



# **Prédiction des compétences émergentes par analyse textuelle**

**Mémoire**

**Charles Roy**

**Maîtrise en économique - avec mémoire**  
Maître ès arts (M.A.)

Québec, Canada

# **Prédiction des compétences émergentes par analyse textuelle**

**Mémoire**

**Charles Roy**

Sous la direction de:

Luc Bissonnette, directeur de recherche  
Marion Goussé, co-directrice de recherche

# Résumé

Dans ce mémoire, l'objectif est de prédire les compétences émergentes en utilisant l'information contenue dans les offres d'emploi. Nous combinons le *Least Absolute Shrinkage and Selection Operator* (LASSO) et l'analyse de texte pour prédire les compétences émergentes. Pour y arriver, nous utilisons la description d'offres d'emploi dans le domaine des animateurs pour le cinéma et l'année de publication des offres. Les offres d'emploi utilisées s'étendent de 2014 à 2021 et proviennent d'un site spécialisé dans le domaine pour les régions de Québec et de Montréal. L'algorithme est en mesure de prédire les compétences émergentes et les compétences moins fréquemment demandées par les employeurs. Nous obtenons respectivement le logiciel Maya et le logiciel Nuke comme compétence émergente et compétence de moins en moins demandée par les employeurs.

# Table des matières

<b>Résumé</b>	<b>iii</b>
<b>Table des matières</b>	<b>iv</b>
<b>Liste des tableaux</b>	<b>v</b>
<b>Liste des figures</b>	<b>vi</b>
<b>Remerciements</b>	<b>vii</b>
<b>Introduction</b>	<b>1</b>
<b>1 Revue de Littérature</b>	<b>3</b>
1.1 Transformation du marché du travail . . . . .	3
1.2 Intelligence artificielle . . . . .	5
1.3 Traitement automatique des langues . . . . .	6
<b>2 Méthodologie</b>	<b>8</b>
2.1 Modèles . . . . .	8
2.2 <i>Least Absolute Shrinkage and Selection Operator (LASSO)</i> . . . . .	9
<b>3 Données</b>	<b>11</b>
3.1 Description des données . . . . .	11
3.2 Manipulations sur les données . . . . .	13
3.3 Analyse des mots fréquents . . . . .	17
<b>4 Résultats</b>	<b>21</b>
4.1 Méthode du LASSO . . . . .	21
4.2 Méthode des moindres carrés . . . . .	28
4.3 Quelques jetons surprenants . . . . .	32
4.4 Limites . . . . .	33
<b>Conclusion</b>	<b>34</b>
<b>Bibliographie</b>	<b>35</b>
<b>A Version longue des tableaux</b>	<b>37</b>
A.1 LASSO . . . . .	37
A.2 Moindres carrés ordinaires . . . . .	49

# Liste des tableaux

3.1	Représentation des données . . . . .	12
3.2	Représentation des données après manipulations . . . . .	16
3.3	Compétences relationnelles . . . . .	18
3.4	Compétences techniques . . . . .	18
4.1	Prédiction par LASSO en utilisant l'année comme variable dépendante avec les compétences techniques et relationnelles en gras . . . . .	22
4.2	Prédiction par LASSO en utilisant l'année comme indicatrice avec les compétences techniques et relationnelles en gras . . . . .	25
4.3	Tableau synthèse du LASSO . . . . .	28
4.4	Tableau synthèse des MCO . . . . .	29
4.5	Estimation par moindres carrés ordinaires en utilisant l'année comme variable dépendante avec les compétences techniques et relationnelles en gras . . . . .	29
4.6	Estimation par moindres carrés ordinaires en utilisant l'année comme indicatrice avec les compétences techniques et relationnelles en gras . . . . .	31

# Liste des figures

3.1	Distribution de la variable année . . . . .	13
3.2	Diagramme de flux du traitement de la matrice $X$ . . . . .	14
3.3	Unigrammes . . . . .	15
3.4	Bigrammes . . . . .	15

# Remerciements

J'aimerais remercier mon directeur de mémoire, Luc Bissonnette, pour m'avoir donné la chance de participer à ce projet très formateur, pour avoir été présent lors de tous les revirements de situation de ce projet ainsi que pour son soutien financier. J'aimerais également remercier la codirectrice de ce projet, Marion Goussé, pour ses commentaires très instructifs. Merci pour votre temps, je me considère très chanceux.

J'aimerais aussi remercier mes parents pour m'avoir supporté dans mes études en ne me mettant jamais de pression. J'aimerais aussi remercier ma copine, qui m'a accompagné tout au long de mon parcours universitaire. Un merci à mes amis et ma famille, avec qui je peux décrocher et délirer durant toutes les conversations.

Finalement, j'aimerais remercier le Ministère de l'Éducation et de l'Enseignement supérieur (MEES) qui a permis à ce projet d'exister.

# Introduction

Le marché du travail est un environnement en constante évolution. Qu'on parle de l'arrivée de la machine à vapeur, de l'implantation de l'électricité ou de l'ordinateur dans les industries, les avancées technologiques changent constamment les tâches des travailleurs. Selon un sondage réalisé conjointement par les entreprises Capgemini et LinkedIn en 2017, 29% des répondants jugeaient que leurs compétences sont ou seront obsolètes dans un horizon de 2 ans (Capgemini (2017)). Les compétences en question diffèrent selon l'industrie et dictent les tendances du marché. Notamment, des logiciels et autres technologies peuvent représenter des compétences recherchées par les employeurs. Plus que jamais, les acteurs économiques de toutes les industries doivent s'adapter individuellement au marché du travail afin de rester compétitifs. En effet, ce sont les compétences émergentes qui déterminent la compétitivité des travailleurs au moment de postuler pour un emploi. De plus, certaines industries nécessitent des compétences techniques précises que doivent posséder les travailleurs en recherche d'un emploi. Le domaine des animateurs pour le cinéma en est un bon exemple, puisque les firmes d'animation exigent un esprit créatif et la maîtrise de plusieurs logiciels auprès de leurs travailleurs. Ce papier utilise une combinaison de méthodes économétriques et d'analyse de texte afin de déterminer les compétences émergentes du domaine des animateurs pour le cinéma en analysant des offres d'emploi.

Les offres d'emploi dénombrent les différentes tâches requises pour le travail demandé, ce qui nous permet de constater en temps réel les différents besoins des employeurs. En effet, l'information contenue dans une offre d'emploi représente exactement ce qu'une compagnie cherche en temps réel. L'utilisation des offres d'emploi nous permet d'extraire plus d'information sur le marché du travail qu'en utilisant les indicateurs traditionnels tels que le taux de chômage, le salaire, le taux d'emploi, etc. Établir des techniques efficaces pour analyser des offres d'emploi peut avoir des conséquences importantes pour les employeurs, mais aussi pour les institutions gouvernementales. En partenariat avec le Ministère de l'Éducation et de l'Enseignement supérieur (MEES), nous voulons établir une procédure permettant de tenir les différents acteurs à jour. En effet, un programme scolaire efficace doit cibler les compétences émergentes sur le marché du travail. La gestion des programmes scolaires nécessite la compréhension des différents milieux économiques. Ne connaissant pas la méthode employée par le MEES pour l'adaptation des programmes, notre but n'est pas de la critiquer,

mais plutôt de faciliter le travail de celui-ci en proposant une méthode automatisée qui cible les compétences émergentes.

Pour atteindre ce but, nous utilisons des techniques de traitement automatique des langues afin de déterminer les compétences émergentes des animateurs pour le cinéma en utilisant la description des offres d'emploi et la date de publication. La description regroupe toute l'information concernant les tâches qui devront être effectuées ainsi que les avantages de travailler à cette compagnie. Cette information devra être convertie en format compréhensible pour un algorithme d'apprentissage automatique. Nous utilisons des modèles qui exploitent la présence des mots dans les offres d'emploi pour déterminer l'émergence des compétences. Une compétence est considérée émergente si sa présence signale une compétence récente.

Nous utilisons un échantillon d'offres d'emploi qui concerne le marché des animateurs pour le cinéma au Québec. L'échantillon contient 456 offres publiées entre 2014 et 2021 qui ont été récoltées sur internet. Malgré une base de données peu volumineuse, nous réussissons tout de même à tirer des conclusions intéressantes. Nous spécifions deux modèles distincts : un modèle où la variable dépendante est déterminée comme l'année de publication de l'annonce et un autre où la variable dépendante est une variable binaire indiquant si l'année de publication est supérieure à 2019. Nos résultats identifient des compétences émergentes dans les offres d'emploi. Bien que nous ne pouvons pas spécifier si un modèle est plus performant que l'autre, les tendances prédites par les deux modèles sont semblables.

La suite du document est organisée dans l'ordre suivant. Au chapitre 2, une revue de la littérature est effectuée. Le chapitre 3 porte sur les données. Le chapitre 4 explique la méthodologie du document et le chapitre 5 présente les résultats. Finalement, nous concluons.

# Chapitre 1

## Revue de Littérature

Dans cette section, nous discutons principalement de trois domaines de littérature : le marché de l'emploi, l'apprentissage automatique (ML) et l'analyse textuelle. Le marché de l'emploi est une composante importante de notre travail, car nous analysons les compétences émergentes présentes dans les offres d'emploi. Bien connaître l'évolution de ce marché est essentiel dans l'analyse textuelle qui sera effectuée. L'analyse du marché du travail par l'apprentissage automatique est pertinente dans le monde d'aujourd'hui. Les ordinateurs ont notamment révolutionné le monde de l'information. Tous les jours, une quantité phénoménale d'information est récoltée sous une panoplie de formats : texte, image et vidéo. Ces données ne peuvent pas toutes être traitées par des méthodes classiques et regorgent d'information jamais étudiée. Les techniques standards économétriques ne sont pas assez efficaces pour traiter cette quantité de données, alors nous devons avoir recours à des techniques d'apprentissage automatique qui seront discutées dans les prochaines sous-sections.

### 1.1 Transformation du marché du travail

Suivre l'évolution du marché du travail est essentiel pour garder la main d'oeuvre compétente et concurrentielle. Cependant, une étude publiée en 2016 par l'OCDE (Rimini et Spiezia (2016)) montre que 55% des adultes au Canada possédaient peu ou pas de compétences informatiques, ou juste assez pour effectuer des tâches simples. Également, elle estime que les métiers de plongeur et de cuisinier sont les deux seuls métiers qui ne nécessitent aucune compétence en informatique. L'arrivée de l'ordinateur sur le marché du travail a entraîné des anomalies sur ce marché. D'ailleurs, Deming et Noray (2020) montrent un niveau élevé de changement dans les compétences recherchées par les firmes. En effet, plusieurs compétences recherchées en 2019 n'existaient même pas en 2017. Certains secteurs sont plus touchés que d'autres, mais en moyenne, pour des compagnies semblables, les offres analysées contenaient environ 29% de nouvelles compétences en 2019. Les professions les plus durement affectées sont reliées à l'informatique et aux mathématiques avec 47% des offres

d'emploi. Les professions les moins touchées proviennent des domaines de l'éducation, du droit, des services sociaux et communautaires avec moins de 20% des offres d'emploi. Ce phénomène est dû aux changements dans les compétences nécessaires pour effectuer un travail (*upskilling*). Les travailleurs doivent posséder des compétences à jour afin de suivre le marché dans son évolution. Un sondage effectué en 2017 aux États-Unis par [Truck et al. \(2017\)](#) indique que près de 60% des employeurs rapportent avoir affiché des offres d'emploi pendant 12 semaines et plus. Les entreprises évaluent cette perte à plus de 800 000\$ par année en moyenne. [Şahin et al. \(2014\)](#) argumentent que le taux de chômage aux États-Unis est en partie causé par la recherche d'emploi dans un mauvais secteur, ce qu'ils attribuent au « *mismatch unemployment* ». Ce phénomène peut se retrouver au sein d'une profession, par exemple le domaine des animateurs pour le cinéma.

Depuis quelques années, des chercheurs ont constaté une nouvelle tendance : la polarisation des emplois par le niveau de compétence. [Acemoglu \(1999\)](#) constate le changement dans la composition de l'emploi aux États-Unis. Quelques années plus tard, [Goos et al. \(2009\)](#) ont montré qu'en Europe le phénomène de la polarisation était tout aussi présent. La polarisation des emplois est caractérisée par l'augmentation du taux d'emploi dans les emplois se trouvant aux extrémités de la courbe de compétence, c'est-à-dire que la croissance des emplois nécessitant des compétences faibles et élevées est plus importante que la croissance des emplois nécessitant des compétences moyennes. [Autor et Dorn \(2013\)](#) attribuent la majorité du changement dans la partie inférieure de la courbe aux métiers de services. Ils observent une augmentation de plus de 50% des heures travaillées pour les travailleurs ne possédant pas de diplômes universitaires entre 1980 et 2005.

Plusieurs professionnels du marché du travail soulèvent l'importance de la qualité des données pour l'analyse du marché du travail dans l'article *Toward understanding the impact of artificial intelligence on labor* ([Frank et al. \(2019\)](#)). Les données « conventionnelles » ne permettent pas de représenter l'évolution très rapide du marché du travail ([Frank et al. \(2019\)](#), [Atalay et al. \(2020\)](#)). La vélocité du marché du travail fait en sorte que les indicateurs traditionnels comme l'éducation, le salaire et le nombre d'heures travaillées ne sont pas des mesures précises pour étudier l'évolution des tendances dans le marché du travail ([Frank et al. \(2019\)](#)). Les auteurs suggèrent plutôt d'utiliser des données qui proviennent d'offres d'emploi, car celles-ci nous informent des changements dans le marché du travail et des compétences recherchées par les employeurs et ce, en temps réel. En effet, les emplois où les tâches ont changé représentaient environ la moitié de la croissance économique liée à l'emploi entre 1980 et 2015 ([Acemoglu et Restrepo \(2019\)](#)). Notamment, [Atalay et al. \(2020\)](#) utilisent les offres d'emploi ainsi que l'analyse de texte pour montrer que le changement se produit au sein des professions plutôt qu'au niveau du taux d'emploi chez les emplois routiniers et non routiniers. L'analyse de [Atalay et al. \(2020\)](#) est réalisable grâce aux nouveaux développements en intelligence artificielle. L'étude des offres d'emploi est propice à l'uti-

lisation de l'intelligence artificielle puisque les offres contiennent beaucoup d'informations qu'on peut combiner avec des techniques d'intelligence artificielle afin d'étudier le marché du travail.

## 1.2 Intelligence artificielle

À proprement parler, l'intelligence artificielle réfère au fait qu'un ordinateur soit capable d'apprendre et d'exécuter des tâches (Marr (2016)). L'apprentissage machine est l'application des méthodes d'intelligence artificielle dans l'objectif que la machine soit en mesure d'apprendre par elle-même (Marr (2016)). Ces techniques permettent le traitement de différents formats de données pour effectuer certaines tâches. La discipline de l'apprentissage machine est très active. Comme plusieurs domaines peuvent l'intégrer, elle se développe sur de multiples fronts. Nous voulons utiliser des techniques d'apprentissage automatique pour bonifier notre analyse du marché du travail. Pour y arriver, nous utilisons le traitement automatique des langues, qui est une branche de l'apprentissage automatique, pour effectuer une analyse de texte.

Dans un autre ordre d'idée, Agrawal *et al.* (2019) considèrent que nous devons distinguer les tâches de décisions des tâches de prédictions, car ces dernières n'ont pas de valeur sans les décisions. Ils argumentent que l'apprentissage machine effectue, dans certains cas, de meilleures prédictions que les humains, car nous n'avons pas la capacité d'analyser certains types de données. Ils ajoutent que les conséquences les plus probables de l'utilisation de l'intelligence artificielle dans le marché du travail toucheront les tâches existantes, qui seront affectées par des prédictions plus efficaces, plus rapides et moins coûteuses. On trouve des applications de l'apprentissage automatique dans une panoplie de contextes différents, mais pour la pertinence de ce document, nous allons nous concentrer sur son application en économique. Contrairement aux techniques économiques standard, l'apprentissage automatique s'intéresse aux prédictions plutôt qu'aux estimations. Il existe deux types principaux de prédictions : les modèles supervisés et les modèles non supervisés, voir par exemple Athey et Imbens (2019). Les modèles supervisés utilisent l'information des intrants  $X_i$  et des sortants  $Y_i$  provenant des observations, tandis que les modèles non supervisés utilisent seulement les intrants  $X_i$  (Athey et Imbens (2019)). Communément, en apprentissage machine, une partie de l'échantillon sert à tester le modèle tandis que l'autre partie est utilisée pour tester la performance du modèle hors-échantillon. De plus, les données massives jouent un rôle primordial dans l'émergence de l'apprentissage automatique puisque la quantité de données à traiter est trop importante pour les techniques économétriques standard (Varian (2014)). En effet, ces dernières ne sont pas réalisables en raison de la dimensionnalité des matrices (Gentzkow *et al.* (2019)). La combinaison des données massives aux techniques d'apprentissage artificiel permet à l'algorithme de trouver par lui-même des relations complexes généralisables qui performant bien hors-échantillon (Mullainathan et Spiess (2017)).

L'objectif du modèle supervisé est de soumettre une bonne prédiction hors échantillon. Pour ce faire, l'algorithme utilise le vecteur des observations  $y$  ainsi que le vecteur des observations prédites  $\hat{y}$  afin de déterminer la fonction de perte associée aux observations  $L(\hat{y}, y)$ . Ensuite, il tente de trouver une fonction  $\hat{f}$  qui minimise  $E_{(y,x)}[L(\hat{f}(x), y)]$  à l'aide d'une nouvelle observation provenant de la même distribution, où  $x$  est le vecteur des caractéristiques correspondant aux observations, comme stipulé par Mullainathan et Spiess (2017). Pour analyser les offres d'emploi et prédire la date de publication reliée à ces offres, nous utilisons une méthode supervisée d'évaluation de modèle linéaire pénalisé : Le *Least Absolute Shrinkage and Selection Operator* (LASSO), développé par Tibshirani (2016). Le LASSO pénalise très fortement à 0 les coefficients qui n'ont pas de pouvoir statistique, ce qui produit un modèle qui s'interprète facilement (Tibshirani (2016), Nowak et Smith (2017)). L'évaluation du modèle crée une matrice sparse, qui contient beaucoup d'éléments valant zéro. Le LASSO permet de sélectionner les variables pertinentes à la prédiction et d'effectuer l'estimation du modèle simultanément. L'ajout du terme pénalisant, qui est proportionnel à la somme des valeurs absolues des paramètres, permet d'être efficace au niveau du temps de calcul (Athey et Imbens (2019)) et de la sélection de variables.

Le LASSO, comme tous les modèles d'apprentissage automatique, doit être ajusté afin de bien prédire hors-échantillon et d'éviter une situation de sur-apprentissage. La technique standard utilisée en économétrie est l'utilisation du *K-fold cross-validation(CV)* (Varian (2014)). Cette technique génère plusieurs itérations du modèle avec différentes valeurs du paramètre que nous devons calibrer, pour finalement déterminer la meilleure valeur du paramètre pour notre modèle (Varian (2014)). Comme le LASSO exécute les sélections de variables et la prédiction en même temps et qu'il performe bien dans un contexte de grande dimension, il s'agit d'un outil idéal pour analyser le texte. Nowak et Smith (2017) utilisent le LASSO pour prédire le prix des maisons à l'aide de la section description des annonces de maisons. Pour y arriver, ils combinent l'intelligence artificielle et le traitement automatique des langues.

### 1.3 Traitement automatique des langues

L'analyse de texte des offres d'emploi nous permet de sortir des tendances du marché de l'emploi. Ces tendances sont importantes dans la mise en place de politiques (Frank et al. (2019)) et pour adapter les programmes des différentes institutions scolaires.

Le traitement automatique des langues est un processus qui permet à l'ordinateur de comprendre le texte. Par exemple, le modèle GPT-3<sup>1</sup>, un produit de OpenAI, permet de continuer une histoire selon la prémisse qu'on lui fournit. Le modèle prédit les prochains mots en se basant sur les mots déjà présents dans le texte. L'article *The Evolution of Work in the United States* (Atalay et al. (2020)), dont nous avons discuté précédemment, est un très bon exemple d'une

---

1. <https://openai.com/blog/better-language-models/>

utilisation puissante de l'analyse textuelle en économie. Les auteurs ont utilisé la fréquence des mots présents dans les offres d'emploi comme indicateur du changement dans les tâches de travail. Cette technique est plus adaptée pour des applications où la nature des données est un critère très important pour être en mesure de suivre l'évolution du marché du travail (Frank [et al.](#) (2019)). L'article de [Atalay et al.](#) (2020) permet d'obtenir une analyse à jour dans un marché très volatile. En effet, certaines classifications de professions sont modifiées tous les 5 ans, mais le marché du travail évolue plus rapidement. En effet, [Anani](#) (2018) argumente que les titres des postes et les systèmes de classifications des professions détiennent de moins en moins d'informations pertinentes concernant les compétences requises pour exécuter les postes.

Le texte peut représenter une source importante d'information et améliorer les prédictions. La discipline d'analyse de texte est relativement jeune en comparaison avec les méthodes classiques ([Gentzkow et al.](#) (2019)). [García](#) (2013) a effectué une analyse portant sur une section financière du *New York Times* et a montré que l'utilisation du texte améliorait la prédiction du rendement des actions, surtout en période de récession. Il utilisait une approche par dictionnaire en dénombrant la quantité de mots à connotation positive et négative dans tous les articles pour estimer le prix des actions. Toutefois, la méthode utilisée par [García](#) (2013) n'intégrait pas l'apprentissage automatique. [Nowak et Smith](#) (2017) ont utilisé l'approche de l'apprentissage automatique pour déterminer les mots pertinents à la prédiction du prix des maisons. Ils ont montré que l'erreur sur le prix pouvait être réduite de plus de 25% en intégrant le texte dans l'analyse.

L'analyse présente dans ce document utilise principalement les techniques décrites par [Nowak et Smith](#) (2017) et [Gentzkow et al.](#) (2019). Ces techniques sont appliquées dans un contexte différent, mais en utilisant une méthodologie semblable. Contrairement au prix des maisons, nous nous intéressons à la prédiction des compétences émergentes en utilisant la description des offres d'emploi.

## Chapitre 2

# Méthodologie

Dans la présente section, la méthode utilisée afin d’obtenir les résultats sera présentée. Pour mettre en place notre modèle, nous avons besoin des informations comprises dans la description des offres d’emploi, ainsi que de l’année de publication de ces offres. Les informations que nous voulons analyser sont les compétences techniques reliées aux emplois. Nous voulons déterminer si les offres d’emploi sont une bonne source pour analyser les tendances du marché du travail dans la détection de compétences émergentes. Pour y arriver, nous altérons la procédure proposée par [Nowak et Smith \(2017\)](#) et [Gentzkow et al. \(2019\)](#) en modifiant le modèle à prix hédonique pour qu’il estime l’année de publication des offres d’emploi.

### 2.1 Modèles

Le modèle hédonique décrit la variable dépendante comme une description de ses paramètres. Les coefficients des paramètres sont associés aux prix implicites qui sont révélés par le prix de la variable qui possède les attributs ([Rosen \(1974\)](#)). [Nowak et Smith \(2017\)](#) utilisent le modèle dans le but de prédire la valeur des maisons selon leurs caractéristiques. Les paramètres du modèle indiquent si les caractéristiques en question augmentent ou diminuent la valeur de la maison.

Cependant, les offres d’emploi mises à notre disposition ne contiennent pas assez de données sur le salaire pour reproduire la méthodologie de [Nowak et Smith \(2017\)](#). Dans notre cas, l’année de publication est la variable dépendante et nous cherchons à expliquer l’année de publication selon les compétences qui se retrouvent dans la section description des offres d’emploi. Nous utiliserons les modèles suivants :

$$Y_1 = X\beta_1 + \epsilon_1 \tag{2.1}$$

$$Y_2 = X\beta_2 + \epsilon_2 \tag{2.2}$$

Où

$Y_1$  : Année de publication de l'offre.

$Y_2$  : Indicatrice valant 1 si l'offre est publiée après 2019.

$X$  : Jetons qui sont présents dans les offres.

Les modèles présentés indiquent quelles sont les compétences émergentes et les compétences qui vont disparaître. La matrice  $X$  est une matrice sparse qui indique la présence des mots dans les offres. Nous discuterons de la création de cette matrice dans la section suivante. Les modèles s'interprètent comme un modèle linéaire standard. Un coefficient positif indique que la présence du mot dans une offre d'emploi augmente en moyenne l'année de publication de l'offre. De la même manière, un coefficient négatif indique que la présence du mot dans une offre d'emploi diminue en moyenne l'année de publication. Nowak et Smith (2017) enrichissent un modèle hédonique classique d'évaluation de propriété où on peut retrouver les caractéristiques communes (nombre de pièces, grandeur des pièces, etc.) par la section description de l'annonce de vente. Parallèlement, nous utilisons les différentes caractéristiques des offres d'emploi pour produire le vecteur  $X$ . Dans ces offres, on retrouve les langages de programmation et la connaissance des logiciels demandés par les employeurs, les compétences recherchées, la localisation de l'emploi, les avantages sociaux et toutes les informations que l'employeur souhaite transmettre aux futurs employés.

Le modèle 2.1 représente la variable dépendante comme l'année de publication des offres d'emploi, tandis que le modèle 2.2 représente la variable dépendante sous forme d'une indicatrice. Cette indicatrice vaut 1 lorsque la date de publication de l'offre d'emploi est supérieure à 2019 et 0 sinon.

## 2.2 *Least Absolute Shrinkage and Selection Operator (LASSO)*

La tendance dans la littérature économique en intelligence artificielle est d'utiliser la technique du LASSO pour ses caractéristiques très puissantes lorsque nous avons accès à une grande quantité d'observations. Le LASSO est une technique de régression pénalisée qui est très intéressante pour notre modèle. Premièrement, le LASSO est reconnu pour ses attributs de sélection de variables. Comme la procédure pénalise durement les coefficients à 0, on se retrouve avec une matrice sparse où les coefficients différents de zéro composent le modèle. Cette caractéristique du LASSO est utile lorsque le nombre de variables est grand. Théoriquement, un modèle de type LASSO n'est pas limité par des modèles de grande dimension avec plus de paramètres que d'observations, où une méthode des moindres carrés ordinaires classiques (MCO) est limitée (Gentzkow *et al.* (2019)). Dans notre cas, le nombre de variables explicatives demeurera inférieur au nombre d'observations, ce qui nous permettra de comparer la méthode classique par MCO avec la méthode de régression pénalisée du LASSO dans la section des résultats. Pour le modèle LASSO, nous cherchons à minimiser :

$$\arg \min_{\beta} \sum_{i=1}^N (Y_i - X_i \beta)^2 + \lambda \sum_{j=1}^J |\beta_j| \quad (2.3)$$

Dans l'équation du modèle, le paramètre  $\lambda$  doit être calibré. Il existe plusieurs techniques pour déterminer le paramètre  $\lambda$  optimal. Premièrement, le paramètre doit fournir des bonnes prédictions hors échantillon. Nous voulons éviter un modèle qui « sur-apprend » les données, soit un modèle qui prédit très bien les observations de l'échantillon, au dépens des observations hors échantillon. De plus, les paramètres engendrés par un  $\lambda$  trop élevé sont sévèrement pénalisés, ce qui signifie que la plupart des coefficients valent 0. À l'opposé, lorsque  $\lambda = 0$ , le modèle devient une estimation des moindres carrés ordinaires. Ainsi, le LASSO permet une sélection très efficace des paramètres pertinents. Dans la section 4, nous utilisons le cas où  $\lambda = 0$  pour comparer la méthode du LASSO avec la méthode des moindres carrés ordinaires.

La calibration du LASSO est exécutée par la technique du *K-fold cross validation (CV)*. Elle s'exécute avec la fonction *LassoCV* du paquet *sklearn* sur Python. En théorie, pour calibrer à l'aide du *K-fold cross validation*, nous devons séparer l'échantillon en  $K$  sous-groupes qui serviront à tester le modèle. Ensuite, nous choisissons la valeur du paramètre pour la première itération, puis avec les  $K - 1$  autres sous-groupes, nous estimons le modèle. Après chaque itération, nous calculons la perte associée à la prédiction en fonction du paramètre  $\lambda$ , puis nous recommençons les étapes précédentes avec un nouveau paramètre jusqu'à avoir utilisé tous les sous-groupes (Varian (2014)). Finalement, nous choisissons  $\lambda$  qui réduit l'erreur moyenne au carré de la procédure. Nous avons effectué un total de 5 validations croisées pour notre modèle.

# Chapitre 3

## Données

L'utilisation des offres d'emploi dans l'étude du marché du travail n'est pas une procédure couramment utilisée dans la littérature. La mise en place de notre modèle passe par le traitement automatique des langues. Nous utilisons l'analyse de texte pour récolter l'information nécessaire à notre prédiction.

### 3.1 Description des données

Ce document utilise des offres d'emploi dans le domaine des animateurs pour le cinéma. Les offres se concentrent sur l'industrie des régions de Montréal et de Québec. Nous avons utilisé la technique du *web scraping*<sup>1</sup>, qui consiste à extraire l'information brute d'une page internet et à la conserver sous forme de base de données. En récoltant l'information sur un domaine précis, nous connaissons la structure exacte du site. Ainsi, nous pouvons indiquer au programme d'aller récolter l'information recherchée à l'adresse exacte. Le tableau 3.1 présente les données. Nous avons récupéré 456 offres d'emploi différentes pour les périodes de 2014 à 2021. Pour chaque offre d'emploi, nous avons sa description en format brut ainsi que sa date de publication.

Pour les modèles économiques, nous avons besoin d'un vecteur  $y$  et d'une matrice  $X$  qui sera créée avec les mots. Nous utilisons l'année de publication pour le vecteur  $y$ . L'année de publication permet de déterminer si une compétence est émergente en exploitant la fréquence des compétences dans les offres d'emploi. Les données extraites avec le *web scraping* nécessitent quelques manipulations afin qu'elles soient conformes à notre algorithme. Nous voulons d'abord isoler l'année de publication de la suite de caractère : Publié le : jour mois année. Nous y arrivons facilement en retirant la séquence « Publié le : » dans toutes les offres d'emploi. L'utilisation des expressions régulières dicte à Python de retirer « Publié le : » de toutes les annonces pour se retrouver seulement avec le format : jour mois année. Avec la même technique, nous demandons à Python d'isoler l'année de publication. Finalement, nous nous

---

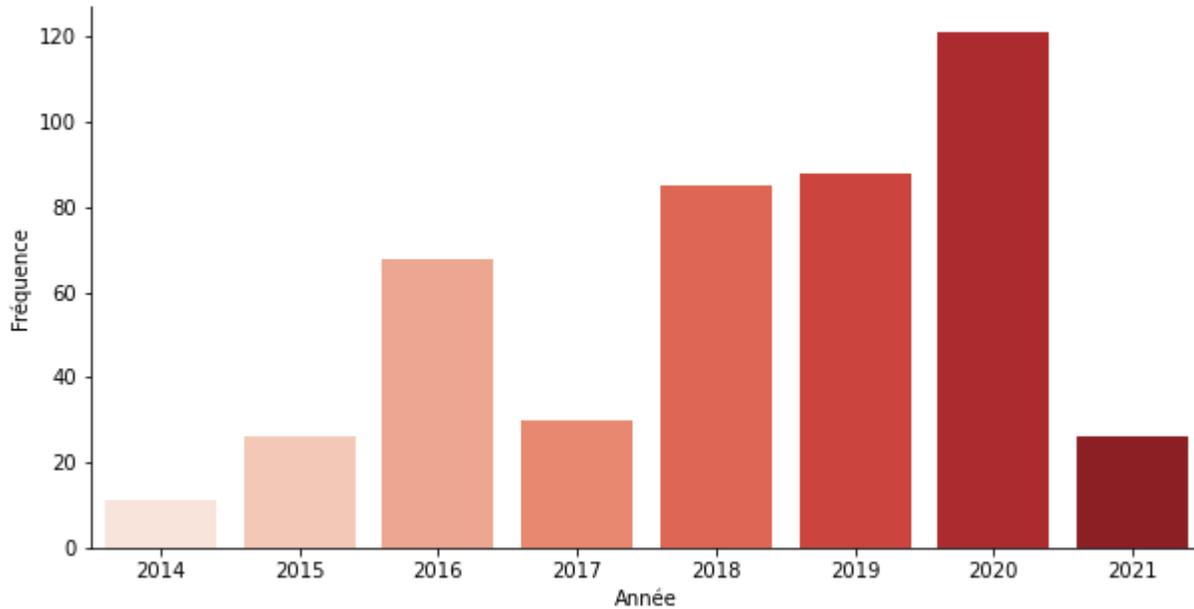
1. Conformément aux consignes du fichier robots.txt de la source

TABLE 3.1 – Représentation des données

Date de publication	Description
0 Publié le : 10 Décembre 2018	Sous la supervision du Lead R&D, le développeur pipeline est responsable de développer les bibliothèques backend nécessaire au développement d'outils par les développeurs intermédiaires et les TD.[...]
1 Publié le : 16 Mars 2018	Une chance unique de se joindre à l'équipe! Nous recherchons un IT débrouillard! Horaire de travail flexible et belle ambiance de travail! Nous priorisons l'embauche de citoyen canadien et résidents permanents.
2 Publié le : 05 Mai 2020	Nous sommes une équipe d'artistes passionnés, geeks à temps plein portés par un environnement sain et stimulant! Nous sélectionnons nos projets avec soin pour mettre notre savoir-faire au service de ce qui nous anime. Nous avons à cœur de maintenir un bel équilibre entre vie professionnelle et vie privée, d'encourager la curiosité, et de favoriser la communication afin que le travail devienne un vrai terrain de jeux coopératif.[...]
3 Publié le : 07 Décembre 2014	Une démo démontrant de fortes aptitudes d'éclairage 3D dans un environnement 3D et sur plusieurs sortes de matériaux 3D est essentielle. L'artiste doit aussi être en mesure de gérer son travail, être en mesure de travailler dans un contexte de production VFX et être en mesure d'assimiler les commentaires et la critique ainsi que de bien communiquer avec les membres de son équipe. L'artiste doit être en mesure de travailler sous pression et respecter les échéances. [...]
4 Publié le : 17 Août 2015	la/le coordonnatrice/eur est une personne ressource clé pour la production qui agit à titre de relais entre les artistes et la production. Effectue le suivi et la mise à jour des documents de suivi de projets; Tient à jour la liste des plans et l'état d'avancement de son département; En accord avec la direction de production et la supervision, fixe et suit les objectifs des artistes; Organise et assiste aux dailies/reviews, prends en note toutes les informations pertinentes et les transmet aux artistes;[...]

retrouvons avec le vecteur  $y$ . La distribution de publication des offres d'emploi est présentée par la figure 3.1.

FIGURE 3.1 – Distribution de la variable année



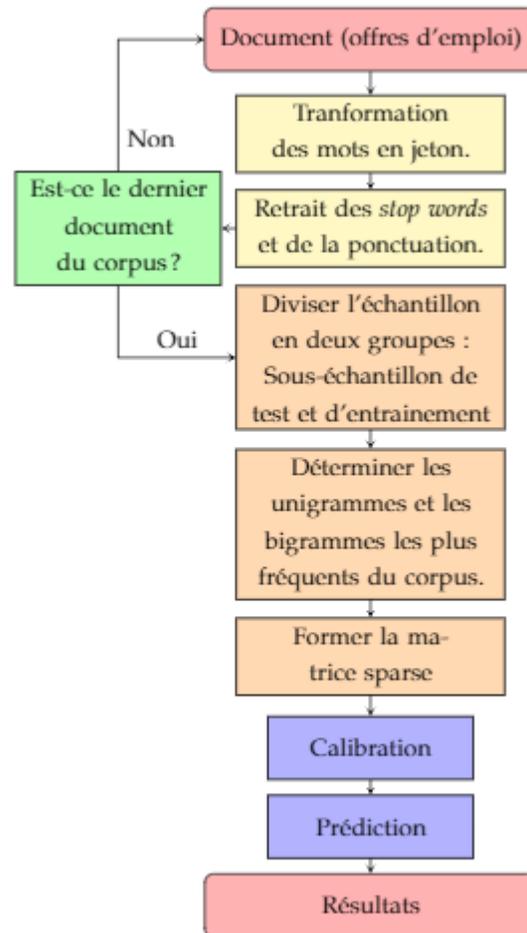
### 3.2 Manipulations sur les données

La matrice  $X$  est formée avec l'information contenue dans la description des offres d'emploi. On y retrouve la description des tâches à effectuer, les qualifications requises tant professionnelles que relationnelles, la description de l'entreprise, etc.. Cependant, la section description des offres d'emploi que nous avons récoltées est en format brut. Nous devons donc y appliquer des modifications, sinon l'algorithme ne fait aucune différence entre les mots et considère toute la description de l'annonce  $i$  comme une seule variable. L'objectif est que chaque mot ou locution d'au plus de 2 mots soit considérée comme une variable. Alors, la matrice indique la présence de chaque mot ou locution d'intérêt dans les offres d'emploi.

Toute l'information contenue dans les offres n'est pas pertinente. Les mots retirés et transformés diffèrent d'une analyse à l'autre et sont en lien avec le sujet. Par exemple, pour analyser l'ensemble d'une oeuvre artistique, nous allons conserver tous les mots qui nous informent de l'état d'âme de l'artiste. Par contre, notre but n'étant pas de faire de l'analyse de sentiment, nous voulons retirer tous les mots superflus qui ne sont pas pertinents pour notre prédiction. Il s'ensuit que nous effectuons les manipulations pour utiliser les mots des offres qui conviennent à notre contexte. Les différentes étapes du processus de prédiction sont représentées par la figure 3.2.

Premièrement, nous devons transformer les mots en jetons. Cette étape est essentielle pour

FIGURE 3.2 – Diagramme de flux du traitement de la matrice  $X$



l’algorithme. Elle permet à l’algorithme de distinguer les différents mots du texte. La mise en jetons consiste à transformer le bloc de texte brut en variables, où chacune représente un mot du texte brut. Les jetons sont des indicatrices représentant la présence des mots dans le texte. Le vecteur des jetons nous indique si un jeton est présent dans une offre et non le compte de ce jeton dans le texte. La procédure se fait à l’aide d’un paquet de code en source libre (*open source*) de traitement automatique des langues (NLP) avancé Spacy<sup>2</sup> qui est disponible avec Python. Nous utilisons ensuite le modèle français<sup>3</sup> disponible sur Spacy. Notre utilisation du paquet Spacy peut être résumé ainsi : nous lui fournissons le paquet d’un texte et il transforme tous les mots en objets de classe jeton. Ceux-ci possèdent ensuite tous les attributs nécessaires à la modification du texte. Le modèle français de Spacy comprend une liste des mots les plus utilisés appelés *stop words* qui n’ont, à priori, aucun pouvoir prédictif dans un contexte d’analyse d’offre d’emploi. On y retrouve, en autres, des prépositions, des

2. <https://spacy.io/>

3. Le modèle `fr_core_news_sm` est entraîné sur WikiNER et UD French Sequoia v2.5.



TABLE 3.2 – Représentation des données après manipulations

Année	Description après manipulations
0 2018	["supervision", "lead", "développeur", "pipelin", "responsable", "développer", "librairie", "back", "end", "nécessaire", "développement", "outil", "développeur", "intermédiaire", "td"...]
1 2018	["chance", "joindre", "équipe", "rechercher", "it", "débrouillard", "horaire", "travail", "flexible", "bel", "ambiance", "travail", "prioriser", "embauche", "citoyen", "canadien", "résident", "permanent"...]
2 2020	["être", "équipe", "artiste", "passionner", "geek", "temps", "porter", "environnement", "sain", "stimuler", "sélectionner", "projet", "soin", "mettre", "savoir", "faire", "service", "anim", "cœur", "maintenir", "bel", "équilibre", "vie", "professionnel", "vie", "priver", "encourager", "curiosité", "favoriser", "communication", "travail", "devenir", "vrai", "terrain", "jeu", "coopératif"...]
3 2014	["démon", "démontrer", "fort", "aptitude", "éclairage", "environnement", "sorte", "matériaux", "essentiel", "artiste", "mesure", "gérer", "travail", "mesure", "travailler", "contexte", "production", "vfx", "mesure", "assimiler", "commentaire", "critique", "communiquer", "membre", "équipe", "artiste", "mesure", "travailler", "pression", "respecter", "échéance"...]
4 2015	["coordonnateur", "eur", "ressource", "clé", "production", "agir", "titre", "relais", "artiste", "production", "effectue", "suivi", "mise", "jour", "document", "suivi", "projet", "tenir", "jour", "liste", "plan", "état", "avancemer", "département", "accord", "direction", "production", "supervision", "fixe", "objectif", "artiste", "organis", "assiste", "dailie", "review", "prendre", "note", "information", "pertinent", "transmet", "artiste"...]

Lorsque toutes les offres d'emplois ont été traitées, nous possédons enfin la structure nécessaire pour procéder aux prédictions. Toutes les observations sont épurées dans un format compréhensible pour l'algorithme, comme représenté dans le tableau 3.2.

Finalement, après avoir déterminé la liste des 200 éléments fréquents, nous pouvons former la matrice sparse où les colonnes correspondent aux unigrammes et aux bigrammes et les lignes correspondent aux mots présents dans les offres d'emploi. Nous devons noter qu'une colonne contenant la constante 1 a été ajoutée au modèle LASSO pour être en mesure de comparer ce modèle avec les MCO. La matrice peut être lue ainsi : l'élément  $ij$  de la matrice prend la valeur 1 si le mot  $j$  fait partie de l'offre  $i$ , et 0 sinon. Voici un exemple de la structure

de la matrice sparse  $X$  avec la description des offres du tableau 3.2. Nous avons fait le choix d'utiliser une matrice sparse binaire. Nous voulons déterminer la présence de l'élément dans les offres, contrairement à la fréquence de sa présence dans l'offre d'emploi.

	supervision	lead	passionner	ambiance	artiste	équipe	projet	outil
0	1	1	0	0	0	0	0	1
1	0	0	0	1	0	1	0	0
2	0	0	1	0	1	1	1	0
3	0	0	0	0	1	1	0	0
4	1	0	0	0	1	0	1	0

### 3.3 Analyse des mots fréquents

Cette section analyse les mots fréquents obtenus à l'aide de Spacy. Tout d'abord, nous avons organisé les 200 mots obtenus dans les tableaux 3.3, 3.4 et 3.5 représentant respectivement les compétences relationnelles, les compétences techniques et les mots qui n'ont pas de signification particulière. On peut y lire les mots fréquents ainsi que leur fréquence entre parenthèses. La fréquence est déterminée comme le nombre d'offres où le mot se retrouve au moins une fois. Ces tableaux faciliteront la compréhension des résultats au chapitre 4. Nous remarquons deux thèmes principaux dans le tableau 3.3 : la communication et la créativité. Ces résultats nous indiquent que, sans tenir compte de l'année de publication, les employeurs cherchent des employés avec ces critères. Sur un total de 456 offres d'emploi récoltées, le mot « artiste » est présent au moins une fois dans 68.64% des annonces et le mot « communication » est présent au moins une fois dans 44.95% des annonces. Les compétences relationnelles ne sont pas la cible de notre estimation, cependant elles peuvent nous informer sur la structure des données et sur l'état du marché du travail.

En faisant l'hypothèse que le coefficient lié au jeton « anglais » de la prédiction par LASSO est négatif, nous pouvons nous attendre à ce qu'une offre qui contient ce jeton ait été publiée avant une annonce qui ne contient pas ce jeton. Nous en tirons deux interprétations. Premièrement, la connaissance de l'anglais n'est plus une compétence à posséder afin d'être compétitif dans le marché du travail. Deuxièmement, la connaissance de l'anglais est devenue une compétence de base pour le poste affiché. Ces interprétations peuvent expliquer plusieurs jetons des compétences relationnelles, car elles ne suggèrent pas de tendances. Les compétences relationnelles recherchées par les compagnies sont ainsi stables.

Les compétences sont au coeur de notre travail. Le tableau 3.4 représente les différentes compétences techniques demandées dans les offres d'emploi. On y retrouve des logiciels tels que « Nuke », « Shotgun » et « Maya » ainsi que des compétences techniques générales telles que « Linux », « Computer Graphics » (CG), « rigging », « compositing », « pipeline », « Python »,

TABLE 3.3 – Compétences relationnelles

anglais (110)	artiste (313)	artistique (206)	communication (205)
communiquer (155)	créatif (179)	créativité (89)	français (122)
travailler équipe (99)			

TABLE 3.4 – Compétences techniques

cg (127)	compositing (86)	connaissance logiciel (98)	informatique (76)
linux (74)	logiciel (250)	logiciel maya (76)	maya (244)
nuke (116)	pipeline (177)	python (131)	résoudre problème (89)
rigging (83)	shotgun (119)	vfx (81)	

« vfx » et « informatique ». Nous voulons déterminer si ces compétences sont retenues dans la prédiction du LASSO et dans l'estimation par MCO. Le signe du coefficient associé à ces compétences pourra nous indiquer si cette compétence cherche à disparaître ou si elle est émergente.

Le tableau 3.5 représente les autres mots qui sont fréquents, mais qui ne fournissent pas d'information sur les compétences recherchées par les entreprises. Cependant, nous pouvons tirer des constats sur notre base de données. La présence des mots qui indiquent des postes nous informent sur le type de données que nous possédons. Par exemple, les jetons tels que « lead », « gestion », « gestionnaire », « producteur », « directeur », « superviseur », « réalisateur » et « chef » nous indiquent qui est concerné par ces postes. L'acteur économique concerné par des postes de gestion n'est pas à sa première année sur le marché du travail et doit posséder un grand nombre de compétences et des compétences différentes qu'une personne à sa sortie de l'école. De plus, nous pouvons faire des liens avec les interprétations précédemment énumérées. Par exemple, les postes de gestion requièrent probablement une maîtrise de la plupart des compétences relationnelles discutées. Donc, la deuxième interprétation du jeton « anglais » pourrait être valide en considérant la nature des postes, c'est-à-dire que la compétence « anglais » est devenue une compétence de base pour le poste affiché.

TABLE 3.5 – Mots sans compétences particulières

amélioration (75)	an (242)	an expérience (197)
animation (346)	animation studio (107)	année (128)
année expérience (111)	aptitude (85)	asset (102)
assurer (159)	atout (263)	baser (90)
bel (77)	besoin (122)	bon (309)
bon connaissance (188)	but (97)	but texte (81)
canada (91)	candidat (93)	candidature (121)
capable (104)	capacité (248)	chef (88)
choix (78)	citoyen (77)	client (83)
collaboration (190)	compréhension (104)	composition (82)
compétence (241)	concept (78)	concevoir (105)
contacter (101)	couleur (75)	création (136)
créer (250)	cœur (94)	date (85)
demande (74)	demander (96)	design (78)
digital (105)	diplôme (189)	diplôme étude (102)
directeur (122)	direction (113)	domaine (73)
dynamique (110)	délai (145)	détenir (80)
déterminer (82)	développement (102)	effectuer (109)
efficace (75)	emploi (103)	envergure (103)
environnement (228)	esprit (113)	excellent (150)
excellente (104)	exigence (118)	exiger (76)
fabrication (79)	faire (199)	film (175)
fonction (137)	fournir (75)	gestion (109)
gestionnaire (81)	grand (113)	générique (80)
générique masculin (80)	gérer (86)	haut (93)
image (158)	industrie (79)	information (73)
jeu (123)	jour (115)	lead (94)
lieu (82)	long (198)	long métrage (175)
maintenir (185)	masculin (99)	masculin utiliser (86)
matière (81)	maîtrise (127)	meilleur (82)
mettre (131)	minimum (163)	montréal (225)
métrage (175)	métrage animation (92)	niveau (93)
numérique (95)	nécessaire (151)	offrir (91)
organisation (115)	organiser (74)	outil (166)
partie (82)	passionner (145)	permanent (94)
personnage (117)	plan (115)	porter (79)
posséder (82)	pouvoir (86)	prendre (92)

*Suite à la page suivante*

TABLE 3.5 – Suite de la page précédente

preuve (90)	principe (95)	priorité (115)
problème (208)	processus (104)	producteur (118)
production (352)	produire (76)	professionnel (86)
profil (79)	projet (316)	propos (120)
qualification (180)	qualité (217)	recherche (129)
requérir (83)	respecter (158)	responsabilité (236)
responsable (153)	réalisateur (92)	réaliser (127)
référence (91)	répondre (111)	résident (91)
résident permanent (79)	résolution (79)	résoudre (106)
rôle (108)	salaire (88)	savoir (131)
savoir faire (87)	sens (133)	service (130)
solide (120)	solution (121)	souci (73)
statut (80)	studio (233)	studio animation (108)
superviseur (236)	supervision (129)	système (76)
série (145)	talent (117)	technique (326)
temps (195)	texte (99)	travail (341)
travaille (85)	travailler (325)	tâche (179)
utiliser (184)	visuel (251)	vouloir (130)
échancier (129)	éclairage (116)	élément (101)
équipe production (80)	établir (151)	étroite (84)
étroite collaboration (81)	étude (113)	

En bref, afin d’obtenir la syntaxe adéquate pour l’évaluation de nos modèles, nous récoltons d’abord les données brutes. Ensuite, nous effectuons des manipulations pour réduire la taille de la matrice sparse en retirant les mots fréquents de la langue et les mots qui ne sont pas assez présents. Finalement, nous obtenons une matrice sparse de taille  $456 \times 201$ <sup>4</sup> qui constitue notre matrice  $X$ .

4. Le nombre 456 représente le nombre d’offres d’emploi et le nombre 201 combine les 200 jetons et la constante qui vaut 1.

# Chapitre 4

## Résultats

Dans cette section, nous effectuons une comparaison entre le modèle où la date de publication est la variable dépendante et le modèle où la date de publication est une indicatrice qui représente si l'observation a été publiée après 2019. Finalement, les limites du modèle sont présentées.

Les tableaux présentés dans cette section contiennent seulement les résultats différents de zéro pour le LASSO et les résultats significatifs à 10% pour la méthode des MCO. Les tableaux complets se trouvent en annexe. Les tableaux présentent des compétences relationnelles, des mots sans signification particulière pour notre analyse et des compétences techniques. Pour interpréter les résultats, nous allons nous fier aux tableaux présentés à la section 3. Les tableaux du LASSO comprennent les mots fréquents, les coefficients de prédiction du LASSO, la fréquence des mots présents dans les offres ainsi que la fréquence relative, soit la proportion du corpus où le mot se retrouve au moins une fois. Les tableaux des résultats des moindres carrés ordinaires comprennent les mots fréquents, les coefficients et la p-value associées au mot, la fréquence et la fréquence relative. Les compétences techniques et relationnelles présentées à la section 3 sont indiquées en police de caractère gras dans les tableaux de cette section.

### 4.1 Méthode du LASSO

Les résultats différents de zéro par la méthode du LASSO sont présentés dans les tableaux 4.1 et 4.2. Le tableau 4.1 représente les résultats pour le modèle 2.1 où la variable dépendante est l'année de publication. Le tableau 4.2 représente les résultats pour le modèle 2.2 où la variable dépendante est une indicatrice. Tout d'abord, on retrouve plusieurs mots faisant partie de la liste des compétences techniques et relationnelles présentées à la section 3. On retrouve les mots tels que « anglais », « créatif » et « travailler équipe » pour les compétences relationnelles ainsi que les mots « computer graphics (CG) », « compositing », « logiciel Maya » et « Nuke » pour les compétences techniques.

Le but de ce document est de tester la méthode afin de savoir si nous pouvons prédire les compétences émergentes à l'aide d'offres d'emploi. Nos résultats montrent que le logiciel « Nuke » est présent dans les offres les moins récentes. Si nous regardons deux offres qui contiennent de l'information semblable, mais qu'une offre contient le mot « Nuke » et l'autre non, alors nous nous attendons à ce que l'offre qui ne contient pas « Nuke » soit 0.112 ans plus récente. Ainsi, l'interprétation est la même pour le mot « logiciel Maya ». Une offre qui contient « logiciel Maya » comme compétence requise est 0.076 ans plus récente qu'une offre ne contenant pas « logiciel Maya ». Aussi, les résultats nous indiquent que le bigramme « computer graphics (CG) » et le mot « compositing » ne sont plus demandés dans les offres d'emploi récentes, avec des coefficients de  $-0.263$  et  $-0.369$  respectivement.

Au niveau des compétences relationnelles, le mot « anglais » et le bigramme « travailler équipe » ne sont plus présents dans les offres récentes avec un coefficient de  $-0.206$  et  $-0.186$  respectivement. Le mot « créatif » est le seul jeton à coefficient positif qui se retrouve dans la liste des compétences relationnelles précédemment établie avec un coefficient de 0.110.

TABLE 4.1 – Prédiction par LASSO en utilisant l'année comme variable dépendante avec les compétences techniques et relationnelles en gras

	Mots fréquents	Importance	Fréquence	Fréquence relative
0	amélioration	0.158	76	0.17
1	<b>anglais</b>	<b>-0.206</b>	<b>111</b>	<b>0.24</b>
2	animation	0.079	347	0.76
3	animation studio	-0.015	108	0.24
4	asset	-0.114	103	0.23
5	assurer	0.188	160	0.35
6	besoin	0.211	123	0.27
7	bon	-0.070	310	0.68
8	candidat	0.001	93	0.20
9	capable	-0.183	104	0.23
10	<b>cg</b>	<b>-0.263</b>	<b>127</b>	<b>0.28</b>
11	chef	0.120	88	0.19
12	citoyen	0.086	77	0.17
13	<b>compositing</b>	<b>-0.369</b>	<b>86</b>	<b>0.19</b>
14	composition	-0.042	82	0.18
15	compréhension	-0.143	104	0.23
16	<b>créatif</b>	<b>0.110</b>	<b>179</b>	<b>0.39</b>
17	créer	-0.011	250	0.55
18	date	-0.265	86	0.19

*Suite à la page suivante*

TABLE 4.1 – Suite de la page précédente

	Mots fréquents	Importance	Fréquence	Fréquence relative
19	demande	-0.138	74	0.16
20	diplôme	0.172	190	0.42
21	directeur	0.040	122	0.27
22	domaine	0.086	73	0.16
23	dynamique	-0.157	110	0.24
24	délai	-0.032	145	0.32
25	détail	0.541	116	0.25
26	déterminer	-0.193	82	0.18
27	développer	0.071	130	0.29
28	effectuer	-0.113	109	0.24
29	efficace	0.108	75	0.16
30	envergure	0.359	104	0.23
31	environnement	-0.007	229	0.50
32	esprit	0.303	113	0.25
33	excellente	-0.079	105	0.23
34	exigence	0.595	118	0.26
35	film	-0.056	176	0.39
36	fournir	0.076	75	0.16
37	grand	0.026	114	0.25
38	image	-0.043	159	0.35
39	industrie	0.086	79	0.17
40	information	-0.074	74	0.16
<b>41</b>	<b>logiciel maya</b>	<b>0.076</b>	<b>76</b>	<b>0.17</b>
42	long métrage	-0.046	176	0.39
43	masculin	0.203	100	0.22
44	matière	0.072	82	0.18
45	maîtrise	0.219	128	0.28
46	métrage	-0.010	176	0.39
47	métrage animation	-0.950	93	0.20
<b>48</b>	<b>nuke</b>	<b>-0.112</b>	<b>116</b>	<b>0.25</b>
49	nécessaire	-0.252	151	0.33
50	organiser	0.066	74	0.16
51	partie	-0.252	82	0.18
52	passionner	0.919	146	0.32
53	permanent	0.211	95	0.21
54	personnage	-0.056	117	0.26

Suite à la page suivante

TABLE 4.1 – Suite de la page précédente

	Mots fréquents	Importance	Fréquence	Fréquence relative
55	prendre	0.042	92	0.20
56	preuve	-0.643	91	0.20
57	processus	-0.109	104	0.23
58	professionnel	0.070	86	0.19
59	profil	0.863	79	0.17
60	propos	0.349	120	0.26
61	qualification	0.551	180	0.39
62	recherche	-0.102	129	0.28
63	responsabilité	0.020	237	0.52
64	responsable	-0.092	153	0.34
65	réalisateur	-0.035	92	0.20
66	résident permanent	0.264	79	0.17
67	sens	-0.004	134	0.29
68	service	0.524	131	0.29
69	solution	-0.019	121	0.27
70	statut	-0.556	81	0.18
71	studio	0.369	234	0.51
72	série	0.661	146	0.32
73	temps	0.445	195	0.43
74	travailler	-0.044	326	0.71
<b>75</b>	<b>travailler équipe</b>	<b>-0.186</b>	<b>99</b>	<b>0.22</b>
76	tâche	-0.148	179	0.39
77	visuel	-0.215	251	0.55
78	vouloir	0.612	130	0.29
79	éclairage	-0.279	116	0.25
80	élément	0.100	101	0.22
81	équipe production	0.051	80	0.18
82	établir	0.135	152	0.33
83	étude	-0.037	114	0.25

La deuxième partie des résultats montre le modèle 2.2 où la variable dépendante est une indicatrice qui indique si l'année de publication de l'offre est supérieure à 2019. Les résultats sont semblables d'une technique à l'autre. On trouve les mots « communiquer », « créatif », « français » et « travailler équipe » comme compétences relationnelles et les mots « computer

graphics (CG) », « compositing », « logiciel Maya », « Nuke », « Python » et « VFX » du côté des compétences techniques.

D’abord, le modèle avec indicatrice prédit un plus grand nombre de compétences relationnelles et techniques que le modèle utilisant l’année de publication comme variable dépendante. Les coefficients reliés aux compétences techniques sont  $-0.099$  pour le mot « computer graphics (CG) »,  $-0.112$  pour le jeton « compositing »,  $0.021$  pour le bigramme « logiciel Maya »,  $-0.039$  pour le mot « Nuke »,  $0.007$  pour le mot « Python » et  $-0.042$  pour la compétence « VFX ». Les compétences relationnelles qui s’y retrouvent ont un coefficient de  $0.019$  pour le mot « communiquer »,  $0.028$  pour le mot « créatif »,  $0.006$  pour le mot « français » ainsi que  $-0.067$  pour le bigramme « travailler équipe ».

Les deux méthodes donnent des résultats semblables, mais le modèle 2.2 prédit un plus grand nombre de compétences techniques (Python et VFX). On retrouve une forme condensée des résultats dans le tableau synthèse 4.3. Les mots sélectionnés correspondent aux compétences relationnelles et compétences techniques précédemment choisies au chapitre 3. L’ampleur des résultats n’est pas égale d’un modèle à l’autre, mais les signes des coefficients correspondent entre les deux modèles pour toutes les variables non nulles. De manière générale, on remarque que les coefficients du modèle 2.1 sont plus élevés en valeur absolue que les coefficients du modèle 2.2. Ceci peut être dû au fait que la variable dépendante est une indicatrice.

TABLE 4.2 – Prédiction par LASSO en utilisant l’année comme indicatrice avec les compétences techniques et relationnelles en gras

	Mots fréquents	Importance	Fréquence	Fréquence relative
0	amélioration	0.043	75	0.16
1	an	-0.023	242	0.53
2	an expérience	-0.018	196	0.43
3	animation	0.026	346	0.76
4	année	0.011	128	0.28
5	baser	0.015	90	0.20
6	besoin	0.033	123	0.27
7	but	-0.131	98	0.22
8	candidat	0.006	93	0.20
9	candidature	-0.030	122	0.27
<b>10</b>	<b>cg</b>	<b>-0.099</b>	<b>126</b>	<b>0.28</b>
11	chef	0.046	88	0.19
12	collaboration	-0.016	190	0.42

*Suite à la page suivante*

TABLE 4.2 – Suite de la page précédente

	Mots fréquents	Importance	Fréquence	Fréquence relative
13	<b>communiquer</b>	<b>0.019</b>	<b>155</b>	<b>0.34</b>
14	<b>compositing</b>	<b>-0.112</b>	<b>86</b>	<b>0.19</b>
15	composition	-0.002	82	0.18
16	compréhension	-0.038	103	0.23
17	compétence	-0.006	241	0.53
18	contacter	-0.093	102	0.22
19	<b>créatif</b>	<b>0.028</b>	<b>179</b>	<b>0.39</b>
20	cœur	-0.066	95	0.21
21	design	-0.064	78	0.17
22	diplôme	0.024	189	0.42
23	directeur	0.023	121	0.27
24	direction	-0.019	113	0.25
25	dynamique	-0.018	110	0.24
26	détail	0.072	116	0.25
27	esprit	0.005	113	0.25
28	excellent	0.010	149	0.33
29	exigence	0.206	118	0.26
30	film	-0.013	175	0.38
31	fournir	0.046	75	0.16
32	<b>français</b>	<b>0.006</b>	<b>122</b>	<b>0.27</b>
33	grand	0.041	113	0.25
34	haut	-0.032	93	0.20
35	information	-0.042	74	0.16
36	<b>logiciel maya</b>	<b>0.021</b>	<b>76</b>	<b>0.17</b>
37	masculin	0.232	100	0.22
38	montréal	0.083	225	0.49
39	métrage animation	-0.294	92	0.20
40	niveau	-0.013	92	0.20
41	<b>nuke</b>	<b>-0.039</b>	<b>116</b>	<b>0.25</b>
42	nécessaire	-0.057	151	0.33
43	organiser	0.008	74	0.16
44	outil	-0.006	166	0.36
45	partie	-0.104	81	0.18
46	passionner	0.196	145	0.32
47	personnage	-0.023	117	0.26
48	pouvoir	-0.048	86	0.19

Suite à la page suivante

TABLE 4.2 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
49	prendre	0.036	92	0.20
50	preuve	-0.116	91	0.20
51	processus	-0.015	104	0.23
52	produire	0.026	76	0.17
53	profil	0.210	79	0.17
54	propos	0.169	120	0.26
<b>55</b>	<b>python</b>	<b>0.007</b>	<b>130</b>	<b>0.29</b>
56	qualification	0.026	179	0.39
57	responsable	-0.041	152	0.33
58	réalisateur	-0.079	92	0.20
59	réaliser	-0.004	127	0.28
60	résolution	0.043	79	0.17
61	savoir	-0.047	131	0.29
62	service	0.124	130	0.29
63	statut	-0.206	80	0.18
64	studio	0.086	233	0.51
65	série	0.130	145	0.32
66	technique	0.059	327	0.72
67	temps	0.105	194	0.43
68	travail	-0.029	341	0.75
<b>69</b>	<b>travailler équipe</b>	<b>-0.067</b>	<b>99</b>	<b>0.22</b>
70	tâche	-0.066	179	0.39
71	utiliser	-0.028	184	0.40
<b>72</b>	<b>vfx</b>	<b>-0.042</b>	<b>80</b>	<b>0.18</b>
73	vouloir	0.367	130	0.29
74	éclairage	-0.029	116	0.25
75	élément	0.028	101	0.22
76	équipe production	0.063	80	0.18
77	établir	0.058	151	0.33
78	étude	-0.009	114	0.25

TABLE 4.3 – Tableau synthèse du LASSO

Mots fréquents	Coefficient		Fréquence	Fréquence Relative
	Modèle 1	Modèle 2		
<b>Compétences techniques :</b>				
Computer graphics (CG)	-0.263	-0.099	127	0.28
Compositing	-0.369	-0.112	86	0.19
Logiciel Maya	0.076	0.021	76	0.17
Nuke	-0.112	-0.039	116	0.25
Python	0	0.007	130	0.29
VFX	0	-0.042	80	0.18
<b>Compétences relationnelles :</b>				
Anglais	-0.206	0	111	0.24
Communiquer	0	0.019	155	0.34
Créatif	0.110	0.028	179	0.39
Français	0	0.006	122	0.27
Travailler équipe	-0.186	-0.067	99	0.22

## 4.2 Méthode des moindres carrés

Les résultats significatifs par la méthode des moindres carrés à 10% sont présentés dans les tableaux 4.5 et 4.6. Le tableau 4.5 indique les résultats du modèle 2.1 avec l'année comme variable dépendante et le tableau 4.6 indique les résultats du modèle 2.2 avec l'année comme indicatrice. Les résultats pertinents sont condensés dans le tableau synthèse 4.4.

Tout d'abord, on remarque que les résultats significatifs des MCO sont moins constants pour les deux modèles que pour la méthode du LASSO. Cependant, les résultats obtenus comprennent des compétences semblables. Pour le modèle 2.1, les compétences techniques résultantes sont « Nuke », « Pipeline » et « Shotgun » avec des coefficients de  $-0.369$ ,  $-0.371$  et  $0.602$  respectivement. Pour ce qui est du modèle 2.2, la seule compétence relationnelle qui est significative à 10% est le jeton « Créatif » avec un coefficient de  $0.691$ . Les compétences techniques nous indiquent que le logiciel Shotgun est une compétence qui émerge dans les offres d'emploi et que le logiciel Nuke n'est plus demandé dans les offres d'emploi.

Les compétences techniques obtenues avec le modèle 2.2 sont « Computer Graphics (CG) » avec un coefficient de  $-0.158$ , « Nuke » avec un coefficient de  $-0.104$ , « Shotgun » avec un coefficient de  $0.086$  et « VFX » avec un coefficient de  $-0.129$ . Les compétences relationnelles sont « Créatif » avec un coefficient de  $0.29$ , « Français » avec un coefficient de  $0.233$  et « Travailler équipe » avec un coefficient de  $-0.121$ . D'après l'estimation du modèle 2.2, le logiciel Shotgun est la seule compétence technique émergente. L'utilisation de CG, Nuke et VFX sont des compétences techniques qui apparaissent moins fréquemment dans les offres d'emploi récentes.

TABLE 4.4 – Tableau synthèse des MCO

Mots fréquents	Coefficient		Fréquence	Fréquence Relative
	Modèle 1	Modèle 2		
<b>Compétences techniques :</b>				
Computer Graphics (CG)	-	-0.158	126	0.28
Nuke	-0.369	-0.104	116	0.25
Pipeline	-0.371	-	177	0.39
Shotgun	0.602	0.086	119	0.26
VFX	-	-0.129	80	0.18
<b>Compétences relationnelles :</b>				
Créatif	0.691	0.229	179	0.39
Français	-	0.233	122	0.27
Travailler équipe	-	-0.121	99	0.22

TABLE 4.5 – Estimation par moindres carrés ordinaires en utilisant l'année comme variable dépendante avec les compétences techniques et relationnelles en gras

	Mots fréquents	Paramètres	p-value	Fréquence	Fréquence relative
0	const	-1.852	0.000	0	0.00
1	année	-0.895	0.025	128	0.28
2	année expérience	1.001	0.016	111	0.24
3	assurer	0.474	0.002	160	0.35
4	bel	0.655	0.051	77	0.17
5	candidat	0.476	0.030	93	0.20
6	capable	-0.316	0.068	104	0.23
7	cg	-0.370	0.024	127	0.28
8	collaboration	-0.332	0.055	190	0.42
9	compréhension	-0.337	0.067	104	0.23
10	compétence	0.299	0.054	242	0.53
11	concept	0.666	0.008	78	0.17
12	connaissance	0.432	0.039	98	0.21
13	<b>créatif</b>	<b>0.691</b>	<b>0.000</b>	<b>179</b>	<b>0.39</b>
14	demande	-0.536	0.023	74	0.16
15	demander	-0.473	0.083	97	0.21
16	domaine	0.427	0.015	73	0.16
17	délai	-0.414	0.015	145	0.32
18	département	0.266	0.094	154	0.34
19	déterminer	-0.432	0.015	82	0.18

*Suite à la page suivante*

TABLE 4.5 – Suite de la page précédente

	Mots fréquents	Paramètres	p-value	Fréquence	Fréquence relative
20	effectuer	-0.344	0.049	109	0.24
21	efficace	0.350	0.074	75	0.16
22	envergure	1.267	0.000	104	0.23
23	esprit	0.579	0.001	113	0.25
24	faire	0.291	0.076	200	0.44
25	industrie	-0.368	0.057	79	0.17
26	information	-0.411	0.083	74	0.16
27	long	0.884	0.002	199	0.44
28	long métrage	-0.448	0.010	176	0.39
29	maintenir	-0.367	0.048	185	0.41
30	masculin	3.034	0.003	100	0.22
31	métrage	-0.448	0.010	176	0.39
<b>32</b>	<b>nuke</b>	<b>-0.369</b>	<b>0.031</b>	<b>116</b>	<b>0.25</b>
33	offrir	0.539	0.008	92	0.20
34	organisation	-0.378	0.049	115	0.25
35	passionner	0.575	0.021	146	0.32
<b>36</b>	<b>pipeline</b>	<b>-0.371</b>	<b>0.053</b>	<b>177</b>	<b>0.39</b>
37	porter	-1.306	0.056	79	0.17
38	preuve	-0.603	0.009	91	0.20
39	problème	-0.399	0.032	208	0.46
40	processus	-0.374	0.027	104	0.23
41	profil	1.031	0.000	79	0.17
42	propos	1.297	0.000	120	0.26
43	qualification	0.700	0.001	180	0.39
44	respecter	0.335	0.062	158	0.35
45	responsabilité	0.537	0.005	237	0.52
46	réalisateur	-0.456	0.035	92	0.20
47	résident	-1.416	0.010	91	0.20
48	rôle	0.383	0.085	108	0.24
49	service	0.518	0.043	131	0.29
<b>50</b>	<b>shotgun</b>	<b>0.602</b>	<b>0.007</b>	<b>119</b>	<b>0.26</b>
51	studio animation	-0.918	0.021	108	0.24
52	série	0.439	0.030	146	0.32
53	temps	0.404	0.030	195	0.43
54	texte	-2.367	0.007	100	0.22
55	travail	0.389	0.071	85	0.19

Suite à la page suivante

TABLE 4.5 – Suite de la page précédente

	Mots fréquents	Paramètres	p-value	Fréquence	Fréquence relative
56	visuel	-0.423	0.004	251	0.55
57	vouloir	1.026	0.000	130	0.29
58	élément	0.425	0.016	101	0.22
59	étude	-0.812	0.089	114	0.25

TABLE 4.6 – Estimation par moindres carrés ordinaires en utilisant l'année comme indicatrice avec les compétences techniques et relationnelles en gras

	Mots fréquents	Paramètres	p-value	Fréquence	Fréquence relative
0	anglais	-0.187	0.089	110	0.24
1	but	-0.379	0.002	98	0.22
2	candidature	-0.309	0.003	122	0.27
3	<b>cg</b>	<b>-0.158</b>	<b>0.009</b>	<b>126</b>	<b>0.28</b>
4	chef	0.107	0.085	88	0.19
5	collaboration	-0.105	0.096	190	0.42
6	compréhension	-0.192	0.005	103	0.23
7	<b>créatif</b>	<b>0.229</b>	<b>0.000</b>	<b>179</b>	<b>0.39</b>
8	design	-0.225	0.008	78	0.17
9	directeur	0.088	0.096	121	0.27
10	délai	-0.125	0.054	145	0.32
11	déterminer	-0.135	0.039	82	0.18
12	exigence	0.233	0.000	118	0.26
13	<b>français</b>	<b>0.233</b>	<b>0.024</b>	<b>122</b>	<b>0.27</b>
14	générique	-0.544	0.061	81	0.18
15	générique masculin	-0.544	0.061	81	0.18
16	information	-0.193	0.027	74	0.16
17	masculin	0.695	0.060	100	0.22
18	masculin utiliser	0.634	0.016	87	0.19
19	métrage animation	-0.478	0.000	92	0.20
20	<b>nuke</b>	<b>-0.104</b>	<b>0.099</b>	<b>116</b>	<b>0.25</b>
21	organisation	-0.125	0.078	115	0.25
22	passionner	0.174	0.055	145	0.32

Suite à la page suivante

TABLE 4.6 – Suite de la page précédente

	Mots fréquents	Paramètres	p-value	Fréquence	Fréquence relative
23	preuve	-0.201	0.017	91	0.20
24	principe	0.204	0.045	96	0.21
25	profil	0.295	0.003	79	0.17
26	propos	0.232	0.065	120	0.26
27	réalisateur	-0.223	0.005	92	0.20
28	salaire	0.237	0.082	88	0.19
29	savoir	-0.171	0.053	131	0.29
30	sens	0.107	0.092	134	0.29
31	service	0.349	0.000	130	0.29
<b>32</b>	<b>shotgun</b>	<b>0.142</b>	<b>0.086</b>	<b>118</b>	<b>0.26</b>
33	solution	-0.130	0.040	121	0.27
34	série	0.166	0.025	145	0.32
<b>35</b>	<b>travailler équipe</b>	<b>-0.121</b>	<b>0.096</b>	<b>99</b>	<b>0.22</b>
36	utiliser	-0.121	0.067	184	0.40
<b>37</b>	<b>vfx</b>	<b>-0.129</b>	<b>0.094</b>	<b>80</b>	<b>0.18</b>
38	vouloir	0.553	0.000	130	0.29
39	équipe production	0.138	0.045	80	0.18
40	étude	-0.311	0.075	114	0.25

### 4.3 Quelques jetons surprenants

Les résultats de nos prédictions et estimations contiennent quelques jetons qui nécessitent une analyse particulière.

Nous remarquons que les deux modèles LASSO prédisent le jeton « diplôme » avec des offres récentes. À priori, « diplôme » est un jeton émergent dans les offres. Nous avons fait une analyse manuelle à l'aide des expressions régulières afin de comprendre le sens du jeton. Les quelques 189 annonces où se retrouve le jeton indiquent plusieurs programmes. Les diplômes d'études demandés sont autant collégiales que universitaires. L'analyse préliminaire révèle que les employeurs ne demandent pas de diplôme en particulier, mais demandent plutôt la possession d'un diplôme dans un domaine artistique ou technique. Aussi, il est souvent mentionné qu'une certaine expérience sur le marché du travail peut être équivalente à la possession d'un diplôme. De plus, les offres d'emploi qui ne contiennent pas la mention du mot diplôme demandent généralement des années d'expériences dans le domaine. Nous

pouvons donc émettre l'hypothèse que ces offres d'emploi sont liées à des postes seniors. Nous ne pouvons pas conclure que le diplôme est un pré-requis émergent dans les offres d'emploi en notre possession.

Nous remarquons aussi le jeton « masculin » dans l'ensemble des tableaux présentés. Les conclusions de l'analyse post-estimation relève que le jeton est toujours employé pour indiquer que les offres sont écrites en utilisant le genre masculin pour alléger le texte. Le coefficient positif relié au jeton dans tous les tableaux indique que les employeurs sont plus inclusifs dans la rédaction des offres. Les offres qui ne contiennent pas le jeton sont majoritairement écrites au masculin, sans aucune mention de l'utilisation du genre.

#### **4.4 Limites**

Nous soulevons quelques limites à la méthodologie employée dans ce document. Dans la littérature, les bases de données utilisées sont plus volumineuses que celle utilisée dans ce document. Ce problème a des impacts au niveau des résultats. Nous aurions préféré avoir une base de données plus volumineuse pour, en autres, obtenir un échantillon plus représentatif du domaine des animateurs pour le cinéma. Notre petite base de données peut sur-représenter des compétences et en sous-représenter d'autres. Par exemple, comme nos données proviennent d'un site d'offres d'emploi d'animateur pour le cinéma et que le domaine se concentre seulement pour les régions de Québec et Montréal, il est possible qu'une compagnie ait décroché un bon contrat qui nécessite une compétence particulière et qu'elle affiche plusieurs offres qui mentionnent cette compétence. Alors, la compétence est sur-représentée dans nos données et peut biaiser la conclusion de nos résultats. De plus, les résultats sont sujets aux paramètres et aux manipulations effectuées dans le modèle. Nous avons fait des manipulations et nous avons choisi nos paramètres selon une méthodologie conforme à la littérature du domaine. Toutefois, nous ne pouvons pas conclure que les résultats présentés ici sont universels pour l'ensemble du domaine des animateurs pour le cinéma. De plus, notre analyse se concentre sur les régions de Québec et de Montréal, d'où les offres proviennent.

La littérature économique s'intéresse aux estimations et à la signficativité des paramètres. Cependant, les méthodes d'intelligence artificielle se concentrent sur la prédiction des paramètres. Elles s'intéressent à prédire hors-échantillon contrairement à trouver le meilleur modèle. Finalement, nous nous concentrons seulement sur un petit segment des résultats. Les résultats comprennent plusieurs jetons qui n'ont aucune signification dans notre contexte.

# Conclusion

Plusieurs recherches ont étudié les tendances du marché du travail en utilisant des indicateurs traditionnels. Cependant, avec la volatilité du marché, ces indicateurs ne sont pas assez précis pour analyser les tendances du marché en temps réel. Dans ce mémoire, nous voulions établir une procédure qui détermine les compétences émergentes à l'aide de données non conventionnelles sous la forme d'offres d'emploi. Pour y arriver, nous avons utilisé des offres d'emploi extraites sur internet concernant le domaine des animateurs pour le cinéma.

Pour déterminer l'émergence des compétences, nous avons utilisé la description des offres d'emploi et comme variable dépendante, leur date de publication. La variable dépendante est représentée de deux façons : comme l'année de publication des offres d'emploi et sous la forme d'une indicatrice qui détermine si l'offre d'emploi a été publiée après 2019. Nous avons utilisé la méthode du *Least absolute shrinkage and selection operator* (LASSO) et les moindres carrés ordinaires pour déterminer quelles compétences parmi celles présentes dans les offres d'emploi représentaient des compétences émergentes. Les deux modèles produisent des résultats semblables et prédisent des compétences techniques et relationnelles qui sont émergentes. Lorsque nous utilisons le modèle du LASSO, les logiciels Maya et Nuke sont les deux principaux résultats de nos modèles. Nos résultats montrent que le logiciel Maya est une compétence émergente, tandis que le logiciel Nuke est de moins en moins demandé par les employeurs. Le modèle des moindres carrés ordinaires indique aussi que Nuke est de moins en moins demandé par les employeurs, mais il n'est pas en mesure d'obtenir de l'information sur Maya. Toutefois, il est en mesure d'indiquer que le logiciel Shotgun est émergent.

Les méthodes économétriques présentées dans ce document pourraient être appliquées à une base de données plus volumineuse et à une base de données qui comporte des salaires affichés par les employeurs. En effet, en possédant des salaires pour les offres d'emploi, nous pourrions comparer la méthode présentée dans ce document avec une méthode qui utiliserait les salaires comme variable dépendante. Ainsi, nous pourrions capter les compétences émergentes dans les offres d'emploi sous forme de primes salariales, indiquées par les coefficients du modèle hédonique. Le document se rapprocherait alors des travaux effectués par Nowak et Smith (2017).

# Bibliographie

Daron ACEMOGLU : Changes in unemployment and wage inequality : An alternative theory and some evidence. American Economic Review, 89(5):1259–1278, 1999. ISSN 00028282.

Daron ACEMOGLU et Pascual RESTREPO : Automation and new tasks : How technology displaces and reinstates labor, 2019. ISSN 08953309.

Ajay AGRAWAL, Joshua S. GANS et Avi GOLDFARB : Artificial intelligence : The ambiguous labor market impact of automating prediction, 2019. ISSN 08953309.

Namir ANANI : Paving the Way for the Future of Work. Canadian Public Policy, pages 1–10, 2018. ISSN 0317-0861.

Enghin ATALAY, Phai PHONGTHIENGTHAM, Sebastian SOTELO et Daniel TANNENBAUM : The evolution of work in the United States, Atalay, E., Phongthientham, P., Sotelo, S., & Tannenbaum, D. (2020). The evolution of work in the United States. American Economic Journal : Applied Economics, 12(2), 1–34. <https://doi.org/10.1257/app.20190070tes>. American Economic Journal : Applied Economics, 12(2):1–34, 2020. ISSN 19457790.

Susan ATHEY et Guido W. IMBENS : Machine Learning Methods That Economists Should Know About. Annual Review of Economics, 11(1):685–725, 2019. ISSN 1941-1383.

David H. AUTOR et David DORN : The growth of low-skill service jobs and the polarization of the US Labor Market. American Economic Review, 2013. ISSN 00028282.

CAPGEMINI : Capgemini and LinkedIn release new report on the digital talent gap : Employees are investing their own resources to remain competitive. 44(0), 2017.

David J DEMING et Kadeem NORAY : Earnings Dynamics, Changing Job Skills, and STEM Careers. The Quarterly Journal of Economics, (May), 2020. ISSN 0033-5533.

Morgan R. FRANK, David AUTOR, James E. BESSEN, Erik BRYNJOLFSSON, Manuel CEBRIAN, David J. DEMING, Maryann FELDMAN, Matthew GROH, José LOBO, Esteban MORO, Dashun WANG, Hyejin YOUN et Iyad RAHWAN : Toward understanding the impact of artificial intelligence on labor. Proceedings of the National Academy of Sciences of the United States of America, 116(14):6531–6539, 2019. ISSN 10916490.

Diego GARCÍA : Sentiment during Recessions. Journal of Finance, 68(3):1267–1300, 2013. ISSN 00221082.

Matthew GENTZKOW, Bryan KELLY et Matt TADDY : Text as data, 2019. ISSN 00220515.

Maarten GOOS, Alan MANNING et Anna SALOMONS : Job polarization in Europe. American Economic Review, 99(2):58–63, 2009. ISSN 00028282.

Bernar MARR : What is the difference between artificial intelligence and machine learning?, Dec 2016. URL <https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/?sh=1aa017872742>.

Sendhil MULLAINATHAN et Jann SPIESS : Machine learning : An applied econometric approach. Journal of Economic Perspectives, 31(2):87–106, 2017. ISSN 08953309.

Adam NOWAK et Patrick SMITH : Textual Analysis in Real Estate. Journal of Applied Econometrics, 2017. ISSN 10991255.

Michele RIMINI et Vincenzo SPIEZA : Skills for a Digital World : Ministerial Meeting on the Digital Economy, Background Report 2016. Rapport technique 250, 2016.

ROSEN : Hedonic Prices and Implicit Markets : Product Differentiation in Pure Competition Sherwin Rosen The. Tetrahedron Letters, 27(50):6079–6082, 1974. ISSN 00404039.

Robert TIBSHIRANI : Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society, 58(1):267–288, 2016.

Tractor-trailer TRUCK, Marketing MANAGERS, Information Security ANALYSTS, Web DEVELOPERS, Industrial ENGINEERS, Sales MANAGERS, Human Resources MANAGERS, Financial MANAGERS et Monthly HIRES : The Skills Gap is Costing Companies Nearly \$1 Million Annually, According to New CareerBuilder Survey - Apr 13, 2017. pages 3–4, 2017. URL <http://press.careerbuilder.com/2017-04-13-The-Skills-Gap-is-Costing-Companies-Nearly-1-Million-Annually-According-to-New-CareerBuilder-Survey>.

Hal R. VARIAN : Big data : New tricks for econometrics. Journal of Economic Perspectives, 28(2):3–28, 2014. ISSN 08953309.

Ayşegül ŞAHİN, Joseph SONG, Giorgio TOPA et Giovanni L. VIOLANTE : Mismatch unemployment. American Economic Review, 104(11):3529–3564, 2014. ISSN 00028282.

# Annexe A

## Version longue des tableaux

La présente section comporte les tableaux complets des prédictions et des régressions obtenues par la méthode du LASSO ainsi que par la méthode des moindres carrés ordinaires.

### A.1 LASSO

Les tableaux obtenus par la méthode du LASSO comprennent les 200 mots choisis par l'algorithme ainsi qu'une constante intégrée au modèle à des fins de comparaison. La colonne « Importance » rapporte le coefficient de prédiction, la colonne « Fréquence » rapporte le nombre d'annonces où le mot est présent et la colonne « Fréquence relative » rapporte la fréquence relative de la présence des mots dans toutes les offres d'emploi. Les tableaux A.1 et A.2 présentent la version complète des tableaux de la méthode du LASSO présentés au chapitre 4.

TABLE A.1 – Prédiction par LASSO avec l'année comme variable dépendante

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
0	const	0.000000	0	0.00
1	amélioration	0.157966	76	0.17
2	an	-0.000000	243	0.53
3	an expérience	-0.000000	197	0.43
4	anglais	-0.206470	111	0.24
5	animation	0.079075	347	0.76
6	animation studio	-0.014992	108	0.24
7	année	-0.000000	128	0.28
8	année expérience	-0.000000	111	0.24
9	aptitude	0.000000	85	0.19

*Suite à la page suivante*

TABLE A.1 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
10	artiste	-0.000000	314	0.69
11	artistique	0.000000	206	0.45
12	asset	-0.113810	103	0.23
13	assurer	0.187881	160	0.35
14	atelier	-0.000000	73	0.16
15	atout	0.000000	264	0.58
16	baser	0.000000	90	0.20
17	bel	0.000000	77	0.17
18	besoin	0.210788	123	0.27
19	bon	-0.070421	310	0.68
20	bon connaissance	-0.000000	188	0.41
21	but	-0.000000	98	0.21
22	but texte	-0.000000	82	0.18
23	canada	0.000000	91	0.20
24	candidat	0.001012	93	0.20
25	candidature	0.000000	122	0.27
26	capable	-0.182741	104	0.23
27	capacité	0.000000	249	0.55
28	cg	-0.262585	127	0.28
29	chef	0.119941	88	0.19
30	choix	0.000000	79	0.17
31	citoyen	0.085796	77	0.17
32	client	0.000000	83	0.18
33	collaboration	-0.000000	190	0.42
34	communication	-0.000000	206	0.45
35	communiquer	0.000000	155	0.34
36	compositing	-0.369155	86	0.19
37	composition	-0.042494	82	0.18
38	compréhension	-0.143301	104	0.23
39	compétence	-0.000000	242	0.53
40	concept	0.000000	78	0.17
41	concevoir	0.000000	105	0.23
42	connaissance logiciel	0.000000	98	0.21
43	contacter	-0.000000	102	0.22
44	couleur	-0.000000	75	0.16
45	créatif	0.109688	179	0.39

Suite à la page suivante

TABLE A.1 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
46	création	-0.000000	136	0.30
47	créativité	0.000000	90	0.20
48	créer	-0.011234	250	0.55
49	cœur	-0.000000	95	0.21
50	date	-0.265016	86	0.19
51	demande	-0.138254	74	0.16
52	demander	0.000000	97	0.21
53	design	-0.000000	78	0.17
54	digital	-0.000000	105	0.23
55	diplôme	0.171588	190	0.42
56	diplôme étude	-0.000000	103	0.23
57	directeur	0.039607	122	0.27
58	direction	-0.000000	113	0.25
59	domaine	0.085950	73	0.16
60	dynamique	-0.157139	110	0.24
61	délai	-0.032021	145	0.32
62	département	0.000000	154	0.34
63	détail	0.541446	116	0.25
64	détenir	0.000000	80	0.18
65	déterminer	-0.193455	82	0.18
66	développement	0.000000	102	0.22
67	développer	0.071016	130	0.29
68	effectuer	-0.113274	109	0.24
69	efficace	0.108420	75	0.16
70	emploi	0.000000	104	0.23
71	envergure	0.358809	104	0.23
72	environnement	-0.006513	229	0.50
73	esprit	0.302625	113	0.25
74	excellent	-0.000000	150	0.33
75	excellente	-0.079253	105	0.23
76	exigence	0.594533	118	0.26
77	exiger	-0.000000	77	0.17
78	fabrication	-0.000000	80	0.18
79	faire	0.000000	200	0.44
80	film	-0.056333	176	0.39
81	fonction	-0.000000	138	0.30

Suite à la page suivante

TABLE A.1 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
82	fournir	0.075539	75	0.16
83	français	-0.000000	123	0.27
84	gestion	-0.000000	109	0.24
85	gestionnaire	0.000000	82	0.18
86	grand	0.026151	114	0.25
87	générique	-0.000000	81	0.18
88	générique masculin	-0.000000	81	0.18
89	gérer	-0.000000	86	0.19
90	haut	-0.000000	93	0.20
91	image	-0.043091	159	0.35
92	industrie	0.085635	79	0.17
93	information	-0.073618	74	0.16
94	informatique	0.000000	76	0.17
95	jeu	0.000000	123	0.27
96	jour	-0.000000	116	0.25
97	lead	-0.000000	94	0.21
98	lieu	-0.000000	83	0.18
99	linux	0.000000	75	0.16
100	logiciel	-0.000000	251	0.55
101	logiciel maya	0.076037	76	0.17
102	long	-0.000000	199	0.44
103	long métrage	-0.045549	176	0.39
104	maintenir	-0.000000	185	0.41
105	masculin	0.203002	100	0.22
106	masculin utiliser	0.000000	87	0.19
107	matière	0.071718	82	0.18
108	maya	0.000000	244	0.54
109	maîtrise	0.219468	128	0.28
110	meilleur	0.000000	82	0.18
111	mettre	0.000000	131	0.29
112	minimum	0.000000	163	0.36
113	montréal	0.000000	226	0.50
114	métrage	-0.009792	176	0.39
115	métrage animation	-0.949526	93	0.20
116	niveau	0.000000	93	0.20
117	nuke	-0.111828	116	0.25

Suite à la page suivante

TABLE A.1 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
118	numérique	-0.000000	95	0.21
119	nécessaire	-0.252427	151	0.33
120	offrir	0.000000	92	0.20
121	organisation	-0.000000	115	0.25
122	organiser	0.066451	74	0.16
123	outil	0.000000	167	0.37
124	partie	-0.251596	82	0.18
125	passionner	0.919000	146	0.32
126	permanent	0.210801	95	0.21
127	personnage	-0.055695	117	0.26
128	pipeline	-0.000000	177	0.39
129	plan	-0.000000	115	0.25
130	porter	-0.000000	79	0.17
131	posséder	-0.000000	82	0.18
132	pouvoir	-0.000000	86	0.19
133	prendre	0.041768	92	0.20
134	preuve	-0.642742	91	0.20
135	principe	0.000000	96	0.21
136	priorité	-0.000000	115	0.25
137	problème	-0.000000	208	0.46
138	processus	-0.108507	104	0.23
139	producteur	0.000000	118	0.26
140	production	0.000000	353	0.77
141	produire	0.000000	76	0.17
142	professionnel	0.070344	86	0.19
143	profil	0.862724	79	0.17
144	projet	0.000000	317	0.70
145	propos	0.349374	120	0.26
146	python	0.000000	131	0.29
147	qualification	0.551460	180	0.39
148	qualité	-0.000000	217	0.48
149	recherche	-0.102070	129	0.28
150	requérir	0.000000	83	0.18
151	respecter	0.000000	158	0.35
152	responsabilité	0.019764	237	0.52
153	responsable	-0.091613	153	0.34

Suite à la page suivante

TABLE A.1 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
154	rigging	0.000000	83	0.18
155	réalisateur	-0.035388	92	0.20
156	réaliser	0.000000	127	0.28
157	référence	-0.000000	92	0.20
158	répondre	0.000000	111	0.24
159	résident	-0.000000	91	0.20
160	résident permanent	0.264035	79	0.17
161	résolution	-0.000000	79	0.17
162	résoudre	0.000000	106	0.23
163	résoudre problème	-0.000000	89	0.20
164	rôle	0.000000	108	0.24
165	salaire	-0.000000	89	0.20
166	savoir	-0.000000	132	0.29
167	savoir faire	-0.000000	88	0.19
168	sens	-0.004282	134	0.29
169	service	0.523847	131	0.29
170	shotgun	0.000000	119	0.26
171	solide	0.000000	120	0.26
172	solution	-0.018663	121	0.27
173	statut	-0.556095	81	0.18
174	studio	0.369220	234	0.51
175	studio animation	-0.000000	108	0.24
176	superviseur	-0.000000	236	0.52
177	supervision	-0.000000	130	0.29
178	système	0.000000	77	0.17
179	série	0.660815	146	0.32
180	talent	-0.000000	118	0.26
181	technique	0.000000	327	0.72
182	temps	0.445289	195	0.43
183	texte	0.000000	100	0.22
184	travail	-0.000000	342	0.75
185	travaille	-0.000000	85	0.19
186	travailler	-0.043758	326	0.71
187	travailler équipe	-0.185594	99	0.22
188	tâche	-0.148206	179	0.39
189	utiliser	-0.000000	185	0.41

Suite à la page suivante

TABLE A.1 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
190	vfx	-0.000000	81	0.18
191	visuel	-0.214636	251	0.55
192	vouloir	0.612185	130	0.29
193	échancier	0.000000	129	0.28
194	éclairage	-0.279321	116	0.25
195	élément	0.100483	101	0.22
196	équipe production	0.051166	80	0.18
197	établir	0.135205	152	0.33
198	étroite	0.000000	84	0.18
199	étroite collaboration	0.000000	81	0.18
200	étude	-0.037112	114	0.25

TABLE A.2 – Prédiction par LASSO avec l'année comme indicatrice

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
0	const	0.000000	0	0.00
1	amélioration	0.040493	76	0.17
2	an	-0.019973	243	0.53
3	an expérience	-0.020839	197	0.43
4	anglais	-0.000000	111	0.24
5	animation	0.024470	347	0.76
6	animation studio	-0.000000	108	0.24
7	année	0.011608	128	0.28
8	année expérience	0.000000	111	0.24
9	aptitude	0.000000	85	0.19
10	artiste	0.000000	314	0.69
11	artistique	0.000000	206	0.45
12	asset	-0.000000	103	0.23
13	assurer	0.000000	160	0.35
14	atelier	-0.000000	73	0.16
15	atout	-0.000000	264	0.58
16	baser	0.014334	90	0.20
17	bel	-0.000000	77	0.17

Suite à la page suivante

TABLE A.2 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
18	besoin	0.033904	123	0.27
19	bon	-0.000000	310	0.68
20	bon connaissance	-0.000000	188	0.41
21	but	-0.130127	98	0.21
22	but texte	-0.000000	82	0.18
23	canada	0.000000	91	0.20
24	candidat	0.005685	93	0.20
25	candidature	-0.030801	122	0.27
26	capable	0.000000	104	0.23
27	capacité	0.000000	249	0.55
28	cg	-0.099408	127	0.28
29	chef	0.045641	88	0.19
30	choix	-0.000000	79	0.17
31	citoyen	0.000000	77	0.17
32	client	0.000000	83	0.18
33	collaboration	-0.015531	190	0.42
34	communication	0.000000	206	0.45
35	communiquer	0.018933	155	0.34
36	compositing	-0.111607	86	0.19
37	composition	-0.002332	82	0.18
38	compréhension	-0.038348	104	0.23
39	compétence	-0.004913	242	0.53
40	concept	-0.000000	78	0.17
41	concevoir	-0.000000	105	0.23
42	connaissance logiciel	-0.000000	98	0.21
43	contacter	-0.094189	102	0.22
44	couleur	-0.000000	75	0.16
45	créatif	0.028686	179	0.39
46	création	-0.000000	136	0.30
47	créativité	0.000000	90	0.20
48	créer	0.000000	250	0.55
49	cœur	-0.062569	95	0.21
50	date	-0.000000	86	0.19
51	demande	-0.000000	74	0.16
52	demander	0.000000	97	0.21
53	design	-0.065059	78	0.17

Suite à la page suivante

TABLE A.2 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
54	digital	-0.000000	105	0.23
55	diplôme	0.023198	190	0.42
56	diplôme étude	-0.000000	103	0.23
57	directeur	0.022608	122	0.27
58	direction	-0.017088	113	0.25
59	domaine	0.000000	73	0.16
60	dynamique	-0.016045	110	0.24
61	délai	0.000000	145	0.32
62	département	-0.000000	154	0.34
63	détail	0.071478	116	0.25
64	détenir	0.000000	80	0.18
65	déterminer	-0.000000	82	0.18
66	développement	-0.000000	102	0.22
67	développer	0.000000	130	0.29
68	effectuer	-0.000000	109	0.24
69	efficace	0.000000	75	0.16
70	emploi	-0.000000	104	0.23
71	envergure	0.000000	104	0.23
72	environnement	-0.000000	229	0.50
73	esprit	0.005840	113	0.25
74	excellent	0.009178	150	0.33
75	excellente	-0.000000	105	0.23
76	exigence	0.205264	118	0.26
77	exiger	-0.000000	77	0.17
78	fabrication	-0.000000	80	0.18
79	faire	-0.000000	200	0.44
80	film	-0.014297	176	0.39
81	fonction	0.000000	138	0.30
82	fournir	0.046111	75	0.16
83	français	0.006182	123	0.27
84	gestion	-0.000000	109	0.24
85	gestionnaire	0.000000	82	0.18
86	grand	0.040371	114	0.25
87	générique	-0.000000	81	0.18
88	générique masculin	-0.000000	81	0.18
89	gérer	0.000000	86	0.19

Suite à la page suivante

TABLE A.2 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
90	haut	-0.032802	93	0.20
91	image	-0.000000	159	0.35
92	industrie	0.000000	79	0.17
93	information	-0.043339	74	0.16
94	informatique	0.000000	76	0.17
95	jeu	-0.000000	123	0.27
96	jour	-0.000000	116	0.25
97	lead	-0.000000	94	0.21
98	lieu	-0.000000	83	0.18
99	linux	-0.000000	75	0.16
100	logiciel	-0.000000	251	0.55
101	logiciel maya	0.020575	76	0.17
102	long	0.000000	199	0.44
103	long métrage	0.000000	176	0.39
104	maintenir	0.000000	185	0.41
105	masculin	0.232056	100	0.22
106	masculin utiliser	0.000000	87	0.19
107	matière	0.000000	82	0.18
108	maya	0.000000	244	0.54
109	maîtrise	0.000000	128	0.28
110	meilleur	0.000000	82	0.18
111	mettre	0.000000	131	0.29
112	minimum	0.000000	163	0.36
113	montréal	0.083657	226	0.50
114	métrage	0.000000	176	0.39
115	métrage animation	-0.292715	93	0.20
116	niveau	-0.014775	93	0.20
117	nuke	-0.038701	116	0.25
118	numérique	-0.000000	95	0.21
119	nécessaire	-0.056598	151	0.33
120	offrir	0.000000	92	0.20
121	organisation	-0.000000	115	0.25
122	organiser	0.006246	74	0.16
123	outil	-0.006349	167	0.37
124	partie	-0.106646	82	0.18
125	passionner	0.196311	146	0.32

Suite à la page suivante

TABLE A.2 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
126	permanent	0.000000	95	0.21
127	personnage	-0.022010	117	0.26
128	pipeline	-0.000000	177	0.39
129	plan	0.000000	115	0.25
130	porter	-0.000000	79	0.17
131	posséder	0.000000	82	0.18
132	pouvoir	-0.047736	86	0.19
133	prendre	0.036513	92	0.20
134	preuve	-0.114455	91	0.20
135	principe	0.000000	96	0.21
136	priorité	0.000000	115	0.25
137	problème	-0.000000	208	0.46
138	processus	-0.014114	104	0.23
139	producteur	0.000000	118	0.26
140	production	-0.000000	353	0.77
141	produire	0.025872	76	0.17
142	professionnel	0.000000	86	0.19
143	profil	0.209637	79	0.17
144	projet	0.000000	317	0.70
145	propos	0.171335	120	0.26
146	python	0.005514	131	0.29
147	qualification	0.022759	180	0.39
148	qualité	-0.000000	217	0.48
149	recherche	0.000000	129	0.28
150	requérir	0.000000	83	0.18
151	respecter	0.000000	158	0.35
152	responsabilité	0.000000	237	0.52
153	responsable	-0.041222	153	0.34
154	rigging	0.000000	83	0.18
155	réalisateur	-0.077880	92	0.20
156	réaliser	-0.002514	127	0.28
157	référence	-0.000000	92	0.20
158	répondre	-0.000000	111	0.24
159	résident	-0.000000	91	0.20
160	résident permanent	0.000000	79	0.17
161	résolution	0.044485	79	0.17

Suite à la page suivante

TABLE A.2 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
162	résoudre	0.000000	106	0.23
163	résoudre problème	-0.000000	89	0.20
164	rôle	0.000000	108	0.24
165	salaire	0.000000	89	0.20
166	savoir	-0.048843	132	0.29
167	savoir faire	-0.000000	88	0.19
168	sens	-0.000000	134	0.29
169	service	0.120543	131	0.29
170	shotgun	-0.000000	119	0.26
171	solide	-0.000000	120	0.26
172	solution	-0.000000	121	0.27
173	statut	-0.207578	81	0.18
174	studio	0.085752	234	0.51
175	studio animation	0.000000	108	0.24
176	superviseur	0.000000	236	0.52
177	supervision	-0.000000	130	0.29
178	système	0.000000	77	0.17
179	série	0.129009	146	0.32
180	talent	0.000000	118	0.26
181	technique	0.061735	327	0.72
182	temps	0.104427	195	0.43
183	texte	0.000000	100	0.22
184	travail	-0.028789	342	0.75
185	travaille	0.000000	85	0.19
186	travailler	0.000000	326	0.71
187	travailler équipe	-0.066548	99	0.22
188	tâche	-0.063653	179	0.39
189	utiliser	-0.028592	185	0.41
190	vfx	-0.046373	81	0.18
191	visuel	-0.000000	251	0.55
192	vouloir	0.367494	130	0.29
193	échéancier	-0.000000	129	0.28
194	éclairage	-0.028756	116	0.25
195	élément	0.028125	101	0.22
196	équipe production	0.063591	80	0.18
197	établir	0.057217	152	0.33

Suite à la page suivante

TABLE A.2 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Importance</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
198	étroite	0.000000	84	0.18
199	étroite collaboration	0.000000	81	0.18
200	étude	-0.008233	114	0.25

## A.2 Moindres carrés ordinaires

Les tableaux obtenus par la méthode des MCO comprennent les 200 mots choisis par l'algorithme ainsi qu'une constante. La colonne « Paramètres » rapporte le coefficient de régression, la colonne « p-value » représente la p-value associée au coefficient de ce mot, la colonne « Fréquence » rapporte le nombre d'annonces où le mot est présent et la colonne « Fréquence relative » rapporte la fréquence relative de la présence des mots dans toutes les offres d'emploi. Les tableaux A.3 et A.4 présentent la version complète des tableaux de la méthode des MCO présentés au chapitre 4.

TABLE A.3 – Estimation par MCO avec l'année comme variable dépendante

	<b>Mots fréquents</b>	<b>Paramètres</b>	<b>p-value</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
0	const	-1.852	0.000	0	0.00
1	amélioration	-0.028	0.921	76	0.17
2	an	-0.043	0.855	243	0.53
3	an expérience	0.023	0.928	197	0.43
4	anglais	-0.466	0.123	111	0.24
5	animation	0.272	0.156	347	0.76
6	animation studio	-0.419	0.466	108	0.24
7	année	-0.895	0.025	128	0.28
8	année expérience	1.001	0.016	111	0.24
9	aptitude	0.118	0.501	85	0.19
10	artiste	-0.011	0.942	314	0.69
11	artistique	-0.133	0.440	206	0.45
12	asset	-0.282	0.161	103	0.23
13	assurer	0.474	0.002	160	0.35
14	atelier	0.458	0.632	73	0.16
15	atout	0.004	0.975	264	0.58

*Suite à la page suivante*

TABLE A.3 – Suite de la page précédente

	Mots fréquents	Paramètres	p-value	Fréquence	Fréquence relative
16	baser	0.101	0.710	90	0.20
17	bel	0.655	0.051	77	0.17
18	besoin	-0.047	0.782	123	0.27
19	bon	-0.136	0.472	310	0.68
20	bon connaissance	0.153	0.368	188	0.41
21	but	0.056	0.865	98	0.21
22	but texte	-0.667	0.611	82	0.18
23	canada	-0.240	0.427	91	0.20
24	candidat	0.476	0.030	93	0.20
25	candidature	-0.345	0.212	122	0.27
26	capable	-0.316	0.068	104	0.23
27	capacité	0.211	0.150	249	0.55
28	cg	-0.370	0.024	127	0.28
29	chef	0.178	0.296	88	0.19
30	choix	-0.259	0.538	79	0.17
31	citoyen	1.183	0.142	77	0.17
32	client	0.295	0.198	83	0.18
33	collaboration	-0.332	0.055	190	0.42
34	communication	-0.118	0.502	206	0.45
35	communiquer	-0.035	0.827	155	0.34
36	compositing	-0.141	0.486	86	0.19
37	composition	-0.235	0.247	82	0.18
38	compréhension	-0.337	0.067	104	0.23
39	compétence	0.299	0.054	242	0.53
40	concept	0.666	0.008	78	0.17
41	concevoir	0.034	0.870	105	0.23
42	connaissance logiciel	0.432	0.039	98	0.21
43	contacter	-0.425	0.301	102	0.22
44	couleur	0.122	0.624	75	0.16
45	créatif	0.691	0.000	179	0.39
46	création	-0.136	0.375	136	0.30
47	créativité	-0.236	0.354	90	0.20
48	créer	-0.038	0.804	250	0.55
49	cœur	0.368	0.252	95	0.21
50	date	-0.248	0.196	86	0.19
51	demande	-0.536	0.023	74	0.16

Suite à la page suivante

TABLE A.3 – Suite de la page précédente

	Mots fréquents	Paramètres	p-value	Fréquence	Fréquence relative
52	demander	-0.473	0.083	97	0.21
53	design	-0.335	0.148	78	0.17
54	digital	-0.319	0.329	105	0.23
55	diplôme	0.036	0.863	190	0.42
56	diplôme étude	0.836	0.109	103	0.23
57	directeur	-0.084	0.564	122	0.27
58	direction	-0.107	0.520	113	0.25
59	domaine	0.427	0.015	73	0.16
60	dynamique	-0.057	0.783	110	0.24
61	délai	-0.414	0.015	145	0.32
62	département	0.266	0.094	154	0.34
63	détail	0.191	0.293	116	0.25
64	détenir	0.242	0.381	80	0.18
65	déterminer	-0.432	0.015	82	0.18
66	développement	-0.110	0.558	102	0.22
67	développer	0.267	0.108	130	0.29
68	effectuer	-0.344	0.049	109	0.24
69	efficace	0.350	0.074	75	0.16
70	emploi	0.368	0.200	104	0.23
71	envergure	1.267	0.000	104	0.23
72	environnement	0.150	0.358	229	0.50
73	esprit	0.579	0.001	113	0.25
74	excellent	-0.033	0.817	150	0.33
75	excellente	0.040	0.820	105	0.23
76	exigence	0.265	0.135	118	0.26
77	exiger	-0.609	0.186	77	0.17
78	fabrication	-0.229	0.526	80	0.18
79	faire	0.291	0.076	200	0.44
80	film	-0.191	0.246	176	0.39
81	fonction	-0.203	0.280	138	0.30
82	fournir	0.100	0.589	75	0.16
83	français	0.410	0.144	123	0.27
84	gestion	-0.294	0.166	109	0.24
85	gestionnaire	0.002	0.995	82	0.18
86	grand	0.182	0.304	114	0.25
87	générique	0.479	0.526	81	0.18

Suite à la page suivante

TABLE A.3 – Suite de la page précédente

	Mots fréquents	Paramètres	p-value	Fréquence	Fréquence relative
88	générique masculin	0.479	0.526	81	0.18
89	gérer	-0.121	0.504	86	0.19
90	haut	-0.088	0.684	93	0.20
91	image	0.074	0.672	159	0.35
92	industrie	-0.368	0.057	79	0.17
93	information	-0.411	0.083	74	0.16
94	informatique	-0.217	0.376	76	0.17
95	jeu	0.058	0.801	123	0.27
96	jour	0.086	0.664	116	0.25
97	lead	-0.058	0.738	94	0.21
98	lieu	0.194	0.602	83	0.18
99	linux	-0.043	0.863	75	0.16
100	logiciel	0.014	0.933	251	0.55
101	logiciel maya	0.156	0.453	76	0.17
102	long	0.884	0.002	199	0.44
103	long métrage	-0.448	0.010	176	0.39
104	maintenir	-0.367	0.048	185	0.41
105	masculin	3.034	0.003	100	0.22
106	masculin utiliser	-0.171	0.812	87	0.19
107	matière	0.141	0.645	82	0.18
108	maya	0.069	0.689	244	0.54
109	maîtrise	-0.007	0.966	128	0.28
110	meilleur	0.349	0.172	82	0.18
111	mettre	-0.006	0.971	131	0.29
112	minimum	-0.180	0.263	163	0.36
113	montréal	-0.137	0.522	226	0.50
114	métrage	-0.448	0.010	176	0.39
115	métrage animation	-0.488	0.142	93	0.20
116	niveau	0.110	0.559	93	0.20
117	nuke	-0.369	0.031	116	0.25
118	numérique	0.124	0.530	95	0.21
119	nécessaire	-0.108	0.502	151	0.33
120	offrir	0.539	0.008	92	0.20
121	organisation	-0.378	0.049	115	0.25
122	organiser	-0.212	0.280	74	0.16
123	outil	0.281	0.131	167	0.37

Suite à la page suivante

TABLE A.3 – Suite de la page précédente

	Mots fréquents	Paramètres	p-value	Fréquence	Fréquence relative
124	partie	-0.217	0.396	82	0.18
125	passionner	0.575	0.021	146	0.32
126	permanent	0.113	0.759	95	0.21
127	personnage	-0.022	0.905	117	0.26
128	pipeline	-0.371	0.053	177	0.39
129	plan	-0.025	0.876	115	0.25
130	porter	-1.306	0.056	79	0.17
131	posséder	-0.040	0.845	82	0.18
132	pouvoir	-0.302	0.143	86	0.19
133	prendre	0.137	0.478	92	0.20
134	preuve	-0.603	0.009	91	0.20
135	principe	-0.099	0.722	96	0.21
136	priorité	0.085	0.684	115	0.25
137	problème	-0.399	0.032	208	0.46
138	processus	-0.374	0.027	104	0.23
139	producteur	-0.106	0.678	118	0.26
140	production	-0.047	0.783	353	0.77
141	produire	0.170	0.396	76	0.17
142	professionnel	0.073	0.783	86	0.19
143	profil	1.031	0.000	79	0.17
144	projet	0.206	0.219	317	0.70
145	propos	1.297	0.000	120	0.26
146	python	0.280	0.116	131	0.29
147	qualification	0.700	0.001	180	0.39
148	qualité	-0.230	0.114	217	0.48
149	recherche	-0.072	0.656	129	0.28
150	requérir	-0.190	0.361	83	0.18
151	respecter	0.335	0.062	158	0.35
152	responsabilité	0.537	0.005	237	0.52
153	responsable	-0.113	0.480	153	0.34
154	rigging	-0.220	0.329	83	0.18
155	réalisateur	-0.456	0.035	92	0.20
156	réaliser	0.170	0.417	127	0.28
157	référence	-0.336	0.212	92	0.20
158	répondre	0.275	0.193	111	0.24
159	résident	-1.416	0.010	91	0.20

Suite à la page suivante

TABLE A.3 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Paramètres</b>	<b>p-value</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
160	résident permanent	1.095	0.290	79	0.17
161	résolution	-0.013	0.949	79	0.17
162	résoudre	0.482	0.204	106	0.23
163	résoudre problème	-0.154	0.697	89	0.20
164	rôle	0.383	0.085	108	0.24
165	salaire	0.146	0.694	89	0.20
166	savoir	-0.211	0.381	132	0.29
167	savoir faire	0.323	0.484	88	0.19
168	sens	-0.047	0.786	134	0.29
169	service	0.518	0.043	131	0.29
170	shotgun	0.602	0.007	119	0.26
171	solide	0.163	0.325	120	0.26
172	solution	-0.167	0.337	121	0.27
173	statut	-0.032	0.933	81	0.18
174	studio	0.314	0.178	234	0.51
175	studio animation	-0.918	0.021	108	0.24
176	superviseur	-0.153	0.359	236	0.52
177	supervision	-0.154	0.436	130	0.29
178	système	0.184	0.303	77	0.17
179	série	0.439	0.030	146	0.32
180	talent	-0.444	0.130	118	0.26
181	technique	-0.192	0.232	327	0.72
182	temps	0.404	0.030	195	0.43
183	texte	-2.367	0.007	100	0.22
184	travail	0.035	0.836	342	0.75
185	travaille	0.389	0.071	85	0.19
186	travailler	-0.063	0.706	326	0.71
187	travailler équipe	-0.099	0.615	99	0.22
188	tâche	0.062	0.714	179	0.39
189	utiliser	-0.219	0.227	185	0.41
190	vfx	0.023	0.913	81	0.18
191	visuel	-0.423	0.004	251	0.55
192	vouloir	1.026	0.000	130	0.29
193	échéancier	0.143	0.466	129	0.28
194	éclairage	-0.093	0.594	116	0.25
195	élément	0.425	0.016	101	0.22

Suite à la page suivante

TABLE A.3 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Paramètres</b>	<b>p-value</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
196	équipe production	0.038	0.839	80	0.18
197	établir	0.053	0.766	152	0.33
198	étroite	0.415	0.609	84	0.18
199	étroite collaboration	-0.294	0.727	81	0.18
200	étude	-0.812	0.089	114	0.25

TABLE A.4 – Estimation par MCO avec l'année comme indicatrice

	<b>Mots fréquents</b>	<b>Paramètres</b>	<b>p-value</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
0	const	0.075	0.409	0	0.00
1	amélioration	0.117	0.255	76	0.17
2	an	0.010	0.907	243	0.53
3	an expérience	-0.082	0.389	197	0.43
4	anglais	-0.185	0.095	111	0.24
5	animation	0.055	0.436	347	0.76
6	animation studio	-0.153	0.471	108	0.24
7	année	0.099	0.497	128	0.28
8	année expérience	-0.042	0.783	111	0.24
9	aptitude	0.016	0.802	85	0.19
10	artiste	0.070	0.198	314	0.69
11	artistique	-0.033	0.604	206	0.45
12	asset	-0.043	0.562	103	0.23
13	assurer	0.046	0.414	160	0.35
14	atelier	-0.232	0.509	73	0.16
15	atout	0.014	0.766	264	0.58
16	baser	0.080	0.420	90	0.20
17	bel	0.053	0.666	77	0.17
18	besoin	0.011	0.861	123	0.27
19	bon	0.079	0.253	310	0.68
20	bon connaissance	0.052	0.405	188	0.41
21	but	-0.362	0.003	98	0.21
22	but texte	0.556	0.249	82	0.18
23	canada	0.055	0.619	91	0.20

Suite à la page suivante

TABLE A.4 – Suite de la page précédente

	Mots fréquents	Paramètres	p-value	Fréquence	Fréquence relative
24	candidat	0.115	0.151	93	0.20
25	candidature	-0.288	0.005	122	0.27
26	capable	-0.033	0.609	104	0.23
27	capacité	-0.026	0.624	249	0.55
28	cg	-0.154	0.011	127	0.28
29	chef	0.108	0.085	88	0.19
30	choix	-0.089	0.566	79	0.17
31	citoyen	-0.309	0.298	77	0.17
32	client	-0.042	0.617	83	0.18
33	collaboration	-0.122	0.055	190	0.42
34	communication	-0.078	0.232	206	0.45
35	communiquer	0.011	0.853	155	0.34
36	compositing	-0.098	0.186	86	0.19
37	composition	-0.009	0.903	82	0.18
38	compréhension	-0.191	0.005	104	0.23
39	compétence	0.062	0.277	242	0.53
40	concept	0.061	0.508	78	0.17
41	concevoir	-0.014	0.853	105	0.23
42	connaissance logiciel	0.007	0.924	98	0.21
43	contacter	-0.170	0.260	102	0.22
44	couleur	0.117	0.200	75	0.16
45	créatif	0.220	0.001	179	0.39
46	création	0.004	0.946	136	0.30
47	créativité	-0.066	0.477	90	0.20
48	créer	0.037	0.511	250	0.55
49	cœur	-0.120	0.309	95	0.21
50	date	0.002	0.973	86	0.19
51	demande	-0.055	0.526	74	0.16
52	demander	0.043	0.664	97	0.21
53	design	-0.230	0.007	78	0.17
54	digital	0.006	0.957	105	0.23
55	diplôme	0.036	0.638	190	0.42
56	diplôme étude	0.199	0.298	103	0.23
57	directeur	0.100	0.062	122	0.27
58	direction	-0.032	0.595	113	0.25
59	domaine	-0.052	0.420	73	0.16

Suite à la page suivante

TABLE A.4 – Suite de la page précédente

	Mots fréquents	Paramètres	p-value	Fréquence	Fréquence relative
60	dynamique	-0.053	0.482	110	0.24
61	délai	-0.156	0.012	145	0.32
62	département	0.054	0.350	154	0.34
63	détail	0.020	0.765	116	0.25
64	détenir	0.117	0.250	80	0.18
65	déterminer	-0.154	0.019	82	0.18
66	développement	-0.090	0.194	102	0.22
67	développer	0.089	0.146	130	0.29
68	effectuer	-0.047	0.467	109	0.24
69	efficace	0.061	0.392	75	0.16
70	emploi	0.031	0.771	104	0.23
71	envergure	0.081	0.507	104	0.23
72	environnement	0.040	0.502	229	0.50
73	esprit	0.119	0.074	113	0.25
74	excellent	-0.015	0.776	150	0.33
75	excellente	-0.017	0.789	105	0.23
76	exigence	0.226	0.001	118	0.26
77	exiger	0.059	0.727	77	0.17
78	fabrication	0.027	0.839	80	0.18
79	faire	0.073	0.228	200	0.44
80	film	-0.036	0.555	176	0.39
81	fonction	-0.029	0.671	138	0.30
82	fournir	0.040	0.562	75	0.16
83	français	0.227	0.028	123	0.27
84	gestion	-0.038	0.623	109	0.24
85	gestionnaire	0.120	0.342	82	0.18
86	grand	0.086	0.189	114	0.25
87	générique	-0.670	0.016	81	0.18
88	générique masculin	-0.670	0.016	81	0.18
89	gérer	0.019	0.779	86	0.19
90	haut	0.024	0.765	93	0.20
91	image	0.019	0.769	159	0.35
92	industrie	-0.101	0.156	79	0.17
93	information	-0.206	0.018	74	0.16
94	informatique	0.041	0.650	76	0.17
95	jeu	-0.074	0.383	123	0.27

Suite à la page suivante

TABLE A.4 – Suite de la page précédente

	Mots fréquents	Paramètres	p-value	Fréquence	Fréquence relative
96	jour	0.081	0.269	116	0.25
97	lead	0.051	0.429	94	0.21
98	lieu	-0.075	0.583	83	0.18
99	linux	0.010	0.917	75	0.16
100	logiciel	0.032	0.607	251	0.55
101	logiciel maya	0.070	0.357	76	0.17
102	long	0.165	0.106	199	0.44
103	long métrage	-0.007	0.909	176	0.39
104	maintenir	-0.052	0.445	185	0.41
105	masculin	0.585	0.120	100	0.22
106	masculin utiliser	0.667	0.012	87	0.19
107	matière	0.076	0.499	82	0.18
108	maya	-0.004	0.944	244	0.54
109	maîtrise	-0.020	0.727	128	0.28
110	meilleur	0.052	0.580	82	0.18
111	mettre	0.020	0.746	131	0.29
112	minimum	-0.068	0.249	163	0.36
113	montréal	0.045	0.570	226	0.50
114	métrage	-0.007	0.909	176	0.39
115	métrage animation	-0.494	0.000	93	0.20
116	niveau	-0.093	0.178	93	0.20
117	nuke	-0.117	0.062	116	0.25
118	numérique	0.029	0.693	95	0.21
119	nécessaire	0.018	0.766	151	0.33
120	offrir	0.107	0.145	92	0.20
121	organisation	-0.113	0.108	115	0.25
122	organiser	-0.076	0.293	74	0.16
123	outil	0.064	0.346	167	0.37
124	partie	-0.130	0.168	82	0.18
125	passionner	0.179	0.050	146	0.32
126	permanent	0.015	0.912	95	0.21
127	personnage	-0.033	0.623	117	0.26
128	pipeline	-0.001	0.986	177	0.39
129	plan	-0.030	0.604	115	0.25
130	porter	-0.067	0.788	79	0.17
131	posséder	-0.026	0.731	82	0.18

Suite à la page suivante

TABLE A.4 – Suite de la page précédente

	Mots fréquents	Paramètres	p-value	Fréquence	Fréquence relative
132	pouvoir	-0.005	0.951	86	0.19
133	prendre	0.068	0.336	92	0.20
134	preuve	-0.202	0.017	91	0.20
135	principe	0.217	0.036	96	0.21
136	priorité	0.118	0.125	115	0.25
137	problème	-0.033	0.627	208	0.46
138	processus	-0.044	0.478	104	0.23
139	producteur	-0.015	0.873	118	0.26
140	production	-0.100	0.110	353	0.77
141	produire	0.030	0.683	76	0.17
142	professionnel	0.018	0.854	86	0.19
143	profil	0.254	0.009	79	0.17
144	projet	0.060	0.331	317	0.70
145	propos	0.213	0.097	120	0.26
146	python	0.059	0.369	131	0.29
147	qualification	0.036	0.627	180	0.39
148	qualité	-0.030	0.579	217	0.48
149	recherche	0.036	0.549	129	0.28
150	requérir	0.036	0.634	83	0.18
151	respecter	-0.028	0.674	158	0.35
152	responsabilité	0.060	0.388	237	0.52
153	responsable	0.007	0.906	153	0.34
154	rigging	-0.080	0.338	83	0.18
155	réalisateur	-0.236	0.003	92	0.20
156	réaliser	-0.086	0.264	127	0.28
157	référence	0.010	0.918	92	0.20
158	répondre	0.031	0.691	111	0.24
159	résident	-0.310	0.122	91	0.20
160	résident permanent	0.541	0.155	79	0.17
161	résolution	0.077	0.299	79	0.17
162	résoudre	-0.002	0.988	106	0.23
163	résoudre problème	0.010	0.946	89	0.20
164	rôle	-0.001	0.991	108	0.24
165	salaire	0.227	0.099	89	0.20
166	savoir	-0.171	0.053	132	0.29
167	savoir faire	-0.027	0.871	88	0.19

Suite à la page suivante

TABLE A.4 – Suite de la page précédente

	<b>Mots fréquents</b>	<b>Paramètres</b>	<b>p-value</b>	<b>Fréquence</b>	<b>Fréquence relative</b>
168	sens	0.110	0.086	134	0.29
169	service	0.350	0.000	131	0.29
170	shotgun	0.158	0.054	119	0.26
171	solide	-0.027	0.656	120	0.26
172	solution	-0.139	0.030	121	0.27
173	statut	-0.104	0.460	81	0.18
174	studio	-0.001	0.992	234	0.51
175	studio animation	0.093	0.525	108	0.24
176	superviseur	0.082	0.182	236	0.52
177	supervision	-0.051	0.486	130	0.29
178	système	0.004	0.948	77	0.17
179	série	0.171	0.022	146	0.32
180	talent	-0.055	0.607	118	0.26
181	technique	0.034	0.566	327	0.72
182	temps	0.054	0.424	195	0.43
183	texte	-0.165	0.608	100	0.22
184	travail	-0.066	0.285	342	0.75
185	travaille	0.070	0.375	85	0.19
186	travailler	0.070	0.250	326	0.71
187	travailler équipe	-0.108	0.136	99	0.22
188	tâche	-0.040	0.523	179	0.39
189	utiliser	-0.116	0.083	185	0.41
190	vfx	-0.122	0.118	81	0.18
191	visuel	0.017	0.752	251	0.55
192	vouloir	0.550	0.000	130	0.29
193	échancier	-0.041	0.569	129	0.28
194	éclairage	-0.084	0.191	116	0.25
195	élément	0.101	0.118	101	0.22
196	équipe production	0.139	0.044	80	0.18
197	établir	0.075	0.252	152	0.33
198	étroite	0.106	0.723	84	0.18
199	étroite collaboration	-0.109	0.725	81	0.18
200	étude	-0.312	0.076	114	0.25