

Tableaux de bord interactifs pour l'analyse de l'attrition à partir des motifs affectant la survie

Youssef Oubelmouh^{*,**}, Frédéric Fargon^{*}, Cyril de Runz^{**},
Arnaud Soulet^{**}, Cyril Veillon^{*}

^{*}Devoteam, Levallois-Perret, prenom.nom@devoteam.com

^{**}BDTLN, LIFAT, Université de Tours, Blois, prenom.nom@univ-tours.fr

Résumé. La rétention des employés est un enjeu majeur pour les entreprises. Nos travaux nous ont permis de développer une méthode permettant d'identifier et d'extraire parmi les motifs fréquents issus de la trajectoire des employés ceux qui amélioreraient leur rétention. Cet article présente la valorisation de ces motifs à travers un outil interactif permettant de les exploiter et de les visualiser. Ces derniers sont présentés sous la forme de tableaux de bord interactifs dans l'objectif d'être exploitables par les responsables RH de l'entreprise. Ils offrent une visualisation globale de l'attrition, une identification des employés à risque de démissionner et une simulation de l'impact d'une intervention.

1 Introduction

L'attrition des employés constitue un défi majeur pour les entreprises, notamment dans le secteur technologique et les services du numérique, où les taux de turnover atteignent parfois 25% selon certains rapports (Morrison, 2021). Ce phénomène implique des coûts importants. Cependant, la majorité des travaux récents dans le domaine de l'analyse de données appliquée à l'attrition négligent souvent la dimension temporelle, comme l'historique des changements vécus par les employés. Afin de mieux comprendre cet aspect temporel, nos recherches récentes ont combiné des techniques d'exploration de motifs séquentiels et d'analyse de survie, permettant d'identifier des motifs séquentiels impactant la survie, appelés les survival-changing patterns (Oubelmouh et al., 2023). L'objectif maintenant est de mettre à disposition ces résultats aux responsables RH. La méthode la plus pertinente selon nous pour les valoriser est de développer des tableaux de bord interactifs. Ces derniers en effet permettent à des non experts en informatique de visualiser les résultats et permettent également de les exploiter de manière concrète facilitant ainsi la prise de décisions en matière de rétention des employés.

La construction de tableaux de bord interactifs peut être réalisée de plusieurs façons, notamment de manière automatique (Soni et al., 2024), mais ces approches négligent souvent l'intégration des motifs séquentiels. Par ailleurs, bien que certains outils de visualisation existent, ils ne prennent pas en compte les informations de survie (Guo et al., 2021), qui sont pour nous une composante essentielle. C'est pourquoi nous proposons un outil complet et innovant permettant de répondre efficacement à ces besoins tout en offrant une vue d'ensemble cohérente.

Dans ce qui suit, nous décrivons d'abord les personas et cas d'usages ciblés, suivis par la conception des tableaux de bord et l'identification des motifs représentant le mieux chaque

Tableaux de bord à partir des motifs affectant la survie

employé. Enfin, nous illustrons l'utilisation de notre approche par un exemple concret. A noter que les données utilisées dans cet article et nos tableaux de bord sont entièrement fictives. Elles ont été créées de toutes pièces et ne reflètent en aucun cas la réalité de l'entreprise afin de respecter le RGPD. Ainsi, toutes les valeurs et toutes les informations mentionnées ne doivent pas être interprétées comme représentant des situations réelles au sein de l'entreprise.

2 Cas d'usages

Les tableaux de bord seront à destination du cabinet de conseil Devoteam. C'est une entreprise internationale spécialisée dans le conseil en technologies et en transformation numérique. Fondée en 1995 et présente dans plusieurs pays à travers l'Europe, le Moyen-Orient et l'Afrique, elle dispose d'un effectif d'un peu plus de 10 000 employés dont 5 000 en France. D'où la nécessité de surveiller et d'analyser l'attrition. Dans le paragraphe suivant sont décrites les fonctions des utilisateurs de Devoteam qui utiliseront les tableaux de bord.

Les utilisateurs des tableaux de bord sont principalement les responsables RH et les managers, répartis en trois groupes :

- **Managers/Directions** : Ils ont besoin d'une version simplifiée des tableaux de bord pour visualiser les principaux indicateurs clés de performance comme le taux d'attrition et prendre des décisions stratégiques. Pour cela, le premier tableau doit donc permettre de filtrer et d'analyser les taux d'attrition globaux.
- **RH Analyste** : Chargés de suivre les tendances globales d'attrition et de satisfaction, ils analysent aussi les indicateurs clés de performance comme l'attrition, mais identifient également les groupes à risque, et évaluent les politiques RH. Le deuxième tableau doit donc permettre de filtrer par groupe d'employés et d'identifier dans ces groupes quels sont les employés ayant le plus de probabilité de démissionner à court ou long terme.
- **RH Suivi des employés** : Directement en contact avec les employés, ils mènent des entretiens réguliers pour identifier les employés à risque et faire des entretiens avec eux dans l'objectif de les retenir. Aussi, un nouveau tableau de bord pourrait être mis en place dans le but de simuler une intervention spécifique afin de visualiser l'effet de cette décision.

Ces différents profils nécessitent des tableaux de bord variés, adaptés à leurs besoins spécifiques, allant de l'analyse approfondie à une vue d'ensemble simplifiée.

3 Identification du motif le plus représentatif pour un employé

Dans nos travaux précédents (Oubelmouh et al., 2023), nous avons plusieurs résultats sur les motifs affectant la survie. Toutefois, pour que ces résultats puissent être exploités de manière pertinente et appliqués efficacement à chaque employé, une étape supplémentaire est nécessaire : il nous faut pour chaque employé identifier, parmi les motifs inclus à la fois dans sa séquence et à la fois dans nos résultats, celui qui représente le mieux cette séquence (ou cet employé).

La sélection du motif le plus représentatif dans une séquence est essentielle car nos résultats concernent uniquement des motifs fréquents, et non les séquences complètes des employés. Il est donc nécessaire de définir des critères objectifs pour hiérarchiser ces motifs et identifier celui qui représente le mieux les événements spécifiques à chaque employé. Par exemple en supposant que l'employé possède la séquence suivante (en remplaçant les événements par des lettres) : $\langle (Age_{35-39}, Male, L5), C, J2, C \rangle$ avec $L5$ représentant la localisation, C un changement de salaire et $J2$ un changement de position dans l'entreprise. Dans notre jeu de données fictif, parmi tous les motifs fréquents qui ont été extraits et qui possèdent au moins un motif affectant la survie, 8 sont contenus dans cette séquence. Donc l'objectif est de déterminer parmi ces 8 motifs, sur lequel l'utilisateur doit se baser pour simuler les futures interventions.

La sélection du motif le plus représentatif repose sur plusieurs critères répartis en deux catégories : critères syntaxiques (portant sur la forme de la séquence) et critères sémantiques (portant sur l'évaluation du motif au sein du jeu de données).

Critères syntaxiques Cette première catégorie vise à favoriser les motifs ressemblant le plus possible à la trajectoire de l'employé considéré avec la longueur et le positionnement du dernier événement. Le premier critère à considérer est la longueur du motif (dénotée par *Length*). Un motif plus long peut capturer plus d'informations sur le comportement de l'employé, incluant des événements successifs ou non. Les motifs plus longs sont souvent plus descriptifs et peuvent contenir une diversité d'éléments reflétant la complexité de la séquence. Cependant, un motif plus long a moins de chance d'être le contexte d'un motif affectant la survie.

Ensuite, un deuxième critère est la position du dernier événement du motif (dénotée par *Latest*), qui met en évidence l'importance des actions récentes. Un motif incluant un événement en fin de séquence peut être particulièrement pertinent pour comprendre les dernières actions influençant le comportement de l'employé. Pour cela nous allons compter le nombre de jours séparant l'apparition du dernier élément du motif avec la date du jour.

Critères sémantiques De manière classique, nous bénéficions aussi des mesures d'intérêt évaluées sur le jeu de données où le motif séquentiel a été extrait : fréquence et moyenne du gain de survie. Premièrement, nous avons la fréquence du motif (dénotée par *Freq*) dans l'ensemble des motifs fréquents qui peut être intéressant à avoir en tant que critère. En effet, elle permet d'indiquer des comportements récurrents au sein de l'entreprise. Un motif fréquent est donc représentatif d'une dynamique typique pour des interventions habituelles.

Deuxièmement, un autre critère important est la fréquence à laquelle un motif joue le rôle de contexte pour des motifs affectant la survie (dénotée par *ContFreq*). Cela permet d'identifier les motifs qui sont régulièrement les bases qui permettent d'identifier les motifs qui influencent fortement la probabilité de rétention ou de départ d'un employé.

Enfin, le dernier critère est la moyenne de gain de survie associée à un motif en tant que contexte (dénotée par *AvgGain*). Cela permet de mesurer à quel point un motif en tant que point de départ contribue à prolonger la durée de rétention d'un employé. Ainsi, les motifs avec un gain de survie élevé sont favorisés, car ils indiquent qu'il existe des motifs améliorant grandement cette durée.

Fonction de score La flexibilité des critères permet d'adapter la sélection des motifs aux cas d'usages. Pour cela, nous avons mis en place un score multi-critères, où chaque motif reçoit

Tableaux de bord à partir des motifs affectant la survie

un score en fonction de sa performance sur les différents critères, standardisés et pondérés selon leur importance. Plus précisément, la fonction de score prend en compte la longueur du motif, sa fréquence, la position de l'événement final, sa fréquence comme contexte de motifs affectant la survie, et la moyenne de gain de survie. La pondération des critères est ajustée selon les besoins de l'analyse, assurant ainsi une solution cohérente et adaptée à chaque contexte d'application.

Definition 1 (Fonction de score) *Compte tenu de tous les critères d'un motif X , son score correspond à la somme pondérée des valeurs des critères ayant été standardisée :*

$$Score(X) = 0.2 \times Length(X) + 0.2 \times Latest(X) + 0.1 \times Freq(X) + 0.2 \times ContFreq(X) + 0.2 \times AvgGain(X)$$

Voici le Tableau 1 qui récapitule les valeurs obtenues pour chacun des motifs candidats pour la représentation de la séquence $\langle (Age35 - 39, Male, L5), C, J2, C \rangle$:

Motifs candidats	<i>Freq</i>	<i>Length</i>	<i>Latest</i>	<i>ContFreq</i>	<i>AvgGain</i>	Score
$\langle Male \rangle$	500	1	474	213	0.068	2.361
$\langle J2 \rangle$	706	1	232	30	0.102	1.804
$\langle C, J2 \rangle$	641	2	232	18	0.092	1.634
$\langle L5, C, J2 \rangle$	388	3	232	25	0.081	1.338
$\langle L5, J2 \rangle$	426	2	232	32	0.075	1.124
$\langle Male, C, J2 \rangle$	317	3	232	19	0.074	1.064
$\langle (Male, L5) \rangle$	295	1	474	114	0.048	0.864
$\langle (Male, L5), C, J2 \rangle$	186	3	232	6	0.036	0.052

TAB. 1 – Exemple de la sélection du motif représentant le mieux une séquence.

En conclusion, la sélection du motif le plus représentatif dans une séquence est un problème crucial pour la transposition des résultats globaux (motifs fréquents) à des cas spécifiques (employé choisi). Grâce à l'approche par fonction de score, nous sommes en mesure d'ajuster notre sélection selon différents critères de performance et de pertinence, tout en restant alignés sur les besoins du métier et les spécificités des séquences observées.

4 Conception de l'interface

Cette section détaille les différents tableaux de bord de l'outil. Il est à noter que les couleurs utilisées sont celles de l'identité visuelle de l'entreprise. Le front-end de l'outil a été développé à l'aide de React, tandis que nous avons utilisé Node.js et Express pour le back-end.

4.1 Suivi de l'évolution de l'attrition

Ce tableau de bord interactif est conçu pour répondre au premier cas d'usage : le suivi de l'évolution de l'attrition pour différents groupes d'employés. Il permet de fournir une vue d'ensemble complète sur l'attrition des employés en fonction de divers critères. La Figure 1 est une capture d'écran du tableau de bord représentant ce cas d'usage.

L'interface dédiée à ce tableau de bord comprend des visualisations claires et intuitives des indicateurs clés de performance (Figure 1(a)), tels que le taux d'attrition global, le nombre



FIG. 1 – Tableau de bord pour le suivi global de l'attrition.

de départs ou encore le nombre d'employés encore actif. Les filtres situés en haut de la page (Figure 1(b)) permettent aux utilisateurs de personnaliser l'analyse en fonction de la période temporelle, de la localisation, de la tranche d'âge, et du genre des employés. Cela leur permet de segmenter précisément les données en fonction des besoins spécifiques de l'analyse.

L'interface comprend différents composants visuels (Figure 1(c)) qui permettent une meilleure compréhension des dynamiques de l'attrition. Par exemple, les graphiques présentent des données segmentées selon les départements, les genres, les tranches d'âge et les types de poste (jobs). Ainsi, il est possible d'identifier, par exemple, quels départements ou quels groupes démographiques présentent des taux d'attrition plus élevés.

Sur la partie droite de l'interface (Figure 1(d)), une liste présente des informations détaillées sur les 5 derniers employés ayant quitté l'entreprise. Elle permet aux utilisateurs d'examiner les caractéristiques des départs récents et de détecter des tendances potentielles qui peuvent nécessiter des actions correctives.

Ce tableau de bord se concentre sur le suivi de l'attrition, offrant des visualisations claires, interactives et granulaires pour analyser les sous-groupes définis par l'utilisateur. Grâce à ses filtres dynamiques, il aide à identifier les groupes avec des taux de départ anormalement élevés. Le tableau de bord permet de répondre à la problématique de la surveillance continue de l'attrition.

4.2 Identification des employés à risque

Ce tableau de bord interactif est conçu pour répondre au deuxième cas d'usage : l'identification des employés présentant le plus grand risque de quitter l'entreprise. La Figure 2 est une capture d'écran de l'interface représentant ce cas d'usage.

La section principale de l'interface (Figure 2(a)) présente une liste d'employés ordonnée par leur probabilité de départ décroissant. Chaque ligne de la table fournit des informations détaillées sur chaque employé, telles que l'identifiant, la probabilité de départ, le rang, le dé-

Tableaux de bord à partir des motifs affectant la survie



ID	Probabilité de départ	Rang	Département	Genre	Âge d'entrée	Pays
202	0.973	Lead consultant	Operations	Male	Age25-29	FR
607	0.965	Lead consultant	IRM 2	Male	Age30-34	FR
727	0.885	Lead consultant	IRM 2	Male	Age20-24	FR
911	0.987	Consultant	Lyon Operations	Male	Age35-39	FR
784	0.981	No rank	Operations	Male	Age20-24	FR

FIG. 2 – Tableau de bord pour l'identification des employés à risque.

partement, le genre, l'âge d'entrée dans l'entreprise, ainsi que le pays. Ces informations permettent aux analystes RH de mieux comprendre le contexte et les caractéristiques des employés à risque. Juste au-dessus à gauche de la liste, des filtres interactifs permettent de sélectionner le nombre d'employés à afficher et la période de calcul pour la probabilité de départ. Cela permet de détecter plus ou moins en amont les risques de départ des employés. Les filtres situés en haut de la page (Figure 2(b)) sont les mêmes que ceux de la Figure 1. Cette interface est conçue pour fournir une visualisation claire et précise des employés les plus à risque, permettant aux responsables RH de prioriser leurs actions et d'intervenir de manière proactive. En s'appuyant sur cette interface, les utilisateurs peuvent identifier les employés nécessitant un suivi particulier et planifier des entretiens ou des interventions adaptées, afin de réduire le risque de départ et améliorer la rétention au sein de l'organisation.

4.3 Simulation d'intervention sur un employé

Ce dernier tableau de bord a été conçu pour répondre au troisième cas d'usage, qui consiste à simuler une intervention sur un employé afin de voir comment cela affecte sa rétention. Il est spécifiquement orienté vers l'analyse détaillée d'un employé individuel, permettant aux utilisateurs de tester différentes actions et de comprendre leur impact potentiel sur la rétention. La figure 3 est une capture d'écran du tableau de bord représentant ce cas d'usage.

L'interface de ce tableau est composée de plusieurs sections qui fournissent un aperçu complet des caractéristiques de l'employé. On y trouve des informations de base telles que l'identifiant, le poste occupé, le salaire, le département, et la localisation géographique (Figure 3(a)). Ces données statiques permettent de contextualiser les résultats des simulations.

Un graphique représentant la séquence d'événements de l'employé est affiché afin de visualiser la trajectoire de cet employé dans l'entreprise (Figure 3(b)). Cette visualisation permet de suivre les différents changements que l'employé a subis au cours de son parcours.

Juste en dessous (Figure 3(c)), nous avons la liste des potentiels motifs fréquents représentant le mieux la trajectoire de l'employé (voir section 3) ordonnée par ordre décroissant de score. Pour chacun de ces motifs, en plus du score, nous trouvons leur aire sous la courbe de survie (AUC) ainsi que leur gain moyen en tant que contexte de motif affectant la survie.

Dans la partie supérieure droite de l'interface (Figure 3(d)), des motifs affectant la survie sont proposés en fonction du motif représentant le mieux l'employé choisi, pour être appliqués à la séquence de l'employé et projeter sa survie. En cliquant sur un motif affectant la survie, l'utilisateur peut visualiser, grâce au graphique de survie situé à droite de l'interface (Figure 3(e)), l'impact de ce motif affectant la survie sur la courbe de rétention de l'employé. Par exemple, la courbe rouge représente la courbe de rétention actuelle tandis que la courbe verte montre la nouvelle courbe de rétention après l'ajout du motif.



FIG. 3 – Tableau de bord pour la simulation d'intervention spécifique.

Enfin, l'interface affiche une estimation de la probabilité de rétention après l'intervention (Figure 3(f)), avec une comparaison entre la valeur actuelle et la nouvelle valeur suite à l'ajout du motif affectant la survie. Ce composant est essentiel pour visualiser de manière immédiate l'impact potentiel d'une intervention, permettant aux responsables RH d'adapter leur stratégie de rétention en conséquence.

Cette interface permet aux responsables RH de simuler différentes actions et d'observer leurs effets en temps réel, tout en offrant une analyse détaillée à l'échelle de chaque employé. Il s'agit d'un outil de simulation qui permet de tester diverses hypothèses, comme l'impact potentiel d'une augmentation de salaire sur la probabilité qu'un employé quitte l'entreprise. Ces fonctionnalités ne se limitent pas à comprendre la situation actuelle, elles facilitent également la prise de décisions éclairées en s'appuyant sur des simulations de scénarios futurs. Ainsi, ce tableau de bord contribue à une gestion proactive et optimisée de la rétention des employés dans l'entreprise.

4.4 Exemple d'utilisation

Nous présentons un exemple étape par étape montrant comment un utilisateur peut interagir avec les tableaux de bord pour obtenir une vue globale puis détaillée, facilitant ainsi la prise de décisions stratégiques en gestion des ressources humaines. L'utilisateur, un RH chargé du suivi des employés, commence par utiliser le premier tableau de bord (Figure 1) pour identifier les pays avec les taux d'attrition les plus élevés. Il passe ensuite au second tableau de bord pour sélectionner l'un de ces pays (par exemple la France) et filtre les employés à haut risque de départ immédiat. Une liste des 5 employés les plus à risque, triée par ordre décroissant, s'affiche alors. L'utilisateur voit que l'employé 202 est celui le plus à risque immédiat (Figure 2).

L'utilisateur clique donc sur l'employé 202 pour accéder au dernier tableau de bord afin de visualiser sa trajectoire, obtenir des informations détaillées, et sélectionner le motif le plus

Tableaux de bord à partir des motifs affectant la survie

représentatif. Le tableau de bord affichera alors la courbe de survie et les motifs affectant la survie associés, permettant de tester différentes stratégies pour retenir l'employé. Dans la Figure 3 est simulé l'ensemble d'événements $\langle(C, J2)\rangle$ comme motif affectant la survie permettant de passer d'une probabilité de 0.154 à 0.309.

5 Conclusion

Nous avons développé des tableaux de bord interactifs permettant la valorisation des motifs affectant la survie issus de la fouille de motifs fréquents intégrant l'analyse de survie. Ils permettront aux utilisateurs d'avoir une vue globale et détaillée de l'évolution de l'attrition, d'identifier les employés ayant une forte probabilité de démissionner et enfin de pouvoir simuler des interventions et voir leurs effets sur la rétention de ces employés. Les responsables RH peuvent ainsi prendre des décisions éclairées pour optimiser la rétention.

Dans les travaux futurs, nous envisageons d'élargir l'outil pour qu'il puisse être utilisé à des utilisateurs non-RH comme la direction ou les managers. Pour cela, il sera nécessaire d'ajouter des outils de restitution pour que les utilisateurs RH puissent communiquer leurs analyses à des non-RH. Nous pouvons également améliorer la fonction de score en y ajoutant un nouveau critère basé sur la durée temporelle entre le premier et le dernier événement du motif considéré. Enfin, nous voulons également implémenter un filtre pour pouvoir analyser l'évolution de l'attrition de l'équipe d'un manager spécifique.

Références

- Guo, Y., S. Guo, Z. Jin, S. Kaul, D. Gotz, et N. Cao (2021). Survey on visual analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics* 28(12), 5091–5112.
- Morrison, G. (2021). How long do people stay at their firms? Technical report, Consulting Point.
- Oubelmouh, Y., F. Fargon, C. de Runz, A. Soulet, et C. Veillon (2023). Identifying survival-changing sequential patterns for employee attrition analysis. In *2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA)*, doi : .
- Soni, P., C. de Runz, F. Bouali, et G. Venturini (2024). A survey on automatic dashboard recommendation systems. *Visual Informatics* 8(1), 67–79.

Summary

Employee retention is a major challenge for companies. Our work has enabled us to develop a method for identifying and extracting from the frequent patterns coming from employees' trajectories those that improve their retention. This publication shows how we have visually valorized our results. These are presented in the form of interactive dashboards, with the aim of being exploited by the company's HR managers. They offer a global visualization of attrition, an identification of employees at risk of resigning, and a simulation of the impact of an intervention.