et tensions sociales comme risques émergents	3
1.1 Les risques émergents dans le monde de l'assurance	3
1.1.1 Qu'est-ce qu'un risque émergent ?	3
1.1.2 Cartographie des risques émergents	3
1.2 Les risques liés aux inégalités sociales	4
1.3 L'écart de rémunération entre les hommes et les femmes dans la presse . .	6
2 Le rôle de l'actuaire dans la gestion des risques sociétaux	9
2.0.1 Les critères environnementaux, sociaux et de gouvernance (<i>ESG</i>) .	9
2.0.2 Engagement de l'IA dans la promotion du métier d'actuaire	10
2.1 Budgétisation RH	10
3 L'index de l'égalité professionnelle	11
3.1 Définition, origine et calcul de l'index	11
3.1.1 Origine et définition	11
3.1.2 Calcul de l'index	11
3.2 Réglementation autour de l'index	13
3.2.1 Nouveautés à venir	13
3.2.2 Sanctions	13
3.3 Taille d'entreprise et Index	13
3.4 Benchmark sur les entreprises du CAC40	14
3.5 Les limites de l'index	14
3.5.1 Sous-indicateur 1	14
3.5.2 Sous-indicateur 2	16
3.5.3 Sous-indicateur 3	16
3.5.4 Sous-indicateur 4	16
3.5.5 Sous-indicateur 5	17

3.5.6	Autres limites	17
II	Méthodes mises en place pour discuter de l'index d'égalité professionnelle	19
III	Base de données utilisée : construction et statistiques descriptives	21
4	Base de données	22
4.1	Construction de la base de données	22
4.1.1	Présentation des données	22
4.1.2	Présentation des variables	24
4.1.3	Corrélations entre variables explicatives	24
4.1.4	Feature Engineering	26
4.2	Statistiques descriptives	27
4.2.1	Répartition Femme-Homme	27
4.2.2	Effectif et Salaire moyen par métier et sexe	29
4.2.3	Comparaison des rémunérations de deux métiers par sexe	30
4.2.4	Masse Salariale	31
4.2.5	Salaire moyen par coefficient hiérarchique	32
4.3	Index pour l'entreprise	32
4.3.1	Note Globale et Tableau récapitulatif	32
IV	Théorie	35
5	Machine Learning : les bases	36
5.1	Quelques généralités	36
5.2	Validation de modèles	40
5.3	Métriques	41
5.3.1	Les métriques dans le cadre d'un problème de régression	41
5.3.2	Le coefficient de détermination linéaire de Pearson (R^2)	42
5.3.3	Métriques dans le cadre d'un problème de classification binaire	42
5.3.4	Courbe ROC et AUC	46
6	Les modèles typiques et optimisation	48
6.1	Generalized Linear Model (GLM)	48
6.2	Classification and Regression Trees (<i>CART</i>)	50
6.3	Extreme Gradient Boosting (Xgboost)	53
7	Interprétabilité des modèles complexes	56
7.0.1	Les méthodes d'interprétabilité	57

V	Présentation des Résultats	60
8	Etude de critère de rémunération	61
8.1	Les objectifs du chapitre	61
8.1.1	GLM	63
8.2	Arbre de régression	65
8.3	Retour sur la budgétisation RH	67
8.4	Limites	69
8.5	Conclusion	70
9	Etude du critère sur la Promotion	71
9.1	Arbre de classification	71
9.2	GLM	72
9.3	Xgboost - modèle de classification sur la Promotion	74
9.3.1	Motivations du modèle	74
9.3.2	Paramétrage du modèle	74
9.3.3	Résultats	75
9.4	Interprétabilité du modèle	79
9.4.1	Interprétabilité Globale	79
9.4.2	Interprétabilité locale	81
9.5	Retour avec les autres modèles	84
9.5.1	Nombre de femmes dans les faux positifs	84
9.5.2	Xgboost avec la variable Sexe	85
9.6	Explication RH	89
9.7	Limite de la méthodologie	90
9.8	Conclusion	90
10	Etude du critère sur l'augmentation	91
10.1	Les objectifs du modèle	91
10.1.1	Rappel sur l'index	91
10.1.2	La démarche	91
10.2	GLM	92
10.2.1	Utilisation d'un GLM et importance de la variable Sexe	92
10.2.2	Création du GLM : ajustement des variables continues	92
10.2.3	Importance des variables et résultats	93
10.3	Xgboost Classification : nombre de revalorisation	93
10.3.1	Creation du Xgboost	93
10.3.2	Résultats	94
10.3.3	Interprétabilité globale	97
10.3.4	Analyse des individus potentiellement discriminés	99
10.4	Introduction de la variable « Sexe »	102
10.5	Xgboost Regression	104
10.5.1	Création d'une revalorisation théorique	105
10.5.2	Conclusion	106

11 Etude du critère sur les plus hauts salaires	107
11.1 Modélisation des salaires par approche ensembliste	108
11.2 Le salaire comme valeur extrême	108
11.2.1 Notions de valeurs extrêmes	109
11.2.2 Fonctionnement des arbres de régression Pareto généralisés	110
11.2.3 Application	112
11.3 Le biais dans les algorithmes	116
11.3.1 "Algorithm Fairness"	116
11.4 Terminologie	118
11.4.1 Introduction du biais dans un algorithme	119
11.4.2 Les différents types d'injustice ("unfairness") notions mathématiques de discrimination	120
11.5 Réduction du biais dans les hauts salaires	124
VI Applications actuarielles : départ à la retraite	127
12 Actuariat des engagements sociaux	128
12.1 Contexte	128
12.2 Les normes existantes	128
12.2.1 Norme IAS 19	129
12.2.2 Identification des avantages sociaux	129
12.3 Comptabilisation des engagements	129
12.3.1 French GAAP	130
12.4 Valorisation des engagements	130
12.4.1 Méthode d'estimation de la valeur des engagements	130
12.4.2 Les hypothèses de projection	131
12.4.3 Méthodes d'étalements des droits	132
12.4.4 Éléments de calculs pour l'approche linéaire (méthode A)	134
12.5 Application	135
12.5.1 Présentation Outil	135
12.5.2 Périmètre	137
12.5.3 Evolution de la dette après rattrapage	138
12.6 Les pensions de retraite en France	140
12.6.1 Retraite Obligatoire	141
12.6.2 Retraite Supplémentaire	141
12.6.3 Impact sur les passifs sociaux	142
12.7 Conclusion	142
Conclusion	143
Bibliographie	145

Table des figures

1	Métriques du modèle de classification	vi
2	Valeur de Shapley du premier individu	vi
3	Pourcentage que la dette vis-à-vis de la dette totale	ix
4	Metrics of the classification model	xii
5	Shapley value of the first individual	xiii
6	Percentage of debt to total debt	xv
1.1	Les 5 plus grands risques émergents à horizon 5 ans en 2022	4
1.2	Cartographie des risques émergents 2020 - à prévision 1 an et 5 ans	5
1.3	Inégalités de salaires par hiérarchie et par département	7
1.4	Ecart de rémunération entre les femmes et les hommes (2012)	8
3.1	Barème de calculs des sous-indicateurs de l'index	12
3.2	Ecart de Salaire non pris en compte dans les hautes CSP	15
3.3	Récapitulatif de l'index de l'égalité professionnelle	18
3.4	Méthodologie mise en place pour approfondir l'index égalité professionnelle	20
4.1	Cheminement de la construction de la base de données	22
4.2	Présentation des variables qualitatives et quantitatives	24
4.3	Matrice de corrélation des variables quantitatives	25
4.4	Matrice de corrélation des variables qualitatives	26
4.5	One hot encoding	27
4.6	Répartition de l'effectif par sexe	27
4.7	Pourcentage de femmes et d'hommes par métier	28
4.8	Effectif et salaire moyen en fonction du type de métiers et du sexe	29
4.9	Rémunération pour deux métiers spécifiques	30
4.10	Masse salariale par sexe et type de métiers	31
4.11	Evolution des salaires moyens par coefficient hiérarchique par sexe	32
4.12	Résultat de l'index et des sous index	33
5.1	Problèmes de prédiction majeurs en Machine Learning	37
5.2	Relation biais variance	38
5.3	Grid search	39
5.4	Cross validation	41

5.5	Illustration du calcul d'une erreur quadratique	42
5.6	Conséquence du seuil de détermination pour un problème de classification	43
5.7	Matrice de confusion	44
5.8	Limite de la métrique « Rappel »	44
5.9	Limite de la métrique « Spécificité »	45
5.10	Limite de la métrique « Précision »	45
5.11	Limite de la métrique « Accuracy »	46
5.12	Courbe ROC	47
6.1	Régression de Poisson	49
6.2	Un exemple d'arbre de classification	50
6.3	Représentation du calcul des résidus	51
6.4	Transcription d'une base de données dans un arbre	52
6.5	La valeur optimale de ω qui minimise l'équation	53
6.6	Algorithmes, pouvoirs prédictifs et interprétabilité	55
7.1	Algorithmes et difficulté d'interprétation	56
7.2	Interprétation avec l'algorithme SHAP	59
8.1	Plan d'action de ce chapitre sur l'étude de la rémunération	61
8.2	Ecart de revalorisation lors du calcul de l'index	62
8.3	Résultat des coefficients du GLM par modalité	63
8.4	Exemple du calcul de l'AIC	64
8.5	Importance des variables GLM par approche stepwise (AIC)	64
8.6	Liste des hyperparamètres choisis	65
8.7	Recherche des individus potentiellement discriminés par la présence de la variable « Sexe »	66
8.8	Récapitulatif des groupes présentant des potentielles discriminations	67
9.1	Plan d'action de ce chapitre sur l'étude de la promotion	71
9.2	Arbre de Classification sur la variable Promotion	72
9.3	Importance des variables du modèle GLM	73
9.4	Paramétrage du modèle	75
9.5	Courbe Précision Recall	75
9.6	Matrice de Confusion	76
9.7	Courbe ROC	77
9.8	Rappel Matrice de confusion	77
9.9	Métriques importantes	78
9.10	Métriques	78
9.11	Courbe Précision Recall	78
9.12	« Summary Plot » ; « Beeswarm Plot »	80
9.13	« Dependance Plot »	81
9.14	Valeur de Shapley du premier individu	82
9.15	Valeur de Shapley du deuxième individu	83

9.16	Valeur de Shapley du troisième individu	84
9.17	Variables importantes du modèle avec la variable « Sexe »	85
9.18	Diagramme en essaim de Shap	86
9.19	Diagramme des dépendances de la variable Coefficient hiérarchique par Sexe	87
9.20	Diagramme des dépendances de la variable Performance par Sexe	88
9.21	Valeur de Shapley d'une femme non promue mais promue par le modèle	89
10.1	Etapas mises en place pour étudier ce sous indicateur	91
10.2	Résultat de l'arbre de régression sur la revalorisation avec la variable Age	92
10.3	Les 5 plus grandes variables contributives du GLM	93
10.4	Hyperparamètre du Xgboost Binaire sur la revalorisation	94
10.5	Matrice de confusion	95
10.6	courbe ROC	95
10.7	Métriques	96
10.8	Précision et Rappel	96
10.9	Diagramme en essaim d'abeille pour la prédiction d'une augmentation	97
10.10	Diagramme de dépendance pour la prédiction d'une augmentation	98
10.11	Diagramme de dépendance de la Promotion pour la prédiction d'une augmentation	99
10.12	1ère valeurs de Shapley	100
10.13	2ème valeurs de Shapley	101
10.14	3ème valeurs de Shapley	101
10.15	Diagramme en essaim d'abeille	102
10.16	Valeurs de Shapley d'un Homme (Faux Positif) avec la Variable 'Sexe' dans le modèle	103
10.17	Valeurs de Shapley d'une Femme (Faux Positif) avec la Variable 'Sexe' dans le modèle	104
10.18	Prédiction du modèle sur la base de test	105
10.19	Prédiction de montant de revalorisation avec et sans le sexe	106
11.1	Répartition des salaires prédits VS théoriques	108
11.2	Découpage en régions	111
11.3	Principe de la méthode	112
11.4	Illustration des résultats	114
11.5	Distribution des salaires	115
11.6	Illustration de l'élément important d'une image (Saliency)	117
11.7	Les différents algorithmes en Algorithm Fairness	120
11.8	Trouver le prédicteur optimal en terme d'égalité de précision et d'égalité des chances	122
11.9	Métriques	124
11.10	Comparaison du pourcentage de résultats favorables (être dans une haute rémunération) dans le groupe des hommes à celui des femmes	125
12.1	Nature des engagements au sens de la norme IAS 19	129

12.2	Méthode comptable linéaire	131
12.3	Rythme d'engagement total de l'employeur par approche	133
12.4	Barème appliqué	137
12.5	Dettes actuarielles en année N pour les individus discriminés	138
12.6	Dettes actuarielles totales de tous les individus	138
12.7	Pourcentage que la dette vis à vis de la dette totale	139
12.8	Proportion de la dette des individus discriminés	140
12.9	Pension en France au fil des années	141

Acronyme

- ***CART*** : Classification And Regression Tree
- ***CSP*** : Catégorie Socio-Professionnelle
- ***ETP*** : Equivalent Temps Plein
- ***FFA*** : Fédération Française de l'Assurance
- ***GLM*** : Generalized Linear Model
- ***H/F*** : Homme / Femme
- ***IA*** : Institut des Actuaire
- ***IAS*** : International Accounting Standards
- ***IFRIC*** : International Financial Reporting Standards
- ***IFRS*** : International Financial Reporting Standards
- ***PIB*** : Produit Intérieur Brut
- ***RF*** : Random Forest
- ***RH*** : Ressources Humaines
- ***RMSE*** : Root Mean Squared Error
- ***XGBoost*** : eXtreme Gradient Boosting

Introduction

Passant de 31 points en 1975 à 8 points en 2018, l'écart de taux d'activité entre les femmes et les hommes s'est considérablement réduit ces dernières décennies. Longtemps admise, la dichotomie entre métiers masculins et féminins est aujourd'hui dépassée. La notion de « métier d'homme » est aujourd'hui une notion désuète avec près de 9 personnes sur 10 considérant que les deux sexes sont à même de travailler dans n'importe quel métier – à l'exception des métiers physiques, qui restent considérés comme plus adaptés aux hommes. Cette perception ne correspond pas toutefois encore à la réalité. Combien même les filles réussissent mieux que les garçons à l'école quels que soient le niveau d'enseignement, la filière ou la discipline, dans les faits, le décalage persiste dans de nombreux domaines. Ainsi, deux fois plus de femmes que d'hommes (27% contre 14% [3]) occupent des postes d'employés au sein des entreprises.

Des progrès dans les politiques de recrutement des entreprises, engendrant une plus grande mixité, ont pu être constatés. Malgré tout, ces évolutions au sein de la société restent incorrectement adressées d'un point de vue rémunération et évolution de carrière et ont incité les pouvoirs politiques à prévoir un cadre législatif qui conduit à mettre en évidence les disparités au sein des entreprises.

L'enjeu reste d'importance : comment correctement comparer des parcours où les interruptions d'activités, liées aux enfants, restent beaucoup plus fréquentes pour les femmes (en 2018, 4 fois plus de femmes à temps partiel que d'hommes) ? Pour l'entreprise, comment prendre en compte l'évolution sociétale et réglementaire, tout en composant avec ses propres spécificités ?

Ce mémoire s'intéressera à l'étude de l'égalité entre les femmes et les hommes à travers le support de l'index de l'égalité professionnelle. Dans un premier temps, il s'agira de présenter l'index et ses limites ainsi que les enjeux des risques sociaux pour les actives. Puis il conviendra d'apporter des solutions plus approfondies et adaptées à chaque entreprise quant aux sous-indicateurs de cet index : les écarts de rémunérations, de promotion, d'augmentation et dans les hautes rémunérations seront abordés avec l'apprentissage automatique, en prenant le temps de discuter de l'éthique dans ces algorithmes. Enfin, un parallèle sur les impacts des campagnes de prévention sur les passifs sociaux des entreprises sera étudié.

Première partie

Les risques liés aux écarts de rémunération Femme-Homme

Chapitre 1

Les risques liés aux inégalités et tensions sociales comme risques émergents

1.1 Les risques émergents dans le monde de l'assurance

1.1.1 Qu'est-ce qu'un risque émergent ?

France Assureurs définit les risques émergents comme "étant les risques en développement ou en évolution qui se caractérisent généralement par une forte incertitude." Le défi que posent les risques émergents tient du manque de données historiques pour caractériser, prévoir et anticiper le danger. L'incertitude peut également venir des "mutations scientifiques, technologiques, sociopolitiques ou les régulations susceptibles de créer des discontinuités."

1.1.2 Cartographie des risques émergents

Chaque année France Assureurs, établit une cartographie des risques émergents pour la profession de l'assurance et de la réassurance [7].

Ces dernières années, le risque climatique, l'environnement économique dégradé (le risque d'une crise du système financier) et le risque cyber occupent tous les trois le haut du classement.

Le classement est toujours en perpétuelle évolution. En effet le risque pandémique était classé dernier (23ème) lors de l'étude réalisée début 2020, puis 3ème pour l'année 2021 et finalement 10ème en 2022.

CLASSEMENT DES RISQUES À 5 ANS			ÉVOLUTION	
RANG	RISQUES	SCORE	RANG	SCORE
1	Cyberattaques majeures	(4,5; 4,3)	(0)	(+0,5; +0,1)
2	Dérèglement climatique	(4,0; 3,9)	(+2)	(+0,4; +0,3)
3	Catastrophes naturelles exceptionnelles	(3,4; 4,1)	Nouveau risque	
4	Poids réglementaire	(3,8; 3,5)	(+7)	(+0,6; +0,4)
5	Environnement économique dégradé	(3,4; 3,7)	(-3)	(-0,8; -0,2)

FIGURE 1.1 – Les 5 plus grands risques émergents à horizon 5 ans en 2022

1.2 Les risques liés aux inégalités sociales

Le risque de croissance des inégalités et tensions sociales occupe quant à lui le milieu de classement (12ème en 2022 et 8ème en 2021) tant d'un point de vue court terme (1 an) que moyen terme (5 ans). Chaque année les définitions sont revues pour correspondre au mieux à la description actuelle des différents risques ; ainsi, suite à la pandémie, la cartographie 2021 a mis davantage l'accent sur la dégradation de la santé mentale.

L'intensification des attentes des populations sur les questions des inégalités sociales, visibles à travers l'accroissement de la justice sociale par le biais des réseaux sociaux tend à accroître le pouvoir de la pression sociale.

C'est pourquoi, le risque de réputation médiatique, et de fait, les incertitudes quant à la pérennité des organisations, se voient classés en tant que risque émergent.

L'augmentation des écarts de revenus entre les différentes professions et catégories socio-professionnelles peut être vue comme une croissance des inégalités et tensions sociales et ainsi un risque émergent significatif selon France Assureurs.

RANG	1 AN		5 ANS	
	RISQUES	SCORE	RISQUES	SCORE
1	(0) Cyber-attaques	(3,8; 3,5)	(0) Cyber-attaques	(4,3; 4,1)
2	(+1) Crise du système financier	(3,1; 3,8)	(+1) Crise du système financier	(3,8; 4,0)
3	(+4) Environnement économique dégradé	(3,3; 3,3)	(-1) Dérèglement climatique	(3,5; 3,7)
4	(+6) Risque de terrorisme	(3,1; 2,5)	(+4) Environnement économique dégradé	(3,5; 3,5)
5	(+3) Qualité des données et leur utilisation	(2,8; 2,8)	(+5) Qualité des données et leur utilisation	(3,4; 3,4)
6	(+10) Augmentation des sanctions réglementaires et du risque de non-conformité	(2,8; 2,6)	(+4) Inadaptation aux nouvelles technologies	(3,2; 3,4)
7	(+5) Poids réglementaire	(2,7; 2,6)	(+7) Augmentation des sanctions réglementaires et du risque de non-conformité	(3,3; 3,1)
8	(0) Risque politique global	(2,9; 2,3)	(+7) Risque de terrorisme	(3,5; 2,8)
9	(-4) Dérèglement climatique	(2,4; 2,8)	(+3) Risque politique global	(3,4; 2,8)
10	(-8) Croissance des inégalités et tensions sociales	(2,8; 2,2)	(-6) Changement de normes comptables et du référentiel prudentiel	(3,1; 3,1)
11	(-7) Changement de normes comptables et du référentiel prudentiel	(2,5; 2,4)	(-2) Croissance des inégalités et tensions sociales	(3,4; 2,8)
12	(-1) Inadaptation aux nouvelles technologies	(2,3; 2,4)	(+1) Dégradation de l'environnement	(3,3; 2,8)
13	(-7) Risque politique européen	(2,3; 2,3)	(-7) Disruption du secteur de l'assurance	(2,9; 3,2)
14	(-1) Impérialisme économique	(2,4; 2,0)	(-7) Poids réglementaire	(3,1; 2,9)
15	(0) Dégradation de l'environnement	(2,2; 2,0)	(+1) Vieillesse de la population	(2,9; 3,0)
16	(+3) Ubérisation de l'économie	(2,2; 1,9)	(-11) Risque politique européen	(3,1; 2,8)
17	(+1) Flux migratoires	(2,6; 1,4)	(+1) Impérialisme économique	(3,0; 2,7)
18	(-2) Judiciarisation et pression sociale	(2,1; 1,9)	(-1) Ubérisation de l'économie	(2,9; 2,5)
19	(-5) Disruption du secteur de l'assurance	(1,8; 2,1)	(+2) Flux migratoires	(3,3; 2,1)
20	(+1) Risque politique français	(1,8; 1,9)	(0) Risque politique français	(2,8; 2,6)
21	(+1) Dégradation de l'habitat	(1,7; 1,8)	(-1) Judiciarisation et pression sociale	(2,7; 2,4)
22	(+1) Vieillesse de la population	(1,6; 1,8)	(0) Dégradation de l'habitat	(2,5; 2,3)
23	(-3) Augmentation du risque épidémique	(1,4; 2,0)	(0) Augmentation du risque épidémique	(2,1; 2,6)

FIGURE 1.2 – Cartographie des risques émergents 2020 - à prévision 1 an et 5 ans

L'actuaire est qualifié pour quantifier, analyser et prévenir les risques en mettant à profit différents outils actuariels et indicateurs.

Le sujet de l'égalité Homme-Femme apparaît ainsi comme un domaine où l'actuaire peut jouer un rôle et a été choisi pour l'étude de ce mémoire.

C'est à travers l'index de l'égalité professionnelle qu'il a été envisagé d'y répondre, abordé à partir du Chapitre 3.

Cependant, avant de commencer, il est intéressant d'avoir quelques chiffres à l'esprit à propos de l'égalité femme-homme.

1.3 L'écart de rémunération entre les hommes et les femmes dans la presse

Quand il s'agit d'écarts de rémunération entre les hommes et les femmes plusieurs chiffres existent dans la presse.

L'étude choisie, est critiquable par son ancienneté mais a été choisie pour son objectivité.

Les chiffres utilisés par la suite sont issus d'une étude de la Direction de l'Animation de la Recherche, des Études et des Statistiques (DARES) [3] de Novembre 2015. "Les résultats sont issus de l'exploitation des déclarations de 2012".

"Ils portent sur l'ensemble des salariés du secteur privé et des entreprises publiques, à l'exclusion du secteur agricole, des agents des trois fonctions publiques (État, territoriale et hospitalière), des salariés des particuliers employeurs ainsi que des apprentis, stagiaires et des postes de travail sans CSP (moins de 0,01 % des postes)."

Le revenu annuel moyen d'une femme est inférieur de 25,7% en 2012 à celui d'un homme. Ce chiffre revient à analyser le pouvoir d'achat d'une femme et d'un homme pris aléatoirement dans la rue.

En prenant en considération l'effet du temps de travail (temps plein, temps partiel) le chiffre est égal à 16,3%, les femmes occupant davantage de temps partiels. D'après une étude de l'INSEE parue en Juillet 2020 , 8 sur 10 des temps partiels sont attribués à des femmes. [9]

Le temps de travail explique à hauteur de 9,4% l'écart de salaire. Ceci signifie qu'à l'heure travaillée, les femmes sont rémunérées 16,3% de moins que les hommes.

L'inégalité horizontale des métiers correspond à une inégalité dans la répartition des individus dans chaque secteur d'activité, c'est-à-dire que certains métiers sont plutôt mixtes, d'autres occupés par des hommes et certains par des femmes. On parle d'inégalité interprofessionnelle ou bien de ségrégation horizontale.

Toutes choses égales par ailleurs, le secteur d'activité explique 3,5% de ces inégalités. Ceci signifie que les métiers plutôt féminins rémunèrent généralement moins que les masculins. Ainsi, il reste 12,8% de différence de rémunération entre les hommes et les femmes.

Quant à l'inégalité intraprofessionnelle, on parle d'inégalité verticale, qui correspond à la position des femmes et des hommes dans la hiérarchie.

Il faut noter deux éléments concernant cette notion.

D'une part il y a la hiérarchie verticale (Salarié, Manager, Directeur...) et d'autre part le département à prendre en considération.

En effet, à niveau semblable de hiérarchie, certains postes sont moins rémunérés.

Par exemple, être au sommet de la hiérarchie du département RH pourrait ne pas être rémunéré autant que la tête du département Actuariat d'une même entreprise.

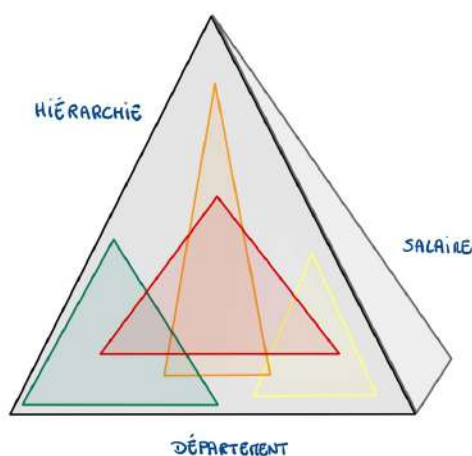


FIGURE 1.3 – Inégalités de salaires par hiérarchie et par département

Visuellement, la pyramide que représente l'entreprise (hiérarchie de salaire en fonction du poste) est composée elle-même de plusieurs pyramides, chacune rémunérant plus ou moins les individus en fonction de la division de l'entreprise (pour le même niveau hiérarchique).

La part des effets de structure au sein des métiers explique 2,3% des inégalités de salaire. Les femmes sont davantage concentrées dans un plus petit nombre de professions qui sont moins rémunératrices que les métiers des hommes.

Finalement à temps de travail, secteur d'activité, département et responsabilité identiques il reste 10,5% d'inégalités non expliquées (ce chiffre reste critiquable, il est difficile de comparer exactement deux individus identiques).

Ce chiffre est à nuancer : il omet d'autres variables potentiellement significatives (par exemple les femmes interrompent davantage leur carrière lors d'une grossesse).

Pour résumer, voici un graphique résumant les différents chiffres :

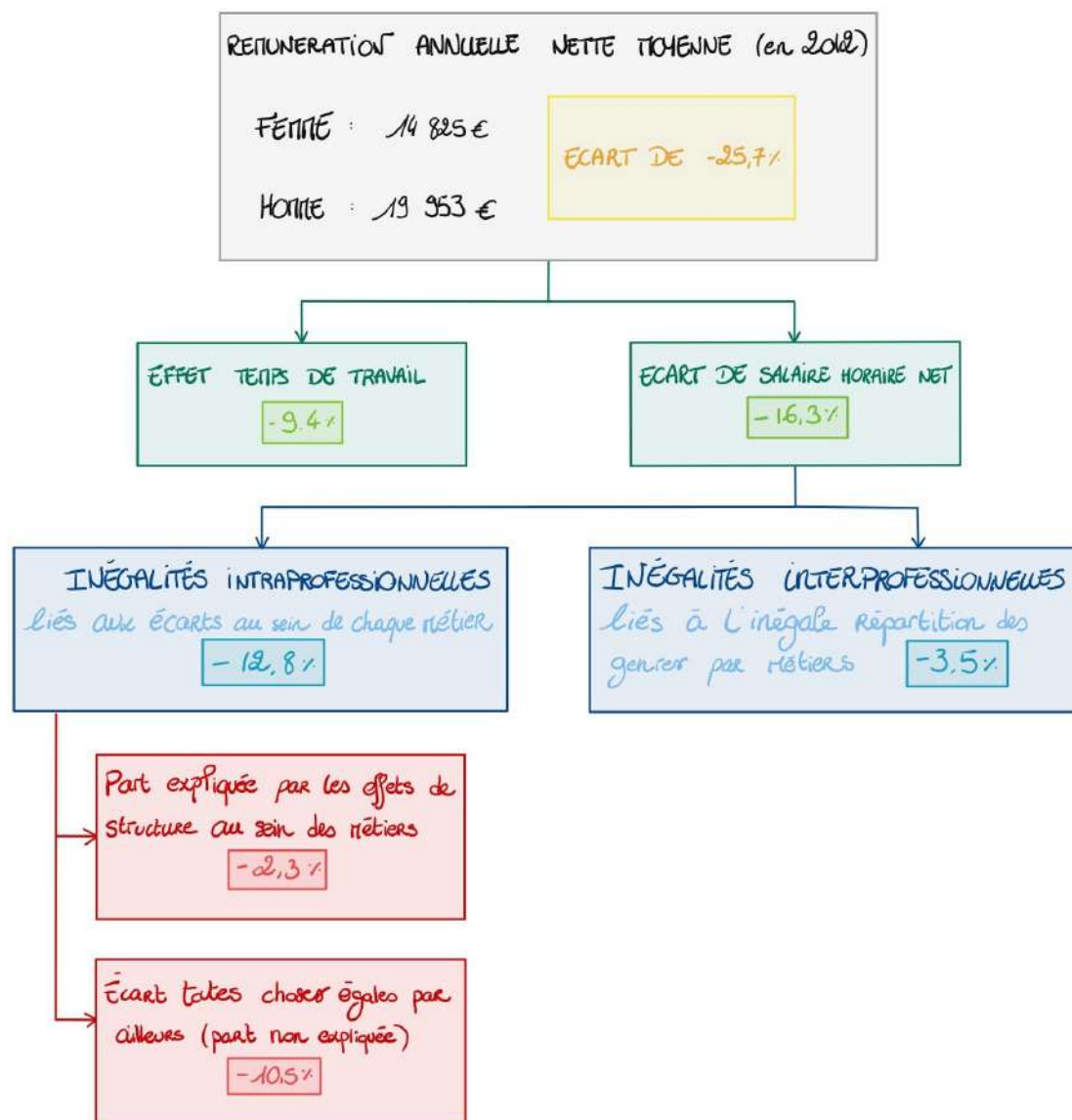


FIGURE 1.4 – Ecart de rémunération entre les femmes et les hommes (2012)

Cet écart de 10,5% reste important c'est pourquoi la mise en place d'une réglementation a été motivée afin de réduire les inégalités constatées.

Chapitre 2

Le rôle de l'actuaire dans la gestion des risques sociétaux

2.0.1 Les critères environnementaux, sociaux et de gouvernance (*ESG*)

Les critères ESG sont des index qui mesurent les activités d'une entreprise pouvant avoir des impacts sur la société ou l'environnement.

Ils servent à évaluer la prise en compte par les différentes instances des enjeux sociétaux et ceux du développement durable. Ils sont utilisés notamment pour les investissements en Finance durable (responsable, verte).

- les critères environnementaux mesurent l'impact de l'activité de l'entreprise sur l'environnement (émissions de CO₂, consommation électrique, recyclage par exemple) ;
- les critères sociaux (ou sociétaux) évaluent l'impact de l'activité de l'entreprise sur les acteurs (salariés, clients...) par référence à des valeurs universelles (droits du travail, éthique, santé des salariés...); par exemple cela peut concerner la qualité du dialogue social.
- les critères de gouvernance traitent de la façon dont l'entreprise est dirigée (la direction, le conseil d'administration, la transparence de la rémunération des dirigeants, la corruption, la féminisation des instances dirigeantes).

De plus en plus d'entreprises prennent acte de ces critères à travers une politique dite RSE, la responsabilité sociétale des entreprises.

Les acteurs économiques intègrent les préoccupations sociales, environnementales, et économiques dans leurs activités de façon volontaire afin d'avoir un impact positif sur la société et de mieux respecter l'environnement. Le risque de baisse de notoriété qu'encourent les entreprises peut avoir des conséquences financières sérieuses allant au-delà du coût de la mesure de prévention. Ils peuvent ainsi se prévaloir de ces risques en adaptant a priori et non a posteriori ces précautions.

2.0.2 Engagement de l'IA dans la promotion du métier d'actuaire

D'une autre manière, l'institut des actuaires s'est engagé en 2021 à travers la création d'une charte sur la parité, projet poussé par Actu'Elles, un groupe de l'institut qui vise à promouvoir la mixité dans l'Actuariat. Ainsi, l'égalité entre les Femmes et les Hommes est désormais inscrite dans les politiques RSE des partenaires de l'IA. Cette charte vise à renforcer la sensibilisation à la mixité dans la profession. D'ailleurs, il est à noter que : L'IA s'est attaché à : «

- Promouvoir l'égalité professionnelle Femme-Homme de ses membres jusqu'au plus haut niveau de responsabilité ;
- Intégrer systématiquement les principes de mixité dans tous les processus de recherche d'intervenants pour des événements organisés pour les membres ;
- Accompagner les membres, femmes et hommes actuaires, avec des programmes dédiés utiles à leur développement professionnel au-delà de leur seule expertise ;
- Diffuser largement la culture de la mixité auprès des membres et auprès des écoles formant les membres, en assurant une démarche de sensibilisation et de formation ;
- Soutenir l'action des réseaux féminins / mixtes des employeurs des membres. »

2.1 Budgétisation RH

Le but de ce mémoire est de proposer un outil d'aide à la décision pour permettre aux RH de pouvoir aller plus loin dans la communication afin d'expliquer l'index de l'égalité professionnelle.

L'idée est de fournir un outil de pilotage plus précis que l'index de l'égalité professionnelle en allant chercher à la maille individuelle des possibles inégalités entre les sexes pour chaque sous-indicateur, toutes choses égales par ailleurs.

L'outil de pilotage pourra présenter : une liste d'individus potentiellement discriminés à étudier, différents indicateurs utiles aux RH (comme des salaires théoriques par profil déterminé à travers les méthodes algorithmiques évoquées par la suite, un dashboard de statistiques globales, un budget théorique de rattrapage salarial ...)

Cette approche vient en prévention des risques sociétaux, en proposant une méthodologie qui permet d'aller plus loin que l'index afin d'en gommer les limites en amont (et d'éviter l'aggravation de celui-ci.)

Pour résumer, les entreprises et notamment les RH sont concernés par ces différents critères et politiques (ESG, RSE, index de l'égalité professionnelle), on vient ici en tant qu'actuaire proposer une aide pour prévenir ces risques sociaux.

L'actuaire est d'autant plus légitime qu'il possède une expertise sur l'étude des conséquences financières des risques que n'aurait pas le data scientist.

Le prochain chapitre abordera la réglementation actuellement en vigueur évoquée dans ce paragraphe, pour traiter des inégalités entre les femmes et les hommes au sein d'une entreprise.

Chapitre 3

L'index de l'égalité professionnelle

3.1 Définition, origine et calcul de l'index

3.1.1 Origine et définition

L'index de l'égalité professionnelle a été mis en place à travers la loi "Avenir Professionnel" du 05 septembre 2018, et vise à évaluer les inégalités entre les femmes et les hommes à travers 4 ou 5 sous-indicateurs en fonction de la taille de l'entreprise.

Chaque année, avant le 1er Mars, les entreprises de plus de 50 salariés doivent calculer et fournir la note de l'index à l'inspection du travail, au Comité Social et Economique (CSE). Cette information doit être visible et lisible sur le site internet de l'entreprise (quand cela est possible).¹

L'index se veut simple et généralisable à toutes les entreprises, c'est pourquoi il est tout à fait contestable comme on pourra le voir dans les sections suivantes. Il s'agit donc d'un indicateur macro qui rend compte de manière globale, il a été créé dans le but d'être facile à mettre en oeuvre au sein de l'entreprise à partir d'informations déjà disponibles dans les SIRH. La critique est donc à atténuer, le calcul se voulant universel à toutes les entreprises. Dans l'idée on pourrait faire un parallèle avec la formule standard de Solvabilité II.

L'objectif de ce mémoire est ainsi de venir *challenge* la robustesse de cet indicateur, ainsi que de mettre en avant des réflexions quant à la mise en place d'un « modèle interne » pour les grandes entreprises. Elles pourront ainsi mieux administrer les inégalités au sein de l'entreprise.

3.1.2 Calcul de l'index

L'index est noté sur 100 points et se décompose ainsi :

1. Cette loi est programmée sur plusieurs années et l'indicateur se verra renforcé d'ici 2030 avec la loi Rixain décrite par la suite.

- Sous-indicateur 1 : Ecart de rémunération [/40 points].
- Sous-indicateur 2 : Ecart de répartition des augmentations individuelles [/35 points] (/20 points pour les entreprises de plus de 250 salariés).
- Sous-indicateur 3 : Ecart de répartition des promotions [/15 points] (concerne seulement les entreprises de plus de 250 salariés).
- Sous-indicateur 4 : Nombre de salariées augmentées à leur retour de congé de maternité [/15 points]
- Sous-indicateur 5 : Parité parmi les 10 plus hautes rémunérations [/10 points].

Sous-indicateur 1 Ecart de rémunération (%)		Sous-indicateur 2 Ecart de taux d'augmentations individuelles (%)		Sous-indicateur 3 Ecart de taux de promotion	
plancher	note	plancher	note	plancher	note
0,0%	40	0,0%	20	0,0%	15
0,1%	39	2,1%	10	2,1%	10
1,1%	38	5,1%	5	5,1%	5
2,1%	37	10,1%	0	10,1%	0
3,1%	36				
4,1%	35				
5,1%	34				
6,1%	33				
7,1%	31				
8,1%	29				
9,1%	27				
10,1%	25				
11,1%	23				
12,1%	21				
13,1%	19				
14,1%	17				
15,1%	14				
16,1%	11				
17,1%	8				
18,1%	5				
19,1%	2				
20,1%	0				

Sous-indicateur 4 Augmentation dans l'année suivant le retour de congé maternité (en % de nombre de salariées)		Sous-indicateur 5 Nombre de salariés du sexe sous-représenté dans les 10 plus hautes rémunérations	
plancher	note	plancher	note
0,0%	0	0,0%	0
100,0%	15	20,0%	5
		40,0%	10

FIGURE 3.1 – Barème de calculs des sous-indicateurs de l'index

EXEMPLE DE CALCULS :

– Avec un écart de salaire entre les femmes et les hommes de 10,5% (après application d'un seuil de pertinence de 2% ou 5%) la note obtenue est de 25 points sur 40 points.

– La méthodologie utilisée pour le calcul de l'index fonctionne, pour l'écart de revalorisation (sous-indicateur 2) et l'écart de promotion (sous-indicateurs 3), en nombre

d'individus promus ou revalorisés.

L'entreprise rentre le nombre d'individus promus ou revalorisés par sexe et peut ainsi connaître sa note. S'il y a plus de 2,1% d'écart en nombre alors la note maximale ne peut être obtenue.

3.2 Réglementation autour de l'index

3.2.1 Nouveautés à venir

À compter de 2022, la loi du 24 décembre 2021 visant à accélérer l'égalité économique et professionnelle, appelée également loi Rixain introduit de nouvelles réglementations quant à cet index.

- En cas d'Index inférieur à 85 points, les entreprises ont jusqu'au 1er septembre 2022 pour fixer et publier des objectifs de progression de chacun des indicateurs. Le décret précise que cela concerne chaque indicateur pour lequel la note n'est pas maximale.
- En cas d'Index inférieur à 75 points, les entreprises ont jusqu'au 1er septembre 2022 pour publier leurs mesures de correction (envisagées ou déjà mises en application) et de rattrapage salarial.

La date butoir du 1er septembre explicitée ci-dessus ne concerne que l'année 2022 qui s'inscrit comme une année transitoire vis à vis de la réglementation de l'index.

Les mesures mises en place sont déterminées par accord ou décision unilatérale de l'employeur après consultation du CSE.

Ces mesures doivent être déclarées au CSE, au ministère du Travail et publiées sur le site internet de l'entreprise.

Le statu quo n'est donc plus d'actualité à partir de 2022 pour les entreprises ayant obtenu une note entre 75 et 85 points.

3.2.2 Sanctions

Chaque entreprise ayant une note inférieure à 75 points (pendant plus de 3 ans d'affilés après la publication du 1er index) sont passibles d'une amende jusqu'à hauteur de 1% de la masse salariale.

Depuis 2019, 32 460 interventions du ministère et de l'inspection du travail ont eu lieu, dont 504 mises en demeure et 32 pénalités financières d'après les résultats du 8 mars 2022 de l'index de l'égalité professionnelle.

3.3 Taille d'entreprise et Index

L'index de l'égalité professionnelle distingue dans son calcul les entreprises de plus de 250 salariés et celles entre 50 et 250 salariés.

Il est intéressant de se pencher sur quelques chiffres concernant la taille des entreprises en France.

D'après l'INSEE [9], il existe en 2015 3,82 millions d'entreprises dans les secteurs principalement marchands non agricoles et non financiers.

- 287 d'entre elles sont qualifiées de "Grandes Entreprises" (GE). Une Grande Entreprise correspond à une entreprise d'au moins 5 000 salariés.
- 5 800 Entreprises sont classées de taille intermédiaire (ETI) : ce sont les entreprises entre 250 et 4 999 salariés.
- Il existe 140 000 Petites et Moyennes Entreprises (PME) : elles sont constituées de moins de 250 personnes.
- 3,67 millions d'entreprise sont qualifiées de microentreprises. Ceci concerne les entreprises de moins de 10 personnes.

(A noter, que Le classement d'une entreprise peut varier en fonction du chiffre d'affaires réalisé par celle-ci).

Pour la suite de ce mémoire, une Grande Entreprise "GE" sera étudiée.

3.4 Benchmark sur les entreprises du CAC40

Les résultats de l'index concernant les entreprises du CAC 40 en 2021 s'élèvent en moyenne à 88 points. Le résultat moyen pour l'année 2021 pour toutes les entreprises évaluées est de 85 points. Le résultat est calculé sur les entreprises du CAC 40 (y compris leurs filiales) et listé dans le fichier distribué sur le Site Index Egapro [\[ici\]](#).

3.5 Les limites de l'index

Si l'index de l'égalité professionnelle a été créé afin de convenir au plus grand nombre d'entreprises, sans dissociation sur leur taille, il présente ainsi un certain nombre de limites qui seront listées dans cette section.

3.5.1 Sous-indicateur 1

L'écart de rémunération se compare soit à travers 4 CSP (uniquement : ouvriers ; employés ; technicien et agents de maîtrise ; ingénieur et cadres) soit par coefficient hiérarchique identique et par tranche d'âge de 10 ans (30-39 ans ; 40-49 ans par exemple). Afin d'avoir le maximum de points il faut que l'écart de rémunération entre Hommes et Femmes soit nul. Une fois tous les écarts de masse salariale entre Femmes et Hommes d'une sous classe (Ouvrier sur la tranche d'âge 30-39 par exemple) calculés, ceux-ci sont sommés afin de donner l'écart global qui servira au calcul du nombre de points.

Cependant un seuil de pertinence est appliqué avant calcul du nombre de points. Si la répartition est faite par coefficient hiérarchique, 2% est retiré de l'écart sinon si la distinction est faite par les 4 CSP explicitées précédemment le seuil est de 5%.

À titre d'exemple :

- à 21% d'écart (19% après le seuil) le nombre de points attribué est de 2.
- à 19% d'écart (17% après le seuil) le nombre de points attribué est de 8.
- à 11% d'écart (9% après le seuil) le nombre de points attribué est de 27.
- à 7% d'écart(5% après le seuil) le nombre de points attribué est de 34.

De même, comme le poids du pourcentage d'inégalité de la sous classe est d'autant plus grand que la sous classe présente un grand nombre d'individus, il est possible d'accroître les inégalités en profitant du maximum de points (ceci afin d'éviter qu'une discrimination marginale n'impacte pas ardemment la note de l'indicateur).

Il suffirait alors d'augmenter le salaire d'un sexe d'une sous population majoritaire légèrement et pouvoir augmenter drastiquement l'écart salarial entre les deux sexes d'une autre sous-classe minoritaire.

Ainsi dès lors qu'il y a une inégalité dans une sous classe pour un des deux sexes, celle-ci peut être compensée par une autre inégalité envers l'autre sexe. En effet, l'inégalité ne sera pas transcrite car en moyenne, celle-ci sera nulle.

Autre défaut, la fracture des sous-classes peut se montrer trop large et ne peut cerner de manière adéquate les inégalités : les choix de carrière significatifs ont davantage lieu entre 30 ans et 39 ans, c'est pourquoi la segmentation devrait être plus fine et permettre des sous-classes davantage homogènes.

Ce sous-indicateur n'a pas pour but de pénaliser les CSP où sont absents des individus. Par exemple, le cas où un des deux sexes ne serait pas représenté dans une des CSP de l'entreprise. Il en résulte qu'il est possible que l'indicateur ne puisse pas déterminer des inégalités dans la répartition par sexe des hautes classes dirigeantes. Donc en cas d'un nombre faible de salariés, la classe ne sera pas étudiée ce qui peut introduire un biais statistique au calcul de l'index.

moins de 30 ans	salaire Femme	salaire Homme	écart	écart après seuil de pertinence	nombre de Femmes	nombre d'Hommes	écart pondéré
30 à 39 ans	-	198 000,00	-	-	-	20	0%
40 à 49 ans	-	313 000,00	-	-	-	20	0%
50 ans et plus	353 200,00	389 000,00	9,39%	7,39%	1	20	0%

FIGURE 3.2 – Ecart de Salaire non pris en compte dans les hautes CSP

3.5.2 Sous-indicateur 2

Une entreprise doit renseigner le nombre de femmes et d'hommes augmentés dans l'année et par un calcul au prorata du nombre d'hommes et de femmes elle se verra accorder : 0 points s'il y'a plus de 10% d'écart en nombre de personnes augmentées.

La perte de points s'effectue à partir de 2% et tombe à 25 points.

La principale limite de ce type de notation est que le calcul s'effectue en nombre et non en montant.

Il suffit d'augmenter à hauteur de 1€ symbolique et il est considéré que l'individu a reçu une augmentation.

Ainsi, n'importe quel montant de revalorisation compte pour la même valeur dans l'indicateur bien qu'ils puissent être sensiblement différents.

3.5.3 Sous-indicateur 3

La définition d'une "promotion" au sens du décret est définie comme le passage à un niveau de classification ou coefficient supérieur, dans la classification de branche ou dans le système de cotation choisi par l'entreprise.

L'entreprise se doit d'être la plus transparente possible sur la méthode de promotion pour que les collaborateurs identifient clairement la notion de promotion.

Le deuxième sous-indicateur ainsi que le troisième sont sensiblement similaires dans leur approche puisqu'ils cherchent à identifier les écarts de promotion et d'augmentation en nombre et non en montant.

De plus, il est important de noter que pour ces deux indicateurs, il n'est pas possible, comme pour l'indicateur 1, de répartir les salariés par niveau ou coefficient hiérarchique. Pour les entreprises de plus de 250 salariés, les salariés sont répartis selon les 4 catégories socioprofessionnelles ouvriers ; employés ; techniciens et agents de maîtrise ; ingénieurs et cadres.

Il est bon de savoir également que le passage à un niveau de classification ou coefficient hiérarchique supérieur n'est pas lié à la répartition des salariés dans les catégories de postes choisie pour le calcul du premier sous-indicateur.

3.5.4 Sous-indicateur 4

Le nombre de femmes augmentées après un congé maternité est similaire à l'autre indicateur de revalorisation puisqu'il ne prend pas en compte le montant de l'augmentation. Il est tout à fait possible d'augmenter d'1€ symbolique une femme de retour de maternité et ne pas se voir pénaliser car seul le nombre compte.

En outre, depuis 2006, il est désormais obligatoire, en loi, d'après l'article L. 1225-26 du Code du travail, d'accorder une augmentation aux salariées à la suite d'un congé de

maternité.

3.5.5 Sous-indicateur 5

Le 5ème indicateur compte le nombre d'individus par sexe au sein des 10 plus grandes rémunérations. Comme son nom l'indique, on regarde le Sexe des 10 plus grandes rémunérations et si le nombre est dans l'intervalle [40%,60%] alors le nombre de points attribués est de 10 sinon 5 points pour la tranche [20%,40%] Le nombre de points attribué est suffisamment faible pour ne pas être impactant au niveau de la note de l'index.

De plus, le sujet de la représentation des femmes dans les fonctions dirigeantes est un sujet majeur, un « plafond de verre » est trop souvent constaté, ce qui signifie que les femmes sont moins nombreuses aux postes hiérarchiques, et moins bien rémunérées. Ce problème sera discuté lorsque l'étude des hauts salaires sera abordée par la suite.

3.5.6 Autres limites

- Le danger de cet indicateur est le statu quo une fois la barre des 75 points passée (qu'il est tout à fait possible d'obtenir tout en maintenant des inégalités). Cependant cette remarque ne sera plus d'actualité à partir de septembre 2022 puisque les entreprises ayant entre 75 et 85 points devront déclarer les mesures mises en place pour avoir le maximum de points dans chaque indicateur.
- Une répartition plus fine de la classification des CSP nécessite que l'entreprise consulte le Comité Sociale et Economique (CSE), afin d'établir avec les représentants élus une segmentation plus juste des salariés (pour le premier indicateur) ce qui peut également être un facteur décourageant.
- Pour chaque classe, un seuil de pertinence existe et vient gommer à hauteur de 2% à 5% les écarts de salaires entre les hommes et les femmes.
Par défaut, le seuil de pertinence est fixé à 5% (pour une catégorisation en 4 CSP). Pour toute autre catégorisation, il est fixé à 2%.
Ainsi s'il existe une différence de salaire de 15%, le sous-indicateur ne prendra en compte que 10%. Comme expliqué ci-dessus, le seuil de pertinence, peut facilement atténuer les écarts et ainsi dissimuler les inégalités. Il est possible de passer d'un score de 15% à 10% et ainsi de doubler le nombre de points obtenus (voir barème ci-dessus Figure 2.1).
- Comparabilité de l'index au fur et à mesure des années : il peut être difficile d'expliquer les variations de l'index notamment suite aux entrées et sorties des salariés. L'index pourrait se voir bonifier sans pour autant faire suite à une amélioration de l'égalité en fonction du sexe.

- Il est nécessaire que les amendes soient suffisamment dissuasives (en terme de coût et d'occurrence) comparé au coût que pourrait représenter une campagne de revalorisation salariale à l'échelle de toute l'entreprise (en prenant en compte que les charges sociales patronales liées au salaire sont proches de 50% du salaire brut).

C'est pourquoi la situation que semble indiquer la note de l'index n'est pas représentative des réalités pouvant exister au sein même de l'entreprise.

La prochaine partie présentera un résumé des méthodes mises en place dans ce mémoire afin de proposer un "modèle interne" à l'index de l'égalité professionnelle. Pour chaque sous-indicateur une autre méthode plus robuste sera proposée.

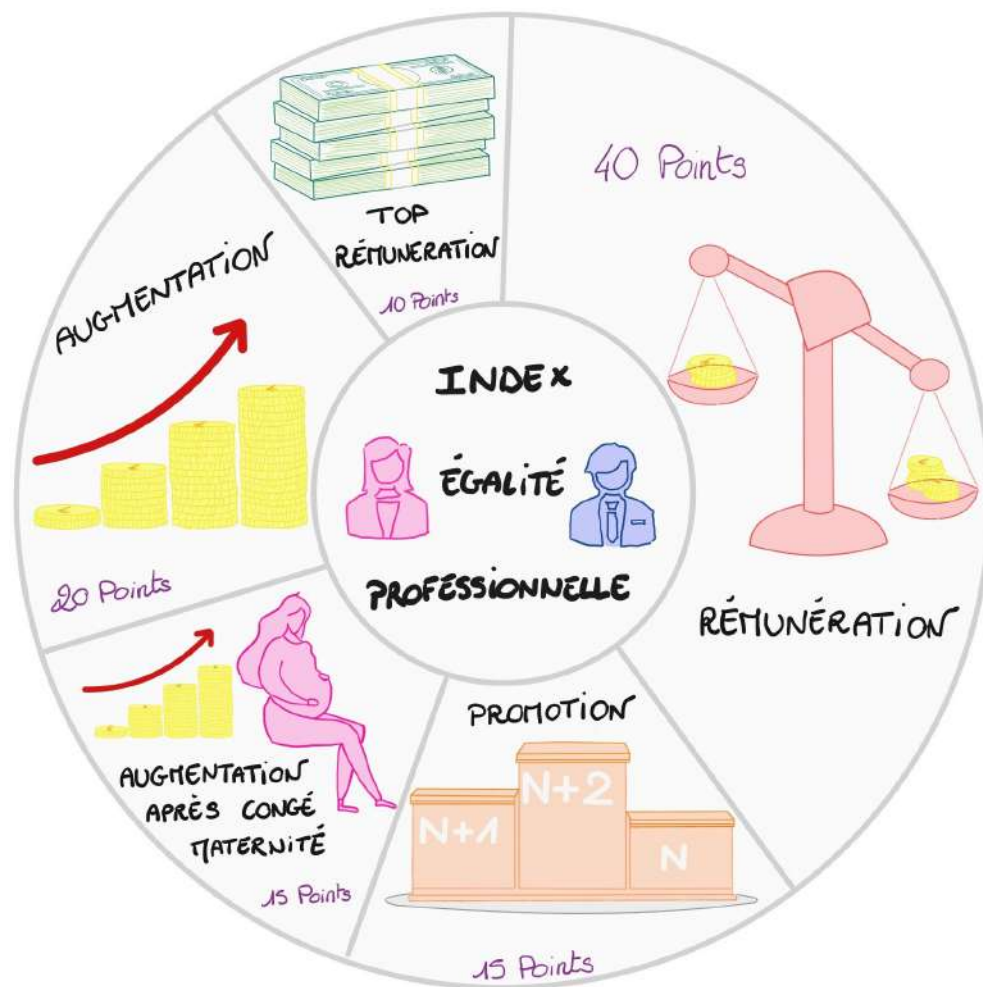


FIGURE 3.3 – Récapitulatif de l'index de l'égalité professionnelle

Deuxième partie

Méthodes mises en place pour discuter de l'index d'égalité professionnelle

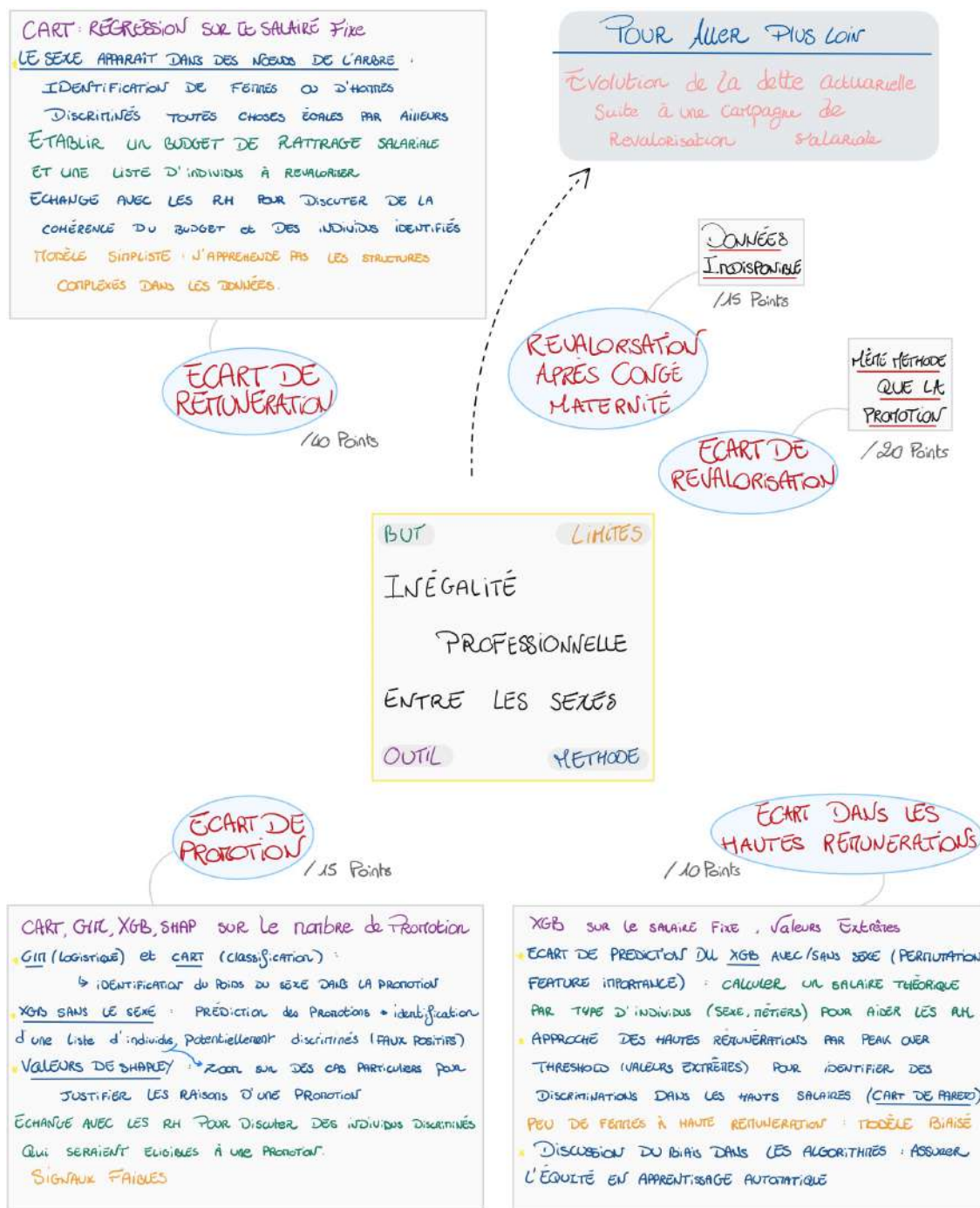


FIGURE 3.4 – Méthodologie mise en place pour approfondir l’index égalité professionnelle

Troisième partie

Base de données utilisée :
construction et statistiques
descriptives

Chapitre 4

Base de données

4.1 Construction de la base de données

Tout d'abord, on rappelle les grandes étapes à suivre pour établir une base de données exploitable.

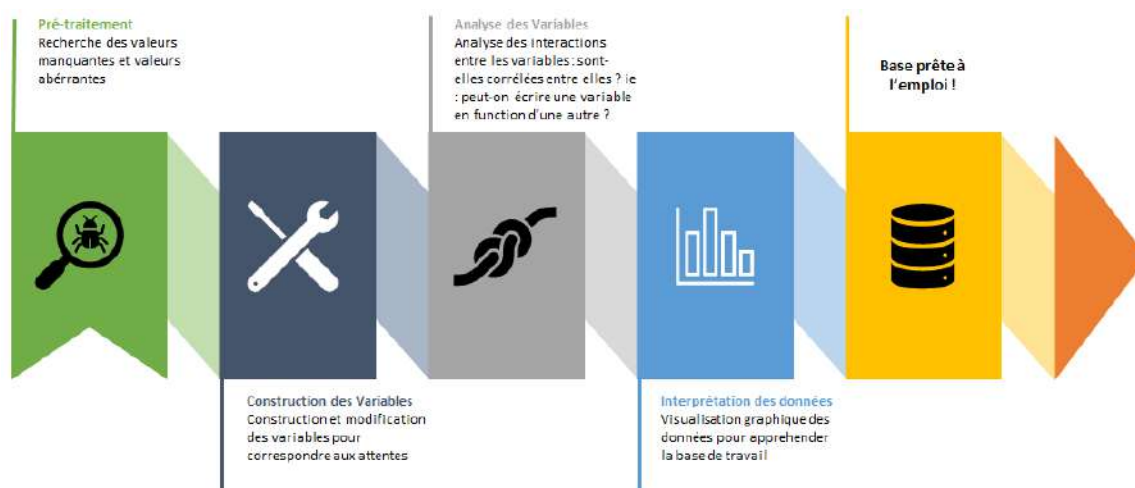


FIGURE 4.1 – Cheminement de la construction de la base de données

4.1.1 Présentation des données

La base de données utilisée contient des données RH.

Celle-ci présente un historique de données allant de 2018 à 2021. 7473 salariés constituent l'entreprise : 2174 femmes et 5299 hommes.

Les individus sont situés dans 4 régions de France : l'entreprise est implantée dans les villes de Nantes, Lyon, Marseille et Paris où réside le siège social. L'entreprise possède 3 filiales.

Les secteurs d'activités des collaborateurs sont très variés, ce qui limite le biais évoqué qui consisterait à considérer un type de métiers déséquilibré dans sa représentation entre les hommes et les femmes.

Globalement il y a peu de retraitements à réaliser. En effet la majeure partie des données étudiées sont issues d'une Déclaration Sociale Nominative (DSN) ce qui permet d'apprécier une bonne qualité de données. La DSN est un processus de communication d'informations sociales entre l'entreprise et divers organismes sociaux. L'utilité de celle-ci réside dans sa facilité de déclaration auprès de l'organisme assureur ainsi que la fiabilisation des données qui sont désormais normées, ouvrant de nouveaux horizons aux deux parties.

Ainsi, très peu de corrections ont été apportées suite à l'étude des données manquantes et des valeurs aberrantes.

Les salaires perçus sont en fixe et correspondent à des temps pleins.

Afin de ne pas biaiser l'étude, il a été convenu avec les RH que les salaires anormalement faibles et élevés soient écartés du périmètre d'étude, puisqu'il s'agit de données erronées. En effet, il ne faudrait pas supprimer des données qui engloberaient justement des discriminations (un homme ou une femme serait trop rémunéré pour un poste et inversement).

Cela concerne des salaires anormalement élevés : plus de 500 000€ à 22 ans dans un métier habituellement peu rémunérateur ; ainsi que des individus ayant perçus moins de 5 000 € par an.

Un indicateur indiquant la performance de l'individu a été créé par l'entreprise ; allant de 0 à 5 (5 étant un individu très performant).

Les indicateurs de performance, ici un indicateur de résultat (efficacité) sont très prisés par les RH puisqu'ils permettent de donner une démarche de progrès et représentent des objectifs à atteindre.

En raison de données manquantes et par souci de fiabilité, cet indicateur a été moyenné sur plusieurs années pour rendre compte au mieux de la performance globale du salarié au fur et à mesure des années.

4.1.2 Présentation des variables

Variable	Description
Âge	Âge de l'individu (de 22 ans à 69 ans)
Ancienneté	Ancienneté de l'individu (de 0 ans à 40 ans)
Coefficient hiérarchique	Échelle hiérarchique (de 1, le plus bas, à 11)
Performance	Historique de Performance (de 0 à 5, 5 étant un individu très performant)
Ville	Paris, Lyon, Nantes, Marseille
Sexe	Binaire ('Homme' ou 'Femme')
Type de métiers	Liste de métiers
Promotion	Binaire ('Promu' ou 'Non promu')
Salaire	Historique de salaire en €
Revalorisation	Historique de revalorisation en %
Filiale	Entité de la filiale auquel appartient l'individu

FIGURE 4.2 – Présentation des variables qualitatives et quantitatives

Le « coefficient hiérarchique » sera abrégé par « Note » dans certains cas.

4.1.3 Corrélations entre variables explicatives

Quantifier la relation entre deux variables est nécessaire afin d'éviter d'implémenter un modèle avec des variables qui pourraient être fortement corrélées entre elles. Cela signifierait que le nombre de variables augmenterait mais n'apporterait pas davantage d'information et donc nuirait à la qualité des prédictions. Il faut tester chacune des variables corrélées et choisir celle qui donne le meilleur modèle.

Pour mesurer la force de la relation linéaire entre deux variables, il est possible d'avoir recours au coefficient de corrélation ainsi décrit :

$$r = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

où $\text{Cov}(X, Y)$ désigne la covariance des variables X et Y , σ_X et σ_Y désignent leurs écarts types.

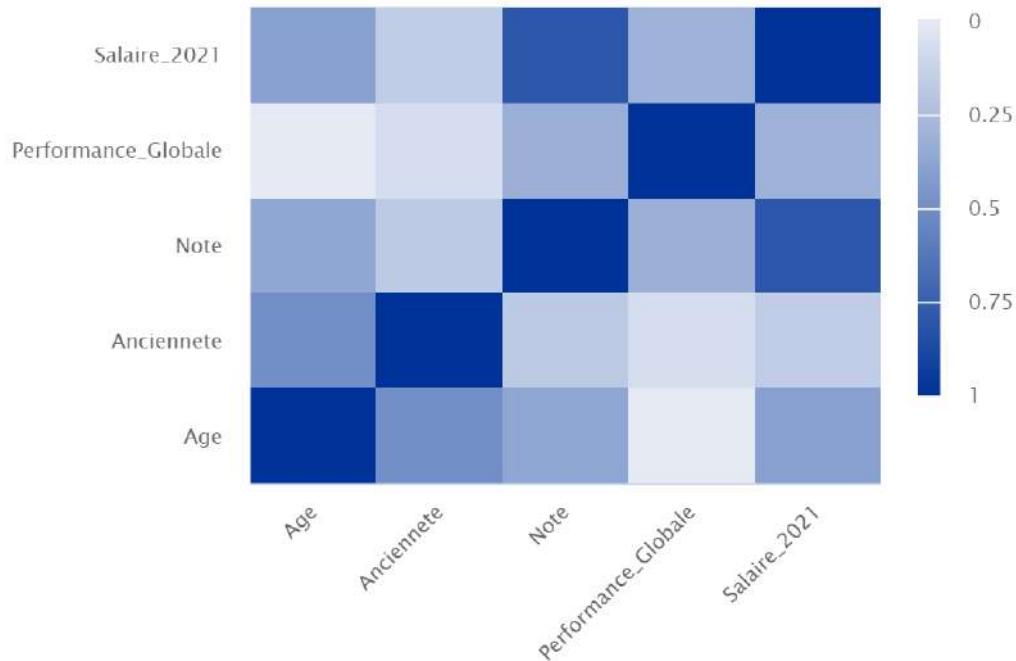


FIGURE 4.3 – Matrice de corrélation des variables quantitatives

Concernant les variables qualitatives, le test V de Cramer, une amélioration du test de Chi 2 de Pearson, peut être utilisé. Il sert à comparer l'intensité du lien entre deux variables. Il est plus fiable que le Chi2 car il est plus stable avec la taille croissante de l'échantillon. Il est défini ainsi :

$$V = \sqrt{\frac{\chi^2}{n \cdot \min(K_1 - 1, K_2 - 1)}}$$

avec, K_1 , K_2 le nombre de modalités des variables X^1 , X^2 respectivement, et n le nombre d'observations.

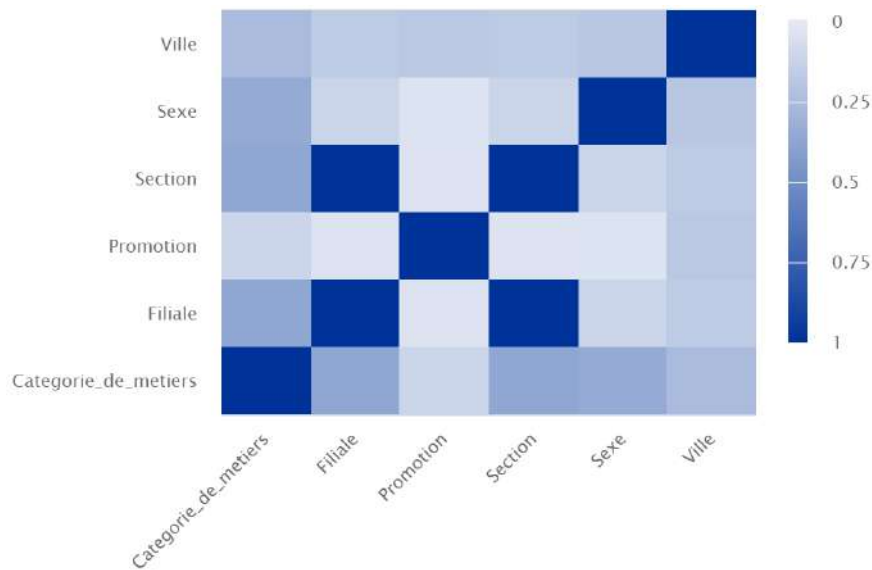


FIGURE 4.4 – Matrice de corrélation des variables qualitatives

Il est observable sur cette matrice de corrélation que la variable "*Filiale*" est très corrélée à la variable "*Section*". En effet, cette variable de la base de données correspond à une autre nomenclature de la variable *Filiale*. Il est donc nécessaire pour la suite de l'étude de supprimer une des deux variables pour que les futurs modèles puissent exercer de meilleures prédictions puisqu'elles donnent essentiellement la même information. La variable "*Section*" est donc écartée de la base de données.

4.1.4 Feature Engineering

Plusieurs transformations sur-mesure des données peuvent être nécessaires afin que le projet de data science soit le plus juste et performant possible.

Il pourrait s'agir de changer d'échelle, par exemple pour traiter des hauts salaires, on peut retraiter en amont les données (normalisation par exemple).

Lorsque le modèle utilisé est un Xgboost il est obligatoire de transformer les données qualitatives en données quantitatives à travers un encodage spécifique. La matrice devant être de rang plein, car la dépendance linéaire entre les variables compromettrait la qualité du modèle, pour " N " modalités, il convient de créer " $N-1$ " nouvelles colonnes.

Sur cette photo il convient de supprimer judicieuse une des trois colonnes.

	Age	Filiale
1	22	A
2	34	A
3	27	C
4	53	B
5	41	C



	Age	Filiale.A	Filiale.B	Filiale.C
1	22	1	0	0
2	34	1	0	0
3	27	0	0	1
4	53	0	1	0
5	41	0	0	1

FIGURE 4.5 – One hot encoding

4.2 Statistiques descriptives

Après avoir construit la base de données, il est intéressant pour mieux appréhender le sujet de visualiser quelques informations concernant cette base.

4.2.1 Répartition Femme-Homme

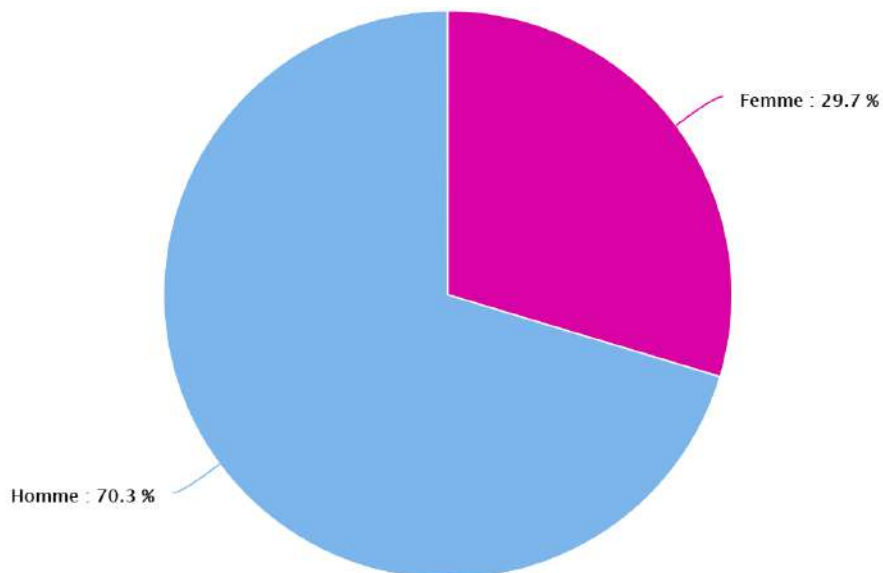


FIGURE 4.6 – Répartition de l'effectif par sexe

La base compte 2174 femmes et 5299 hommes, environ 29% de l'effectif travaillant dans cette entreprise sont des femmes.

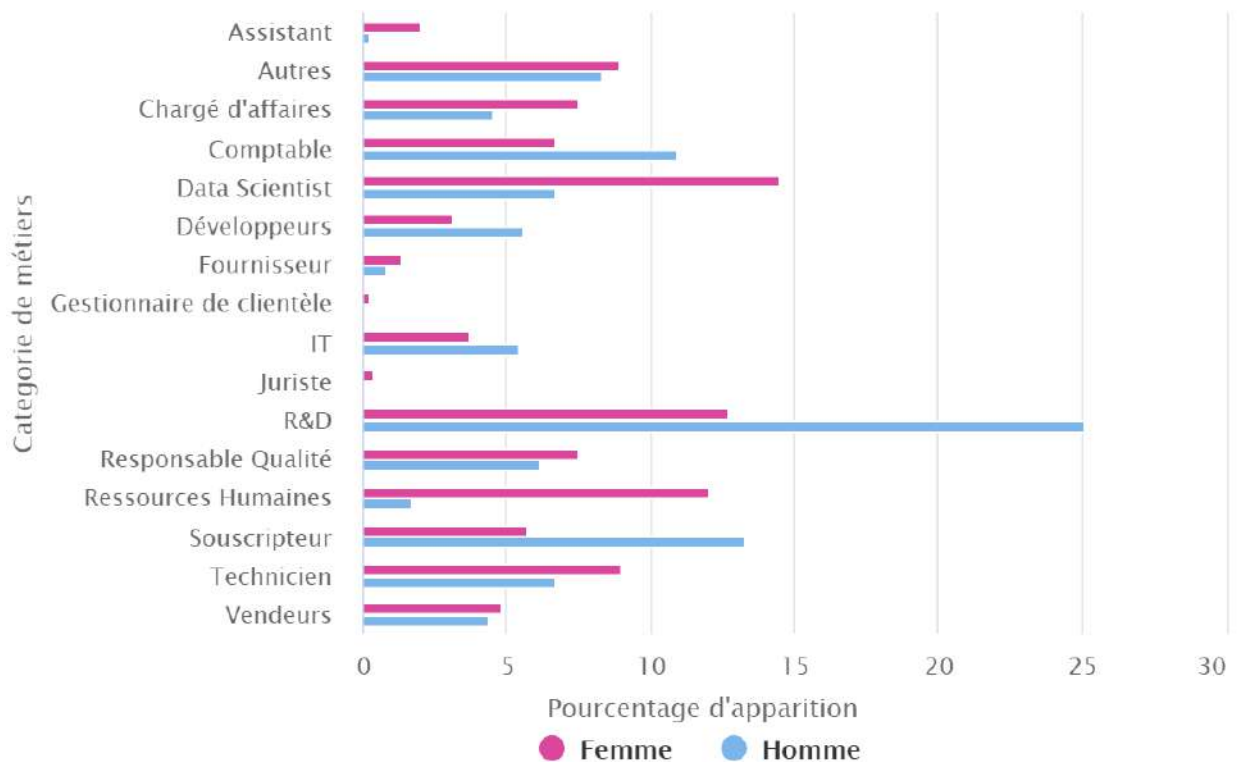


FIGURE 4.7 – Pourcentage de femmes et d’hommes par métier

Dans un souci d’anonymisation des données, les noms de métier ont été changés. Certains métiers sont davantage féminisés (Assistant, data science) et d’autres sont majoritairement occupés par des hommes (R&D).

En proportion du nombre de femmes, le métier de data scientist présente un plus grand nombre de femmes que d’hommes. Parmi les 2174 femmes de la base, 14,5% d’entre elles occupent un poste de data scientist. Parmi les 5299 des hommes de la base, 6,7% occupent ce poste.

A contrario, le métier de Recherche et développement est très populaire chez les hommes, 1 homme sur 4 y travaille.

4.2.2 Effectif et Salaire moyen par métier et sexe

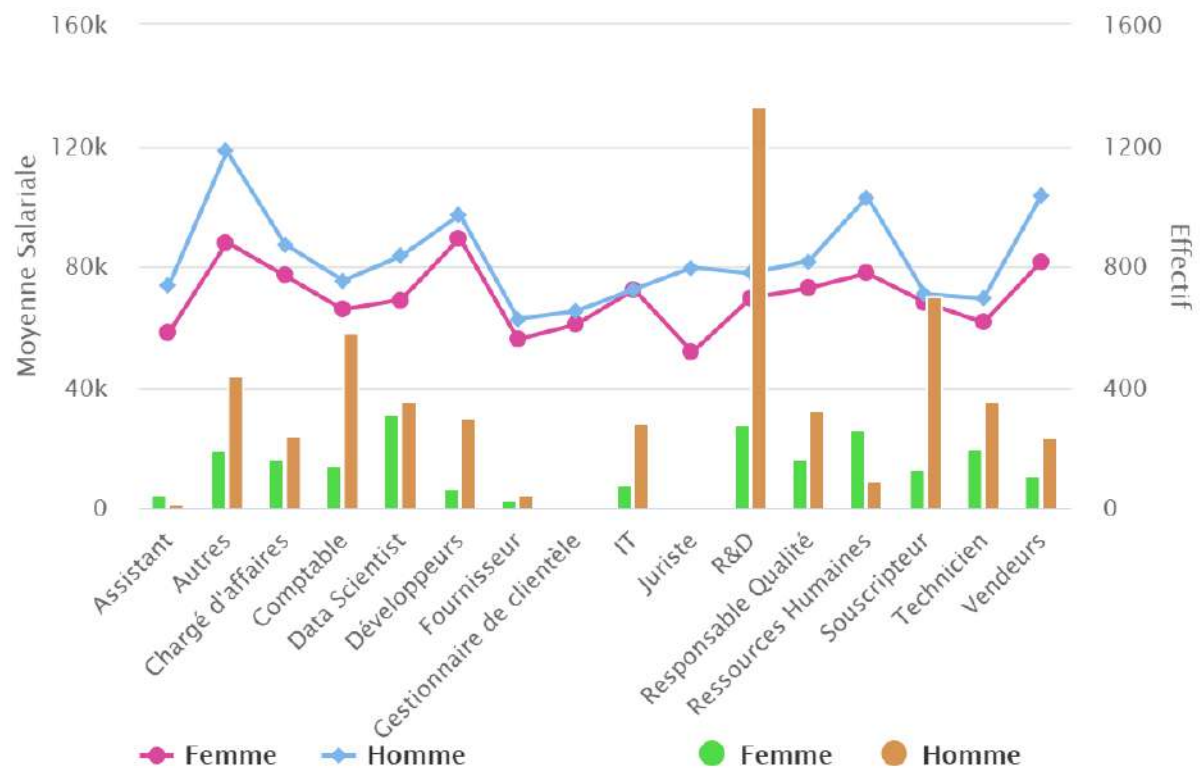


FIGURE 4.8 – Effectif et salaire moyen en fonction du type de métiers et du sexe

Quel que soit le métier le salaire des hommes est en moyenne supérieur à celui des femmes. Des métiers concentrent davantage d'hommes que de femmes, c'est le cas de la Recherche et Développement (R&D), des développeurs, des souscripteurs et de la technologie de l'information (IT).

Les data scientist masculins sont mieux rémunérés en moyenne que les femmes, bien qu'ils soient équirépartis en nombre (bien qu'en proportion il y a davantage de femmes).

4.2.3 Comparaison des rémunérations de deux métiers par sexe

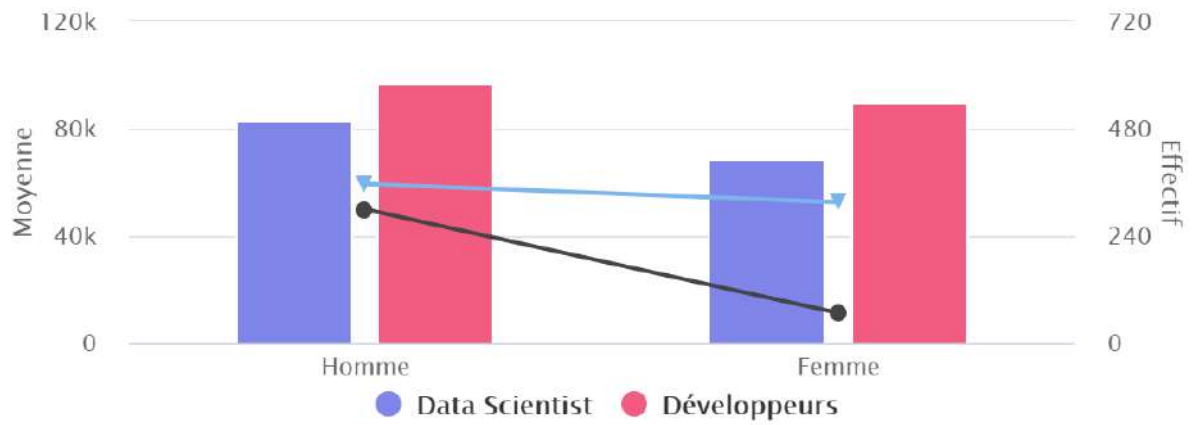


FIGURE 4.9 – Rémunération pour deux métiers spécifiques

Deux métiers similaires dans leur domaine, en terme de rémunération et d'effectif ont été isolés.

Les data scientist et les développeurs sont mieux rémunérés en moyenne chez les hommes, tout âge, ancienneté et coefficient hiérarchique confondus.

4.2.4 Masse Salariale

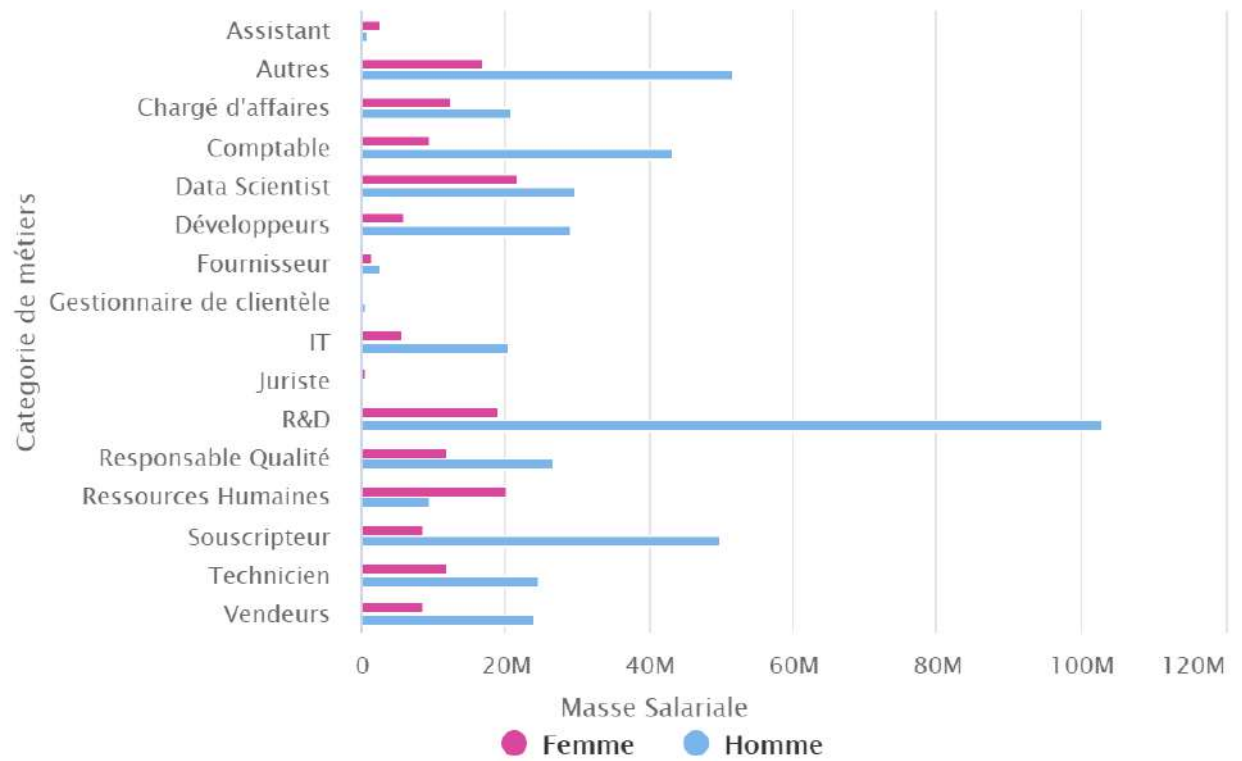


FIGURE 4.10 – Masse salariale par sexe et type de métiers

Il y a moins de femmes en terme d'effectif et en moyenne moins bien payées. La masse salariale totale des femmes est donc sensiblement plus faible.

4.2.5 Salaire moyen par coefficient hiérarchique

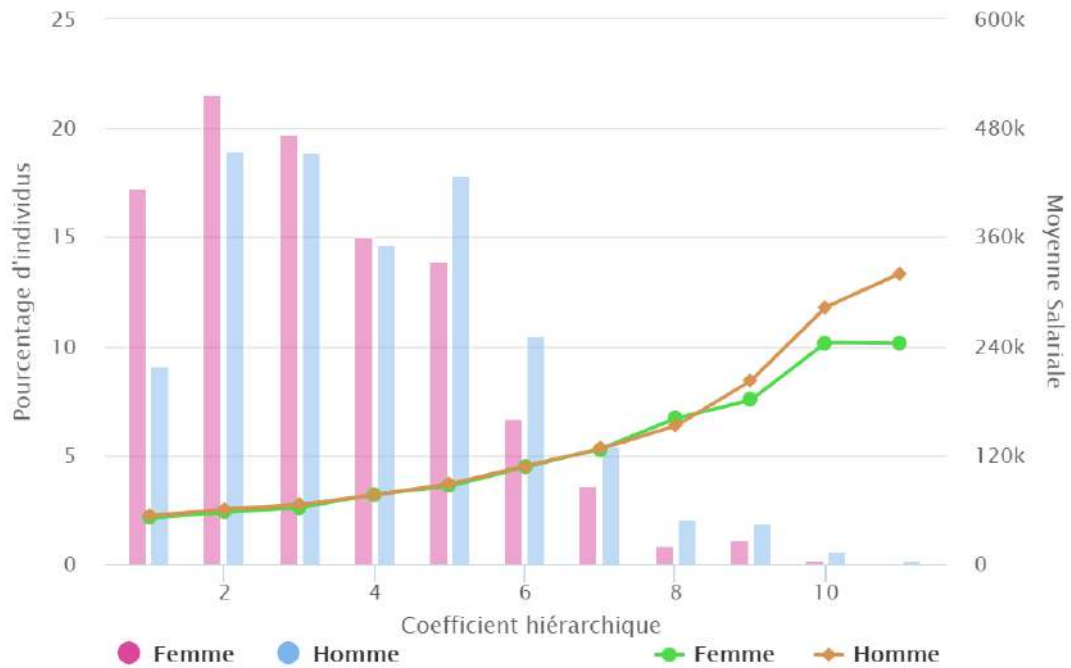


FIGURE 4.11 – Evolution des salaires moyens par coefficient hiérarchique par sexe

Les hommes sont davantage présents dans les métiers à grands coefficients hiérarchiques.

La proportion des femmes est plus grande que celle des hommes dans les faibles coefficients hiérarchiques et a fortiori dans les plus faibles rémunérations.

A coefficient hiérarchique équivalent, le salaire des hommes tend à être légèrement supérieur à celui des femmes dans les coefficients hiérarchiques.

Plus les coefficients hiérarchiques sont élevés, plus la proportion de femmes diminue et l'écart de salaire en moyenne augmente.

Le plafond de verre désigne le fait que, dans une structure hiérarchique, les niveaux supérieurs ne sont pas accessibles à certaines catégories de personnes.

Un plafond de verre semble être illustré dans cette base à travers ce graphique.

4.3 Index pour l'entreprise

4.3.1 Note Globale et Tableau récapitulatif

Le score obtenu pour cette entreprise est de 89 points au titre de l'index. Le détail des points obtenus pour chacun des sous-indicateurs est indiqué dans le tableau récapitulatif

ci-après. Pour rappel (cf. Sanctions), le score de 75 points est le seuil fixé par l'Etat en-deçà duquel les entreprises sont sujettes à une sanction.

Ici, aucune sanction n'est applicable, seul un plan d'amélioration est obligatoire.

L'écart de rémunération a été réalisé à partir d'un classement par coefficient hiérarchique.

La colonne « valeur du sous-indicateur » correspond à l'écart constaté pour l'entreprise pour chaque observation.

Ici, l'écart de rémunération (calculable) constaté après application du seuil de pertinence (2%) est de 0.9%.

Le nombre de revalorisation et de promotion accordés entre les deux sexes est quasi identique ; autrement dit femmes et hommes perçoivent proportionnellement autant d'augmentations et de promotions.

Il y a 2 femmes parmi les 10 meilleures rémunérations, ce nombre est donc inférieur à la répartition [40% ;60%] ou [50% ;50%] nécessaire pour obtenir le nombre maximal de points.

	valeur du sous-indicateur	points obtenus	nombre de points maximum de l'indicateur
1- écart de rémunération (en %)	0,90%	39	40
2- écart de taux d'augmentation individuelles (%)	0,10%	20	20
3- écart de taux de promotion (%)	0,10%	15	15
4- pourcentage de salariées augmentées au retour d'un congé maternité (%)	100%	15	15
5- nombre de salariés du sexe sous-représenté parmi les 10 plus hautes rémunérations	2	0	10
Total des indicateurs calculables		89	100

FIGURE 4.12 – Résultat de l'index et des sous index

Comme abordé à travers les statistiques descriptives, en moyenne les femmes ont tendance à être moins rémunérées, quels sont alors les facteurs discriminants et comment justifier le niveau obtenu par l'index réglementaire ?

Au regard des statistiques descriptives, il serait possible que des disparités puissent exister au sein de l'entreprise étudiée. Cependant ce sont des statistiques globales, il est probable que cette vision soit biaisée. Ainsi une étude plus approfondie sera menée pour

y répondre.

C'est pourquoi l'utilisation d'outils plus complexes et plus puissants seront utilisés par la suite pour rendre compte au mieux des disparités constatées.

La partie suivante présentera ces outils théoriques utilisés pour *challenge* l'index d'égalité professionnelle.

Quatrième partie

Théorie

Chapitre 5

Machine Learning : les bases

5.1 Quelques généralités

Le principe de base de l'apprentissage automatique (Machine Learning) consiste à la création d'algorithmes, lesquels à partir de données fournies en entrée, sont capables de prédire une sortie à l'aide d'une fonction.

Pour se faire, il va falloir que l'algorithme s'entraîne à prédire, puis qu'il vérifie si ses prédictions sont les bonnes.

Il faut donc séparer les données d'intérêt en deux jeux de données : la base d'apprentissage sur laquelle l'algorithme va s'entraîner et une base de test afin de confirmer le pouvoir prédictif de la machine et de vérifier sa faculté de généralisation à des données non observées lors de la phase d'apprentissage.

Il existe plusieurs types d'apprentissage cependant il convient d'en distinguer deux en particulier : l'apprentissage supervisé et non supervisé.

Apprentissage supervisé

L'apprentissage supervisé, utilisé ici dans le cadre de ce mémoire, est, comme son nom l'indique, un apprentissage guidé. La machine va apprendre une fonction à partir d'exemples étiquetés, annotés de la bonne réponse à obtenir (on parle de données labellisées). Il est important d'y distinguer deux types de problèmes :

Si le résultat attendu est constitué d'une ou plusieurs variables continues (quantitative) il s'agit d'un **problème de régression**. Par exemple, ce serait la prédiction de la taille en fonction du poids d'un individu.

Lorsque le résultat attendu appartient à deux ou plusieurs classes, il s'agit d'un **problème de classification**. A partir de données déjà labellisées le but est de prédire la classe de données non étiquetées. L'idée consiste à savoir classer correctement les données. Un exemple de problème de classification serait de savoir prédire si un individu est atteint d'une maladie grave ou non.

Pour résumer, cela peut se voir visuellement ainsi : pour la classification il s'agit de distinguer les données roses des données bleues.

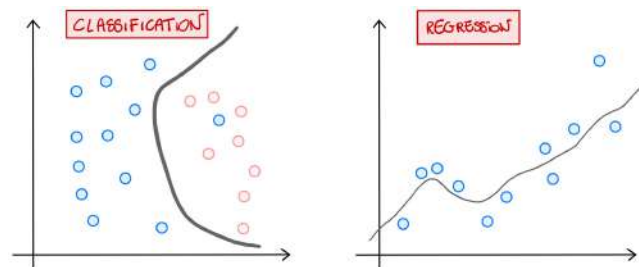


FIGURE 5.1 – Problèmes de prédiction majeurs en Machine Learning

On pose le cadre mathématique suivant dans le cas de l'apprentissage supervisé :

Soit $(\Omega, \mathbb{F}, \mathbb{P})$ un espace probabilisé. On note $X \in \mathbb{R}^p$ le vecteur aléatoire composé de p variables aléatoires explicatives, appelé également l'entrée (ce sont les observations de la base de données). On note $Y \in \mathbb{R}^1$ la variable aléatoire à expliquer (également appelé sortie de l'algorithme, label, étiquette, cible).

Le but est de déterminer une fonction (mesurable) de prédiction de la sortie Y notée f .

Soit \hat{y} les valeurs prédites par l'algorithme et y les valeurs réellement observées en sortie. Soit une fonction de coût qui évalue la perte entre les y et les \hat{y} , autrement dit l'écart entre les valeurs prédites et les valeurs observées.

On définit la fonction de coût par :

$$\mathcal{L}(Y, f(X)) = (Y - f(X))^2 :$$

En particulier, on peut définir une fonction de perte par :

$$\text{dans le cas d'une régression } Y \in \mathbb{R} : \forall (y, \hat{y}) \in \mathbb{R}^2, \mathcal{L}(y, \hat{y}) = (y - \hat{y})^2$$

$$\text{dans le cas de la classification } Y \text{ a un nombre de valeurs possible fini : } \forall (y, \hat{y}) \in \{0, 1\}^2, \mathcal{L}(y, \hat{y}) = \mathbb{1}_{y \neq \hat{y}}$$

Le but est de déterminer la fonction de prédiction f qui minimise la fonction de coût \mathcal{L} . On définit le risque comme la quantité : $\mathcal{R} = \mathbb{E}[\mathcal{L}(Y, f(X))]$ qui permet d'évaluer la qualité de la fonction de prédiction.

La meilleure fonction de prédiction f est celle qui minimise le risque \mathcal{R} c'est-à-dire la quantité :

$$\operatorname{argmin}_f \mathbb{E}[\mathcal{L}(Y, f(X))]$$

L'optimisation de ces fonctions conduisent également à certains dilemmes qu'il convient de rappeler. On va donc maintenant s'intéresser à quelques notions de base à bien avoir en tête lorsque l'on mène un projet de machine learning.

On parle de biais quand l'espérance des prédictions d'un algorithme est éloignée de sa cible. Ce problème apparaît lorsque le modèle est trop simple (ou ne fait pas intervenir assez de paramètres) ce qui entraîne une erreur de prédiction.

D'un autre côté, on parle de variance quand le modèle est sensible à la variation des données d'entrée. Ce problème apparaît lorsque le modèle est trop complexe et peut entraîner ce qu'on appelle le surapprentissage : c'est-à-dire le fait que, lorsqu'un modèle trop complexe est entraîné sur les données d'apprentissage, les résultats seront bons sur la base d'entraînement mais il aura du mal à généraliser ses résultats sur la base de test. Biais et variance sont fondamentalement liés et le levier de contrôle est la complexité du modèle : lorsque l'on enrichit le modèle, le biais diminue mais la variance augmente et lorsque l'on simplifie le modèle, la variance diminue mais le biais augmente. C'est ce que l'on appelle le dilemme biais-variance. Il s'agit de déterminer le bon modèle qui permet d'obtenir des résultats satisfaisants tout en évitant le sur-apprentissage.

Cela peut se résumer avec le schéma suivant :

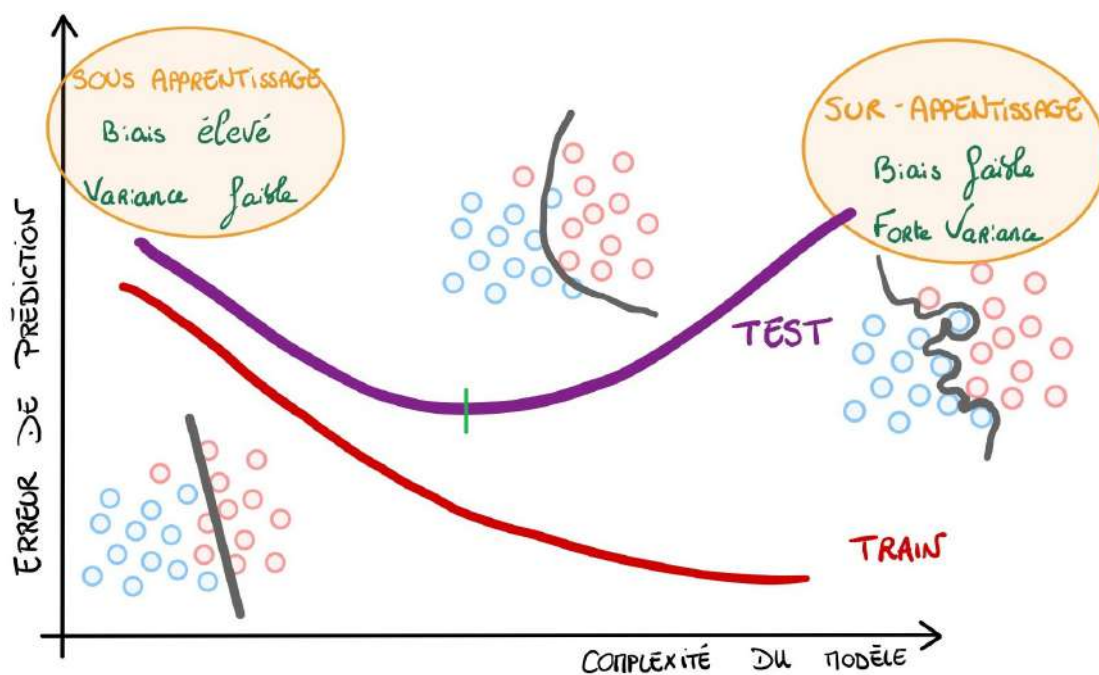


FIGURE 5.2 – Relation biais variance

L'idéal est donc d'être au milieu du graphique là où l'erreur sur la base de test est la

plus faible.

Mathématiquement, si l'on suppose un ensemble d'apprentissage constitué d'un ensemble de points x_1, \dots, x_n avec $\forall i \in \{1, \dots, n\} x_i \in \mathbb{R}^p$ et de valeurs réelles y_i associée à chaque point x_i où $\forall i \in \{1, \dots, n\} y_i \in \mathbb{R}$ et on suppose qu'il est possible d'écrire :

$$y_i = f(x_i) + \epsilon_i \text{ où le bruit, } \epsilon_i \text{ a une moyenne nulle et une variance } \sigma^2$$

Soit \hat{f} la fonction associée au modèle utilisé où $\mathbb{E}[(Y - \hat{f}(X))^2]$ est l'erreur attendue. Celle-ci s'écrit :

$$\mathbb{E}[(Y - \hat{f}(X))^2] = [\text{Biais } [\hat{f}(X)]]^2 + \text{Var } [\hat{f}(X)] + \sigma^2$$

avec :

$$\text{Biais } [\hat{f}(X)] = \mathbb{E}[\hat{f}(X) - f(X)] \quad \text{Var } [\hat{f}(X)] = \mathbb{E}[(\hat{f}(X) - \mathbb{E}[\hat{f}(X)])^2]$$

Les modèles plus complexes de machine Learning fonctionnent avec des paramètres définis avant le lancement de l'algorithme : appelés *hyperparamètres*.

Avant d'analyser les résultats des différents modèles mis en place dans un projet de Data Science, il convient d'optimiser les différents paramètres qui les composent afin d'obtenir un modèle qui convient au mieux au problème étudié.

Ces paramètres ont des valeurs par défaut mais il convient de choisir leur valeur qui donnent les meilleures prédictions. Dans la littérature, on parle d'optimisation d'hyperparamètres. Ils ont pour fonction de contrôler l'apprentissage de l'algorithme.

Une méthode classique consiste à tester toutes les valeurs possibles pour trouver celle qui minimisera l'erreur des prédictions (sur la base de test).

L'idée est de créer une grille de paramètre et d'essayer toutes les combinaisons possibles : cela s'appelle un *Grid search* littéralement *Grille de Recherche*. Visuellement cela peut se voir ainsi :

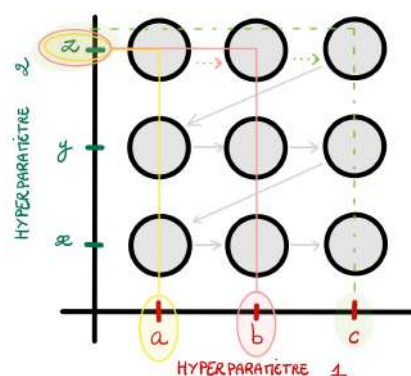


FIGURE 5.3 – Grid search

L'objectif est donc de trouver les meilleurs paramètres pour le modèle afin de minimiser l'erreur sur la base de test et d'éviter les problèmes d'overfitting (surapprentissage).

5.2 Validation de modèles

Pour les modèles les plus simples, les bases de train et de test peuvent suffirent cependant pour des modèles plus complexes avec des hyperparamètres, il est nécessaire d'avoir une base spécifique pour les sélectionner.

Cependant, disposer de trois bases de données (apprentissage, validation et test) peut être compliqué c'est pourquoi le moyen idéal est d'avoir recours à ce qu'on appelle la cross validation à k-couches.

Cross Validation

La validation croisée consiste à entraîner un modèle sur une base d'apprentissage pour différents paramètres choisis, et de tester les résultats sur un échantillon de données d'apprentissage.

La cross validation est utilisée pour sélectionner les meilleurs modèles parmi un jeu d'hyperparamètres. Autrement dit, la meilleure valeur de ce paramètre est celle qui minimisera l'erreur.

La base de données d'apprentissage est divisée en couche de même taille ; par exemple pour une CV-10 folds, cela correspond à une segmentation en 10 couches soit en 10 parties ou en 10 sous bases.

Le modèle est entraîné sur 9 parmi 10 des sous couches et validé sur la 10ème. Ce processus est répété pour toutes les décompositions possibles.

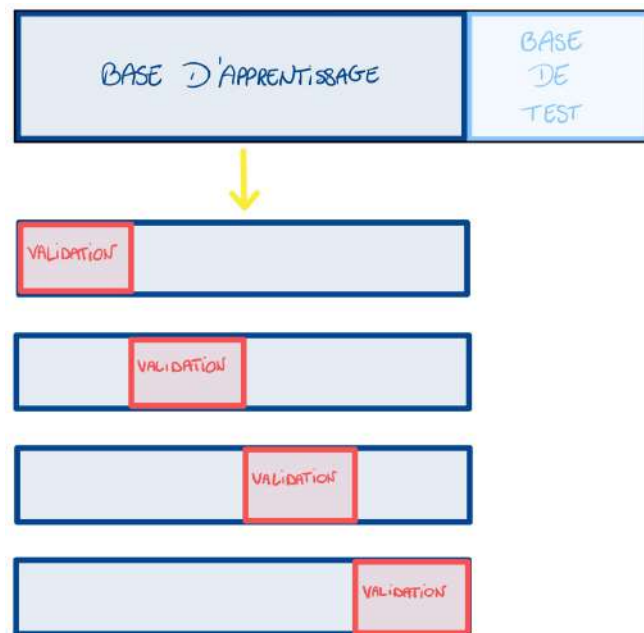


FIGURE 5.4 – Cross validation

La phase d'apprentissage a lieu sur la base d'apprentissage ; l'optimisation des hyper-paramètres est faite à travers la base de validation. La base de Test est utilisée comme expliquée précédemment pour évaluer l'apprentissage réalisé et donc "tester" à quel point l'algorithme est généralisable sur d'autres données.

Calculer l'erreur de prédiction d'un algorithme sur la base de test est un bon moyen pour savoir si l'algorithme est efficace ou non. Il existe ainsi plusieurs métriques utilisés pour évaluer les performances des modèles.

5.3 Métriques

Il convient de distinguer les métriques à utiliser dans le cadre d'un problème de régression et de classification.

5.3.1 Les métriques dans le cadre d'un problème de régression

L'erreur quadratique moyenne (Root Mean Square Error - RMSE)

Fréquemment utilisé, l'erreur quadratique moyenne comme son nom l'indique est la racine carrée de la moyenne des erreurs (c'est-à-dire la différence entre les valeurs prédites

par le modèle et les valeurs observées) au carré.
$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}}$$

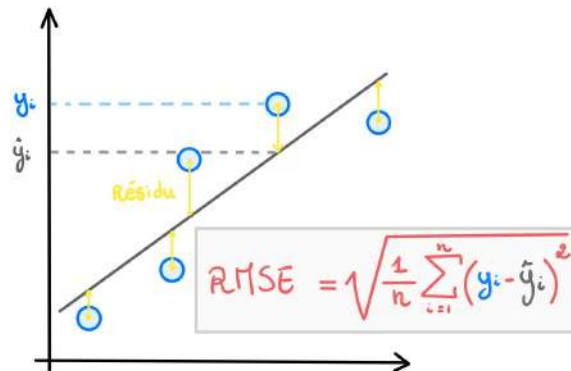


FIGURE 5.5 – Illustration du calcul d'une erreur quadratique

5.3.2 Le coefficient de détermination linéaire de Pearson (R^2)

Le coefficient de détermination linéaire de Pearson, R^2 , est une mesure de la qualité de la prédiction d'une régression linéaire. Il fournit une mesure de l'efficacité avec laquelle les résultats observés sont reproduits par le modèle. Il est défini par la variance expliquée par le modèle divisée par la variance totale.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Le R^2 peut être interprété comme la proportion de la variation des réponses "expliquée" par les régresseurs du modèle.

Ainsi, un $R^2 = 1$ indique que le modèle explique toute la variabilité de la réponse.

Tandis que $R^2 = 0$ indique l'absence de relation "linéaire".

Un $R^2 = 0,8$ signifie que : 80% la variance de la variable de réponse peuvent être expliqués par les variables explicatives.

Les 20% restants peuvent être attribués à des variables inconnues, cachées ou à la variabilité inhérente.

5.3.3 Métriques dans le cadre d'un problème de classification binaire

Dans un problème de classification binaire, les modèles de machine learning prédisent une probabilité d'appartenance à une classe (1 ou 0 dans le cas binaire). Il convient donc à l'utilisateur de choisir un seuil de classification pour lequel le modèle va trier les données dans chacune des deux classes par rapport aux probabilités prédites. Par exemple, l'utilisateur peut dire à la machine que toutes les probabilités prédites au-dessus du seuil = 0.3 appartiennent à la classe 1.

Le modèle peut donc se tromper, il est nécessaire de distinguer 4 issues possibles. Dans le cas où le modèle prédit correctement (0) la valeur déclarée comme négative (0)

il s'agit d'un **vrai négatif (VN)**.

Dans le cas où le modèle prédit correctement (1) la valeur déclarée comme positive (1) il s'agit d'un **vrai positif (VP)**.

Dans le cas où le modèle ne prédit pas correctement la valeur déclarée comme négative (0), (il prédit positivement : 1), il s'agit d'un **faux positif (FP)**.

Dans le cas où le modèle ne prédit pas correctement la valeur déclarée comme positive (1), (il prédit négativement : 0), il s'agit d'un **faux négatif (FN)**.

Voici un exemple illustré pour un seuil de classification (ou de détermination) de 0.5.

CIBLE	SORTIE MODÈLE		
0	0.24		0
1	0.53	SEUIL	1
1	0.36	DE	0
0	0.14	DÉTERMINATION	0
0	0.18	= 0.5	0
1	0.41		0

VN = 3
FP = 0
VP = 1
FN = 3

FIGURE 5.6 – Conséquence du seuil de détermination pour un problème de classification

Le seuil de classification à son importance. Trop faible il augmentera le nombre de faux positifs (et abaissera le nombre de faux négatifs), trop élevé il augmentera le nombre de faux négatifs (et abaissera le nombre de faux positifs).

Matrice de confusion

La matrice de confusion fait une synthèse de ces différents résultats de prédiction pour un problème de classification. C'est un outil de mesure de la performance des modèles de classification. Elle compare les valeurs réelles de la variable à expliquer à celles prédites par le modèle. Voici le cas binaire où la matrice de confusion est un tableau à 4 valeurs.

		CLASSE ESTIMÉE			
		VRAI	FAUX		
CLASSE RÉELLE	VRAI	VRAIS POSITIFS (VP)	FAUX NEGATIFS (FN)		
	FAUX	FAUX POSITIFS (FP)	VRAIS NEGATIFS (VN)		

FIGURE 5.7 – Matrice de confusion

De ces différentes valeurs, des métriques sont construites pour définir la qualité d'un modèle de classification. Il est nécessaire de prendre en compte toutes les métriques dans leur ensemble pour vérifier la qualité d'un modèle. Il n'y a pas de meilleure métrique, cela dépend de l'objectif visé ! Il faut connaître les limites de chaque métrique pour ne pas les utiliser dans des cas inappropriés. En effet certaines peuvent être biaisés à cause de la distribution des classes. En voici quelques exemples :

$$\text{Rappel (Recall)} = \frac{VP}{VP+FN}$$

Rappel = $\frac{50}{50+0} = 1$

		CLASSE ESTIMÉE	
		VRAI	FAUX
CLASSE RÉELLE	VRAI	50	0
	FAUX	50	0

FIGURE 5.8 – Limite de la métrique « Rappel »

La prédiction d'échantillons positifs avec un seuil de confiance élevé pourrait faire ressortir un tel cas. De plus, lorsque les échantillons positifs sont disproportionnellement plus nombreux que les négatifs, les faux positifs seront statistiquement plus rares. Par conséquent, le rappel aura tendance à être élevé.

$$\text{Spécificité} = \frac{VN}{VN+FP}$$

$$\text{Spécificité} = \frac{50}{50+0} = 1$$

		CLASSE ESTIMÉE	
		VRAI	FAUX
CLASSE RÉELLE	VRAI	0	50
	FAUX	0	50

FIGURE 5.9 – Limite de la métrique « Spécificité »

Correspond au taux de vrais négatifs. Contrairement au rappel, la spécificité se concentre sur la classe négative. Par conséquent, nous sommes confrontés à ce problème lorsque les échantillons négatifs sont disproportionnellement plus nombreux.

$$\text{Précision} = \frac{VP}{VP+FP}$$

$$\text{Précision} = \frac{1}{1+0} = 1$$

		CLASSE ESTIMÉE	
		VRAI	FAUX
CLASSE RÉELLE	VRAI	1	49
	FAUX	0	50

FIGURE 5.10 – Limite de la métrique « Précision »

Lorsque les échantillons positifs sont disproportionnellement plus élevés, le classifieur sera généralement biaisé vers les prédictions de classes positives afin de réduire le nombre d'erreurs.

La spécificité et la précision sont utilisées lorsque la classe des vrais négatifs a peu d'importance (cas de données déséquilibrées) c'est la justesse à déterminer les "1" qui primera dans l'évaluation du modèle.

Le rappel et la précision sont des métriques importantes lorsque les données sont déséquilibrées.

$$\text{Accuracy} = \frac{VP+VN}{VP+VN+FP+FN}$$

Puisque l'accuracy attribue un coût égal à tous les types d'erreurs, le fait d'avoir beaucoup plus de labels positifs que de négatifs fera pencher l'accuracy vers les labels positifs (la classe majoritaire).

$$\text{Accuracy} = \frac{99+0}{0+0+99+1} = 0.99$$

		CLASSE ESTIMÉE	
		VRAI	FAUX
CLASSE RÉELLE	VRAI	1	49
	FAUX	0	50

FIGURE 5.11 – Limite de la métrique « Accuracy »

Une grande précision mais un rappel faible signifie que la classe n'a pas été bien détectée cependant si le modèle le détecte il le fera de manière très fiable.

Si l'on préfère explorer de plus grandes zones en diminuant le seuil, on prendra davantage de risques (le classifieur se trompera), le rappel sera plus faible mais la précision plus grande.

Une précision et un rappel élevé signifie que la classe a bien été traitée par le modèle.

La précision et le rappel sont importants lorsque les données sont déséquilibrées. Une courbe Précision - Rappel permet d'illustrer ces deux métriques à travers un graphique avec des valeurs de précision $\frac{VP}{VP+FP}$ sur l'axe des ordonnées et des valeurs de rappel $\frac{VP}{VP+FN}$ sur l'axe des abscisses.

$$\mathbf{F1-Score} = 2 * \frac{\text{Précision} + \text{Rappel}}{\text{Précision} * \text{Rappel}}$$

Le F1-score permet de mesurer la précision et le rappel.

Le F1 score (qui est la moyenne harmonique des deux, est une métrique qui permet l'optimisation jointe de ces 2 critères.

5.3.4 Courbe ROC et AUC

La courbe ROC renseigne les performances d'une classification binaire. Pour un ensemble de seuils donnés : un point correspond à un seuil.

La courbe correspond à la variation du taux de vrais positifs rapport au taux de faux positifs pour un seuil de détermination donné.

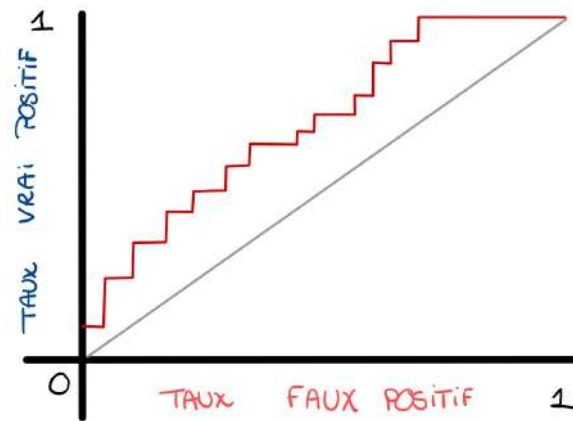


FIGURE 5.12 – Courbe ROC

L'AUC (Area Under Curve) correspond à l'aire sous la courbe ROC. Maximiser l'AUC est une bonne façon d'améliorer un modèle.

A titre d'exemple, l'aire sous la courbe ROC peut être vue comme la probabilité que le modèle ordonne correctement deux individus issus de deux classes distinctes.

La partie suivante présentera les différents modèles de Data Science utilisés dans ce mémoire.

Chapitre 6

Les modèles typiques et optimisation

6.1 Generalized Linear Model (GLM)

Les modèles *GLM* sont une extension du modèle linéaire. Le modèle linéaire est relié à la variable cible par une fonction de lien. Son but est de contraindre les prédictions à être dans une échelle de valeurs précises.

Mathématiquement, le modèle s'écrit ainsi :

$$\mathbb{E}(Y|X) = g^{-1}(X)$$

où $E(Y|X)$ est la valeur attendue de Y (variable réponse) sachant X ; $X\beta$ est le prédicteur linéaire : une combinaison linéaire des variables explicatives et g est la fonction de lien. Également,

$$\mathbb{V}(Y|X) = \mathbb{V}(g^{-1}(X))$$

où \mathbb{V} est une variance qui dépend de la loi de Y .

Ce modèle est donc composé d'un prédicteur linéaire, d'une fonction de lien et d'une distribution supposée de la variable de sortie. Pour illustrer voici le cas d'une régression de poisson.

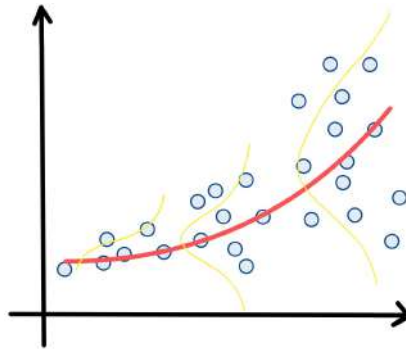


FIGURE 6.1 – Régression de Poisson

Les courbes jaunes illustrent les masses de probabilités de la distribution de Poisson. La courbe rouge représente la prédiction par une régression de Poisson.

Une multitude de fonctions de liens est possible (identité, log ...) associée chacune à une distribution particulière (normale, Poisson ...). Le type de distribution retenue caractérise le GLM.

L'avantage de ce type de modèle est qu'ils sont simples à mettre en place. Le fait qu'ils soient basé sur un prédicteur linéaire les rend plus facile à interpréter. L'avantage du GLM est que pour certaines distributions de probabilité, le paramètre de moyenne est lié au paramètre de variance ce qui n'est pas le cas pour la régression linéaire. Cependant, ils sont sensibles aux effets de corrélations entre les données c'est pourquoi un traitement en amont est nécessaire.

Par exemple, si la fonction de lien est un *logit* le modèle prend la forme de :

$$\mathbf{X}\boldsymbol{\beta} = \ln\left(\frac{\mu}{1-\mu}\right)$$

d'où :

$$E(\mathbf{Y}|\mathbf{X}) = \mu = \frac{\exp(\mathbf{X}\boldsymbol{\beta})}{1+\exp(\mathbf{X}\boldsymbol{\beta})} = \frac{1}{1+\exp(-\mathbf{X}\boldsymbol{\beta})}$$

avec Y qui suit une distribution multinomiale.

Un exemple d'usage de cet exemple pourrait concerner le comptage du nombre d'occurrence de différents types (1, 2, 3 ... K) sur un total de N occurrences.

Comment sont calculés les paramètres de ce modèle ?

Les coefficients sont estimés par la méthode du maximum de vraisemblance.

Cette méthode fournit des estimateurs de ces coefficients noté $\hat{\boldsymbol{\beta}}$, ils sont calculés en maximisant la fonction de log vraisemblance suivante :

$$\hat{\beta} = \arg \max_{\beta} [\ln \mathcal{L}(\beta)] = \arg \max_{\beta} \left[\sum_t \left(y_t \ln \left(\frac{\exp(x_t' \beta)}{1 + \exp(x_t' \beta)} \right) + (1 - y_t) \ln \left(\frac{1}{1 + \exp(x_t' \beta)} \right) \right) \right]$$

6.2 Classification and Regression Trees (CART)

Il existe deux types d'arbre de décision :

- Les arbres de classification (Classification Tree) prédisent à quelle classe la variable-cible appartient.
- Les arbres de régression (Regression Tree) prédisent une valeur numérique.

Un arbre est construit par des branches et des feuilles. Les noeuds des branches correspondent aux variables d'entrée, "la descente dans l'arbre" mènent aux feuilles qui représentent les valeurs de la variable-cible.

L'avantage des arbres est qu'ils sont très *visuels*, voici donc un arbre de qui classe selon l'âge et la taille si l'individu est un enfant, un adolescent ou bien un adulte.

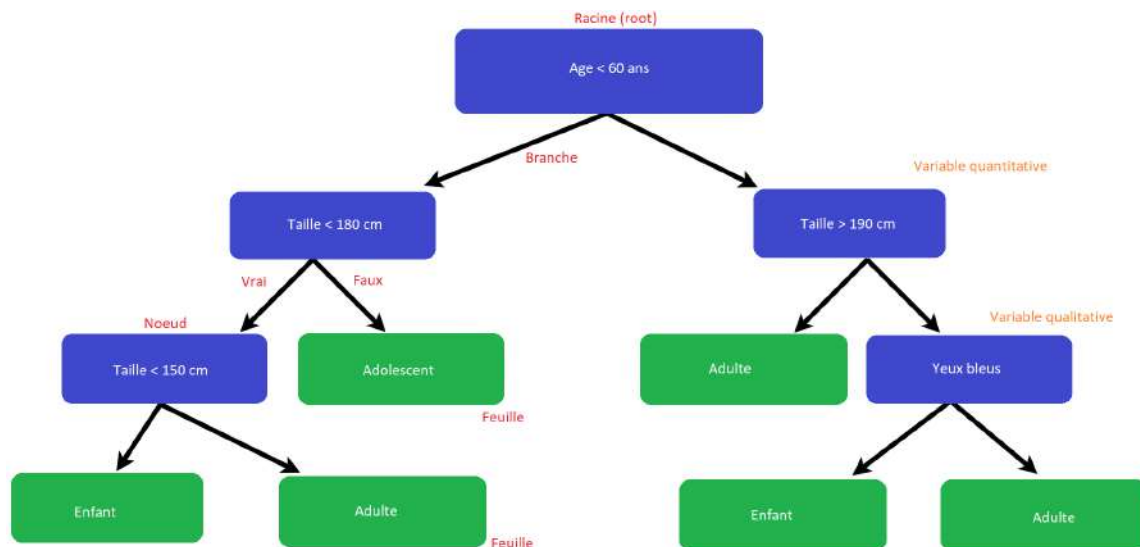


FIGURE 6.2 – Un exemple d'arbre de classification

Par exemple, voici comment l'arbre fonctionne. Si un individu est âgé de moins de 60 ans, la descente dans l'arbre s'effectue à gauche puis si sa taille est inférieure à 1m80, la descente continue à gauche, et si sa taille est supérieure à 1m50 mais moins qu'1m80 alors l'arbre classe cet individu en tant qu'adulte.

Ce type de modèle est très limpide et intuitif : il s'agit de segmenter les données à travers un choix binaire (oui/non) à chaque branche jusqu'à s'arrêter à la feuille de l'arbre. Bien qu'ils soient facilement interprétables (car il est facile de visionner l'arbre

et les choix réalisés), ce sont des modèles simples et donc faiblement performants. Ils ont une variance élevée : un petit changement dans les données peut renvoyer un arbre complètement différent, ils ne sont pas stables.

Les arbres sont généralement construits en deux étapes : d'abord, un très grand arbre précis est construit, appelé *arbre maximal*, puis l'arbre est *élagué*, c'est-à-dire que certains noeuds sont écartés, (*pruning*) pour ne garder que les parties intéressantes et ainsi éviter le surapprentissage.

Pour connaître la séparation idéale à chaque noeud, l'arbre va calculer la somme des résidus au carré (*sum of squared residuals* - RSS) pour chaque configuration possible et choisir celle qui minimise cette somme.

Mathématiquement :

en posant p_i la prédiction et y_i la valeur réellement observée,

$$\text{RSS} = \sum_{i=1}^n (y_i - p_i)^2$$

Pour illustrer cette notion, voici un exemple de calcul de RSS pour une configuration (trait cyan).

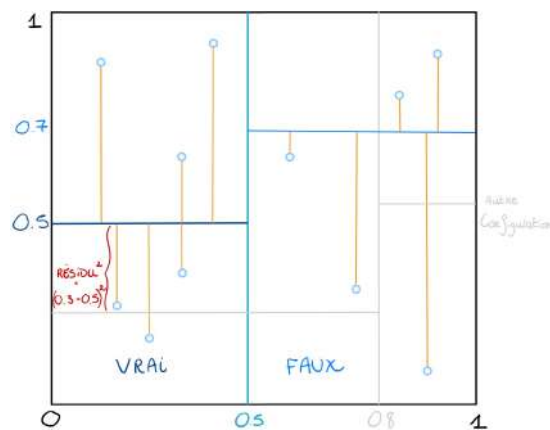


FIGURE 6.3 – Représentation du calcul des résidus

Mathématiquement, il s'agit de calculer une fonction de perte "Loss Function" qu'il convient de minimiser. Une fonction de perte est ici utilisée pour estimer des paramètres (ici la bonne segmentation dans l'arbre). La fonction de perte quadratique est la fonction utilisée pour la technique des moindres carrés.

Pour résumer, voici une illustration de la segmentation d'une base de données transcrite dans un arbre.

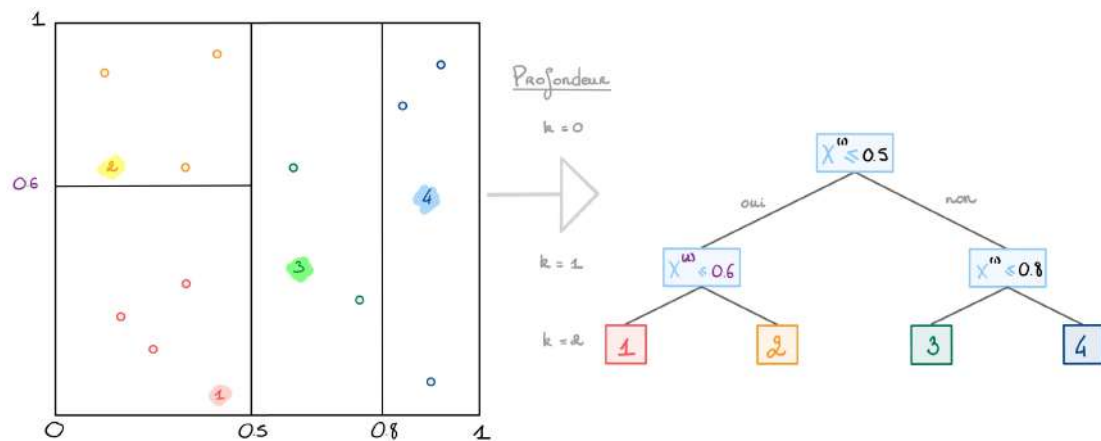


FIGURE 6.4 – Transcription d'une base de données dans un arbre

$X^{(2)}$ représente la variable selon l'axe des ordonnées et $X^{(1)}$ la variable selon l'axe des abscisses.

On voit bien que si $X^{(1)} \leq 0.5$ et que $X^{(2)} \leq 0.6$ cela représente le groupe des données rouges. Dans l'arbre cela correspond à aller une fois à gauche ($X^{(1)} \leq 0.5$ est VRAI) et une nouvelle fois à gauche ($X^{(2)} \leq 0.6$ est VRAI).

Comme évoqué, le problème des arbres de régression et de classification est le manque de stabilité du modèle. Il serait souhaitable de les rendre plus robuste en les agrégant. C'est ainsi que sont nés les modèles ensemblistes. Il existe deux méthodes possibles : le bagging et boosting.

Dans ce mémoire le bagging (Bootstrap Aggregation) n'est pas abordé. Le bagging a pour but de réduire la variance de l'estimateur c'est-à-dire de réduire l'instabilité en moyennant les résultats d'un ensemble d'estimateurs. Les forêts aléatoires utilisent cette technique.

Le boosting est aussi une technique utilisée pour améliorer l'apprentissage d'algorithmes considérés comme «faibles» (weak-learners).

Par itérations successives, la connaissance d'un algorithme faible est ajoutée au «meta» algorithme, l'objectif étant de corriger les estimateurs précédents.

Le boosting a pour but de réduire le biais de l'estimateur c'est à dire l'écart à la réalité. Le principe est différent du bagging : les estimateurs sont construits de manière séquentielle, à chaque itération l'algorithme «apprend de ses erreurs».

Un exemple d'algorithme utilisant cette technique est le Xgboost qui sera présenté par la suite.

6.3 Extreme Gradient Boosting (Xgboost)

Le Xgboost est une technique ensembliste qui est construit à partir d'un arbre de décision. Xgboost signifie : extreme gradient boosting.

C'est un algorithme d'apprentissage supervisé qui combine les résultats d'un ensemble de modèles plus simple (*weak-learners*) (agrégat d'arbre de décisions) permettant de fournir une meilleur prédiction.

A chaque itération d'un nouveau modèle, l'algorithme va utiliser les résultats précédents pour s'améliorer, et "apprendre de ses erreurs" (bien qu'il introduit également des sources d'aléa).

Il agit de manière séquentielle ce qui joue sur la complexité temporelle du modèle.

Voici à présent comment est construit cet algorithme :

1) Création d'une première prédiction initiale (construction d'un premier arbre).

2) Calcul à quel point la prédiction est de bonne qualité avec une fonction de perte. Le Xgboost va créer des arbres en minimisant une fonction de perte et un terme de régularisation, mathématiquement cela s'écrit :

$$\operatorname{argmin}_{\omega} \sum_i^n \mathcal{L}(y_i, p_i) + \gamma T + \frac{1}{2} \lambda \|\omega\|^2$$

où $\sum_i^n \mathcal{L}(y_i, p_i)$, fonction de perte ; $\frac{1}{2} \lambda \omega^2$, terme de régularisation
 γT , T le nombre de feuilles et γ un terme de pénalité qui encourage l'élagage et ω représente le poids de la i-ème feuille

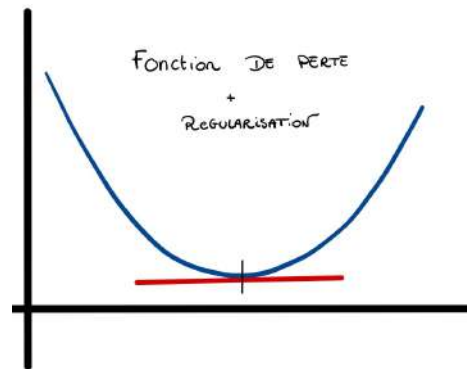


FIGURE 6.5 – La valeur optimale de ω qui minimise l'équation

Nouvelle Prédiction = Ancienne Prédiction + ω (celle qui minimise la fonction décrite précédemment).

L'algorithme Xgboost construit des arbres en trouvant le ω qui minimise cette équation : (mathématiquement, on peut montrer qu'il revient à minimiser cette équation).

$[\sum \mathcal{L}(y_i, p_i^0 + \omega)] + \frac{1}{2}\lambda\omega^2$ une somme d'une fonction de perte et un terme de régularisation. (p_i^j correspond à la prédiction de l'observation i de la j -ème occurrence de l'algorithme).

On résout cette équation afin d'obtenir le meilleur ω .

Le Score de Similarité et le gain servent à déterminer comment découper les données. Ces valeurs sont calculables car la valeur optimale de ω a été calculée précédemment.

Pour résumer on a :

La valeur optimale de ω s'écrit : $\omega = \frac{-\sum g_i}{\sum h_i + \lambda}$

Score de similarité = $\frac{(\sum g_i)^2}{\sum h_i + \lambda} = \frac{\text{Somme des résidus}^2}{\text{Nombres de résidus} + \lambda}$

avec g_i le gradient (première dérivée) et h_i la hessienne (deuxième dérivée) de la fonction de perte.

Gain = Score de Similarité_(feuille gauche) + Score de Similarité_(feuille droite) - Score de Similarité_(racine)

4) On élague l'arbre (*pruning*) en calculant la différence entre le Gain et un paramètre défini : le Gamma (c'est donc un paramètre de complexité de l'arbre).

Si ce différence est négative alors l'algorithme élague.

5) Calcul d'une output Value pour les feuilles restantes. λ est bien un paramètre de régularisation car si c'est positif alors le pruning est plus important par réduction du Similarity score and cela résulte en une output value plus faible pour les feuilles

L'algorithme d'apprentissage XGBoost comporte une méthode de Cross validation possible à chaque itération. Voici les différents hyperparamètres qu'il convient d'optimiser :

nrounds : est le nombre d'arbres total sur lequel l'algorithme va s'entraîner. Un grand nombre d'arbre influence la vitesse d'implémentation de l'algorithme.

max depth : correspond, lui à la profondeur des arbres, un modèle profond conduit à un surapprentissage ;

eta : le taux d'apprentissage, c'est le pourcentage d'apprentissage à chaque étape de l'algorithme. Plus sa valeur est grande, plus l'algorithme convergera rapidement. Sa valeur est comprise entre 0 et 1 ;

gamma : la régularité du modèle. Il correspond à la valeur minimale de la fonction de coût pour choisir d'effectuer une nouvelle partition sur une feuille.

subsample : le pourcentage d'observations prises en compte pour chaque arbre. Sa valeur comprise entre 0 et 1 ;

colsample bytree : le pourcentage de variables prises en compte dans la construction de chaque arbre. Sa valeur comprise entre 0 et 1.

Les algorithmes décrits dans ce chapitre ont chacun leur atout, comme l'illustre ce tableau, bien que le CART soit moins performant que le XGboost, sa force réside dans sa facilité d'interprétation.

Algorithme	Explicabilité de l'algorithme	Vitesse d'apprentissage	Interprétabilité des résultats	Pouvoir prédictif
GLM	★★★	★★★★★	★★★★	★
CART	★★★★	★★★★★	★★★★★	★★
Random Forest	★★	★	★★	★★★
XGBoost	★	★	★★	★★★★

FIGURE 6.6 – Algorithmes, pouvoirs prédictifs et interprétabilité

Mais alors comment faire pour interpréter ces modèles complexes ? C'est ce qui va être abordé dans la partie suivante.

Chapitre 7

Interprétabilité des modèles complexes

Les algorithmes de Machine Learning ont au fur et à mesure gagnés en performance devenant ainsi de plus en plus efficaces mais ce faisant au prix d'une plus grande difficulté de compréhension, d'interprétation.

En effet, les algorithmes les moins robustes comme les arbres sont aussi les plus faciles à interpréter ; visuellement il est aisé de lire un arbre ce qui en fait aussi sa popularité.

Pour résumer,

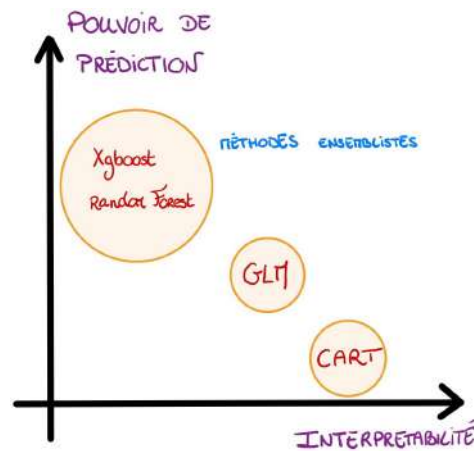


FIGURE 7.1 – Algorithmes et difficulté d'interprétation

Généralement, plus un modèle est précis plus il sera difficile d'interpréter et comprendre les décisions prises par les algorithmes.

Ces algorithmes se veulent de plus en plus complexes, il est pourtant nécessaire de pouvoir expliquer les choix qu'ils font et justifier les résultats qu'ils produisent.

Ces puissants modèles sont souvent cités sous le nom de boîte noire ("black box model") pour souligner la difficulté à expliquer pourquoi ils prennent une décision précise.

Définition

Le but de l'interprétabilité est de pouvoir démystifier l'effet boîte noire des algorithmes complexes, leur apporter de la transparence et permettre d'expliquer comment ils produisent leurs résultats.

Murdoch and Singh [13] définissent l'interprétabilité en donnant un cadre appelé PDR : Precision, Description, Relevance.

Miller [12] propose la définition suivante : «l'interprétabilité comme le degré à partir duquel un humain peut comprendre la cause d'une décision.»

L'interprétabilité s'évalue suivant ces 5 critères : la confiance ; la causalité ; la transférabilité ; l'informativité ; la prise de décision juste et éthique.

En effet, par rapport au 5ème critère évoqué, la réglementation actuelle essaie d'apporter un cadre et une législation autour de ces algorithmes qui prennent de plus en plus de place dans le pouvoir décisionnel des entreprises ; « une personne ne doit pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé et émanant uniquement de la décision d'une machine », article 22 du Règlement Général sur la Protection des Données (RGPD).

Par exemple ce serait le cas d'un individu qui se verrait refuser une opération chirurgicale par la seule décision d'un algorithme. Si aucun chirurgien n'est en mesure d'expliquer la décision de l'algorithme cela constituerait un véritable problème. D'où la nécessité d'avoir des méthodes d'interprétation de modèle.

7.0.1 Les méthodes d'interprétabilité

Différentes méthodes existent en interprétabilité :

INTRINSEQUE VS POST-HOC

- Méthode intrinsèque : c'est l'interprétabilité liée à la simplicité du modèle.
- Méthode Post-hoc : ce sont les méthodes d'interprétabilité pour les modèles complexes. Ici, pour interpréter un modèle il est nécessaire d'avoir des outils annexes.

LOCALE VS GLOBALE

- Méthode locale : interprète pour un faible nombre d'observations.
- Méthode globale : afin d'explicitier toutes les observations en même temps.

A PRIORI VS A POSTERIORI

- Méthode a priori : a lieu avant le lancement de la prédiction (pas d'hypothèse sur les données).
- Méthode a posteriori : s'établit après la création du modèle.

AGNOSTIQUE VS SPECIFIQUE

- Méthode agnostique : s'applique à tous les modèles.
- Méthode spécifique : s'applique qu'à une certaine catégorie d'algorithmes.

Plusieurs algorithmes existent pour interpréter les résultats il s'agit de choisir celui adapté au problème étudié.

Dans le cadre de ce mémoire une méthode post-hoc, locale, a posteriori et agnostique a été choisie. Il s'agit de l'algorithme "*SHAP*" pour SHapley Additive exPlanations qui est basé sur la théorie des jeux.

SHAP est une méthode d'interprétation locale d'un modèle mais il est possible de déduire des tendances globales.

SHapley Additive exPlanations (*SHAP*)

En théorie des jeux, des joueurs se disputent un gain à travers un jeu, l'idée est donc de maximiser ses gains (et dans une certaine mesure surtout minimiser ses pertes).

Dans le cas d'un jeu coopératif, les joueurs collaborent pour obtenir un certain gain. Le problème est alors la répartition du gain entre les différents joueurs. Lloyd Shapley a proposé une répartition « équitable » des gains de la coalition de n joueurs.

En supposant que chaque joueur entre de manière aléatoire dans chaque coalition possible, on peut interpréter la valeur de Shapley comme la valeur de l'accroissement de gain que la coalition peut réaliser avec entrée du joueur. La valeur de Shapley est alors attribuée à ce joueur.

La valeur de Shapley correspond à la valeur marginale moyenne, pour les différentes coalitions, du joueur i .

L'algorithme SHAP établit un parallèle avec le machine learning dans lequel le jeu est le résultat du modèle (la prédiction) et les joueurs sont les variables.

L'algorithme SHAP va calculer la contribution de chaque joueur au jeu, l'importance de chaque joueur au gain total.

Donc, ici la contribution de chaque variable à la prédiction finale faite par le modèle. Certaines variables ont plus d'impact que d'autres sur la décision prise par le modèle, toutes ne jouent pas le même rôle.

L'idée de SHAP est que pour connaître l'implication d'un joueur : il est nécessaire d'étudier chaque coalition (combinaison) de joueurs pour déterminer l'importance de ce joueur. Pour n variables : on a 2^n choix possibles d'ensembles.

Tout d'abord il faut calculer les valeurs de Shapley pour un individu en particulier (simuler différentes combinaisons de valeurs pour les variables d'entrée), puis pour chaque combinaison, calculer la différence entre la valeur prédite et la moyenne des prédictions.

La valeur de Shapley d'une variable correspond alors à la moyenne de la contribution de sa valeur en fonction des différentes combinaisons.

La valeur de Shapley d'une variable (ou plusieurs) :

est sa contribution à la différence entre la valeur prédite par le modèle et la moyenne des prédictions de tous les individus.

Grâce à son fondement mathématique, SHAP pourrait ainsi répondre aux exigences de la réglementation RGPD évoqué ci-dessus (le droit à l'explication).

Pour étudier la prédiction d'un individu, d'une ligne, le Waterfall plot est généralement utilisé pour sa lisibilité :

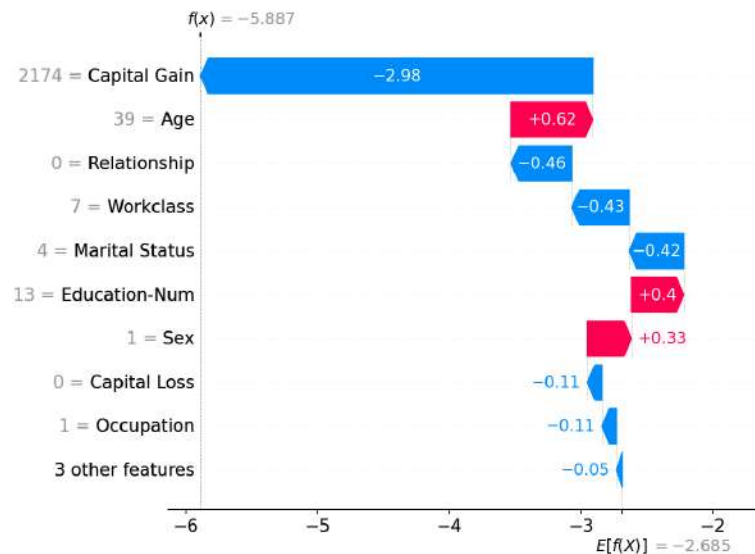


FIGURE 7.2 – Interprétation avec l'algorithme SHAP

Sur ce graphique, on découvre que la variable *Capital Gain* est très importante dans la prédiction, elle contribue fortement (-2.98) à la prédiction finale de -5,887.

L'étude de la distribution des valeurs de shapley peut permettre de déduire **l'influence globale** des variables et donc de passer d'une vision locale à vision globale.

Ainsi, il est possible d'utiliser SHAP dans plusieurs circonstances :

- dans le cas de l'étude d'une prédiction (avec un « *waterfall plot* ») : il s'agit d'une méthode locale.
- dans le cas de l'étude de la distribution des valeurs de Shapley associés à une variable spécifique pour l'ensemble des lignes de la base (avec un « *dependence plot* »)
- dans le cas de l'étude de l'ensemble des valeurs de Shapley pour l'ensemble des variable de la base (importance des variables au sens de Shapley).

La partie suivante présentera les résultats suite à l'utilisation de ces outils mentionnés ci dessus.

Cinquième partie

Présentation des Résultats

Chapitre 8

Etude de critère de rémunération

8.1 Les objectifs du chapitre



FIGURE 8.1 – Plan d'action de ce chapitre sur l'étude de la rémunération

Rappel sur l'index de l'égalité professionnelle

L'index de l'égalité professionnelle présente de manière globale par coefficient hiérarchique et par tranche d'âge, les rémunérations moyennes des salariés.

La note obtenue était de 33 points sur 40 points et un écart global de 6% était constaté par l'index. Cet index ne permet pas d'identifier des individus en particulier, mais permet essentiellement de cibler des CSP et des tranches d'âge.

L'idée ici de venir enrichir le résultat de l'index en détectant de manière plus fine des possibles inégalités dans la base.

Le choix du CART permet de grouper des individus potentiellement discriminés afin de pouvoir lister les salariés et budgétiser un coût de reprise salariale pour les RH.

De plus les modèles de Machine Learning utilisés serviront à améliorer la véracité de l'écart de salaire entre les hommes et les femmes en comparant ce qui est comparable (par métier, par âge, par coefficient hiérarchique de même valeur par exemple).

CSP (coefficient hiérarchique)	tranche d'âge	rémunération annuelle brute moyenne par EQTP		écart de rémunération moyenne	écart après seuil de pertinence et pondération
		femmes	hommes		
1	<i>moins de 30 ans</i>	42 357,97	55 160,92	23,2%	1,18%
	<i>30 à 39 ans</i>	44 296,96	56 975,21	22,3%	0,67%
	<i>40 à 49 ans</i>	47 839,11	59 690,68	19,9%	0,36%
	<i>50 ans et plus</i>	58 282,72	63 909,83	8,8%	0,11%
2	<i>moins de 30 ans</i>	44 985,84	54 931,99	18,1%	0,72%
	<i>30 à 39 ans</i>	47 944,63	60 153,22	20,3%	1,18%
	<i>40 à 49 ans</i>	54 668,92	62 690,29	12,8%	0,52%
	<i>50 ans et plus</i>	64 189,68	68 677,01	6,5%	0,23%
3	<i>moins de 30 ans</i>	47 531,73	57 088,30	16,7%	0,37%
	<i>30 à 39 ans</i>	60 168,34	61 601,77	2,3%	0,02%
	<i>40 à 49 ans</i>	62 939,41	67 082,36	6,2%	0,22%
	<i>50 ans et plus</i>	67 301,06	74 373,64	9,5%	0,40%
4	<i>moins de 30 ans</i>	54 957,19	67 594,75	18,7%	0,09%
	<i>30 à 39 ans</i>	73 248,14	74 298,16	1,4%	0,00%
	<i>40 à 49 ans</i>	75 573,79	75 795,77	0,3%	0,00%
	<i>50 ans et plus</i>	84 460,18	84 475,71	0,0%	0,00%

FIGURE 8.2 – Ecarts de revalorisation lors du calcul de l'index

8.1.1 GLM

Estimation d'écart de salaire par GLM

Paramètres estimés par l'analyse du maximum de vraisemblance								
Paramètre	DDL	Valeur estimée	Erreur	Wald 95% intervalle de confiance		Khi-2 de Wald	Pr > Khi-2	
Intercept	1,00	12,00	0,17	12,12	12,81	5 095,00	<,0001	
Classe_Age	+56ans	1,000	0,076	0,011	0,054	0,097	47,14	<,0001
Classe_Age	-30 ans	1,000	- 0,382	0,013	- 0,408	- 0,356	843,88	<,0001
Classe_Age	[31;35]	1,000	- 0,279	0,013	- 0,304	- 0,255	498,56	<,0001
Classe_Age	[36;40]	1,000	- 0,192	0,011	- 0,215	- 0,170	289,17	<,0001
Classe_Age	[41;45]	1,000	- 0,129	0,010	- 0,149	- 0,109	159,72	<,0001
Classe_Age	[46;50]	1,000	- 0,065	0,010	- 0,084	- 0,046	46,41	<,0001
Classe_Age	[51;55]	-	-	-	-	-	.	.
Coefficient Hiérarchique	1	1,000	- 1,577	0,093	- 1,759	- 1,395	288,94	<,0001
	2	1,000	- 1,546	0,092	- 1,727	- 1,365	279,83	<,0001
	3	1,000	- 1,469	0,092	- 1,649	- 1,288	253,08	<,0001
	4	1,000	- 1,355	0,092	- 1,536	- 1,174	215,62	<,0001
	5	1,000	- 1,221	0,092	- 1,402	- 1,040	175,03	<,0001
	6	1,000	- 0,996	0,092	- 1,177	- 0,815	116,28	<,0001
	7	1,000	- 0,785	0,093	- 0,967	- 0,603	71,67	<,0001
	8	1,000	- 0,627	0,094	- 0,811	- 0,444	44,82	<,0001
	9	1,000	- 0,417	0,093	- 0,599	- 0,235	20,19	<,0001
	10	1,000	- 0,131	0,093	- 0,313	0,052	1,98	0
	11	-	-	-	-	-	.	.
Sexe	Femme	1,000	- 0,031	0,007	- 0,044	- 0,018	21,09	<,0001
Sexe	Homme	-	-	-	-	-	.	.

FIGURE 8.3 – Résultat des coefficients du GLM par modalité

Il est possible d'avoir une approximation de la hausse nécessaire des salaires des femmes pour être équivalent à ceux des hommes.

$$\text{On obtient } \frac{1 - \exp(-0.031)}{\exp(-0.031)} = 3,15\% \quad \exp(\beta_{\text{Sexe}}) = 96.94\%$$

ce pourcentage signifie que le salaire des femmes représente 96.94% du salaire des hommes toutes choses égales par ailleurs.

Importance des variables

Afin de déterminer le modèle le plus adéquat on recherche le modèle qui va minimiser l'AIC. Avec la méthode de sélection stepwise, une variable (prédicteur) est ajoutée (ou supprimée) itérativement. Ainsi une multitude de modèles est créée et le modèle le plus performant peut ainsi en être déduit.

Le critère d'information d'Akaike (AIC) mesure la qualité d'un modèle à travers la vraisemblance en pénalisant le nombre de paramètre. Un modèle à faible AIC est à

privilégier.

$AIC = 2k - 2\ln(L)$; avec "k" le nombre de paramètre du modèle et "L" le maximum de vraisemblance du modèle.

Une variable est rajoutée à chaque étape puis l'AIC est calculée comme dans l'exemple suivant :

Model	R ² ajusté	AIC	Variables associées
1	0,663	78 069	Coefficient Hiérarchique
2	0,694	77 766	Categorie_de_metiers + Coefficient Hiérarchique
3	0,711	59 170	Age + Categorie de métiers + Coefficient Hiérarchique

FIGURE 8.4 – Exemple du calcul de l'AIC

Le modèle GLM travaillant avec des coefficients et la variable qualitative *Age* et *Ancienneté* n'ayant pas des comportements linéaires sur toute la planche de valeurs, il est nécessaire de segmenter par classe ces variables.

Pour trouver la meilleure segmentation de celle-ci un arbre de décision peut être créé sur chacune des deux variables en question. Les feuilles de l'arbre indiqueront une proposition de segmentation de la variable (c'est pourquoi le graphique précédent comportait des classes sur l'âge).

Le calcul de l'AIC permet de déterminer l'importance des variables :

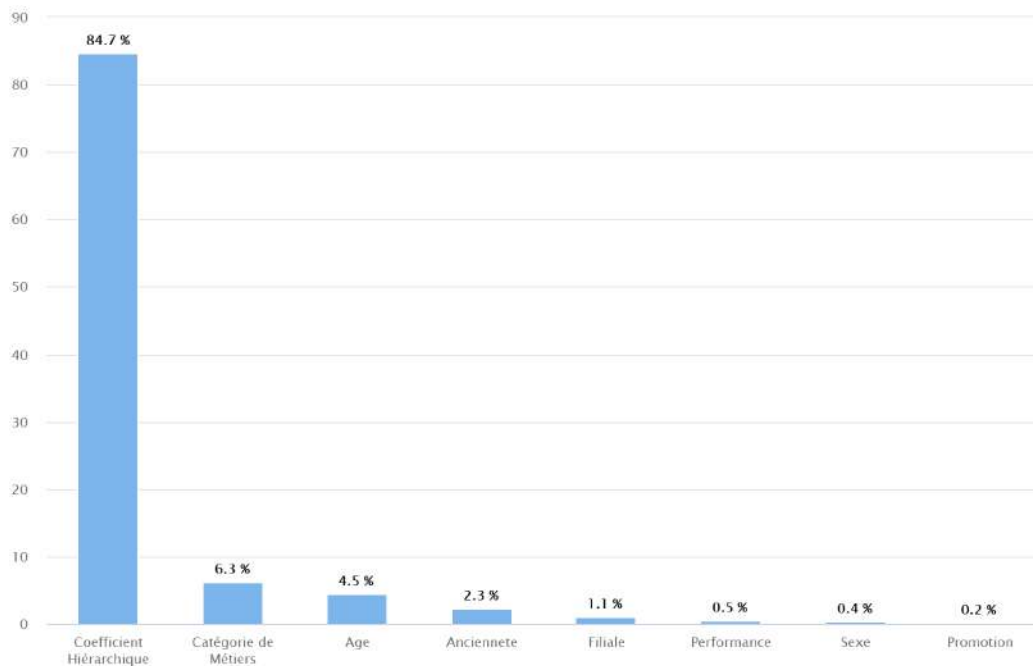


FIGURE 8.5 – Importance des variables GLM par approche stepwise (AIC)

A première vue, le poids du sexe est faible cependant le GLM considère quand bien

même pertinent de garder cette variable, toute chose égale par ailleurs le GLM considère un écart de 3% de rémunérations.

Toujours est-il qu'il peut subsister à une maille plus fine (individuelle) des écarts de salaire à isopérimètre. C'est ce qui va être étudié dans la suite de ce chapitre à travers des arbres de décisions.

8.2 Arbre de régression

Un arbre de régression construit sur la variable Salaire 2021 est d'abord élaboré pour estimer des premières inégalités existantes dans la base. Cet arbre de régression aura pour utilité la classification d'individus discriminés.

Si le sexe apparaît dans une branche de l'arbre, ceci signifie que le sexe importe dans la rémunération obtenue par le collaborateur.

Hyperparamètre	Description
<i>cp</i>	0,0001
<i>minsplit</i>	20
<i>minbucket</i>	7
<i>maxdepth</i>	10
<i>méthode</i>	anova

FIGURE 8.6 – Liste des hyperparamètres choisis

On calibre ainsi l'arbre de décision à travers l'élaboration d'une grille de recherche des hyperparamètres (*grid search*). L'arbre sera d'autant détaillé que le paramètre de complexité *cp* est grand. *Minbucket* correspond au nombre minimum d'observations dans chaque feuille. *Minsplit* correspond au nombre d'observations qui doivent exister dans un noeud pour qu'une division en deux branches distinctes ait lieu.

La méthode anova (qui n'est pas en soit un hyperparamètre) a pour critère de division : $RSS_{\text{noeud}} - (RSS_{\text{branche gauche}} + RSS_{\text{branche droite}})$; pour rappel, $RSS = \sum_{i=1}^n (y_i - \bar{y})^2$.

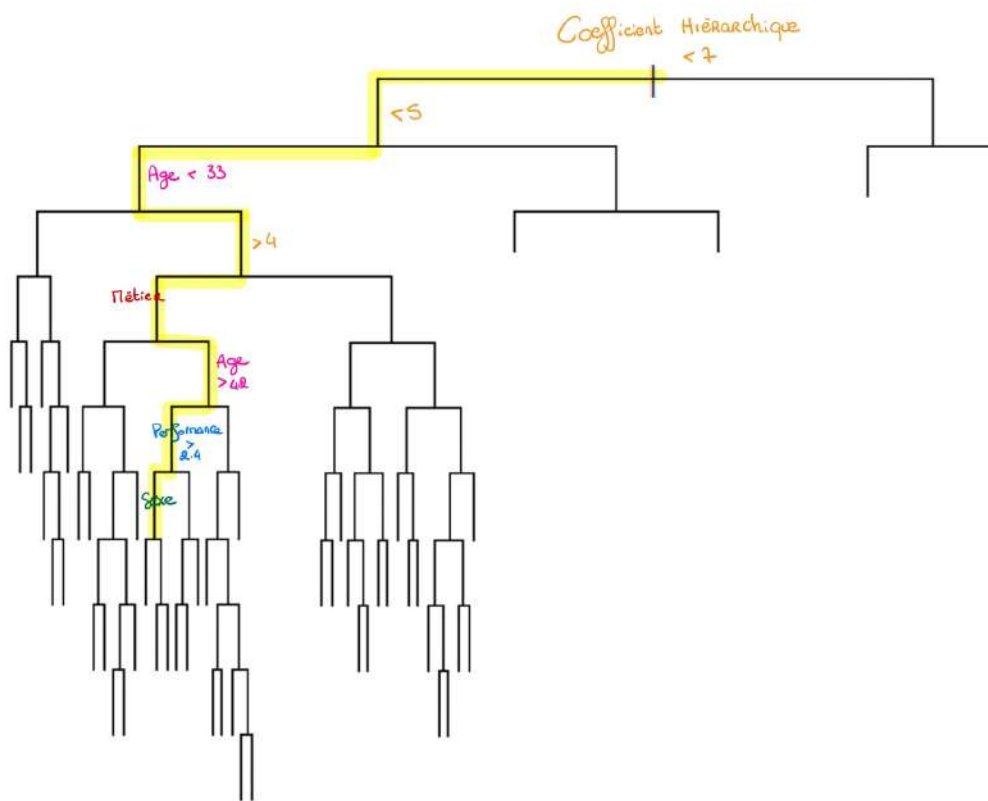


FIGURE 8.7 – Recherche des individus potentiellement discriminés par la présence de la variable « Sexe »

Un noeud ayant pour variable le sexe apparaît dans l'arbre ; le sexe apparaît donc significatif pour certains individus quant à l'obtention d'un salaire plus élevé.

On remarque que l'algorithme détermine des différences de salaire bien plus importante que l'écart global des 3% du GLM et des 6% de l'index, en effet en s'intéressant à une maille plus fine on peut observer des différences de salaires plus importantes qui pourraient être cachés en moyenne dans un calcul global d'écart de salaire. L'algorithme distingue trois groupes de femmes discriminées dans la base de données :



FIGURE 8.8 – Récapitulatif des groupes présentant des potentielles discriminations

Premier groupe

14 femmes sont concernées et touchent 16.2% de moins.

- Le coefficient hiérarchique est = 4
- L'Age est compris entre $\in [32.5, 42.5]$
- La Performance Globale est ≥ 2.46
- Les individus travaillent dans les secteurs suivants : Data Scientist, Développeur, Souscripteur, Technicien, RH

Deuxième groupe

18 femmes sont concernées et touchent 9.2% de moins.

- Le coefficient hiérarchique est = 4
- L'Age est compris entre $\in [42.5, 59.5]$
- La Performance Globale est ≤ 2.38
- Les individus travaillent dans les secteurs suivants : Data Scientist, Développeur, Souscripteur, Technicien, RH

Troisième groupe

9 femmes sont concernées et touchent 15.2% de moins.

- Le coefficient hiérarchique est = 5
- L'Age est compris entre $\in [36.5, 46.5]$
- Les individus travaillent dans les secteurs suivants : IT, RH, Autres

8.3 Retour sur la budgétisation RH

La force de cette méthodologie est de pouvoir passer d'une vision macroscopique à une maille individuelle plus fine. Cependant, l'actuaire délivre une liste d'individus potentiellement discriminées selon le modèle mais l'algorithme n'est là qu'en support. C'est

un outil d'aide à la décision qui doit venir appuyer les RH dans leurs analyses. En effet eux, seuls connaissent vraiment la base de données et peuvent avoir le dernier mot sur si oui ou non l'individu est anormalement sous payé.

Ici, on dépose une idée de budget de rattrapage salarial selon les individus. Ce rattrapage pourra avoir lieu en une fois (augmentation fixe) ou bien étalée sur plusieurs années (augmentation de 2% annuelle supplémentaire) pour éviter que l'impact du passif soit trop important et inattendu (d'une année à l'autre) entre deux bilans.

En faisant un calcul grossier :

- il faudrait rattraper les salaires de 16,2% de 14 femmes. Ce qui correspond à un montant de 10 045 € par tête. Soit un total de **140 630 €** pour ce groupe d'individus. Le code R est :

- Sexe=Homme 14 8.833904e+08 62058.71 *
- Sexe=Femme 16 2.024145e+09 72103.19 *

- il faudrait rattraper les salaires de 9,2% de 18 femmes. Ce qui correspond à un montant de 5 783 € par tête. Soit un total de **104 094 €** pour ce groupe d'individus. Le code R est :

- Sexe=Femme 18 2.188633e+09 63080.22 *
- Sexe=Homme 30 6.092757e+09 68893.47

- il faudrait rattraper les salaires de 15,2% de 9 femmes. Ce qui correspond à un montant de 11 139 € par tête. Soit un total de **100 251 €** pour ce groupe d'individus. Le code R est :

- Sexe=Femme 9 5.907428e+08 73231.67 *
- Sexe=Homme 16 4.417061e+09 84370.00 *

Dépendamment de la taille de l'entreprise cet impact peut être très important. Ici, la masse totale de rattrapage serait de : $140\,630 + 104\,094 + 100\,251 = \mathbf{344\,975\ €}$.

Cependant la volatilité du budget peut être très importante due à la volatilité de l'arbre ce qui pose une première limite à cette méthode. En effet, en considérant un arbre un peu plus complexe. Ici, le surapprentissage de l'arbre n'est pas un problème pour l'étude, au contraire il est plus intéressant d'aller le plus loin possible pour segmenter le plus possible et ainsi de pouvoir déterminer des très petits groupes d'individus discriminés. D'autant plus que, plus la segmentation est fine, plus on est assuré que les individus soient très similaires (même profil de carrière).

L'algorithme découvre un nouveau groupe d'individus discriminés :

Quatrième groupe

18 femmes sont concernées et touchent 9.2% de moins.

- Le coefficient hiérarchique est = 5
- L'Age est compris entre $\in [49, 57]$
- La Performance Globale est ≤ 2.38
- Les individus travaillent dans les secteurs suivants : Autres, Chargé d'affaires, Vendeurs, IT, Data Science
- Sexe=Femme 10 7.936958e+08 76823.30 *
- Sexe=Homme 27 3.445567e+09 83646.33

- il faudrait rattraper les salaires de 8.9% de 10 femmes. Ce qui correspond à un montant de 6 823 € par tête. Soit un total de **68 230 €** pour ce groupe d'individus.

Ce qui ferait passer le budget de 344 975 € à 413 205 € soit une augmentation de quasi 20%. C'est pourquoi l'analyse RH est primordiale puisque seule la connaissance de l'entreprise permettra de conclure. Il convient cependant de soumettre une liste d'individus issus d'un arbre suffisamment complexe pour rendre compte de tous les collaborateurs potentiellement discriminés. Après reprise salariale sur ces individus, l'arbre identifie correctement aucun écart de salaire.

8.4 Limites

- Une des principales critiques qu'il est possible de faire à cette approche est la comparabilité des individus. Les profils dont l'arbre groupe sont-ils suffisamment proches et comparable ?

L'algorithme identifie des groupes d'individus de même sexe discriminés par rapport à leur homologue de l'autre sexe (par rapport à leur performance, métiers ...). Si l'algorithme groupe des profils de métiers ensemble, est-ce acceptable dans l'état pour prétendre à une différence de salaire significative entre les deux groupes ? Quelle est la répartition par métier dans ce sous-groupe ?

C'est pourquoi il est important que les groupes soient assez petits.

Ici, les résultats sont les suivants à coefficient hiérarchique égale à 4 (où DS signifie "Data Science") :

Sexe	Age	Métier	Ancienneté	Promotion	Performance	Salaire
Homme	38	RH	3	Non promu	2.5	78 444
Femme	38	RH	5	Non promu	2.5	66 510
Homme	35	RH	9	Non promu	3	66 897
Femme	35	RH	10	Non promu	4	48 374
Homme	39	DS	3	Non promu	3	89 663
Femme	41	DS	4	Non promu	3	56 781
Homme	36	DS	3	Non promu	4	72 990
Femme	36	DS	3	Non promu	4	64 300

On remarque que les individus ne sont pas dans l'absolu identique cependant le modèle les regroupe ensemble.

- Une deuxième limite que peut poser les arbres surapprentis concerne la prédiction, comme ici il s'agit d'un problème de classification, le surapprentissage n'est pas dangereux. Cependant, c'est la volatilité des arbres qui va être problématique pour l'explication des groupes identifiés d'une année à l'autre notamment avec les entrées et sorties de collaborateurs. La volatilité des arbres pose problème pour expliquer les évolutions des groupes d'une année à l'autre.

8.5 Conclusion

Cette étude a permis de venir affiner les résultats de l'index de l'égalité professionnelle qui ne permettait pas de déterminer une liste d'individus discriminés et restait trop vague sur des potentielles inégalités. L'actuaire est acteur de l'identification et de la correction d'un risque social.

Cependant il est nécessaire de faire ce travail d'échange avec les RH en amont de la modélisation du modèle pour que les choix lors de la création de celui-ci soient les plus pertinents et adaptés à la base de données.

De plus, l'avantage est que l'actuaire vient soumettre des points d'attention et de réflexion à mener, pour faciliter les RH dans leurs missions à travers une budgétisation et un *listing* d'individus potentiellement discriminés.

Cet échange avec les RH est essentiel puisqu'il vient compléter par une approche "terrain" (une approche qualitative) l'analyse des différents groupes potentiellement dévalorisés par l'algorithme.

Dans le prochain chapitre, le critère de promotion sera abordé, il s'agira de déterminer notamment des individus comme potentiellement discriminés en terme de promotion.

Chapitre 9

Etude du critère sur la Promotion

Voici les étapes qui vont être mené dans le cadre de ce chapitre; le but sera de déterminer des individus potentiellement discriminés en terme de promotion.



FIGURE 9.1 – Plan d'action de ce chapitre sur l'étude de la promotion

9.1 Arbre de classification

Afin d'appréhender l'étude de la promotion au sein de la base de données, une première approche en utilisant l'algorithme CART est effectuée. La motivation d'utiliser l'arbre de classification est de révéler en un premier coup d'œil si le sexe aura tendance à jouer dans l'attribution de promotion.

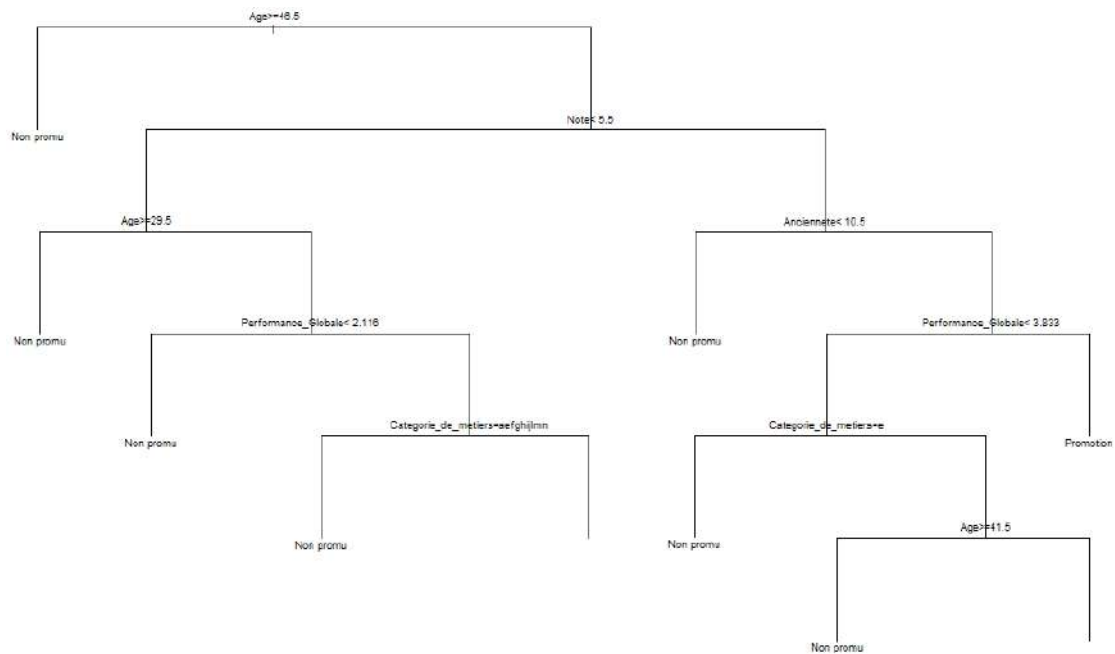


FIGURE 9.2 – Arbre de Classification sur la variable Promotion

Le variable « Sexe fg n'apparaît dans aucune segmentation de l'arbre ce qui laisse à penser que le sexe n'est pas significatif à l'obtention d'une promotion.

Ce modèle étant peu robuste il convient d'étayer cette hypothèse avec d'autres modèles. Dans un second temps un modèle logistique est établi pour étudier l'importance du « Sexe » à la promotion.

9.2 GLM

Un traitement des variables quantitatives est nécessaire afin d'avoir plusieurs coefficients pour chaque modalité des variables Age et Ancienneté étant donné que leur influence n'est pas linéaire.

On effectue un modèle GLM sur la variable « Promotion ». C'est une classification binaire ('Promotion' == 1 et 'Non promu' == 0).

La fonction de lien la plus adéquate est le logit :

$$\log\left(\frac{p}{1-p}\right) = \sum_{j=1}^n \beta_j X_{ij}$$

avec les coefficients β_j des log odds ratio où le « odd » est le rapport de deux probabilités complémentaires l'une de l'autre.

On va sélectionner les variables importantes à travers une méthode stepwise, en ajoutant (cas « forward » de la méthode) au fur et à mesure les variables au modèle. Le modèle va continuer d'ajouter des variables selon un critère de précision, et s'arrêter si le fait d'ajouter une variable supplémentaire n'apporte plus d'information utile. Comme le Critère d'Information d'Akaike pénalise les modèles avec le nombre grandissant de variables et que l'AIC augmente en ajoutant la variable *Ville* il convient de supprimer cette variable.

La formule du modèle est le suivant :
 « Promotion \sim Coefficient hiérarchique + Age + Ancienneté + Catégorie de métiers + Filiale + Performance »

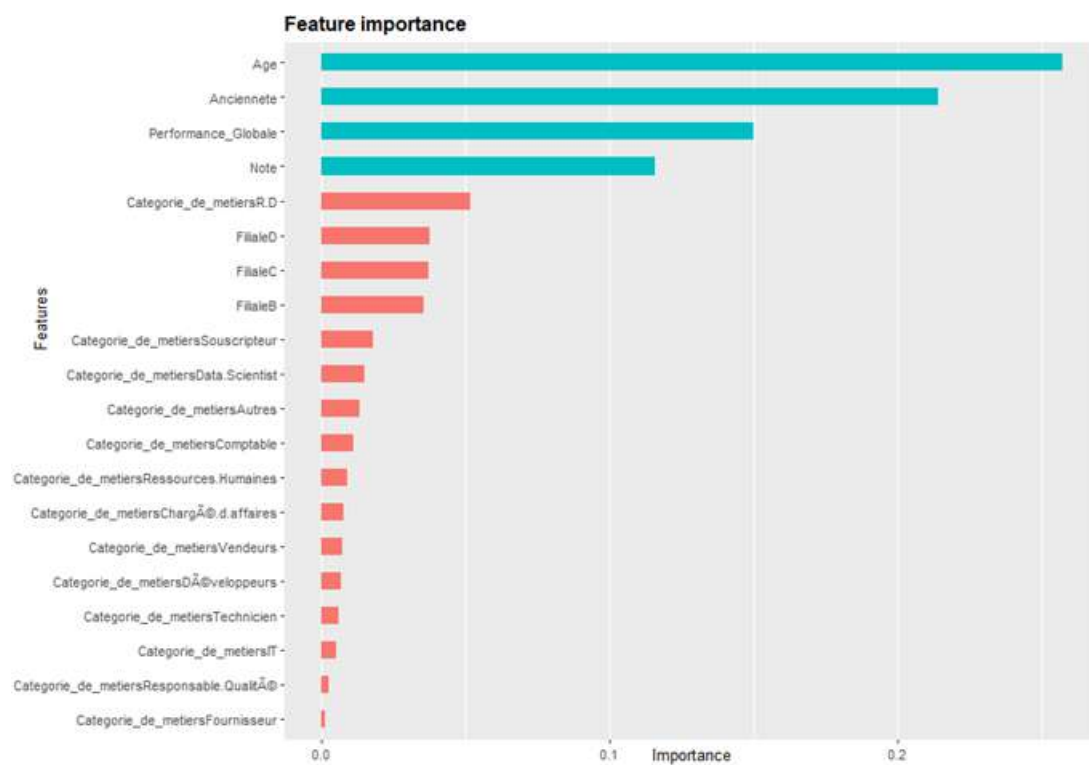


FIGURE 9.3 – Importance des variables du modèle GLM

Les résultats sont similaires à ce qui a été vu précédemment, les variables les plus importantes sont le coefficient hiérarchique (noté « Note » pour abrégé), l'âge et la performance qui prime à expliquer la promotion.

Les modèles utilisés jusqu'à présent ne permettent pas d'étudier une maille plus fine. C'est pourquoi, la partie suivante abordera l'utilisation de modèles plus complexes.

9.3 Xgboost - modèle de classification sur la Promotion

9.3.1 Motivations du modèle

L'idée de cette partie est de pouvoir identifier des individus discriminés dans l'obtention d'une promotion corrigée du sexe (pour supprimer un potentiel biais lié au sexe). Le modèle va prédire une probabilité de promotion pour chaque individu basé sur les variables explicatives de la base de données.

La nécessité de comprendre ce que prédit ce type de modèle « boîte noire » est essentielle; d'autant plus que la portée des conclusions s'adressent à des RH dans un but de prévention. Le but n'est pas dire si une inégalité est présente ou non mais de venir directement étudier des cas à l'échelle individuelle pour fournir une liste d'individus comme potentiellement discriminés au niveau du sexe dans le cadre d'une promotion.

L'idée est de contraster la sortie du modèle avec la promotion réelle. Autrement dit il s'agit d'étudier pourquoi le modèle explique qu'un individu devrait recevoir une promotion tandis que dans la réalité la promotion n'a pas été donnée. On cherche donc à étudier les faux positifs du modèle.

9.3.2 Paramétrage du modèle

Comme évoqué précédemment, il est nécessaire de créer un modèle qui ne sur-apprend pas mais qui soit suffisamment robuste pour que les prédictions soient généralisables à tout type de base.

Etant donné que le modèle va être entraîné sur une partie de la base de données qui sera par la suite testée sur toute la base, il est nécessaire de savoir si les prédictions sont bonnes et consistantes sur de nouvelles données.

En séparant la base de données en deux bases : une base d'apprentissage qui représente 65% de la base totale et une base de test qui représente 35% de la base totale.

Un modèle est construit puis sera utilisé pour faire des prédictions sur toute la base de données pour identifier les promotions potentielles. Contrairement à de l'actuariat classique, comme nous disposons seulement d'une base de salariée, un nouveau modèle sera à refaire l'année prochaine avec les entrées et sorties des salariés.

Ces hyperparamètres ont été ajustés à travers une validation croisée et une grille de recherche.

Les hyperparamètres et variables choisis pour ce modèle sont les suivants :

Variables	Type	Hyperparamètres	Valeurs
Age	Quantitative	eta	0,50
Ancienneté	Quantitative	max_depth	12
Coefficient Hiérarchique	Quantitative	min_child_weight	1
Filiale	Qualitative	subsample	1
Catégories de métiers	Qualitative	colsample_bytree	1
Performance	Quantitative	gamma	0

FIGURE 9.4 – Paramétrage du modèle

ainsi la ville d'origine de l'emploi a été exclue du cadre puisqu'elle n'apportait pas d'information complémentaire utile.

Le Xgboost prédit des probabilités de promotion. Il convient d'établir un seuil à partir duquel on considère qu'un individu est promu.

9.3.3 Résultats

Le seuil de pertinence est déterminé pour qu'il maximise les métriques du modèle.

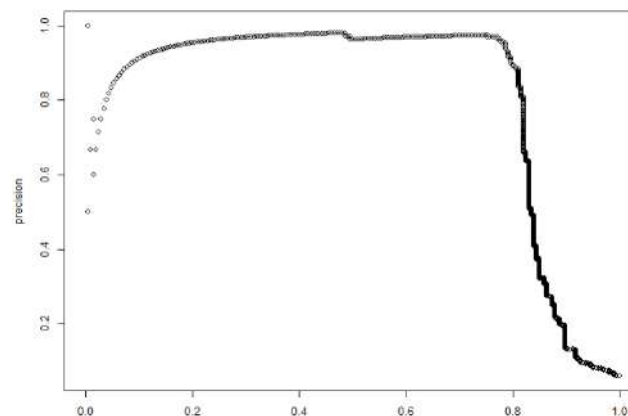


FIGURE 9.5 – Courbe Précision Recall

Sur ce graphique on peut voir que le seuil idéal se situerait à partir de 0.3 points. Il est souhaitable que l'algorithme ait à la fois une précision élevée et un rappel élevé. Cependant, la plupart des algorithmes d'apprentissage automatique impliquent souvent un compromis entre les deux. En utilisant une grille de recherche qui maximise ces métriques la valeur obtenue est la suivante ; 0.36. On fixe le seuil à 0.36 et on regarde les résultats de la matrice de confusion ;

Les résultats de prédiction du modèle sont résumés dans une matrice de confusion :

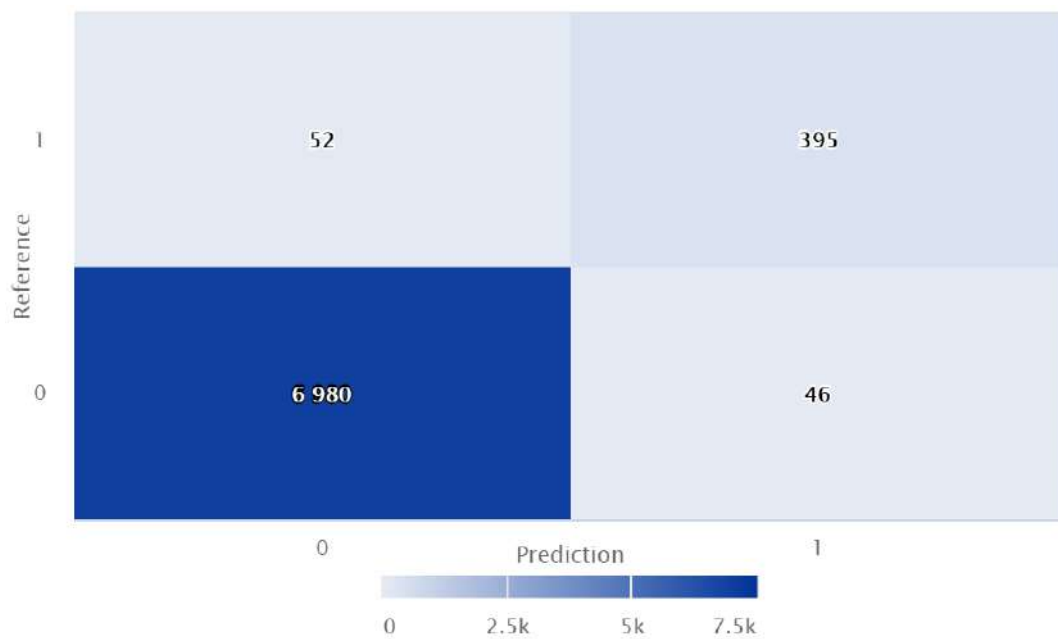


FIGURE 9.6 – Matrice de Confusion

Le nombre de faux positifs est de : 46. Ce sont ces individus qui seront étudiés par la suite.

Le nombre de vrais positifs est de : 395

Le nombre de vrais négatifs est de : 6980

La courbe ROC (pour “Receiver Operating Characteristic”) est une courbe qui représente le comportement du classifieur pour tous les seuils de détection possibles. Un bon modèle est un modèle qui maximise l’AUC (tendre vers 1).

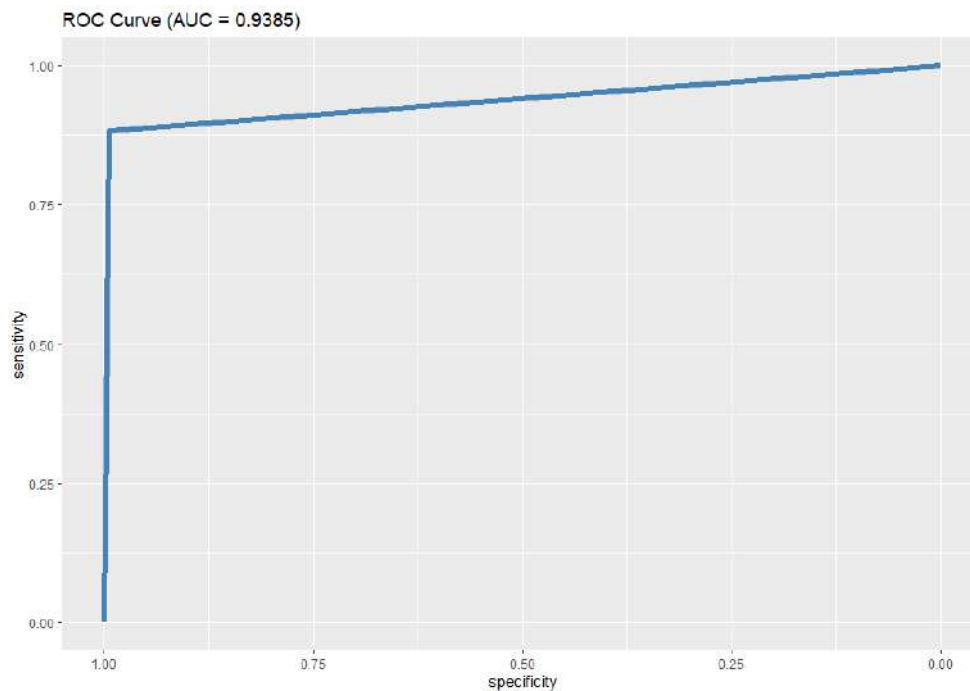


FIGURE 9.7 – Courbe ROC

Le modèle semble plutôt performant cependant cette courbe n'est pas adaptée dans le cas de données déséquilibrées.

Les données présentes ici étant mal balancées dans la classe 1. Autrement dit, beaucoup sont dans la classe des "Non promu". On souhaite minimiser le nombre de faux positifs en choisissant un bon seuil, en effet, le but est de choisir ceux pour qui la probabilité de promotion est quasi certaine.

Comme les données sont mal balancées avec des vrais négatifs en sur nombre il convient d'utiliser et maximiser les métriques adaptées à ce problème. Les métriques de précision et rappel car ils ne prennent pas en compte le nombre de vrais négatifs, sont plus informatifs que les diagrammes ROC qui inclue le nombre de vrais négatifs. C'est pourquoi les métriques de précision et de rappel seront préférés dans cette situation.

		Réal	
		0	1
Prédit	0	TN	FN
	1	FP	TP

FIGURE 9.8 – Rappel Matrice de confusion

Le rappel $\frac{TP}{TP+FN}$; signifie que parmi tous ceux prédit positivement quelle est la proportion de vraiment positifs.

Et la précision $\frac{TP}{TP+FP}$; correspond au taux de vrai positif.

Un compromis est obligatoire car on ne peut maximiser les deux en même temps.

On cherche donc à optimiser cette métrique le plus possible pour que le modèle soit performant.

Précision	Recall
89,7%	88,3%

FIGURE 9.9 – Métriques importantes

Les métriques de précision et de rappel donnent de bons résultats indiquant un modèle plutôt performant.

Les autres métriques que donnent le modèle sont les suivantes : Le FPR correspond au taux de faux positifs, le FNR correspond au taux de faux négatif, le TPR correspond au taux de vrai positif et le TNR correspond au taux de vrai négatif.

TPR (Sensibilité)	FNR	TNR (Spécificité)	FPR	BALANCED ACCURACY
88,3%	11,7%	99,4%	0,6%	93,9%

FIGURE 9.10 – Métriques

Même si les données sont déséquilibrées, nous pouvons déterminer si notre modèle fonctionne bien ou pas. Pour cela, les valeurs de TPR et TNR doivent être élevées, et FPR et FNR doivent être aussi basses que possible.

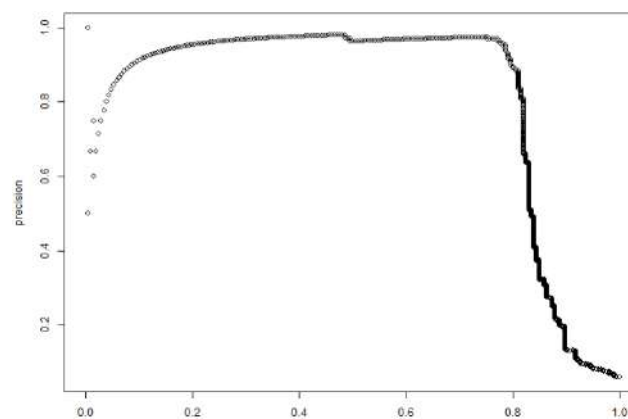


FIGURE 9.11 – Courbe Précision Recall

On préférera ainsi ces métriques lors de l'évaluation de classifieurs binaires sur des données déséquilibrées. Les diagrammes ROC peuvent être visuellement trompeurs en ce qui concerne les conclusions sur la fiabilité des performances de classification.

Il existe d'autres métriques qui ne corrigent les problèmes liés à la distribution de la matrice de confusion. Par exemple, la précision équilibrée (bACC) normalise les prédictions positives et négatives vraies par le nombre d'échantillons positifs et négatifs, respectivement, et divise leur somme par deux.

9.4 Interprétabilité du modèle

Bien que le modèle distingue plusieurs dizaines d'individus potentiellement discriminés toutes proportions gardées, il a été choisi de se fixer un seuil d'une vingtaine de personnes qui ont la plus grande probabilité de promotion.

Sur ces individus, il est nécessaire de comprendre pourquoi le modèle indique qu'ils devraient recevoir une promotion bien qu'en réalité ça n'a pas été le cas.

L'avantage de l'interprétabilité est de pouvoir comprendre simplement en un coup d'oeil pourquoi l'algorithme considère les individus comme à promouvoir.

9.4.1 Interprétabilité Globale

Dans un premier temps il est nécessaire de comprendre globalement comment fonctionne le modèle.

Le graphique en essaim d'abeilles est conçu pour afficher un résumé de la manière dont les principales variables d'une base de données influencent la prédiction du modèle.

Chaque ligne de la base de données est représentée par un seul point pour chaque variable.

La position du point selon l'axe des abscisses est déterminée par la valeur de Shapley de cette variable.

Les points s'empilent le long de chaque variable pour montrer la densité.

La couleur est utilisée pour afficher la valeur originale de la variable.

Par défaut, les variables sont ordonnées, par la valeur absolue moyenne des valeurs Shapley pour chaque variable. Cet ordre met davantage l'accent sur l'impact moyen général, et moins sur les impacts rares mais de grande ampleur.

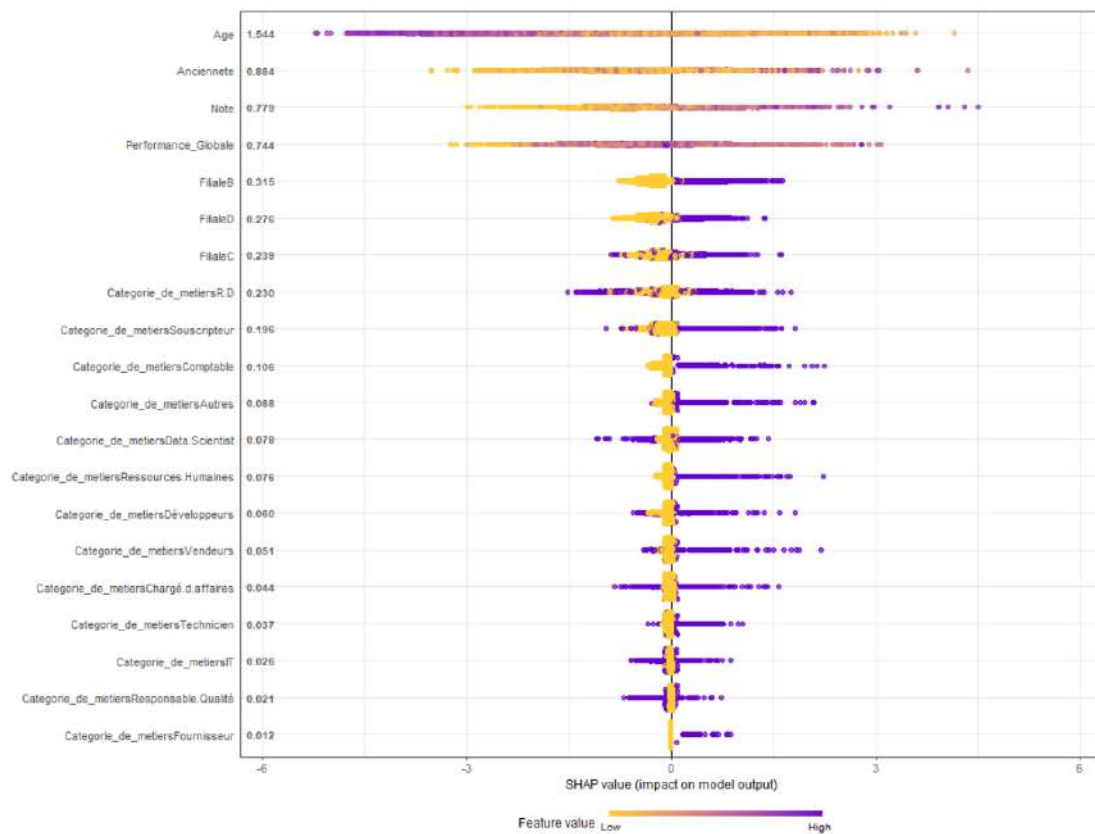


FIGURE 9.12 – « Summary Plot » ; « Beeswarm Plot »

Ce graphique illustre l'importance de l'âge, du coefficient hiérarchique et de la performance dans l'obtention d'une promotion.

Par exemple, les faibles valeurs de l'âge (en jaune) ont une plus grande valeur de Shapley et ont une grande importance sur la prédiction d'une promotion par le modèle. A contrario, de faibles valeurs (en jaune) du coefficient hiérarchique (abrégé en 'Note') impliquent de moins grande chance de promotion par le modèle. Cette notion de couleur peut également être plus facile à lire à travers un les diagrammes de dépendance.

Un autre type de graphique utile en matière d'interprétabilité est le **diagramme de dépendance** de SHAP qui est un diagramme de dispersion qui montre l'effet d'une seule variable sur les prédictions faites par le modèle.

Chaque point est une prédiction unique de la base de données (une ligne).

L'axe des abscisses est la valeur de la variable.

L'axe des ordonnées est la valeur Shapley pour cette variable, qui représente la mesure dans laquelle la connaissance de la valeur de cette variable modifie la sortie du modèle pour la prédiction de cet échantillon.

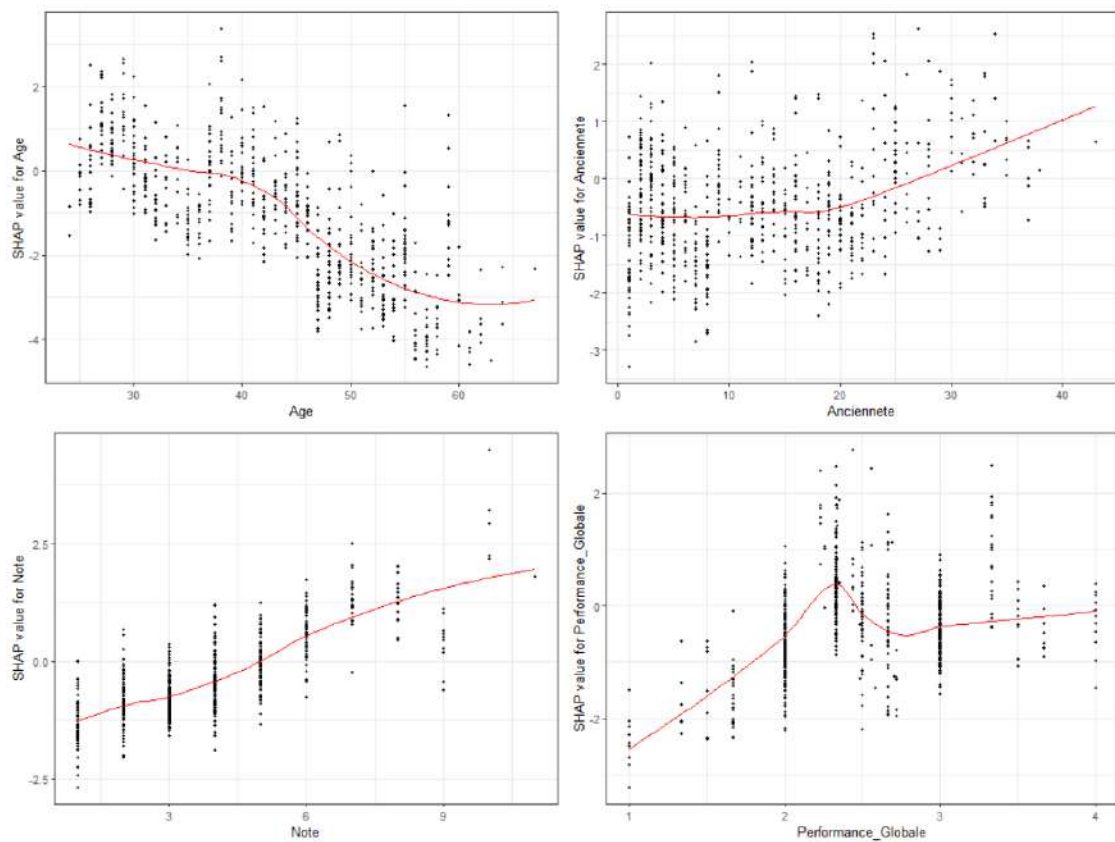


FIGURE 9.13 – « Dependence Plot »

Par exemple, le diagramme de performance indique qu'à faible performance la valeur de Shapley est d'autant plus faible et contribue moins à la promotion d'un individu. En somme un individu très performant aura d'autant plus de chance d'être promu. Le coefficient hiérarchique égal à 6 contribue davantage à obtenir une promotion que celui égal à 3.

En particulier, il serait intéressant de connaître les raisons pour lesquelles le modèle prédit des promotions aux faux positifs. Trois individus avec de fortes probabilités de promotions sont sélectionnés.

9.4.2 Interprétabilité locale

Le trait fin bleu vertical correspond à la prédiction moyenne. Comme ici le modèle prédit des promotions et que les données sont déséquilibrées la prédiction moyenne est très faible mais n'est pas égale à 0.

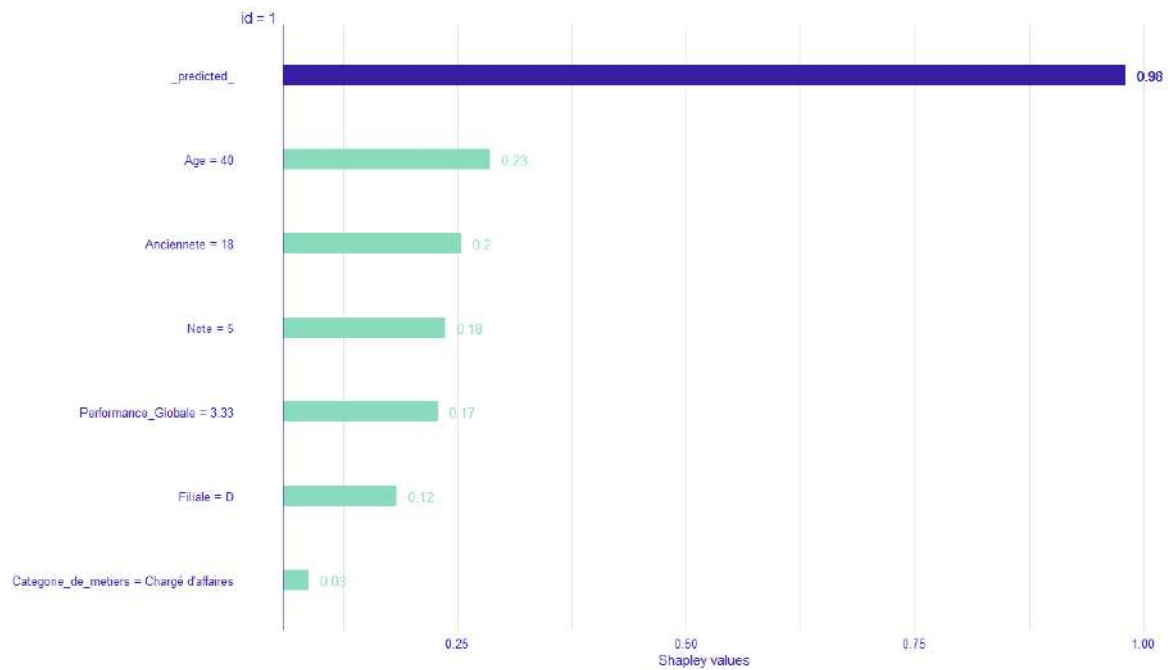


FIGURE 9.14 – Valeur de Shapley du premier individu

Le modèle donne une prédiction de promotion de 0.98, cette probabilité de prédiction de promotion est très élevée. Pourtant dans les faits, cet individu n'a pas été promu. Il serait quand bien même intéressant de connaître les raisons selon lesquelles cet individu a été promu par le modèle et ainsi lorsque cet individu sera remonté aux RH, il sera possible de se justifier.

Dans la configuration de cet individu, son ancienneté et son âge (il est donc rentré à l'âge le plus bas possible de 22 ans) contribue fortement à sa promotion selon le modèle. Si l'on se réfère aux diagrammes de dépendance cela coïncide bien avec ce qui a été observé.

A noter que le métier ici n'explique quasi en rien la potentielle promotion de cet individu ce qui est accord avec ce qui a été souligné dans le diagramme en essaim d'abeille : les métiers impactent faiblement la promotion d'un individu.

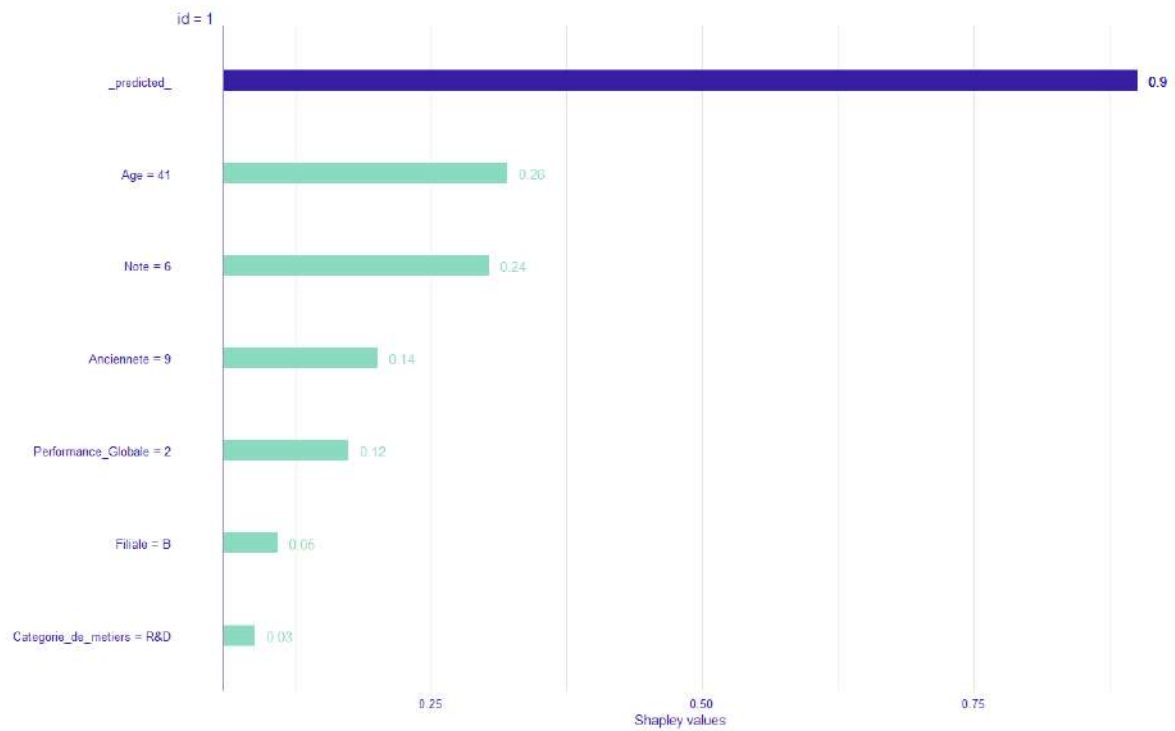


FIGURE 9.15 – Valeur de Shapley du deuxième individu

Ici, on a un individu avec un coefficient hiérarchique plus élevé, avec moins d'ancienneté et de performance. Sa performance contribue donc moins à la prédiction d'une promotion par le modèle cependant son coefficient hiérarchique supérieur et son âge contribuent grandement à sa promotion. Son manque d'ancienneté semble impacter davantage la prédiction du modèle, et contribue ainsi moins à la prédiction d'une promotion. La filiale dans laquelle travaille cet individu ne semble pas expliquer une potentielle promotion à l'opposé de l'individu étudié précédemment.

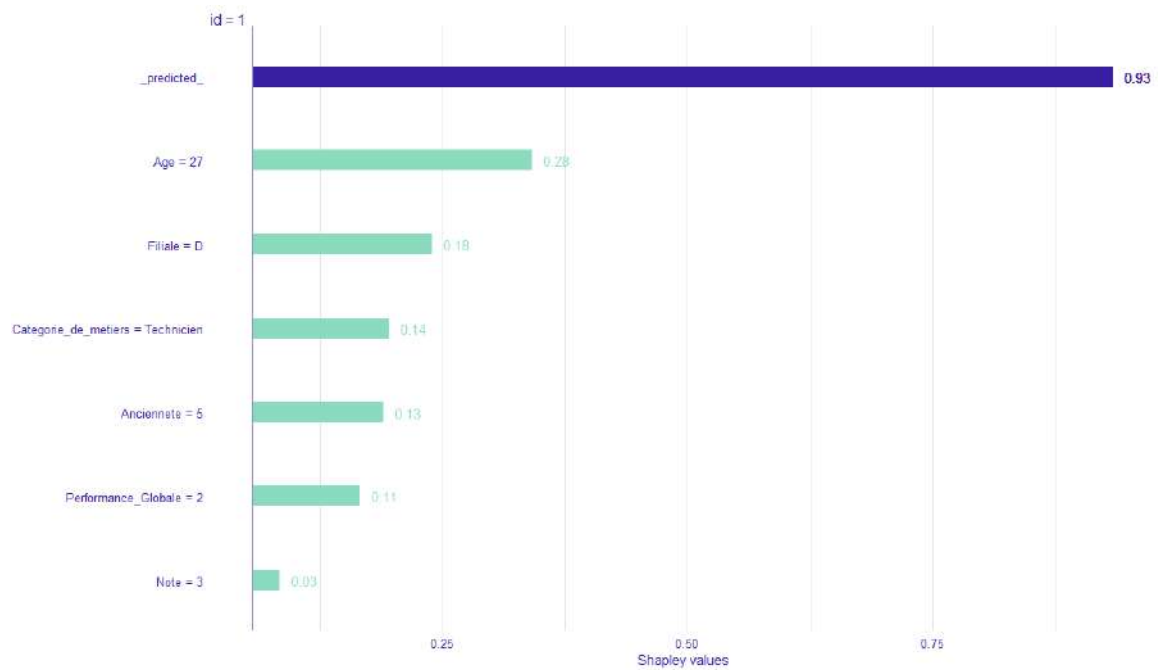


FIGURE 9.16 – Valeur de Shapley du troisième individu

On remarque à travers l'analyse des valeurs de Shapley que les raisons de prédiction de Promotion peuvent être tout à fait distinctes les unes des autres, en effet, pour cet individu plus jeune, la promotion par le modèle est expliquée par son métier et la filiale dans laquelle il travaille. Sur le graphique générale en essaim, ces variables expliquaient moins la promotion d'un individu cependant cela ne veut pas dire que ce n'est jamais le cas.

9.5 Retour avec les autres modèles

9.5.1 Nombre de femmes dans les faux positifs

On réintroduit la variable Sexe dans le modèle et on regarde le nombre de femmes dans les faux positifs. On observe 15 femmes sur 41 individus. Soit 36,5% d'individus c'est légèrement supérieur à la répartition de femmes dans notre base de données.

Le modèle a les variables les plus significatives suivantes :

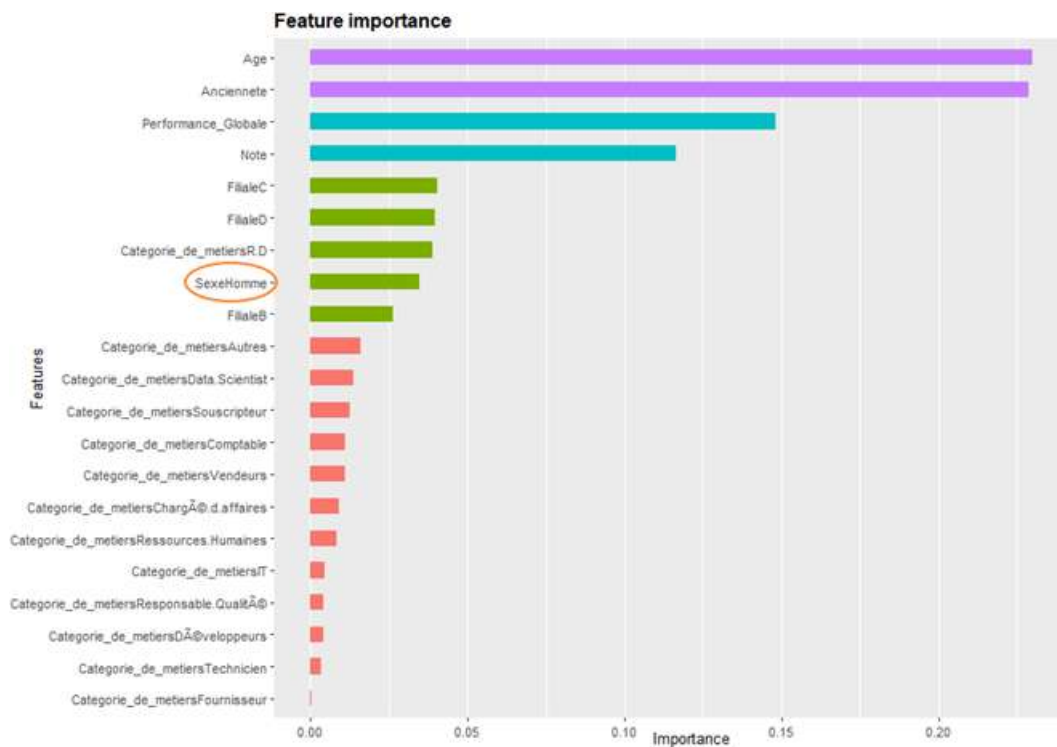


FIGURE 9.17 – Variables importantes du modèle avec la variable « Sexe »

Prendre en compte le « Sexe » pourrait biaiser le modèle, il pourrait ainsi ne pas promouvoir un individu en raison du sexe.

9.5.2 Xgboost avec la variable Sexe

On s'intéresse à la significativité de la variable Sexe dans les faux positifs en prenant cette fois en compte la variable Sexe. Le modèle ne prédit pas un nombre différent de faux positifs lorsqu'on prend en compte cette variable, on compte un nombre de femmes légèrement supérieur (34,7%) à la proportion de femmes dans la base de données. On considère une femme ayant une forte probabilité de promotion selon le modèle mais ne l'est pas en pratique.

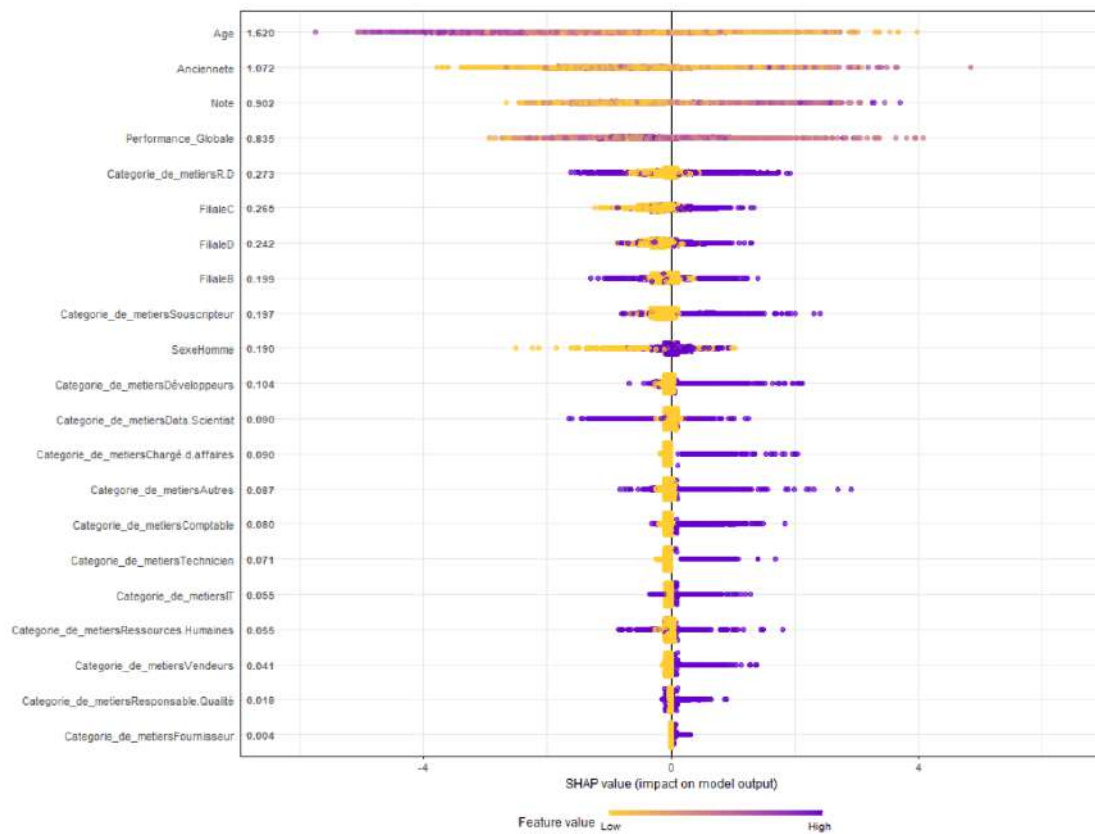


FIGURE 9.18 – Diagramme en essaim de Shap

Le graphique des importances selon Shapley indique que le Sexe peut jouer en partie à l'attribution d'une promotion.

Les prédictions de promotions peuvent donc être biaisées en faveur d'un sexe. Il est donc nécessaire de se préserver de modèles biaisés.

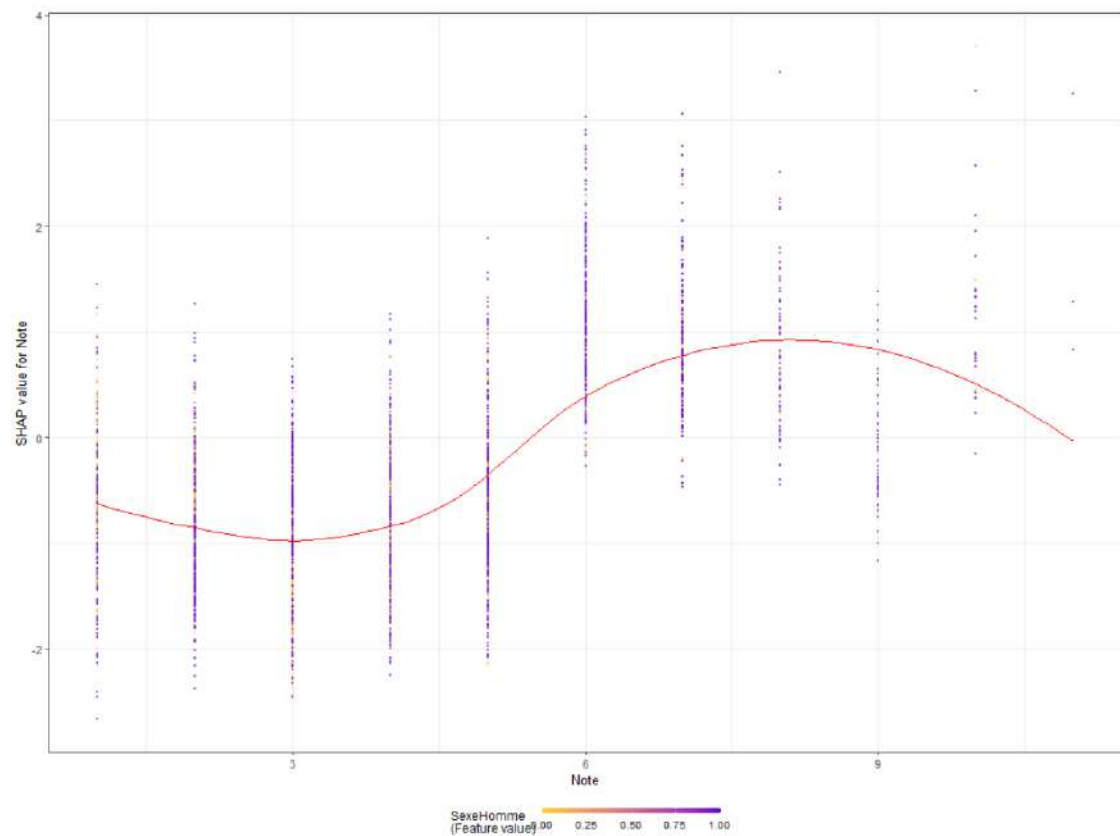


FIGURE 9.19 – Diagramme des dépendances de la variable Coefficient hiérarchique par Sexe

La couleur présente ici indique le Sexe de l'individu : la couleur jaune est une Femme et la couleur mauve un homme ($\text{SexeHomme} == 1$). Il présente l'influence de la performance sur la valeur de Shapley. Une plus faible performance indique est liée à une plus faible valeur de Shapley et contribue donc moins à la promotion d'un individu.

Voici le cas d'un autre diagramme de dépendance.

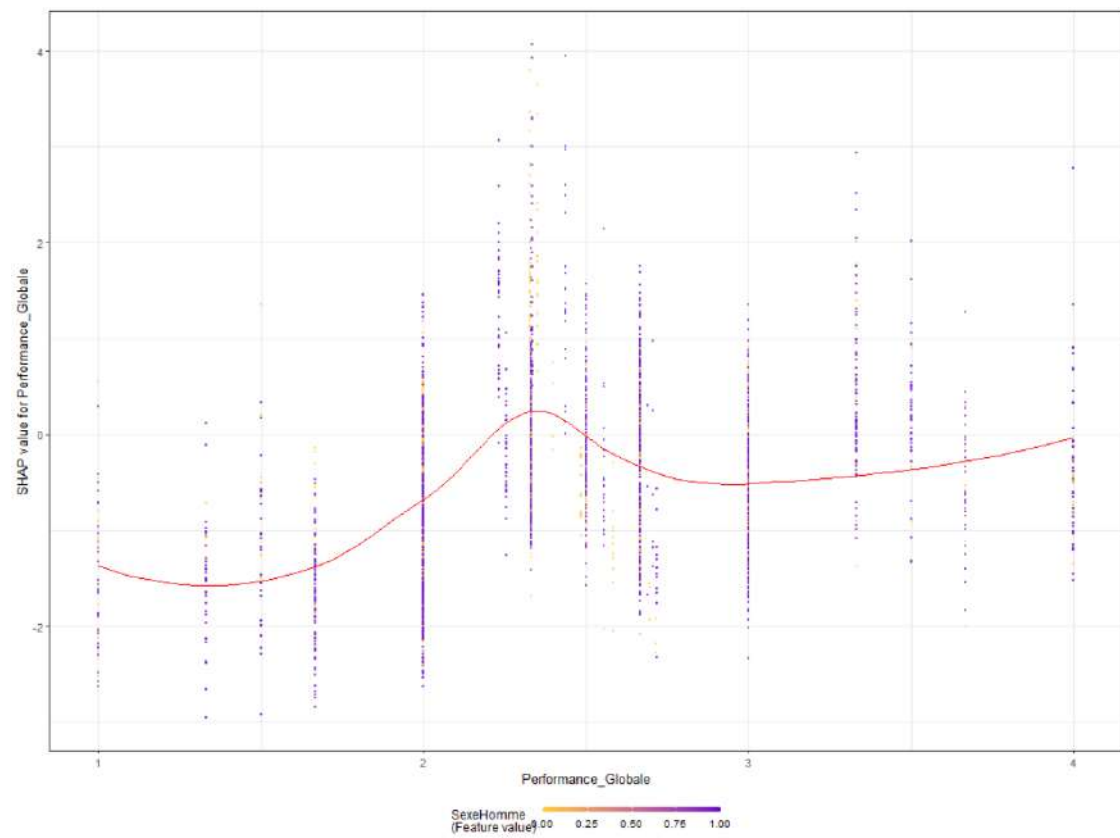


FIGURE 9.20 – Diagramme des dépendances de la variable Performance par Sexe

Voici le cas d'un faux positif où l'individu est une femme.

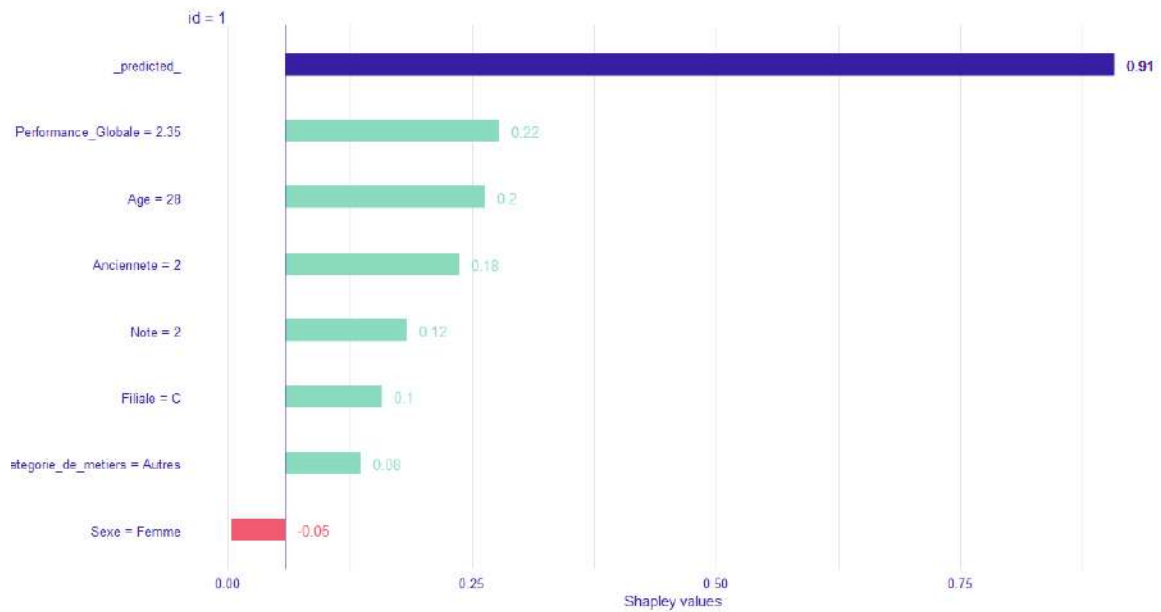


FIGURE 9.21 – Valeur de Shapley d’une femme non promue mais promue par le modèle

Le modèle prédit qu’il est fort probable que cet individu ait été promu. Les valeurs de Shapley révèle une contribution négative quant à la promotion par rapport à la variable Sexe. Selon le modèle, le fait que l’individu soit une femme contribue négativement à la promotion de cette femme. Le modèle va diminuer sa probabilité de promotion à cause de son Sexe.

Il est donc nécessaire de faire remonter cet individu au RH pour se renseigner et prévenir une potentielle inégalité de traitement.

9.6 Explication RH

Ce chapitre a permis d’établir un moyen de déterminer des individus potentiellement discriminés qu’il convient de faire remonter aux RH. Il a été mis en lumière les biais qui peuvent être introduits dans les modèles et dont il est nécessaire de s’en prémunir.

Seul quelqu’un ayant une connaissance approfondie de l’entreprise est à même de pouvoir étayer les conclusions et les explications quant aux individus que les modèles ont permis de détecter comme potentiellement discriminés.

Ainsi, les RH pourraient expliquer à travers des informations qui ne sont pas prises en compte dans les modèles : peut-être qu’un individu en particulier n’a pas été promu en raison d’une période d’absence ou cause de maladie ? peut-être a-t-il déjà été promu l’an passé ? peut-être va-t-il quitter prochainement l’entreprise ...

9.7 Limite de la méthodologie

Le fait que les données utilisées soient une base de salariés avec un nombre de variables explicatives réduites peut limiter les conclusions qu'il est possible de faire vis-à-vis de la prédiction d'une promotion. Cependant il convient de noter que ce chapitre présentait une méthodologie généralisable à un plus grand nombre de variables.

Un feature engineering avec des techniques d'oversampling tel que (Synthetic Minority Oversampling Technique (SMOTE)) pour augmenter le nombre d'individus de la classe promotion aurait pu être intéressant à aborder.

Pour les individus arrivant dans l'année la promotion indiquée est "non" ce qui introduit un biais au modèle.

De plus comme l'entreprise a par le passé commencé des mesures pour lutter contre les discriminations, des inégalités ont déjà été gommées.

9.8 Conclusion

Là où l'index de l'égalité professionnelle permettait seulement de compter le nombre de promotions par sexe, qui par conséquent ne présentait aucune discrimination dans cette base (avec d'ailleurs un taux de promotion légèrement supérieur pour les femmes), cette méthodologie permet de déterminer des individus potentiellement discriminés en donnant les raisons d'une promotion potentielle. Il a été vu dans ce chapitre comment déterminer des individus potentiellement discriminés quant à une promotion. Une liste d'une vingtaine de faux positifs à grande probabilité de promotion a été choisie pour être remontée aux ressources humaines.

La partie suivante traitera du critère d'augmentation, dans un premier temps à l'instar de l'index de l'égalité professionnelle, le nombre de promotion sera étudié puis un modèle sur le montant d'augmentation sera discuté.

Chapitre 10

Etude du critère sur l'augmentation

10.1 Les objectifs du modèle

Les étapes suivies dans ce chapitre sont les suivantes :



FIGURE 10.1 – Etapes mises en place pour étudier ce sous indicateur

10.1.1 Rappel sur l'index

L'objectif de cette partie est d'étudier les inégalités de revalorisation salariale de notre base de données et de mettre en parallèle nos résultats avec ceux calculés par le sous-indicateur de l'index de l'égalité professionnelle. Pour rappel, l'index ne déterminait aucune inégalité.

10.1.2 La démarche

L'objectif de cette partie est d'identifier l'écart entre l'augmentation moyenne des hommes et des femmes.

10.2 GLM

10.2.1 Utilisation d'un GLM et importance de la variable Sexe

Afin d'avoir une première vision globale de l'étude menée. Un GLM est implanté afin de connaître l'importance des variables dans l'explication de la variable revalorisation.

10.2.2 Création du GLM : ajustement des variables continues

Pour y parvenir, il est nécessaire de retravailler certaines variables de la base de données.

En effet, l'âge et l'ancienneté sont actuellement des variables continues, cependant le comportement de la revalorisation n'est pas linéaire avec les années qui passent.

Les revalorisations peuvent être plus élevées en début de carrière, là où les salaires sont plus faibles ; il est possible d'observer un pic de revalorisation aux âges où généralement les postes à hautes responsabilités sont obtenus.

Il est donc nécessaire que le GLM traite la variable Age et Ancienneté comme des catégories d'âge et d'ancienneté.

Afin d'y remédier, il va être établi deux modèles CART (Classification and Regression Tree) ; l'avantage des arbres est que les noeuds de nos arbres vont donner directement la segmentation la plus adéquate au modèle.

Il convient donc de faire deux arbres de régression sur la variable revalorisation avec respectivement uniquement la variable Age et uniquement la variable d'ancienneté pour chacun des deux modèles.

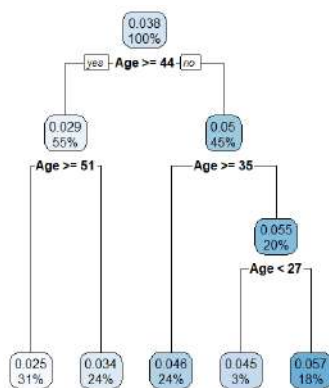


FIGURE 10.2 – Résultat de l'arbre de régression sur la revalorisation avec la variable Age

Une fois le GLM modélisé, l'importance des variables permet de connaître la contribution de chaque variable au modèle.

Pour les modèles linéaires, la contribution de chaque variable du modèle est estimée avec

la valeur absolue du test de Student pour chaque paramètre du modèle.
(le test de Student permet de tester si l'un des paramètres est égal à une valeur précise.
En particulier, la nullité de chacun des paramètres est souvent testée).

10.2.3 Importance des variables et résultats

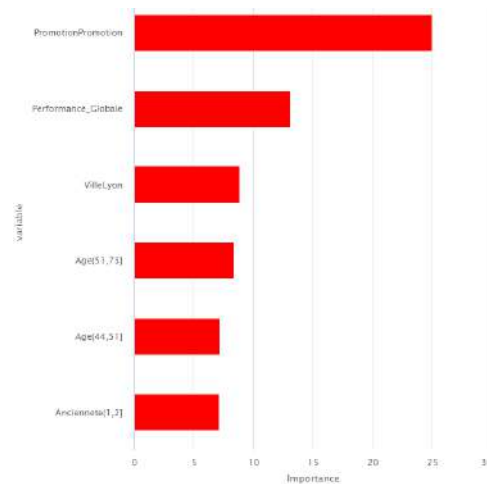


FIGURE 10.3 – Les 5 plus grandes variables contributives du GLM

Ici, il en ressort assez logiquement que les variables Promotion et Performance sont les plus contributives.

Le sexe est affiché contribuer à hauteur de 1.9 tandis que la promotion à hauteur de 25. Le sexe apparaît parmi les variables les plus faiblement contributives. A première vue, le sexe n'est pas significatif dans l'intensité de la promotion obtenue.

A travers la même méthodologie, la réception d'une promotion ne semble pas être expliquée par la variable sexe.

10.3 Xgboost Classification : nombre de revalorisation

10.3.1 Creation du Xgboost

Pourquoi réaliser utiliser un Xgboost ?

Plus puissant qu'un CART classique, le xgboost est ici utilisé pour essayer de capturer de faibles signaux étant donné sa meilleure robustesse.

Deux Xgboost sont réalisés sur la variable réponse : "est revalorisé (Oui/Non)", sur la base des individus présents / présents de l'année 2020 à 2021.

Un premier avec la variable sexe dans le modèle et un second sans cette variable.

Chaque modèle est entraîné sur une base d'apprentissage et évalué par cross validation

répétée : le but étant d'estimer si les prédictions des modèles se rapprochent aux mieux des données observées.

Après calibration du modèle par grille de recherche ; il en résulte le choix des hyperparamètres suivant :

Hyperparamètre	Description
Nrounds	70
Eta	1
Max_depth	12
Gamma	0
Colsample_bytree	1
Subsample	1
Min_child_weight	1

FIGURE 10.4 – Hyperparamètre du Xgboost Binaire sur la revalorisation

Le xgboost étant un modèle de régression, pour performer une classification, il convient de calculer manuellement à partir d'un seuil les prédictions choisies comme 1 ou 0. Dès lors que la prédiction dépasse le seuil, on attribue ceci à la prédiction d'une revalorisation. Il est nécessaire d'appliquer un seuil assez élevé pour optimiser les métriques : on cherche à minimiser le nombre de faux positifs pour ne garder que ceux qui sont potentiellement vraiment discriminés.

Ainsi, le nombre de faux négatifs ne nous intéresse pas car il s'agirait d'individus déjà revalorisés (on ne va donc pas chercher à leur retirer leur augmentation de salaire).

10.3.2 Résultats

La matrice de confusion permet d'identifier visuellement les prédictions correctement établies par le modèle et celles faussement prédites.

Les Faux positifs sont les résultats déclarés positifs mais qui s'avèrent en réalité négatifs. Ici, il s'agit d'individus que le modèle prédit comme devant recevoir une revalorisation mais en pratique ne la reçoit pas.

A contrario, les faux négatifs (Prédiction : 0 ; Observé : 1) correspondent aux individus non revalorisés par le modèle mais qui ont été revalorisés en pratique.

La matrice de confusion du modèle sans la variable sexe est la suivante.

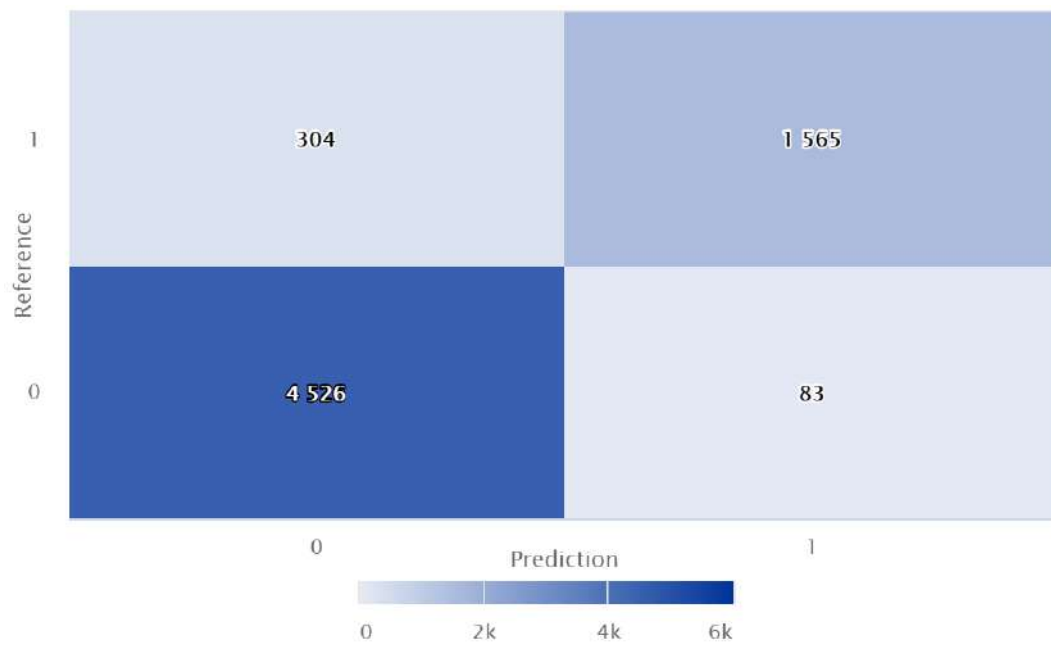


FIGURE 10.5 – Matrice de confusion

De cette matrice, des métriques sont déterminées, les données ne sont pas aussi mal balancées que pour l'étude de la promotion. La courbe ROC a davantage son intérêt ici.

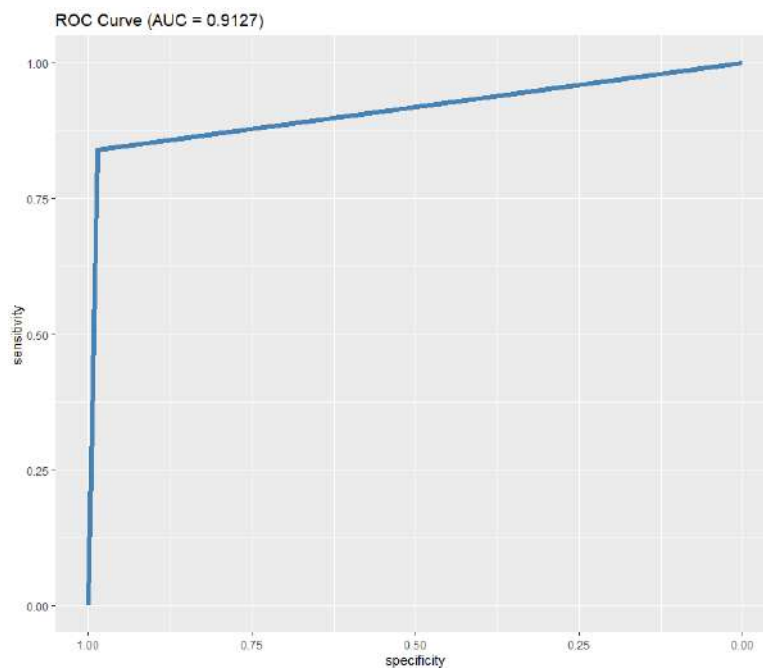


FIGURE 10.6 – courbe ROC

Le taux de vrais positifs (TPR) est légèrement moins bon ; le modèle est un peu moins précis que les modèles créés au chapitre précédent.

TPR (Sensibilité)	FNR	TNR (Spécificité)	FPR	BALANCED ACCURACY
83,7%	16,3%	98,2%	1,8%	91,0%

FIGURE 10.7 – Métriques

La précision est meilleure mais cela est au prix d'un rappel plus faible.

Précision	Recall
95,0%	83,7%

FIGURE 10.8 – Précision et Rappel

Il est possible de proposer aux RH, la liste des faux positifs, autrement dit la liste de potentiels discriminés ; cependant pour comprendre au mieux pourquoi ces individus sont potentiellement discriminés il est nécessaire de calculer leurs valeurs de Shapley qui vont permettre d'interpréter quantitativement.

Dans un premier, on va s'intéresser à l'interprétabilité globale du modèle en considérant tous les individus, voici l'importance des variables au sens de Shapley.

10.3.3 Interprétabilité globale

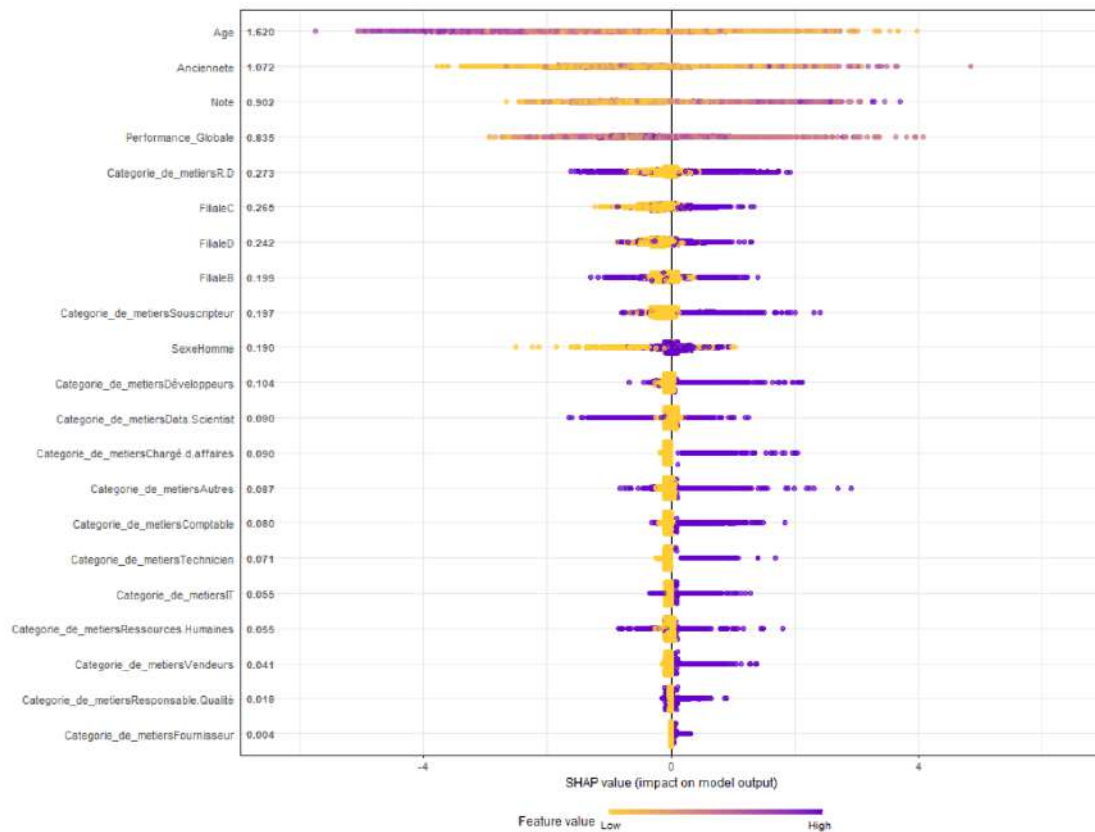


FIGURE 10.9 – Diagramme en essaim d’abeille pour la prédiction d’une augmentation

Les variables qui contribuent le plus à l’obtention d’une augmentation sont l’âge, l’ancienneté, le coefficient hiérarchique et la performance.

Les diagrammes de dépendance sont les suivants :

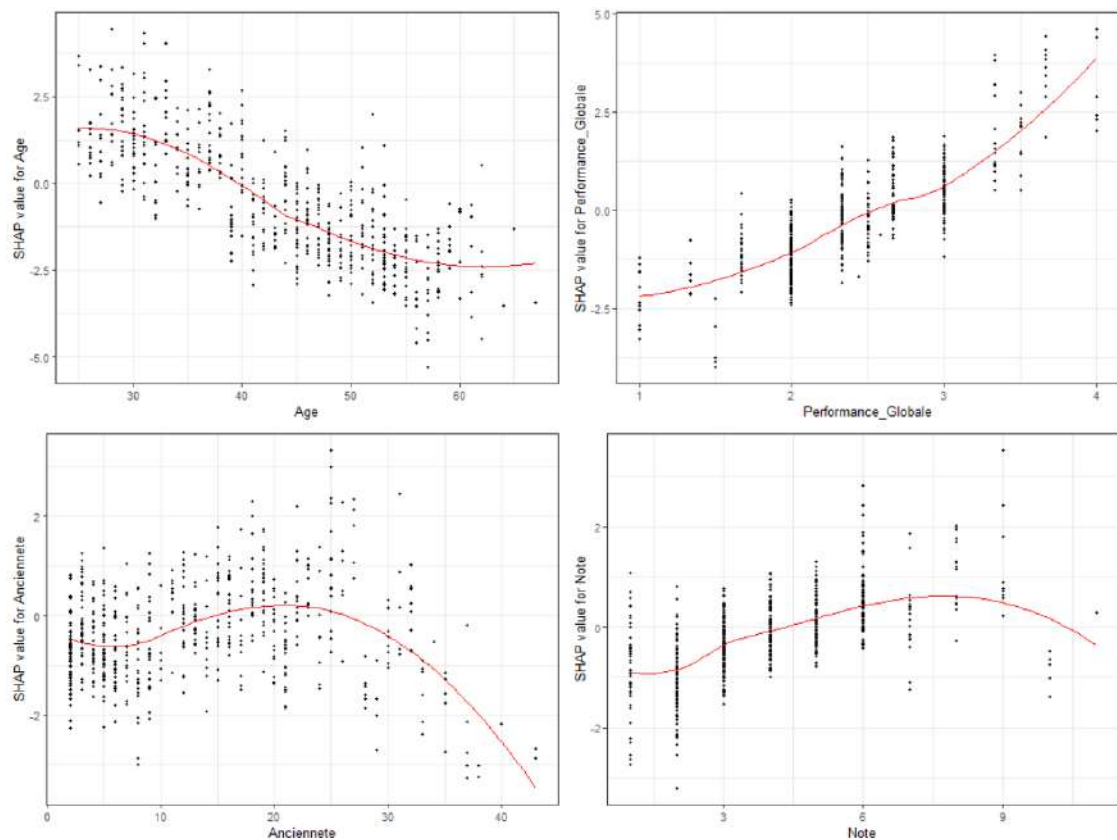


FIGURE 10.10 – Diagramme de dépendance pour la prédiction d'une augmentation

Le coefficient hiérarchique semble être moins déterminant dans l'obtention d'une augmentation à l'inverse de ce qui avait été vu pour l'obtention d'une promotion. Plus une personne a de l'ancienneté (et donc est âgée) moins celle-ci contribue à l'obtention d'une augmentation. Les très bonnes performances contribuent d'autant plus à l'obtention d'une augmentation.

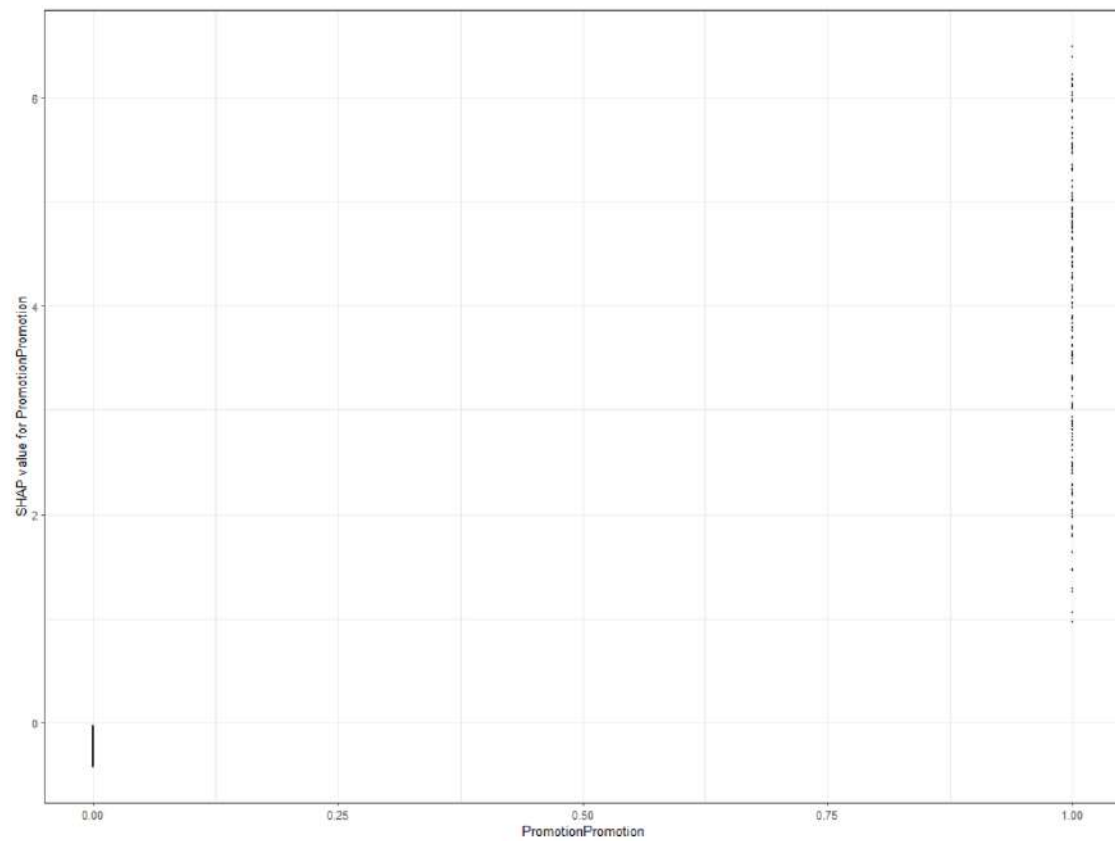


FIGURE 10.11 – Diagramme de dépendance de la Promotion pour la prédiction d’une augmentation

La promotion d’un individu semble contribuer de manière conséquente à l’obtention d’une augmentation.

Dans un second temps, quelques individus en particulier sont analysés :

10.3.4 Analyse des individus potentiellement discriminés

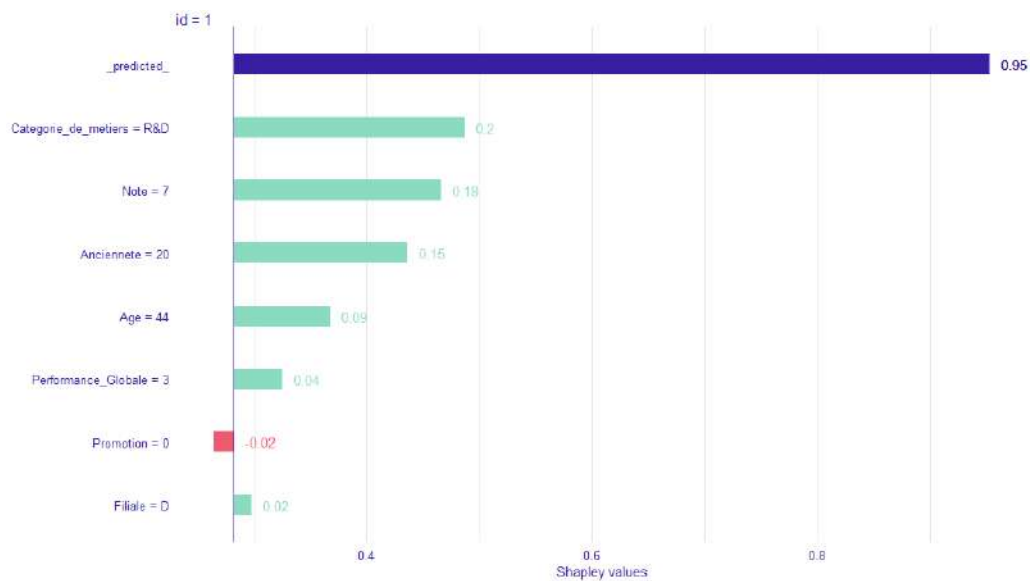


FIGURE 10.12 – 1ère valeurs de Shapley

Cet individu a été prédit comme recevant une augmentation par le modèle notamment par le fait qu'il travaille dans le secteur R&D de l'entreprise. Cette variable n'était pas présentée sur le diagramme en essaim d'abeille comme importante à l'attribution d'une augmentation pour les individus de manière général.

Bien qu'il possède une grande performance, cette variable n'est pas importante à l'obtention d'une augmentation. Le fait qu'il n'est pas été promu dans l'année contribue négativement à l'obtention d'une augmentation.

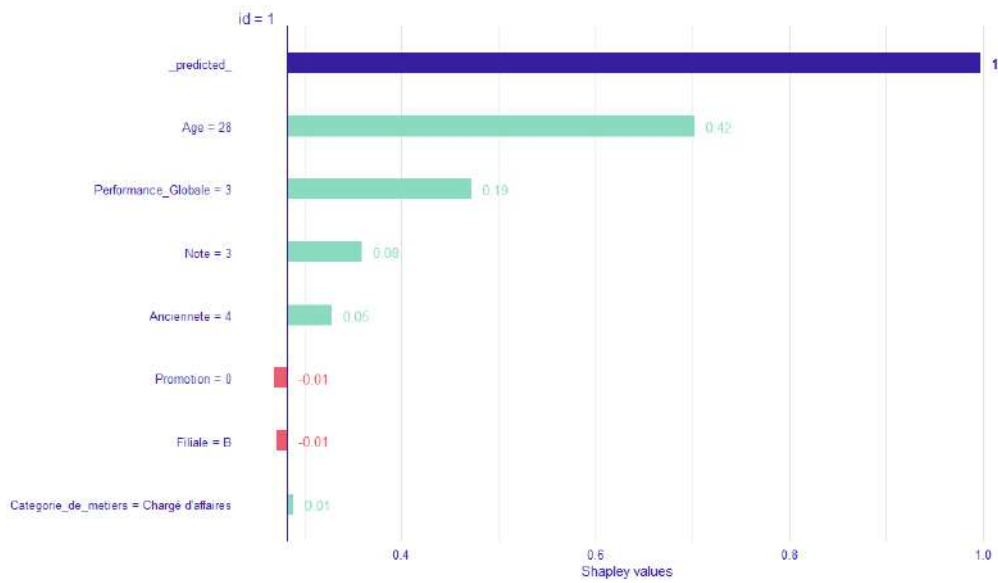


FIGURE 10.13 – 2ème valeurs de Shapley

Son jeune âge et ses bonnes performances expliquent selon Shapley la revalorisation attribuée par le modèle. Logiquement un jeune individu avec un salaire plus faible mais qui performe bien se verra attribuer une augmentation. De plus son coefficient hiérarchique augmente également les chances d'une revalorisation.

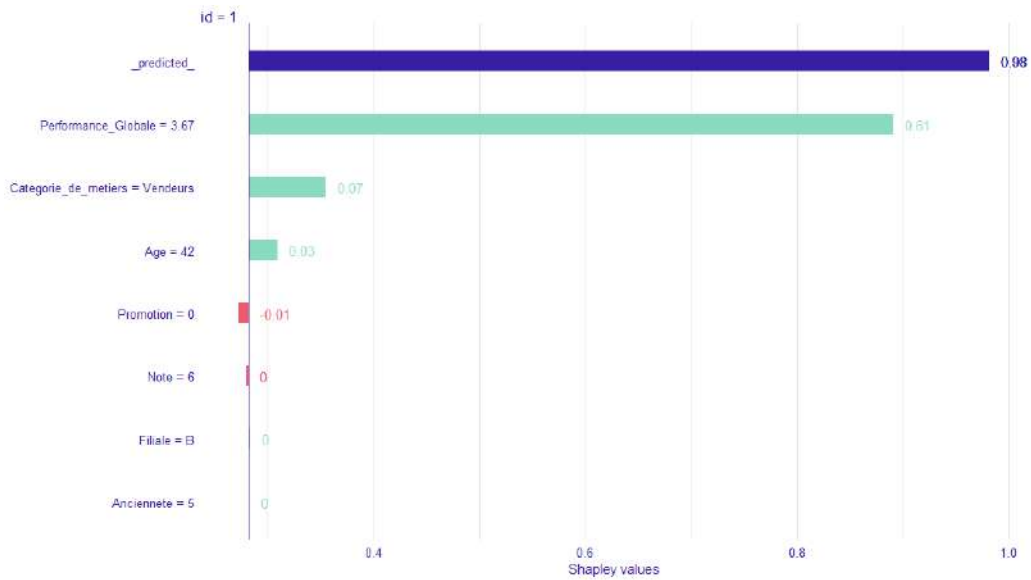


FIGURE 10.14 – 3ème valeurs de Shapley

Cet individu présente d'excellentes performances en tant que 'Vendeur' ce qui lui vaut

une revalorisation d'après le modèle. On peut observer que toutes les autres variables ne comptent plus pour l'interprétation de l'obtention d'une revalorisation.

10.4 Introduction de la variable « Sexe »

Les statistiques descriptives de la base de données indiquent : une revalorisation significative (supérieur à 3%) pour 32,8% des femmes contre 27,6% pour les hommes. En introduisant la variable dans le modèle précédent les métriques ne varient pas, le diagramme en essaim d'abeille est le suivant.

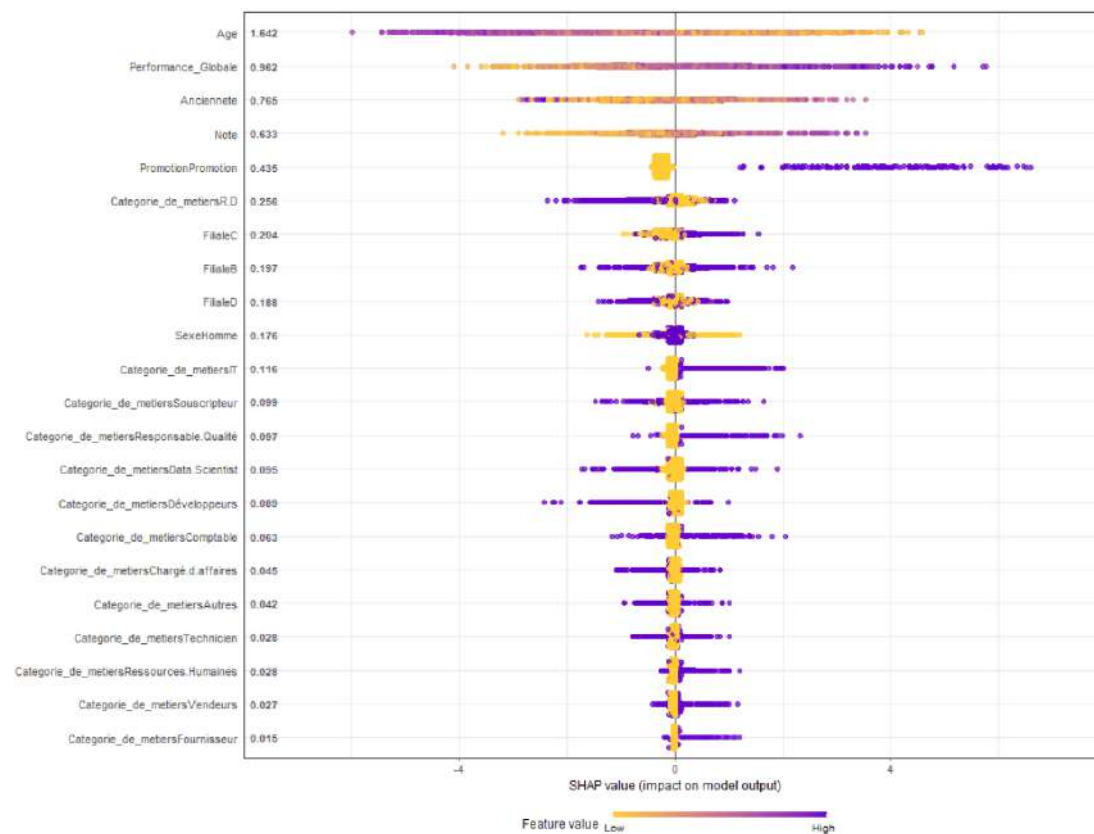


FIGURE 10.15 – Diagramme en essaim d'abeille

Il est à observer que la variable 'Sexe' semble contribuer selon le modèle à l'obtention d'une augmentation, c'est pourquoi il été nécessaire de l'enlever pour éviter ce biais dans le modèle.

Deux exemples d'individus 'Faux positifs' sont étudiés. On se place dans le cadre de faux positifs avec une forte probabilité de promotion.

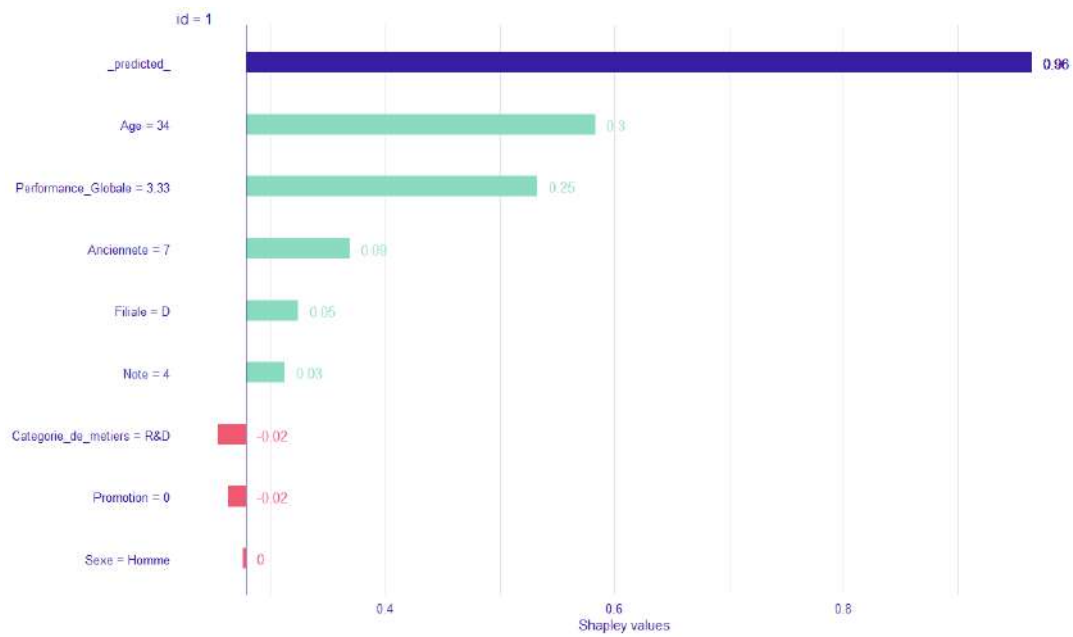


FIGURE 10.16 – Valeurs de Shapley d’un Homme (Faux Positif) avec la Variable ‘Sexe’ dans le modèle

Pour ce premier individu, les valeurs de Shapley suivantes indiquent que le Sexe ne contribue pas à l’obtention d’une revalorisation salariale.

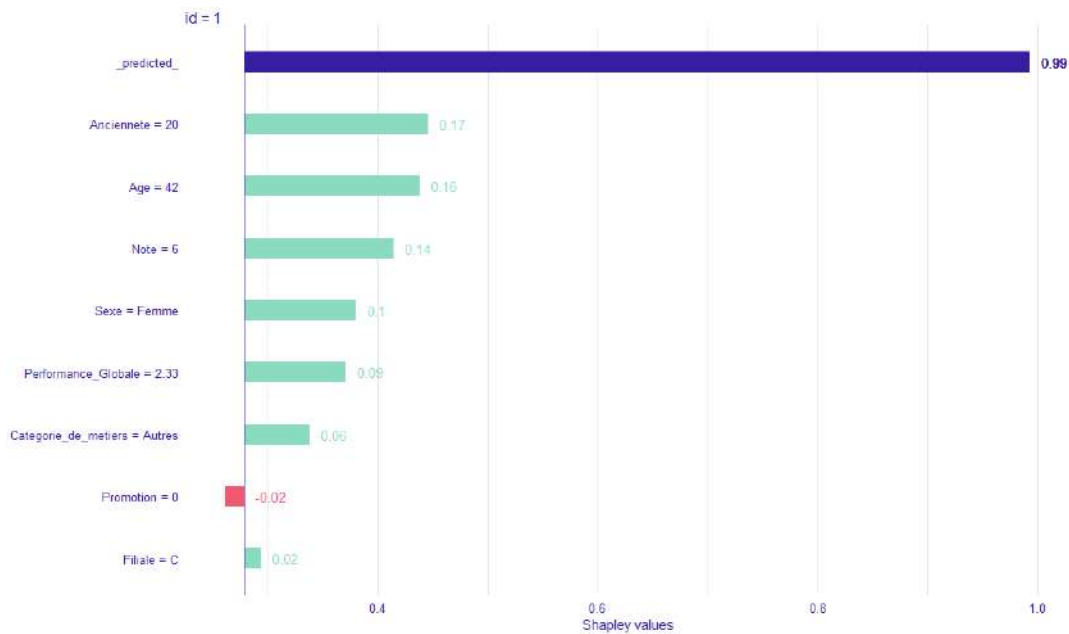


FIGURE 10.17 – Valeurs de Shapley d’une Femme (Faux Positif) avec la Variable ‘Sexe’ dans le modèle

Pour un autre individu, on observe que le Sexe aide positivement vers une prédiction d’une revalorisation. Il semblerait que le modèle la discrimine positivement quant à l’attribution d’une augmentation. Il est nécessaire de connaître les différents biais d’une base de données (abordé dans la partie suivante).

Dans les faits comme l’entreprise a déjà commencé une campagne de revalorisation des employés discriminés, certains se voient donc discriminés positivement, chose que le modèle semble arriver à détecter.

Il est logique de se demander si en plus de l’obtention d’une revalorisation, le montant attribuée est biaisée en fonction du sexe. Le but de cette partie est de s’interroger sur la création d’un indicateur de revalorisation théorique et de se demander si le montant de revalorisation n’est pas discriminant par rapport à un sexe.

10.5 Xgboost Regression

Cette section vise à modéliser le montant de revalorisation attribué en fonction du profil du salarié et d’étudier si le sexe a une importance dans l’obtention d’un montant de revalorisation plus grand.

10.5.1 Création d'une revalorisation théorique

Deux modèles Xgboost sont créés; l'un avec la variable Sexe et l'autre sans cette variable.

Le but étant d'analyser si le sexe est une variable prépondérante à la création d'une revalorisation théorique.

La RMSE du premier modèle, qui prédit une revalorisation théorique en % est de 4.652386.

Voici quelques résultats de prédictions du modèle sur la base de test.

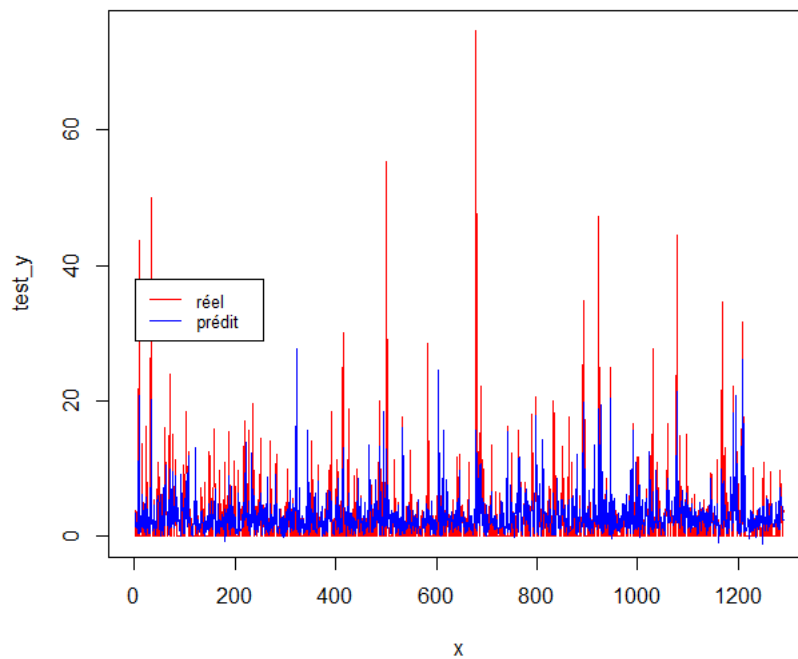


FIGURE 10.18 – Prédiction du modèle sur la base de test

On observe que le modèle prédit relativement mal les fortes augmentations.

Quelque soit le sexe des individus, il ne semble pas possible de conclure quant à l'existence de discrimination sur le montant de revalorisation.

En effet, les prédictions de revalorisation sont quasi identiques avec et sans cette variable.

revalorisation	revalo_accorde	pred_avec_sexe	pred_ss
3.449859	1	3.0148790	2.9413905
3.449538	1	3.5820401	3.4945388
3.449364	1	5.2878680	5.2487564
3.448369	1	4.4152393	4.4737048
3.448369	1	4.2783346	4.3433223
3.430616	1	2.3012493	2.1662056
3.429142	1	1.8561869	1.7365127

FIGURE 10.19 – Prédiction de montant de revalorisation avec et sans le sexe

10.5.2 Conclusion

L'index de l'égalité professionnelle étudie le nombre de revalorisation entre les sexes. Cette partie était elle, dédiée à déterminer à une maille plus fine des individus potentiellement discriminés quant à l'obtention d'une augmentation. Il a été observé un biais discriminant en faveur d'un sexe qu'il convient de supprimer. L'étude des valeurs de shapley a révélé que le modèle peut prédire des augmentations pour des raisons très diverses. Il convient de soumettre aux ressources humaines ces individus pour une analyse supplémentaire. Ce chapitre a essayé de créer des indicateurs théoriques de revalorisation et de détecter des discriminations dans le montant de la revalorisation. Les résultats observés n'ont pas permis de conclure quant à l'existence de discrimination.

Ces deux derniers chapitres ont permis de révéler l'existence de biais en faveur du sexe, la partie suivante s'intéressera à l'algorithmie éthique qui tente de lutter contre les biais discriminants en apprentissage automatique. Dans un premier temps, les hautes rémunérations seront étudiées en utilisant la théorie des valeurs extrêmes.

Chapitre 11

Etude du critère sur les plus hauts salaires

On considérera dans cette partie, un haut salaire comme toute rémunération au-dessus de 127.000€. Ce seuil a été déterminé en s'intéressant aux quantiles de distributions des salaires.

La modélisation des hauts salaires est problématique : d'une part les modèles ont tendance à saisir difficilement un faible nombre de données, d'autre part cela est couplé par un biais a priori dans la donnée émanant de discriminations sexistes. Il est nécessaire de disposer de méthodes plus poussées pour étudier ce sujet.

Plusieurs idées de modélisations des hauts salaires ont été considérées. Dans un premier temps, il va être expliqué en quoi la première méthodologie à travers des algorithmes ensemblistes est limitée et ne peut garantir des résultats consistants pour la modélisation des très hauts salaires.

Dans un second temps, une approche théorique à travers des valeurs extrêmes pour la modélisation des salaires sera abordée. Il s'agit de mettre en lumière un cadre théorique possible pour ce sujet et d'en dégager les forces et les faiblesses dans notre cas d'étude.

En 3ème lieu, sera abordée la notion de biais en algorithmie. Il s'agira dans cette section de s'interroger sur comment mesurer, comprendre et enfin limiter le biais non désiré induit par le sexe dans l'algorithme. Cela permettra notamment d'expliquer en quoi la première démarche était limitée pour discuter du sujet.

En effet étant donné le déséquilibre important des sexes dans les données émanant de la présence d'un plafond de verre au-dessus duquel presque aucune femme n'est présente, la modélisation des hauts salaires peut engendrer un biais défavorable aux femmes.

Le modèle saisira donc assez mal les salaires extrêmes, il serait judicieux de faire un traitement plus spécifique pour étudier ces valeurs.

11.1 Modélisation des salaires par approche ensembliste

On implémente un modèle Xgboost sur le salaire. On vient en première instance regarder la répartition du sexe dans les hauts salaires. On remarque que dans les hautes rémunérations les femmes sont très peu présentes (points de couleur noire dans la figure 11.1). De plus, les prédictions des hauts salaires ont tendance à avoir une assez grande variance, ce qui pose un problème de fiabilisation de la démarche puisqu'au départ on souhaitait étudier ces valeurs. On aimerait avoir un modèle où l'erreur est plus faible sur les haut salaires.

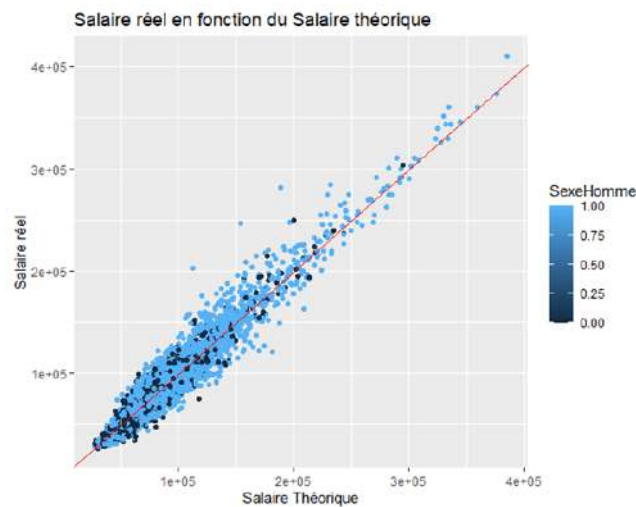


FIGURE 11.1 – Répartition des salaires prédits VS théoriques

L'idée de la partie suivante est d'essayer d'approcher les hauts salaires d'une autre manière en utilisant la théorie des valeurs extrêmes, largement utilisée en assurance pour la modélisation des sinistres extrêmes.

11.2 Le salaire comme valeur extrême

Un sinistre extrême est donc ici assimilé à un salaire extrême.

Le but est d'étudier le sexe dans les hauts salaires de la même manière qu'au Chapitre 8.

Il s'agit de déterminer l'importance de la variable « Sexe » dans le processus de prise de décision de l'algorithme et de pouvoir identifier une liste d'individus à hauts salaires potentiellement discriminés.

Cette partie se base sur les travaux [6] sur les arbres de régression de Pareto généralisés pour l'analyse des événements extrêmes.

Des extensions basées sur le principe de gradient boosting ont déjà été développées et décrites dans la littérature scientifique [14] mais nous avons choisi de nous limiter aux approches par arbres afin d'étudier le processus de prise de décision des algorithmes, ce qu'un modèle ensembliste rend beaucoup plus complexe. Ici, la compréhension du modèle prime sur la performance.

11.2.1 Notions de valeurs extrêmes

La théorie des valeurs extrêmes concerne un pan des statistiques qui traite des queues de distributions des mesures de probabilité.

Elle étudie le comportement asymptotique des maximas des valeurs prises par les valeurs de variables aléatoires indépendantes et identiquement distribuées (iid).

Deux approches existent :

- La méthode des maximas : une approche par blocs de tailles identiques de données qui ne sera pas abordée dans ce mémoire.
- La méthode des excès : une approche qui s'intéresse à la distribution des données au-delà d'un seuil fixé. Elle consiste à ajuster une distribution généralisée de Pareto aux excédants des données au-dessus de ce seuil.

La popularité de cette théorie tient notamment à l'un de ses résultats fondamentaux : pour toute variable aléatoire, la queue de la distribution d'une série de tirages iid de cette variable aléatoire peut être approximée par une distribution paramétrique qui est une distribution généralisée de Pareto.

Ce théorème permet donc de connaître la loi des excès (si elle converge). Voici le résultat en question :

Balkema et de Haan (1974)

Soit $Y_1, Y_2 \dots \in \mathbb{N}$ une série de variables indépendantes et identiquement distribuées

Soit un seuil (assez grand) $u \in \mathbb{R}$

On appelle événement extrême tout Y_i qui excède u ,

étant donné $Y_i > u$, l'excès est défini par $X_i = Y_i - u$

s'il existe $(a_u) > 0, (b_u)$ et une fonction de distribution non dégénérée H tel que :

$$\mathbb{P}[Y_i - u \geq a_u x + b_u | Y_i] \xrightarrow[u \rightarrow \infty]{d} 1 - H(x)$$

alors H est nécessairement de la forme :

$$H_{\sigma,\gamma}(x) = \begin{cases} 1 - (1 + \frac{\gamma}{\sigma}x)^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0 \\ 1 - \exp(-\frac{x}{\sigma}) & \text{si } \gamma = 0 \end{cases}$$

La famille des lois limites possibles pour la loi des excès est une famille paramétrique de lois qui s'appelle les lois de Pareto Généralisées.

Le σ est un paramètre d'échelle. Le γ (paramètre de forme) indique la sévérité des évènements extrêmes (appelé aussi indice de valeurs extrêmes). Ce paramètre de forme présente un intérêt particulier car il va décrire l'épaisseur de la queue de distribution ie la vitesse à laquelle la queue de la distribution va tendre vers 0.

Le but sera d'estimer γ conditionnellement aux caractéristiques du salaire (ce sera l'indice des valeurs extrêmes de l'observation $Y|X$).

11.2.2 Fonctionnement des arbres de régression Pareto généralisés

Le but de l'algorithme CART est d'estimer une fonction de régression m^* définit comme :

$$m^* = \arg \min_{m \in \mathcal{M}} \mathbb{E} [\mathcal{L}(Y, m(X))]$$

Y est la variable d'intérêt ; le salaire d'un collaborateur.

$X \in \mathcal{X} \subset \mathbb{R}^d$ un ensemble de covariables.

\mathcal{M} une classe de fonctions cibles dans \mathbb{R}^d

\mathcal{L} une fonction de perte qui dépend de la quantité qu'on souhaite estimer.

Le but est d'estimer ce m^* . Ceci va être fait de manière itérative.

A chaque pas dans l'arbre, à chaque étape, on va diviser les données en deux sous-groupes plus homogènes.

Comment sépare-t-on alors les données ?

Les observations sont séparés selon des règles.

Chaque observation est affectée à une classe en fonction des valeurs de ses covariables.

Les règles de séparation s'écrivent :

$$x = (x^{(1)}, \dots, x^{(d)}) \longrightarrow R_j(x)$$

$$\text{avec : } \begin{cases} R_j(x) & \neq 0 \text{ ou } 1 \\ R_j(x)R_{j'}(x) & = 0 \text{ pour } j \neq j' \\ \sum_j R_j(x) & = 1 \end{cases}$$

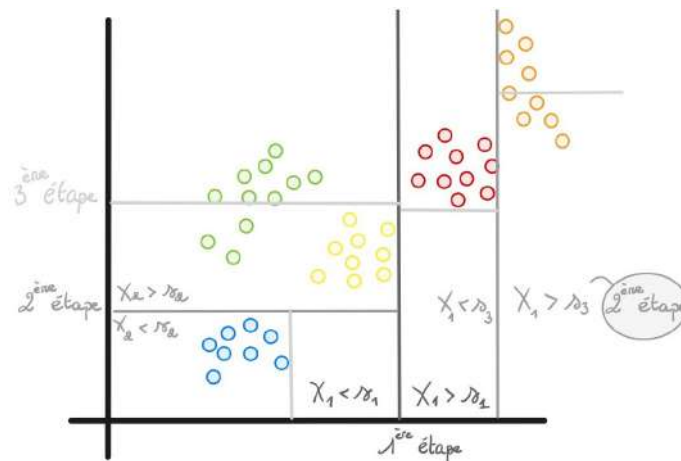


FIGURE 11.2 – Découpage en régions

Sur ce graphique, il est possible de voir qu'à la première étape, si la première covariable est plus petite que le seuil s_1 alors l'observation sera classée à gauche sinon à droite.

Ces choix sont faits tel qu'ils minimisent une version empirique du risque $\mathbb{E}[\mathcal{L}(Y, m(X))]$. Dans chaque sous-groupe, une estimation de m^* est faite (qui va donc dépendre de X) et permet de recommencer itérativement la méthode. En fonction des caractéristiques il y a donc une différente estimation de m . Il faut choisir les meilleurs seuils qui divisent au mieux les données en minimisant le risque.

L'estimateur \hat{m} de m^* :

$$\hat{m}^{\mathcal{R}}(x) = \sum_{j=1}^s \hat{m}(R_j) R_j(x)$$

$$\text{où } \hat{m}(R_j) = \arg \min_{m \in \mathcal{M}} \sum_{i=1}^n \mathcal{L}(Y_i, X_i) R_j(X_i)$$

Il faut choisir la covariable pour lequel un seuil permet de séparer au mieux les données.

A chaque fois il faut minimiser le risque sur chaque sous-groupe. Comme le but est de minimiser un risque, il faut choisir et le seuil et la variable sur laquelle la séparation est faite de façon à minimiser le risque empirique.

La fonction de perte choisie dépend essentiellement de ce qui doit être estimée. Par exemple, une fonction de perte quadratique revient à chercher l'espérance conditionnelle de Y sachant X dans chaque sous-groupe (dans chaque sous-groupe s'opère une régression L2).

Dans le cas de Pareto, dans chaque sous-groupe une loi de Pareto va être ajustée. De telle façon que dans chaque sous-groupe la vraisemblance soit maximale ce qui revient à minimiser la $-\log(\text{vraisemblance})$.

La fonction de perte considérée est la $-\log$ vraisemblance où la vraisemblance est une vraisemblance de Pareto généralisée définie ainsi :

$$\mathcal{L}(y, m(x)) = -\log(\sigma(x)) - \left(\frac{1}{\gamma(x)} + 1\right) * \log\left(1 + \frac{y\gamma(x)}{\sigma(x)}\right)$$

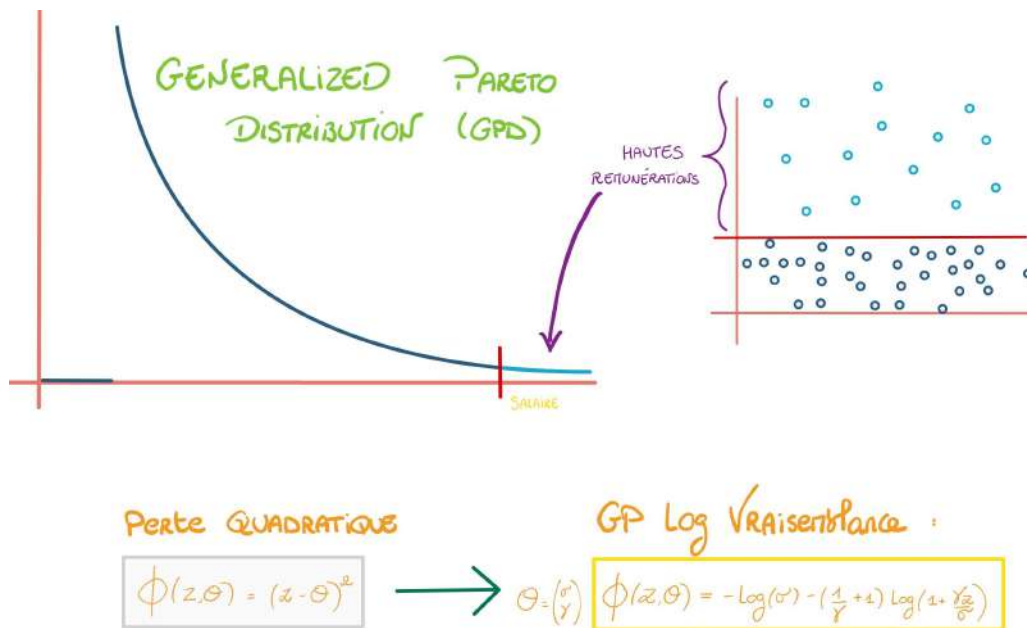


FIGURE 11.3 – Principe de la méthode

Dans chaque sous-groupe, une loi de Pareto généralisée va être ajustée. Et les sous-groupes sont construits de façon à maximiser la vraisemblance.

Il s'agit là donc de remplacer le critère de moindre carré actuellement utilisé comme fonction de perte par une fonction de perte alternative pour performer une régression de valeurs extrêmes. La partie suivante concerne l'application de cette théorie aux salaires des collaborateurs pour la base de données de ce mémoire.

11.2.3 Application

L'idée consiste à implémenter des arbres de régressions de Pareto généralisées sur les excès. L'évènement extrême est considéré comme des salaires qui dépassent un certain seuil u .

Le Cart de Pareto nécessite d'adapter la fonction de perte du Cart. Une méthode sur le logiciel R se décompose en 3 parties :

- une fonction initialisation
- une fonction d'évaluation. La fonction d'évaluation est appelée une fois par nœud.
- une fonction de fractionnement ("split"). C'est dans cette fonction que réside l'essentiel du travail à faire. Elle sera appelée une fois pour chaque covariable à chaque fractionnement potentiel.

Il est nécessaire de choisir un seuil pour appliquer la théorie des valeurs extrêmes, ce seuil doit être suffisamment grand pour isoler les valeurs extrêmes des autres valeurs (problème de biais). Il doit être assez grand pour être dans la queue de distribution (être suffisamment dans les événements extrêmes). Cependant, il convient de ne pas fixer un seuil trop élevé qui exclurait quasiment toutes les données (problème de variance).

Après application d'un seuil de : 127 000€

Le seuil est retiré du salaire pour se ramener à l'excès.

L'algorithme est appliqué sur le salaire sur les variables suivantes toutes les variables de la base de données. Les résultats obtenus sont les suivants :

```
Age< 45.5 108 1218.0230 0.230331000
Métiers=Chargé d'affaires,Développeurs,R&D,Souscripteur, 29 310.8643 0.13924
Métiers=Autres,Comptable,Data Scientist,IT 79 903.5094 0.17024
Sexe=Homme 55 627.0951 0.18519
  Age< 42.5 25 285.5596 0.14788
  Age>=42.5 30 341.5088 0.21065
Sexe=Femme 24 276.3128 0.12999
```

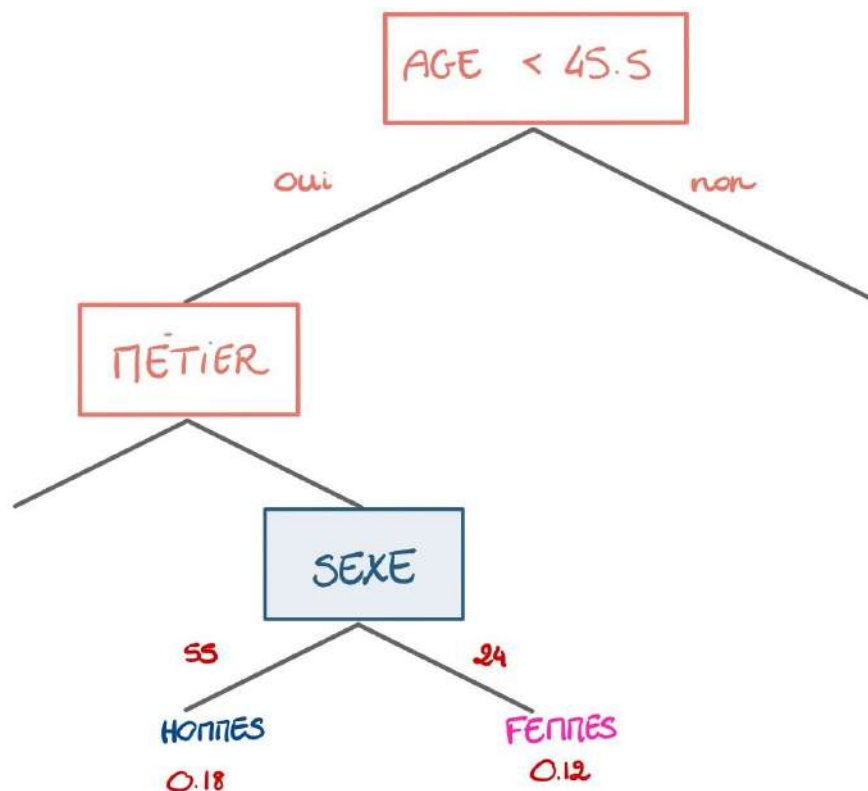


FIGURE 11.4 – Illustration des résultats

Le sexe est présent dans un nœud de l'arbre ce qui signifie que le sexe a une importance selon le modèle dans le montant du salaire extrême. Il est possible d'identifier dans les feuilles de l'arbre 24 individus potentiellement discriminés. Par la suite un dialogue avec les RH est nécessaire pour discuter si les individus identifiés sont discriminés ou non.

Pour résumer, Un haut salaire est assimilé à une valeur extrême. Une séparation des données selon le sexe s'opère dans les hautes rémunérations. La sévérité de l'évènement extrême est plus faible chez les femmes. On a plus de risque d'avoir des salaires plus extrêmes chez les hommes.

Les évènements extrêmes sont moins sévères chez les femmes autrement dit elles touchent des hautes rémunérations moins grandes.

Le problème quant à ces deux modèles est la fréquence de ces évènements extrêmes. Le nombre de sinistre des individus féminins est bien plus faible, autrement dit on observe une sous représentativité des femmes dans les hautes rémunérations.

En effet, on souhaite construire des modèles pour étudier des potentielles discrimina-

tions entre le sexe mais le problème sous jacent est qu'il n'y a quasiment aucune femme à étudier dans les hautes rémunérations.

On veut donc quantifier ce biais et analyser ce problème de fréquence, chose qui sera abordé dans la prochaine partie, nous allons pour l'instant étudier les différentes limites de cette approche.

Limites

L'algorithme est basé sur des lois de Pareto qui traite d'évènements extrêmes. Une limite possible à cette méthode est que ce parallèle peut être un peu exagéré dans le sens où les salaires étudiés ne sont pas si extrêmes. Ce graphique illustrant la répartition des salaires de la base de données montre une queue de distribution légère et fine.

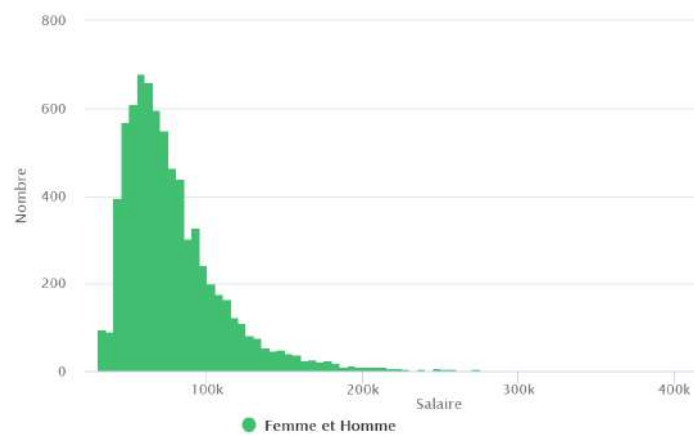


FIGURE 11.5 – Distribution des salaires

De plus, le groupe de femmes identifiées comme discriminées est limité par le fait que dans les hauts salaires les femmes sont peu représentées.

Les deux approches précédentes posent problèmes puisqu'elles sont construites autour d'un biais. Comment l'algorithme peut-il étudier les discriminations dans les hauts salaires puisque presque aucune femme n'est présente (plafond de verre mentionné auparavant)?

L'idéal serait d'avoir des outils pour étudier les biais dans les bases de données afin que les modèles ne les incluent pas au moment de la phase d'apprentissage et engendrent des erreurs dans les résultats d'une étude. Notamment il serait intéressant de savoir si l'attribution d'une haute rémunération, n'est pas biaisé (pour rappel l'index de l'égalité professionnelle considère le nombre de femmes et d'hommes dans les hautes rémunérations). Au lieu de retirer la variable Sexe du modèle pour éviter le biais dans les modèles une approche sur le biais dans les algorithmes sera abordée par la suite. En effet, cette

approche permettra corriger le biais là où la suppression de la variable ne réglait pas ce problème. L'obtention d'un haut salaire sera étudiée par la suite, en effet avant même de pouvoir étudier des inégalités en terme de montant de revalorisation, il serait nécessaire de savoir s'il n'y a pas déjà un biais dans l'attribution d'un haut salaire.

Pour résumer, l'index identifie un problème dans les 10 plus grandes rémunérations mais n'illustre pas un problème de fond sur le nombre de femmes à partir d'un certain seuil hiérarchique qui concerne plus que 10 individus.

Il est compliqué de déterminer des discriminations entre les sexes dans les hautes rémunérations car les femmes y sont peu présentes.

Il est cependant possible d'identifier des individus potentiellement discriminés en montant dans les hautes rémunérations et ainsi budgétiser des montants de rattrapage salariale pour les RH (on corrige le problème de sévérité).

Mais cela ne résout pas le problème de sous représentativité des femmes dans les hauts salaires.

Pour ce qui est de la représentativité, les actuaires ne peuvent pas conclure sur ce sujet là.

La prochaine partie fait un tour d'horizon sur le traitement du biais dans les algorithmes et tente d'identifier les biais présents dans les hauts salaires.

11.3 Le biais dans les algorithmes

Compte tenu du biais intrinsèque à la base de données dans l'étude des hauts salaires, cette partie est dédiée à l'analyse du biais en apprentissage automatique.

11.3.1 "Algorithm Fairness"

Un exemple d'algorithme biaisé

L'algorithme de recadrage de photos de twitter en 2021. Twitter avait dévoilé un algorithme de recadrage automatique de photos qui permettait de désencombrer les fils d'actualités en réduisant la taille des photos circulants, pour afficher l'image en entier il fallait donc appuyer sur cet aperçu. Le problème était que l'algorithme présentait certains biais.

L'algorithme choisissait l'élément le plus important d'une image avec une méthode de prépondérance ("saliency"), l'élément ressortant le plus important sur l'image.



FIGURE 11.6 – Illustration de l'élément important d'une image (Saliency)

Ainsi comme on peut le voir sur cette image, la tête du chien est déterminée comme élément important de l'image.

Le recadrage par Machine Learning avait tendance à considérer les poitrines des femmes comme élément prépondérant de l'image au lieu du visage. Les visages blancs étaient également privilégiés aux visages noirs pour l'aperçu de l'image.

Présentation

Le champ d'étude sur la « fairness » qu'on pourrait définir par "équité" en français est le champ qui vise à comprendre et éviter le biais en Machine Learning afin d'obtenir la plus grande neutralité et équité dans la sélection de l'information par l'algorithme et dans ses prédictions.

Les algorithmes de machine learning sont de simples fonctions mathématiques, elles agissent de manière objective et amoral. Bien que le choix du modèle peut amplifier le biais, ce ne sont pas les modèles vu ci-dessus qui comme on peut le lire dans la presse qui sont "racistes", ce sont les données biaisées introduites à l'entrée qui conduisent à un modèle biaisé.

Les raisons pour lesquelles les modèles sont biaisés peuvent être liées à un déséquilibre dans une classe, les femmes sont minoritaires dans notre cas de l'étude des hauts salaires. Amazon a par le passé fait face à un problème avec ses algorithmes de tri de CV (afin de fluidifier les processus de recrutement), la compagnie avait entraîné ses algorithmes sur la base de CV ayant réussi les entretiens, comme le nombre d'hommes dans le milieu de la Technologie était supérieur à ceux des femmes, le fait d'être une femme

était pénalisé par l'algorithme et le tri n'était donc pas neutre vis à vis des femmes. Les modèles faisaient des prédictions discriminantes car les données étaient biaisées en entrée.

L'interprétabilité abordé précédemment dans ce mémoire va généralement de paire, puisque comprendre comment les modèles font des prédictions permettra de savoir si les modèles sont "équitables".

En Algorithm Fairness les approches ne sont pas essentiellement quantitatives, les raisons peuvent aller au delà.

Cela signifie qu'il faut comprendre la cause plus profonde de l'injustice. Dans notre base de données, les femmes sont moins présentes dans les hauts salaires, cette raison est multi factorielle mais elle peut être lié à un phénomène de décrochage (carrière fragmentée) à cause des temps partiels liés à une grossesse (notion de plafond de verre). Le but des algorithmess sera d'aller chercher la partie d'écart sur la variable d'intérêt réellement due à la discrimination «non explicable».

La section suivante présente les termes spécifiques employées dans la littérature sur le Machine Learning éthique.

11.4 Terminologie

Un "*label*" est une valeur correspondant à un résultat.

Une "*favorable label*" est une variable dont la valeur correspond à un résultat qui procure un avantage au destinataire. Par exemple, être embauché pour un emploi dans l'exemple d'Amazon. (c'est la variable d'intérêt ou la variable à prédire).

Un "*protected attribute*" est un attribut qui divise une population en groupes dont les résultats devraient être paritaires (la parité en termes d'avantages reçus). C'est un critère qui devrait être paritaire et ne devrait pas influencer les résultats du modèle. Par exemple : la race ou le sexe.

Concrètement il s'agit de **décorrél**er au maximum l'attribut protégé de la sortie. L'éthique est "obtenu" si la sortie n'est pas corrélée à l'attribut protégé.

La sortie peut-être corrélée aux autres variables explicatives mais cela reste éthique.

On définit le "*Group Fairness*" par le fait que les groupes définis par des attributs protégés reçoivent des traitements ou des résultats similaires.

L'"*Individual Fairness*" correspond au fait que des individus similaires reçoivent des traitements ou des résultats similaires.

Une "*Fairness metric*" est une métrique qui quantifie le biais dans les données ou les modèles. On entend par biais une erreur systématique. En Algorithm Fairness on étudie les préjugés non désirés qui donnent un avantage systématique aux groupes privilégiés et

un désavantage systématique aux groupes non privilégiés.

Un "*algorithme d'atténuation des biais*" est une procédure de réduction des biais indésirables dans les données ou les modèles.

Avant de commencer, il est nécessaire de définir les notions de biais mathématiquement.

11.4.1 Introduction du biais dans un algorithme

Les biais peuvent être introduits dans chacune des **3 phases** : "*Pre-processing*", "*In-processing*", "*post-processing*". Il faut ainsi avoir des algorithmes adaptées à chaque étape.

"*Algorithme de pré-traitement*" c'est un algorithme d'atténuation des biais qui est appliqué aux données d'apprentissage.

Les données d'apprentissage peuvent être biaisés en faveur de certains types de classe.

"*Algorithme de traitement en cours (in-processing)*" c'est un algorithme d'atténuation des biais qui est appliqué à un modèle pendant sa phase d'entraînement.

Il est possible d'utiliser des algorithmes d'in-processing s'il est possible de modifier la phase d'apprentissage d'un modèle d'apprentissage automatique. Cela revient donc à modifier l'algorithme qui est entraîné.

L'algorithme qui crée le modèle peut être biaisé dans la mesure où il peut générer des modèles qui sont pondérés en fonction des caractéristiques particulières des données en entrée. Cela revient par exemple à changer la fonction de coût d'un modèle.

"*Algorithme de post-traitement*" c'est un algorithme d'atténuation des biais qui est appliqué aux prédictions. S'il n'est pas possible de modifier les données ou l'algorithme, il faut utiliser ces algorithmes de post-traitement.

L'ensemble de données de test peut être biaisé dans la mesure où les attentes concernant les réponses peuvent être biaisées.

Algorithmes de pré-traitement

Le biais inhérent à la classification est le plus souvent causé par les données elles-mêmes plus que par les classifieurs.

Le moment le plus opportun pour atténuer le biais est donc de le faire en amont c'est-à-dire lors du pré traitement, comme les biais sont généralement intrinsèquement liés aux données. Le biais est ainsi corrigé avant d'entraîner les modèles.

L'algorithme utilisé pour mitiger le biais dans ce mémoire sera le :

"*Reweighing Algorithm*" L'avantage de cette approche est qu'au lieu de modifier la variable à prédire, elle attribue des poids différents en fonction des catégories d'attributs protégés et de résultats, de sorte que les biais sont supprimés de la base d'apprentissage.

Les pondérations sont basées sur les fréquences d'apparition dans chaque groupe afin de garantir l'équité avant la classification.

Après avoir calculé ces poids sur la base d'apprentissage, ils peuvent être utilisés en entrée de classifieurs (un modèle qui prédit des réponses catégorielles) comme des régressions logistiques ou des Xgboost.

Cependant, comme cette technique est conçue pour fonctionner uniquement avec des classifieurs capables de gérer les poids au niveau des observations cela peut limiter les choix de modèles.

D'autres algorithmes de pré processing existent pour mitiger le biais : "*Optimized Pre-Processing*"; "*Learning Fair Representations*"; "*Disparate impact remover*".

Voici un récapitulatif des différents algorithmes existants pour chaque étape.

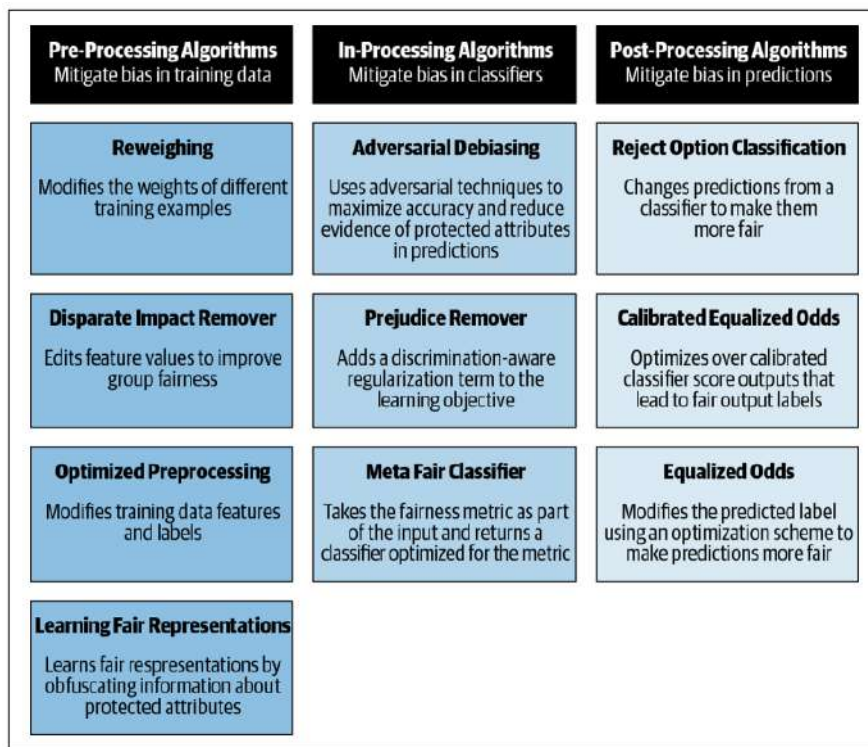


FIGURE 11.7 – Les différents algorithmes en Algorithm Fairness

11.4.2 Les différents types d'injustice ("unfairness") notions mathématiques de discrimination

La discrimination en matière de classification s'exprime à travers différentes notions qu'il convient de définir.

Tout d'abord le cadre mathématique est le suivant :

Soit (Y, X, A) où Y est la cible (variable à expliquer, la sortie) ; X les variables explicatives et A l'attribut protégé (défini par la suite).

L'objectif de l'apprentissage supervisé est de construire un prédicteur $\hat{Y} = f(X, A)$ qui prédit Y (mesuré par une fonction de perte). Le but est d'obliger que \hat{Y} ne soit pas discriminant par rapport à A .

La non-discrimination du point de vue de l'apprentissage supervisé, consiste donc à prédire un résultat Y à partir de variables X basées sur des données d'apprentissage étiquetées en s'assurant qu'elles ne sont pas discriminantes par rapport à un attribut protégé A .

Un premier concept de non discrimination est la **parité démographique** : une décision se doit d'être indépendante de l'attribut protégé.

Dans le cas d'une décision binaire, $\hat{Y} \in \{0, 1\}$ et l'attribut binaire protégé est tel que $A \in \{0, 1\}$ autrement dit : $\mathbb{P}(\hat{Y} = \neq | A = \neq) = \mathbb{P}(\hat{Y} = \neq, A = \neq)$. La connaissance de A ne change pas la prédiction. Cette définition a ses limites (elles seront abordés dans chacune des sous parties suivantes), c'est pourquoi de nouvelles définitions ont été introduites.

L'égalité de précision de la procédure conditionnelle

"Equalized odds" en anglais également appelé "disparate mistreatment".

Un prédicteur \hat{Y} satisfait l'égalité de précision de la procédure conditionnelle par rapport à l'attribut protégé et la variable à expliquer Y , si \hat{Y} et A sont indépendants conditionnellement à Y .

L'égalité de précision de la procédure conditionnelle s'applique aux cibles et aux attributs protégés prenant des valeurs dans n'importe quel espace (binaire, continu...).

Le cas des variables aléatoires binaires Y , \hat{Y} et A trouve son application dans de nombreux domaines (égalité des chances à l'inscription dans une université par exemple) ; ce qui revient au cas suivant :

$$\mathbb{P}(\hat{Y} = 1 | A = 0, Y = y) = \mathbb{P}(\hat{Y} = 1 | A = 1, Y = y) \text{ avec } y \in \{0, 1\}$$

Dans le cas binaire, l'issue $Y = 1$ est généralement considéré comme l'issue "favorisée" (qui donne un avantage) ou la classe sous représentée par exemple "recevoir une promotion".

C'est une métrique d'équité qui vérifie si, pour une sortie et un attribut particulier, un classifieur prédit cette sortie équitablement pour toutes les valeurs de cet attribut.

Contrairement à la parité démographique, l'égalisation de précision de la procédure conditionnelle permet à \hat{Y} de dépendre de A mais uniquement par l'intermédiaire de la variable cible Y .

Ainsi, la définition encourage l'utilisation de variables qui permettent de prédire directement Y , mais interdit l'utilisation abusive de A comme variable de substitution pour Y .

Pour le résultat $y = 1$, la contrainte exige que \hat{Y} ait des taux de vrais positifs égaux dans les deux groupes démographiques $A = 0$ et $A = 1$.

Un classifieur répond à cette définition si les sujets des groupes protégés et non protégés ont un taux de vrais positifs et un taux de faux positifs égaux.

L'égalité des chances

Un assouplissement de la notion est possible, en exigeant la non-discrimination uniquement au sein du groupe de résultats "favorisé" (positifs). Cela conduit à un assouplissement de la notion qu'il convient d'appeler "égalité des chances".

Par exemple, exiger que les personnes qui remboursent leur prêt aient les mêmes chances d'obtenir le prêt en premier lieu (sans spécifier aucune exigence pour ceux qui remboursent leur prêt, c'est-à-dire de faire défaut ou non).

C'est une mesure d'équité qui vérifie si, pour une sortie (qui confère un avantage à une personne) et un attribut donné, un classifieur prédit cette sortie de manière égale pour toutes les valeurs de cet attribut.

$$\mathbb{P}(\hat{Y} = 1 | A = 0, Y = 1) = \mathbb{P}(\hat{Y} = 1 | A = 1, Y = 1)$$

En d'autres termes, l'égalité des chances mesure si les personnes qui devraient se qualifier pour une opportunité ont les mêmes chances de le faire, indépendamment de leur appartenance à un groupe.

Pour résumer,

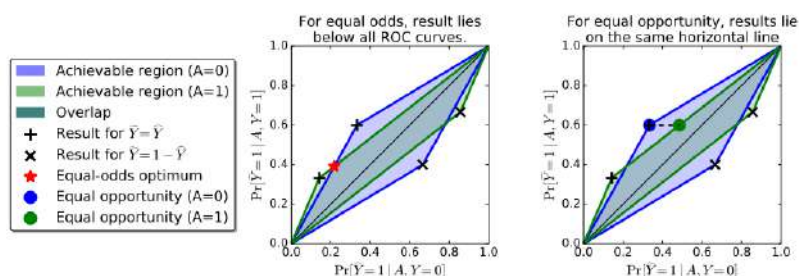


FIGURE 11.8 – Trouver le prédicteur optimal en terme d'égalité de précision et d'égalité des chances

La partie suivante présente les façons dont le biais peut intervenir dans un modèle.

Fairness métriques

Voici quelques métriques qui sont utilisées pour quantifier le biais.

Statistical Parity Difference Il s'agit de la différence de probabilité de résultats favorables entre les groupes non privilégiés et privilégiés.

Une valeur de 0 implique que les deux groupes ont un avantage identique.

Une valeur inférieure à 0 implique un bénéfice pour le groupe privilégié, et une valeur supérieure à 0 un avantage plus important pour le groupe non privilégié.

Disparate Impact Il s'agit du rapport de la probabilité de résultats favorables entre les groupes non privilégiés et privilégiés.

Une valeur de 1 implique que les deux groupes ont un avantage identique.

Une valeur inférieure à 1 implique un avantage supérieur pour le groupe privilégié, et une valeur supérieure à 1, un avantage supérieur pour le groupe non privilégié.

Equal Outcome Difference Il s'agit de la différence de taux de vrais positifs entre les groupes non privilégiés et privilégiés.

Une valeur de 0 implique que les deux groupes ont un avantage identique.

Une valeur inférieure à 0 implique un avantage supérieur pour le groupe privilégié et une valeur supérieure à 0, un avantage supérieur pour le groupe non privilégié.

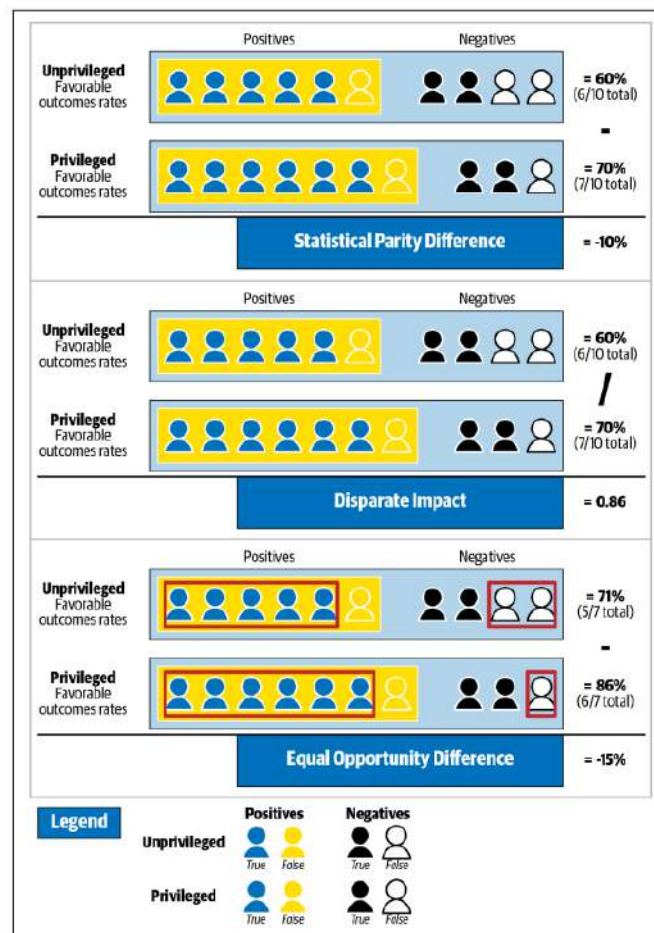


FIGURE 11.9 – Métriques

A présent, voici une illustration de ces notions pour la base de données de ce mémoire.

11.5 Réduction du biais dans les hauts salaires

Les méthodes d'algorithm Fairness sont appliquées sur la base de données à travers l'algorithme AIF360.

Le but est de déterminer si l'attribution d'une haute rémunération est biaisée et trouver une façon de réduire le biais dans les hauts salaires.

On va appliquer un algorithme de pré-processing.

Le favorable label est le **Salaire Extrême**.

Le protected attribute est le **Sexe**.

Le privilege value est le **Sexe Homme**.

La métrique utilisée est une métrique basée sur une sortie binaire : l'attribution d'une haute rémunération.

Après avoir défini les différents arguments pour la réduction du biais (priviledge value, protected attribute ...) on utilise AIF360, un package dédié à "l'algorithme fairness" pour détecter les biais dans les données concernant les femmes dans les hautes rémunérations (unbalanced class).

On va comparer le pourcentage de résultats favorables (être dans une haute rémunération) dans chacun des groupes (priviledge et non priviledge) en soustrayant le premier pourcentage du second.

Si on obtient une valeur négative signifie des résultats moins favorables pour les groupes non priviledge.

On utilise la méthode "mean difference" de "BinaryLabelDatasetMetric".

The image shows a handwritten formula in red and blue ink, enclosed in a red rectangular box. At the top, it says "JE SUIS DANS UNE HAUTE RÉMUNÉRATION". Below that, the formula is: $P(Y=1 | D = \text{NON PRIVILÉGIÉ}) - P(Y=1 | D = \text{PRIVILÉGIÉ})$. Underneath the first term, it says "SACHANT JE SUIS UNE FEMME" and under the second term, "SACHANT JE SUIS UN HOMME".

FIGURE 11.10 – Comparaison du pourcentage de résultats favorables (être dans une haute rémunération) dans le groupe des hommes à celui des femmes

On obtient en sortie -0.1462622.

Autrement dit, la différence moyenne de résultats entre les deux groupes est de -**0.1462622**.

Le groupe privilégié Homme obtient **14,6%** de résultats positifs en plus vis à vis de l'obtention d'une haute rémunération. On vient donc de prouver un biais en faveur des hommes dans l'obtention d'une haute rémunération.

On peut donc essayer d'atténuer ce biais, pour éviter que les algorithmes soient biaisés vis à vis des hauts salaires, on vient donc en amont avant la création du modèle traiter les données. On utilise l'algorithme évoqué plus haut (Reweighing Algorithm) pour transformer les données et ainsi avoir plus d'équité dans les deux groupes.

Après transformation des données on peut de nouveaux utiliser la métrique "mean difference" pour vérifier qu'on a bien réduit le biais ; l'algorithme affiche 0 en sortie ce qui signifie qu'on est passé d'avoir un biais qui avantage les hommes dans la présence dans les hauts salaires à une égalité en terme de résultat moyen.

L'importance du biais dans les algorithmes est un sujet capital notamment dans des

études sur des sujets sensibles, il est nécessaire d'être à même de savoir ce que font les algorithmes, bien qu'il soit de plus en plus compliqué de comprendre les choix des modèles boîtes noires qui sont de plus en plus complexes. C'est pourquoi il est nécessaire de s'équiper de diverses méthodes comme celles vues dans ce mémoire : interprétabilité, réduction du biais ... Ces enjeux peuvent être d'autant plus majeurs quand des questions d'éthiques sont de mises.

Pour résumer cette cinquième partie

Un index a été mis en place afin de donner un cadre commun à l'ensemble des entreprises. Ceci implique une méthodologie « macro », applicable par tous types de sociétés.

Ce mémoire répond aux besoins des RH d'avoir une analyse spécifique à la structuration de leur entreprise, avec des modèles de machine learning venant challenger cet index :

- L'écart de rémunération entre les femmes et les hommes : une identification plus segmentée, plus fine avec l'identification d'un budget spécifique pour gommer ces discriminations.
- Le nombre de promotions entre les femmes et les hommes : une identification des salariés considérés par les modèles de machine learning avec une forte probabilité de promotion, mais non promu.
- Le nombre d'augmentations entre les femmes et les hommes : une identification des salariés considérés par les modèles de machine learning avec une forte probabilité d'être augmentés, mais non augmentés.
- Le poids des femmes dans les plus hautes rémunérations : problématique du plafond de verre, avec une complexité d'utiliser des modèles de machine learning par manque de représentativité des femmes dans les plus hauts salaires.

Une communication primordiale avec les RH :

- Un besoin d'expliquer et d'interpréter ces modèles « boîtes noires »,
- Un besoin de retranscrire la connaissance des RH sur les salariés, dans les données ou a posteriori des modèles.

Sixième partie

Applications actuarielles : départ à
la retraite

Chapitre 12

Actuariat des engagements sociaux

12.1 Contexte

Cette partie est dédiée à l'analyse de l'impact de reprises salariales sur le montant des engagements d'une entreprise vis à vis de ses salariés au titre de régimes à prestations définies.

L'étude se limite ici aux avantages postérieurs à l'emploi qui sont généralement intimement liés au salaire et dont l'évaluation dépend donc des revalorisations annuelles successives attendues : la retraite supplémentaire se base sur le dernier salaire perçu, les indemnités de fin de carrière dépendent du salaire au moment du départ à la retraite.

L'analyse portera ici sur un régime en particulier celui des indemnités de fin de carrière.

Le but est de mesurer l'impact sur les passifs sociaux des entreprises selon les différentes normes applicables et selon le profil de chaque salarié.

Définitions d'un avantage social

Un avantage social naît d'une promesse d'avantage faite par l'employeur à ses salariés. Cet avantage est susceptible de faire courir un risque à la société qui l'a formulé.

Il y a donc nécessité d'une norme afin d'encadrer les périmètres et les méthodes d'évaluation, ainsi que les méthodes de comptabilisation de l'engagement au titre de cet avantage.

12.2 Les normes existantes

Plusieurs normes encadrent les avantages du personnel : les normes IAS 19 de l'IASB (International Accounting Standards Board) pour les comptes consolidés et les FRENCH GAAP pour les comptes sociaux d'une entreprise française.

12.2.1 Norme IAS 19

La norme IAS 19 est spécifiquement relative aux avantages du personnel. Cette norme indique comment **identifier**, **valoriser** et **comptabiliser** les avantages sociaux accordés aux salariés d'une entreprise.

12.2.2 Identification des avantages sociaux

La norme IAS 19 classe ainsi les différents régimes.

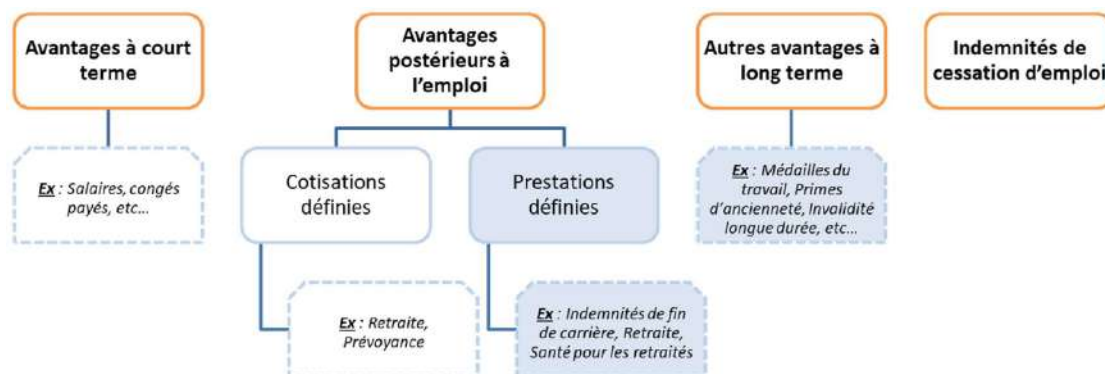


FIGURE 12.1 – Nature des engagements au sens de la norme IAS 19

12.3 Comptabilisation des engagements

La norme IAS 19 impose, pour les "avantages postérieurs à l'emploi" et pour les régimes de retraite à prestations définies (discutés par la suite) :

Pour le bilan

La provision à constituer au passif du bilan de l'entreprise au titre des avantages du personnel correspond à un montant égal à la Valeur Actuelle Probable (VAP) des prestations futures versées aux salariés en prenant en considération l'ancienneté dans l'entreprise au moment des évaluations. Il est nécessaire de diminuer la valeur de cet engagement du Best Estimate ("juste valeur") des actifs de couverture lorsque l'entreprise en possède.

Pour le compte de résultat

La comptabilisation en charges du coût annuel des régimes de retraite, pour rattacher la charge à l'exercice comptable concerné. Ce coût est défini par l'IAS 19 par :

- Le coût des services rendus.

- Le coût des services passés.
- La charge d'intérêt.
- Le rendement réel/attendu des actifs de couverture.
- Les écarts actuariels (pour les autres avantages à long terme).
- Les prestations réellement versées/attendues.
- Les cotisations au régime (employeur ou bénéficiaire).

Éléments à reconnaître en autre élément du résultat global (OCI) :

Les écarts actuariels (pour les avantages postérieurs à l'emploi) d'hypothèses démographiques, financières, d'expérience et les écarts actuariels sur le rendement des actifs de couverture.

12.3.1 French GAAP

Provision sous les normes comptables françaises

La comptabilisation des engagements au bilan est la méthode préférentielle, mais elle n'est pas obligatoire

En cas de comptabilisation au **bilan** :

la provision est égale à la valeur des engagements diminuée de la valeur des actifs de couverture lorsque l'entreprise en possède et augmentée ou diminuée des écarts actuariels non reconnus et des coûts des services passés non reconnus.

En compte sociaux, dans le cadre de la méthode préférentielle, il est possible de se rapprocher des dispositions de la norme IAS 19 :

- Reconnaissance immédiate en compte de résultat des coûts des services passés.
- Rendement attendu des actifs de couverture égal au taux d'actualisation.

La différence entre les normes françaises et la norme IAS 19 se situe au niveau de la méthode de reconnaissance des écarts actuariels.

- pour l'IAS19 : les écarts actuariels sont reconnus intégralement en capitaux propres (OCI)
- pour les normes françaises il existe deux possibilités : soit les écarts actuariels sont reconnus progressivement en compte de résultat (méthode du corridor) ou bien immédiatement en compte de résultat.

12.4 Valorisation des engagements

12.4.1 Méthode d'estimation de la valeur des engagements

La norme IAS 19 impose une méthode d'estimation de la valeur des engagements qui s'intitule la méthode des "Unités de Crédit Projetées" (Projected Unit Credit Method). Cette méthode actuarielle repose sur le principe qu'à chaque période de service rendu le

salarié a droit à une unité supplémentaire de droits à prestation.

L'estimation des prestations futures se fait à partir des salaires projetés au moment du départ à la retraite (**d'où la motivation de cette étude dans le cadre des inégalités de salaire et de revalorisation**).

Le but est d'attribuer à chaque année passée la quotepart de droits revenant au salarié.

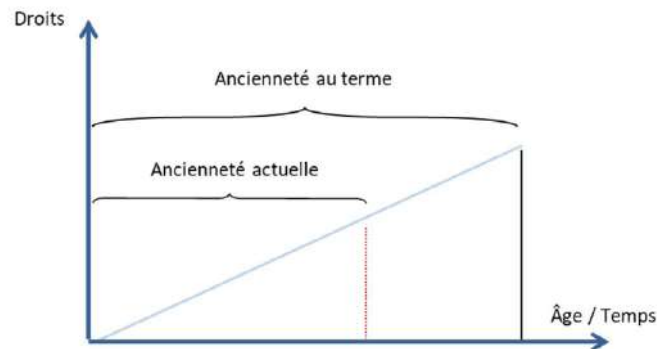


FIGURE 12.2 – Méthode comptable linéaire

12.4.2 Les hypothèses de projection

L'estimation de la valeur des engagements nécessite un ensemble d'hypothèses actuarielles « exemptes de parti pris et mutuellement compatibles » (§75 de la norme IAS 19). Les engagements sont valorisés tête par tête, c'est pourquoi des données individuelles (types RH) sont nécessaires :

- Entité
- Matricule
- Sexe**
- Dates de naissance, d'entrée dans l'entreprise
- Catégorie socio-professionnelle**
- Salaire annuel brut**

Ces hypothèses peuvent être de nature :

- Démographique :
Table de mortalité, Table de turnover, Âge de départ à la retraite
- Financière :
Taux d'actualisation, Taux de revalorisation des salaires, Taux d'inflation.

Pour certaines hypothèses long terme (turnover, profil de carrière, ...) la mise à jour complète annuelle n'est pas nécessaire si des indicateurs prouvent que des évolutions significatives ne sont pas survenues au cours de la période de service écoulée.

12.4.3 Méthodes d'étalements des droits

Pour rappel, on se place dans le cas d'un régime IFC.

L'approche A :

Consiste en un étalement linéaire de l'acquisition des droits entre la date à laquelle l'employé commence à travailler avec l'entité (date d'entrée dans le régime) et la date de départ à la retraite.

L'approche B :

Consiste en un étalement linéaire de l'acquisition des droits entre la date d'entrée dans le régime et la date d'atteinte du niveau plafond des droits (si un tel plafond existe).

L'approche C : la recommandation 2021 de l'IFRIC (Comité d'interprétation des Normes internationales d'information financière)

Elle consiste à linéariser l'acquisition des droits sur la période précédant l'âge de départ à la retraite, sur une durée équivalent à l'ancienneté requise pour atteindre le plafond de droits (période de services rendus considérée dans leur analyse comme réellement attributive des droits).

C'est une nouvelle approche de rattachement des droits aux périodes de services rendus pour un régime d'avantages du personnel à prestations définies.

Les sociétés cotées soumises aux normes IFRS et uniquement pour l'établissement de leurs comptes consolidés (maison-mère et filiales) sont concernés par cette méthode comptable.

Les comptes sociaux et les comptes consolidés des sociétés non cotées pourront faire application de cette nouvelle méthode de comptabilisation pour les engagements s'ils répondent aux conditions suivantes :

les droits dépendent de l'ancienneté dans l'entreprise, ils sont plafonnés après un certain nombre d'années d'ancienneté et ils sont payables à compter de la date de départ en retraite.

En cas de barème avec plafonnement des droits : l'approche C s'applique obligatoirement aux comptes consolidés.

Pour les comptes sociaux : l'entreprise a le choix avec la méthode A ou C.

Un exemple : Âge actuel : 40 ans Ancienneté actuelle : 10 ans
Âge de retraite : 62 ans Ancienneté au terme : 32 ans Ancienneté plafond : 16 ans

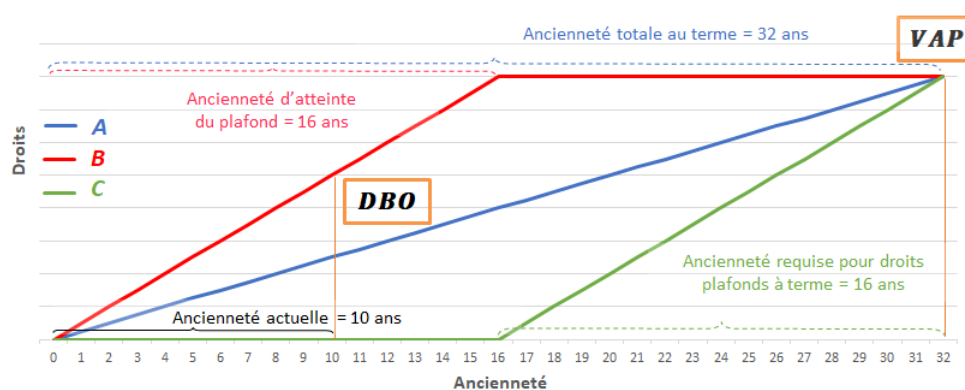


FIGURE 12.3 – Rythme d'engagement total de l'employeur par approche

Dans cet exemple, le passage de l'approche A à l'approche C se traduirait par un relâchement immédiat de la DBO mais par une charge future plus importante et non linéaire à VAP constante.

L'approche adoptée concernant la période d'acquisition des droits a son importance puisqu'elle va se traduire pour l'entreprise par un étalement plus ou moins lissé de la charge annuelle au titre des droits en cours d'acquisition.

La conséquence pour le provisionnement des engagements en compte sociaux va dépendre de la norme appliquée laquelle dépend du type de structure que représente l'entreprise.

Ainsi, comme discuté en première partie, comme la méthodologie mise en place dans ce mémoire se veut généralisable à tout type d'entreprise, les conclusions avancées par la suite peuvent être influencées par le type de normes (internationales ou françaises) auxquelles sont soumises les entreprises.

En effet, discuter d'une reprise de salaire pour des jeunes collaborateurs avec la méthode comptable C n'aura aucun impact sur la DBO mais un léger impact avec l'approche A.

Cependant, les individus plus âgés sont concernés par un plus fort impact avec l'approche C recommandée par l'IFRIC que l'approche A classique.

L'approche IFRIC ouvre l'opportunité de réduire le poids des passifs sociaux de retraite au bilan, et dans une certaine mesure renforcer les fonds propres de l'entreprise.

Les impacts vus par la suite seront discutés selon les différentes approches vues pour comptabiliser l'engagement, puisqu'ils peuvent être plus ou moins conséquents selon le profil concerné (jeune / âgé).

L'enjeu de cette partie est d'étudier la sensibilité de la dette actuarielle à des revalorisa-

tions salariales dépendamment des différentes méthode de comptabilisation.

Voici quelques éléments de calculs des différentes notions évoquées précédemment pour l'approche A.

12.4.4 Éléments de calculs pour l'approche linéaire (méthode A)

Valeur Actuelle Probable (VAP)

La VAP correspond au montant de l'engagement total de l'entreprise vis à vis de ses salariés.

$$VAP = \sum_i (F_i * \nu^i * \mathbb{P}_i)$$

avec : i : la date du i -ème flux F_i : la valeur du i -ème flux
 ν : facteur d'actualisation entre la date de calcul et la date de versement du flux
 \mathbb{P}_i : la probabilité de versement du i -ème flux ; les régimes concernés ici sont des avantages postérieurs à l'emploi, la probabilité de versement correspondra donc à une probabilité de présence dans l'entreprise et de survie au moment du départ à la retraite.

Dette actuarielle réelle

La dette actuarielle réelle (Defined Benefit Obligation, DBO) correspond au montant de l'engagement réel de l'entreprise vis-à-vis de ses salariés au sens de la norme IAS 19, et est calculée toujours au sens de cette norme au prorata temporis de la *Valeur Actuelle Probable (VAP)*.

$$DBO = VAP * \frac{\text{Ancienneté actuelle}}{\text{Ancienneté au terme}}$$

Les autres éléments nécessaires à la projection

- Coût des services rendus (Service Cost, ou SC) : valeur des droits acquis durant la période de service à venir.
- Charge d'intérêt (Interest Cost, ou IC) : correspond à la « désactualisation » de l'engagement sur la période de service à venir, du fait du rapprochement de l'échéance.
- Prestations probables : prestations probables à payer au cours de la période à venir.

Charges de la période N+1

$$SC = VAP * \frac{VAP}{\text{Ancienneté totale}} = \frac{DBO}{\text{Ancienneté actuelle}}$$

$$IC = (DBO + SC - \frac{\text{Prestations probables}}{2}) * i$$

Engagement projeté

Engagement déterminé en N, projeté en N+1 =

Engagement réel en N + SC + IC – Prestation probables en N+1

12.5 Application

La recommandation 2021 de l'IFRIC d'appliquer la méthode C nécessite de répondre aux critères déjà évoqués (régime à prestations définies prévoyant une prestation en fonction de l'ancienneté au terme et avec un plafonnement des droits à partir d'une ancienneté donnée).

Pour les comptes sociaux, les entreprises ont le choix entre les deux méthodes A et C. Pour les comptes consolidés, la méthode C est adoptée.

Pour les sociétés françaises qui provisionnent leurs avantages sociaux de retraite et font partie d'un groupe, cela permet qu'une même méthode d'étalement pour les deux référentiels de comptes soit appliquée (et ainsi permet d'éviter des écarts de méthode).

Nous allons étudier, les différentes méthodes de valorisation des engagements : la méthode dite IFRIC concerne la méthode "C" du graphique ci-dessus et la méthode "Hors IFRIC" sera la méthode "A".

L'impact sera faible car la revalorisation restera faible vis à vis de la masse salariale totale, de plus seul un nombre faible d'individus est concerné par rapport par les reprises de salaire.

Vouloir analyser l'impact total d'une revalorisation est liée à beaucoup d'autres facteurs externes.

Il n'y a pas la prétention ici de faire une analyse fine de l'impact sur les fonds propres de l'entreprise, par exemple moins de provisions implique aussi plus d'impôt.

Cette étude est là pour réfléchir sur les sujets auxquels les actuaires pourraient être confrontés dans le cadre d'évaluation des impacts d'une campagne revalorisation salariale.

Ici on va seulement évaluer de manière macro la provision figurant dans le bilan de l'entreprise vis à vis des engagements de l'entreprise envers ses salariés.

12.5.1 Présentation Outil

Afin de calculer les Dette actuarielles individuelles (DBO) de l'année en cours, un outil actuariel est utilisé.

Celui-ci a été adapté afin de prendre compte les différentes méthodes de calcul des engagements.

Cet outil nécessite une base de données RH (CSP, Age, salaire, ancienneté ...) ainsi que plusieurs hypothèses qu'il est nécessaire de définir et d'obtenir avec l'entreprise concernée.

Le turnover représente le taux de renouvellement du personnel, généralement établis par âge et par CSP.

Ici, dans le cas d'évaluations d'engagements sociaux, le turnover désigne en réalité une probabilité de sortie (démission, et parfois aussi licenciement, rupture conventionnelle), on ne tient pas compte d'une probabilité d'entrée puisqu'on l'applique à quelqu'un qui est déjà dans l'entreprise.

Les tables de turnover utilisées sont bâties sur l'expérience dès que cela est possible : effectifs suffisants et recul sur les années antérieures disponible.

Le taux d'actualisation, d'après la norme IAS 39, doit être choisi « par référence à un taux de rendement, à la fin de la période de présentation de l'information financière, du marché des obligations de sociétés de haute qualité ».

Celui-ci correspond à la prise en compte par avance d'une partie des revenus financiers futurs obtenus sur la gestion des provisions.

Le taux de revalorisation des salaires doivent refléter la politique salariale de l'entreprise à long terme, ils sont généralement supérieurs à l'inflation long terme.

Par la suite, l'outil va calculer pour chaque individu tous les éléments nécessaires à la valorisation des engagements.

Pour la méthode IFRIC, il est nécessaire de faire des calculs intermédiaires afin de savoir quand sera atteint le dernier palier de barème possible pour l'individu et ainsi calculer une provision que si nécessaire.

Par exemple, là où l'entreprise pour un individu de 27 ans avec 2 ans d'ancienneté va commencer à comptabiliser ses engagements envers ce salarié, avec la méthode IFRIC, elle ne commencera à ne provisionner que dans 8 ans.

L'outil va calculer selon la table de mortalité et de turnover des probabilités de survie au terme et de présence au terme (des $L(x)$ dans le jargon actuariel).

Puis, il va utiliser ces probabilités ainsi que le facteur d'actualisation, une durée résiduelle et le montant de l'IFC calculé en fonction du profil pour déduire une Valeur actuelle probable (VAP N) (avec et sans charge sociale comprise) pour chaque individu de manière séquentielle.

Puis l'outil en déduira les différents éléments de calcul (SC, IC, DBO N, DBO N+1) et compilera le tout pour connaître la dette actuarielle totale de la base.

12.5.2 Périmètre

Hypothèses retenues au 31/12/2021	
<i>Date d'évaluation</i>	31/12/2021
<i>Taux d'actualisation</i>	1,00%
<i>Taux d'inflation</i>	2,00%
<i>Profil de carrière</i>	3,00%
<i>Taux de charges sociales</i>	50,00%
<i>Table de mortalité</i>	INSEE 2015 - 2017 - de 35 ans : 8% entre 35 ans et 45 ans : 4% entre 45 ans et 55 ans : 2% entre 55 ans et 60 ans : 1% 60 ans et + : 0%
<i>Turnover</i>	
<i>Âge de départ à la retraite</i>	63 ans
<i>Âge de début de carrière</i>	22 ans

Droits IFC en mois de salaire	
Ancienneté	Palier
[0 an ; 2 ans [0,00
[2 ans ; 5 ans [0,50
[5 ans ; 10 ans [1,00
[10 ans ; 20 ans [2,00
[20 ans ; 30 ans [3,00
[30 ans ; 40 ans [5,00
[40 ans ; 50 ans [6,00

FIGURE 12.4 – Barème appliqué

Une première idée pourrait consister à étudier différentes façons de revaloriser les individus en fonction de l'âge et analyser l'impact sur la dette actuarielle à l'année N ; cependant, ce qui est important dans le calcul d'une IFC est le salaire en fin de carrière ainsi que la revalorisation soit instantanée ou bien étalée sur 5 ans importe peu.

Ici, sont revalorisés 82 individus à hauteur d'une dizaine de milliers d'euros ; les individus ont été choisis suite à une analyse plus profonde de l'arbre vu en première partie.

12.5.3 Evolution de la dette après rattrapage



FIGURE 12.5 – Dette actuarielle en année N pour les individus discriminés

On observe que dépendamment de la méthode comptable retenue, la dette actuarielle en année N varie. Comme attendu, comme la méthode préconisée par l'IFRIC prévoit un étalement des droits uniquement sur un certain nombre d'années précédant le départ à la retraite, la dette diminue comme la moyenne des individus discriminés est de 44 ans. La revalorisation augmente le salaire en fin de carrière et donc vient augmenter l'engagement de l'entreprise envers le salarié.

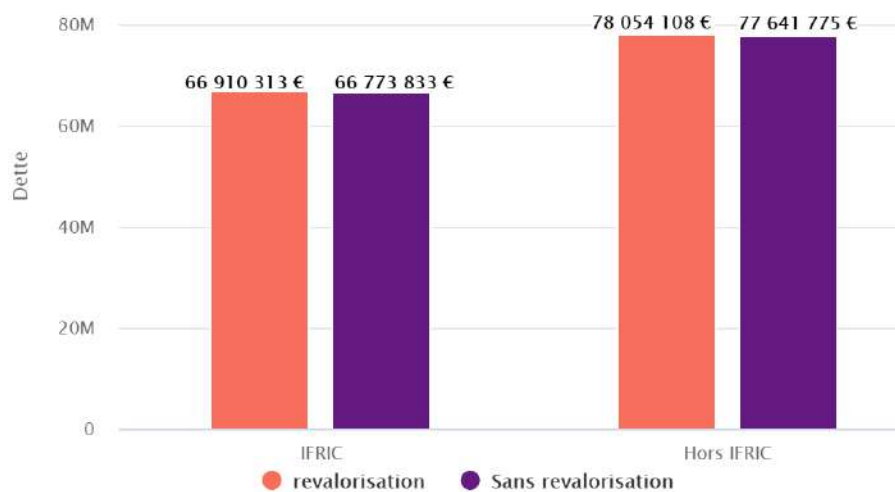


FIGURE 12.6 – Dette actuarielle totale de tous les individus

Le montant de la dette actuarielle est impacté par le changement de méthode comptable, en effet étant donné que les individus sont assez jeunes dans l'entreprise, l'impact sera d'autant plus grand. Le changement de méthode permet de réduire de 14% la dette. Comme le nombre total d'individus est important et celui des individus revalorisés faible en proportion, cette reprise de salaire a une faible répercussion sur la dette totale. On observe une variation de 20 points de base concernant l'IFRIC contre 53 points de base en méthode comptable A. Comme évoqué précédemment, une revalorisation en méthode comptable IFRIC aura moins d'importance sur la dette actuarielle à l'année N. Bien que le salaire soit revalorisé, le provisionnement pour l'individu n'aura pas encore débuté.

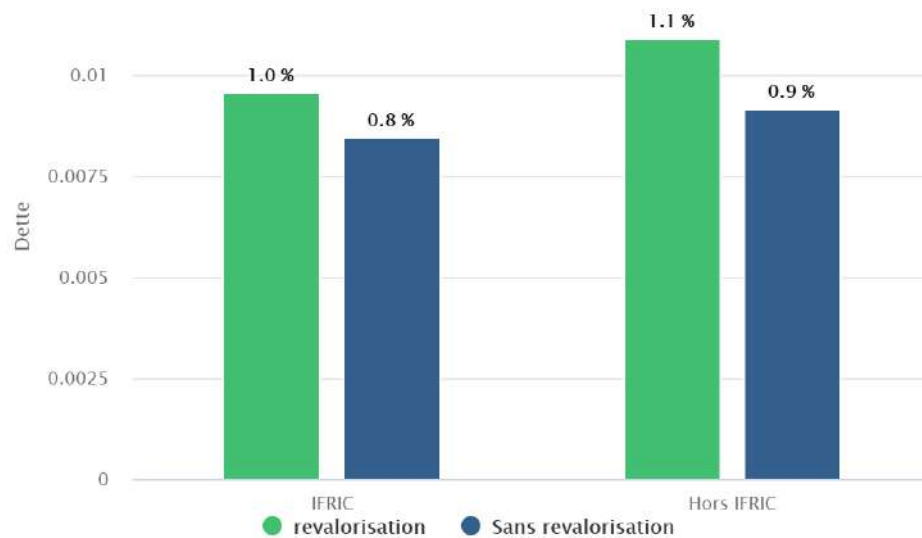


FIGURE 12.7 – Pourcentage que la dette vis à vis de la dette totale

La dette actuarielle des individus discriminés est logiquement plus importante dans la méthode comptable A des comptes sociaux. La part de la dette actuarielle des individus discriminés varie de 10 points de base entre les deux méthodes. L'impact de la revalorisation sur la dette totale est de 20 points de base. Ceci signifie que le montant de la dette des individus discriminés par rapport à la dette totale a augmenté de 0.2%. On observe une différence de 30 points de base entre aucune revalorisation en IFRIC par rapport à une revalorisation comptabilisé dans la méthode A.

Visuellement cela se représente ainsi :

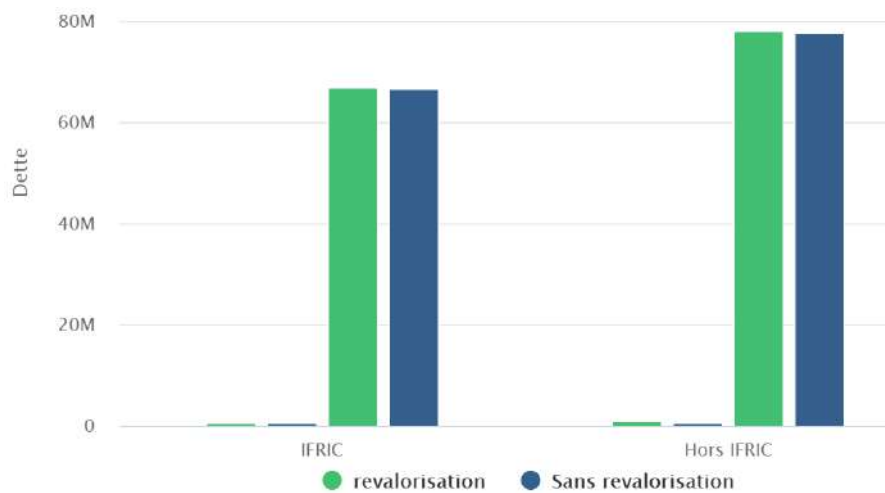


FIGURE 12.8 – Proportion de la dette des individus discriminés

L'index de l'égalité professionnelle se veut généralisable à toutes les entreprises, il convient donc d'étudier les méthodes comptables existantes pour ces différentes structures, puisque le discours pourrait être différent en fonction de l'acteur économique. L'impact des méthodes choisies a un impact très divers sur la dette à l'instant N pour l'entreprise.

Ainsi, si l'entreprise établit ses comptes selon une méthode comptable plutôt qu'une autre l'impact des revalorisations salariales en sera affecté.

Une campagne de revalorisation ne va pas risquer de menacer l'entreprise, étant donné le faible nombre d'individus concernés et le montant de la revalorisation mais elle pourrait être plus significative si la structure concernée est plus petite.

Dans la partie suivante, l'impact sur le montant de retraite obtenue sera discuté.

12.6 Les pensions de retraite en France

Comme expliqué en première partie : « L'écart non négligeable entre le montant des pensions des retraités et des retraitées s'explique par divers facteurs. Pour commencer, les femmes occupent davantage d'emplois précaires. Elles occupent plus rarement des postes de cadre. Des emplois peu qualifiés, avec une rémunération relativement faible, donnant droit à une pension de retraite basse. »

A poste égal, toutes choses égales par ailleurs, les femmes peuvent être moins bien payées que les hommes. De plus, les femmes ont aussi des parcours professionnels fragmentés, notamment par la vie de famille (temps partiel, congé maternité ..) impliquant souvent des temps réduits de travail.

Ces arrêts ont un impact sur la pension de retraite car la durée de cotisation sera moindre.

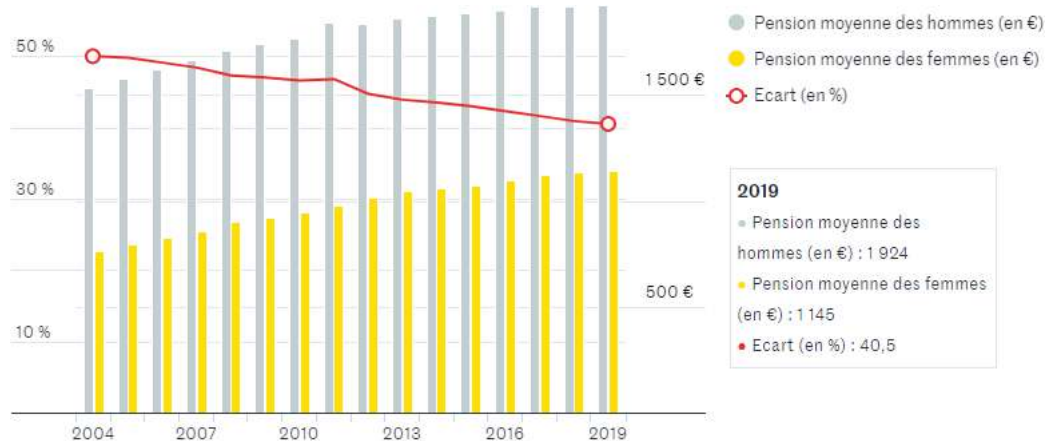


FIGURE 12.9 – Pension en France au fil des années

12.6.1 Retraite Obligatoire

Le calcul de la retraite obligatoire est le suivant, avec Taux $\in [37,5\%, 50\%]$:

$$\text{Moyenne des 25 meilleures années} * \text{Taux} * \frac{\text{Trimestres inscrits sur le relevé de carrière}}{\text{Nombre de trimestres qui varie selon l'année de naissance}}$$

Le montant de la pension de retraite de base versée par le régime général de la Sécurité sociale ne peut pas dépasser 50% du plafond de la Sécurité sociale applicable l'année du départ à la retraite.

Dans le cas de la base de données étudiée, les salaires étant importants le montant de retraite obligatoire atteint le plafond de la Sécurité sociale. Les salariés sont susceptibles de recevoir une retraite supplémentaire contractée par l'entreprise qu'elle devra provisionner.

12.6.2 Retraite Supplémentaire

La retraite supplémentaire complète la pension perçue au travers des régimes obligatoires de la Caisse Nationale d'Assurance Vieillesse (CNAV) et de l'Agirc-Arrco. Contrairement à la retraite complémentaire qui fait partie des régimes obligatoires, elle est facultative.

Les régimes de retraite supplémentaires peuvent être classés dans deux catégories : les régimes à cotisations définies et les régimes à prestations définies.

Les régimes à prestations définies (articles 39) sont des contrats qui garantissent au salarié le versement au départ en retraite d'un montant de prestation déterminé à l'avance en pourcentage du **dernier salaire versé** (ce qui motive l'étude des régimes à prestations définies dans le cadre de potentielles discriminations salariales).

A titre d'exemple, 4% du dernier salaire par année d'ancienneté.

Ils peuvent être *additifs* dans le cas où la rente s'ajoute aux pensions des régimes obligatoires. Ces régimes garantissent un pourcentage fixe du dernier salaire, indépendamment des autres régimes de retraite. Ces régimes ont un poids sur le passif social indépendant des rendements des régimes obligatoires.

Exemple : 5% du dernier salaire lors du départ à la retraite.

Ou bien *différentiels* (ou "chapeau") si la rente correspond à la différence entre un objectif global de retraite et des droits constitués par ailleurs (régimes obligatoires, régimes à cotisation définies). Ces régimes garantissent au bénéficiaire, au moment de sa mise à la retraite, un niveau global de revenus qui inclut le régime de base et les régimes complémentaires de retraite. Ces régimes ont une forte dépendance à l'évolution du rendement des régimes obligatoires.

Exemple : 50% du dernier salaire, y compris les retraites acquises par ailleurs.

12.6.3 Impact sur les passifs sociaux

L'impact sur les passifs sociaux pourrait être bien plus important qu'un régime IFC dans le cas de régimes à prestations définies octroyant des droits plus élevés.

Ces engagements sociaux pouvant être très élevés, ils conduisent les entreprises à les remplacer par des régimes à cotisations définies avec, éventuellement, un régime différentiel à prestations définies (régime transitoire).

12.7 Conclusion

Il a été vu dans ce chapitre combien les impacts de revalorisations salariales peuvent être importants relativement au choix des méthodes d'évaluation des engagements. Ces choix sont généralement encadrés et dépendent de la structure de l'entreprise (présence de filiale, entreprise située en France...).

Ces impacts sont d'autant plus faibles que la proportion de salariés est faible.

Cependant, il est nécessaire de prendre en considération que le type d'avantages sociaux (ici postérieurs à l'emploi) peut avoir une grande influence sur les passifs sociaux de l'entreprise comme c'est le cas pour les régimes à prestations définies par rapport à des IFC.

Conclusion

Les nouvelles évolutions réglementaires constituent de nouveaux challenges pour le monde de l'actuariat et la data science ; du suivi des indicateurs prévu par la loi, à la mise en place de nouveaux indicateurs et alertes, le rôle de l'actuaire peut être profitable compte tenu de l'outillage dont il dispose pour exploiter les données disponibles.

Ce mémoire avait pour ambition de présenter des solutions à destination des RH dans le but de prévenir les risques sociaux liés aux potentielles discriminations entre les femmes et les hommes. L'idée était d'aller plus en profondeur, à une maille plus fine, par rapport à ce que propose la réglementation actuelle en terme de l'égalité professionnelle entre les sexes en proposant une méthodologie de type modèle interne à une entreprise.

A travers les outils de Data Science que les sont le CART, GLM, XGboost, les différents indicateurs sur l'égalité entre les sexes étaient étudiés en profondeur afin de soumettre aux RH différents collaborateurs qu'il conviendrait d'analyser de leur côté. Les outils que constituent l'apprentissage automatique motivaient également la construction de différents indicateurs théoriques (rémunération théorique par individu, budget théorique) dans le but de venir fortifier les moyens des ressources humaines des entreprises.

Bien qu'il ne soit plus à prouver la puissance des modèles d'apprentissage automatique il est aujourd'hui plus que jamais essentiel d'être en mesure de justifier les décisions qu'entreprennent ces algorithmes. Une dimension éthique doit également être intégrée aux processus d'explicabilité des modèles que pourraient utiliser les entreprises, notamment sur le sujet des discriminations des genres. Ce mémoire présente des outils techniques permettant de révéler la présence de biais discriminatoires induits par les données, favorisant la prédiction des hautes rémunérations aux hommes par les modèles.

Afin de pousser plus loin les réflexions, une étude sur les impacts des rattrapages salariaux sur les passifs sociaux des entreprises a été menée, c'était l'occasion de faire un point sur les méthodes d'évaluations actuarielles des engagements sociaux qui concernent le provisionnement des différents types d'entreprises.

Bibliographie

- [1] R. K. E. BELLAMY, K. DEY, M. HIND, S. C. HOFFMAN, S. HOUDE, K. KANNAN, P. LOHIA, J. MARTINO, S. MEHTA, A. MOJSILOVIC, S. NAGAR, K. N. RAMAMURTHY, J. RICHARDS, D. SAHA, P. SATTIGERI, M. SINGH, K. R. VARSHNEY ET Y. ZHANG, *AI Fairness 360 : An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias*, oct. 2018.
- [2] T. CHEN ET C. GUESTRIN, *Xgboost : A scalable tree boosting system*, in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, p. 785–794.
- [3] DARES, *Ségrégation professionnelle et écarts de salaires femmes-hommes*. <https://dares.travail-emploi.gouv.fr/sites/default/files/pdf/2015-082.pdf>, 11/2015.
- [4] D. DELCAILLAU, A. LY, F. VERMET ET A. PAPP, *Interprétabilité des modèles : état des lieux des méthodes et application à l'assurance*. <https://arxiv.org/pdf/2007.12919.pdf>, 2020.
- [5] S. DIDO, *Les risques psychosociaux : impact sur l'absentéisme et prévention en entreprise*, 2020.
- [6] S. FARKAS, A. HERANVAL, O. LOPEZ ET M. THOMAS, *Generalized Pareto Regression Trees for extreme events analysis*. working paper or preprint, déc. 2021.
- [7] FFA, *Cartographie des risques émergents*. <https://www.apref.org/wp-content/uploads/2022/01/Cartographie-des-risques-France-assureurs.pdf>, 2022.
- [8] M. HARDT, E. PRICE ET N. SREBRO, *Equality of opportunity in supervised learning*, CoRR, abs/1610.02413 (2016).
- [9] INSEE, *Temps partiel*. https://www.insee.fr/fr/statistiques/fichier/4501614/ECRT2020_F4-2.pdf, 2020.
- [10] J. KUNST, *highcharter : A Wrapper for the 'Highcharts' Library*, 2022. <https://jkunst.com/highcharter/>, <https://github.com/jbkunst/highcharter>.
- [11] A. LY, *Algorithmes de machine learning en assurance : solvabilité, textmining, anonymisation et transparence*, theses, Université Paris-Est, nov. 2019.
- [12] T. MILLER, *Explanation in artificial intelligence : Insights from the social sciences*, Artificial Intelligence, 267 (2019), p. 1–38.

-
- [13] W. J. MURDOCH, C. SINGH, K. KUMBIER, R. ABBASI-ASL ET B. YU, *Definitions, methods, and applications in interpretable machine learning*, Proceedings of the National Academy of Sciences, 116 (2019), p. 22071–22080.
- [14] J. VELTHOEN, C. DOMBRY, J.-J. CAI ET S. ENGELKE, *Gradient boosting for extreme quantile regression*, (2021).
- [15] J. VELTHOEN, C. DOMBRY, J.-J. CAI ET S. ENGELKE, *Gradient boosting for extreme quantile regression*, 2021.