

Employee Turnover Analysis

BEM2031 Introduction To Business Analytics - Final Report

Student Number - 720029134

Table of contents

1. Analysis of Ghouzam's Report	3
1.1 Business Understanding	3
1.2 Data Understanding	4
1.3 Data Preparation	5
1.4 Modelling	6
1.5 Evaluation	7
1.6 Deployment	7
2. Improved Analysis - XGBoost Model	9
3. References	12

GenAI Statement

Responsible and Ethical Use of GenAI Tools in the Business School

Within the Business School, we support the responsible and ethical use of GenAI tools, and we seek to develop your ability to use these tools to help you study and learn. An important part of this process is being transparent about how you have used GenAI tools during the preparation of your assignments. Information about GenAI can be found [here] and guidance on the responsible use of GenAI tools can be found [here]. The declaration below is intended to guide transparency in the use of GenAI tools and to assist you in ensuring the appropriate citation of those tools within your work.

GenAI Declaration

I *have* used GenAI tools in the production of this work.

The following GenAI tools have been used: *[please specify]* ChatGPT*

I have used GenAI tools for the following purposes:

- ☐ I have used GenAI tools to assist with research or gathering information.
- ☐ I have used GenAI tools to help me understand key theories and concepts.
- ☐ I have used GenAI tools to help me analyse data.
- ☐ I have used GenAI tools to create code.
- ☐ I have used GenAI tools to suggest a plan or structure of my assessment.
- [Yes] I have used GenAI tools to give me feedback on a draft.
- ☐ I have used GenAI tools to generate images, figures, or diagrams.
- ☐ I have used GenAI tools to generate creative content for my work.
- ☐ I have used GenAI tools to proofread and correct grammar or spelling errors.
- [Yes] **Other** (*please specify*) Debugging Code

Declaration of Citation

- [Yes] I declare that I have referenced the use of GenAI tools and outputs within my assessment in line with the University guidelines for referencing GenAI in academic work.

1. Analysis of Ghouzam’s Report

1.1 Business Understanding

Ghouzam’s report aimed to understand the reasons behind employees leaving a company, using historical human resources (HR) data to identify patterns in employee behaviour that might indicate dissatisfaction or disengagement. Employee turnover has significant impacts on businesses, especially when high-performing individuals leave unexpectedly (Sajjadi et al., 2023). This could cause financial, operational, and cultural costs (Hancock et al., 2013). By analysing structured HR data, the project aimed to support the development of more effective retention strategies, making HR decision-making more proactive.

The dataset consists of 15,000 individual employee records, each with ten features including both behavioural and employment-related metrics. These involve *satisfaction_level*, *last_evaluation*, *number_project*, *average_monthly_hours*, *time_spend_company*, *Work_accident*, *promotion_last_5years*, *sales*, *salary*, and the target variable *left*. These variables offer a good insight into internal drivers of employee turnover. The report applies a structured methodology, incorporating exploratory data analysis (EDA), clustering, and classification models like decision trees and random forests. These techniques uncover patterns in employee exits, including low satisfaction, extended working hours, and a lack of progression.

However, while Ghouzam’s report is technically well-structured, it frames employee turnover as a narrow classification problem, with limited consideration of its real-world business application. The analysis does not address how the model’s outputs will be used by HR teams or what the potential risks and consequences of these predictions might be. It overlooks the broader organisational context and ethical considerations that should accompany predictive HR analytics.

One of the key limitations is the assumption that internal HR data alone is sufficient to predict turnover. The model fails to incorporate external factors such as labour market competitiveness, macroeconomic pressures, or opportunities elsewhere, all of which are known to influence employee decisions (Ayodele et al., 2020). Furthermore, the analysis does not distinguish between voluntary and involuntary turnover, an important distinction in HR strategy, as highlighted by Maertz et al. (2022). The lack of demographic variables also limits the depth of the analysis, since features like age, gender, and education are often essential for identifying key retention trends (Trevor, 2001).

There are clear benefits to this type of analysis. It allows HR and leadership teams to intervene early, refine retention strategies, and better allocate resources. However, there are also risks. If misused, predictive analytics can result in unfair treatment or excessive monitoring. Indarapu et al. (2023) stress that a lack of algorithmic transparency can undermine employee trust, reiterating the importance of clear communication and ethical complicity in deploying such models.

The outcomes of this analysis have implications for multiple stakeholders such as HR teams, line managers, and employees. When implemented accurately, it can improve engagement, reduce turnover, and support a more stable workforce. But to achieve these outcomes, the model must be continuously validated, contextually interpreted, and governed by ethical principles.

1.2 Data Understanding

The dataset used in Ghouzam’s paper is a reasonable representation of the factors that might influence an employee’s decision to leave the company. Variables such as *satisfaction_level*, *last_evaluation*, *average_monthly_hours*, and *number_project* provide valuable insight into key aspects like employee satisfaction, performance, and workload. However, with only three categorical levels, the *salary* variable lacks sufficient detail and limits its relevance in the analysis. Furthermore, the dataset does not differentiate between voluntary and involuntary turnover—an important omission that restricts the organisation’s ability to derive meaningful insights into employee turnover (Maertz et al., 2022). Additionally, the absence of demographic data may reduce the effectiveness of the analysis, as these features often help discover meaningful patterns in employee turnover (Trevor, 2001). Including variables such as age, gender, and educational background could improve both the accuracy and interpretability of the model. Moreover, incorporating external factors such as labour market trends, economic conditions, and competitor compensation offers could enhance the model’s predictive power by accounting for influences beyond the internal work environment. Ayodele et al. (2020) provide evidence that these factors have a significant relationship with employee turnover.

The analysis presented in the report effectively utilised visualisations to explore key patterns in the data. For example, the radar chart provided a concise summary of the mean differences between employees who stayed and those who left. This visualisation highlighted satisfaction level as the most prominent distinguishing factor, with departing employees generally reporting lower satisfaction. However, reliance on average values in radar plots can obscure important variation within the data, particularly among subgroups. This limitation is especially relevant given that employees might leave for different reasons, and aggregating the data in this way may hide meaningful variations in reasons for employee turnover.

To explore underlying patterns in the dataset, dimensionality reduction techniques such as Principal Component Analysis (PCA) and Isomap were applied. PCA did not reveal clear separation between employees who stayed and those who left, which is likely due to its limitation in capturing non-linear relationships. In contrast, the Isomap projection was more effective in highlighting distinct groupings within the data, suggesting that different types of employees may have left due to a combination of reasons rather than a single dominant factor. However, the target variable (*left*) was not removed prior to conducting PCA and Isomap, which may have introduced label bias into the visualisations. Consequently, this could

have exaggerated the distinction between the two groups, potentially resulting in misleading interpretations of the underlying structure.

This finding was further reinforced by the KMeans clustering analysis, which identified three distinct subgroups among employees who left the company. The first group included high-performing but overworked individuals who, despite receiving strong evaluations and handling heavy workloads, still chose to leave. The second cluster looked at employees who were both dissatisfied and underperforming, suggesting possible issues with motivation, role fit, or engagement. The third group consisted of employees who reported high satisfaction and strong performance yet still exited the organisation—potentially due to external opportunities or a lack of perceived growth within the company. This type of segmentation highlights the importance of recognising variation within the workforce, as employee turnover is rarely driven by a single cause. Research suggests that segmenting employees based on performance, engagement, or experience can significantly improve the impact of HR interventions and help prevent employee turnover (Kwon & Rupp, 2013).

The radar plots were then revisited using these clusters, visualising the differences between each group. This made the analysis much more compelling and actionable, since it can be used to suggest specific solutions for each type of employee. For example, overworked employees might benefit from workload management, while disengaged employees might need better support or clearer progression paths.

While the dataset lacks depth in key areas like employee motivation, external influences, and demographics, the analysis makes good use of the data available. The visualisations, especially when combined with clustering, go beyond surface-level summaries and start to build a deeper understanding of the problem. However, to advance the analysis further, the dataset would need to be expanded. Including feedback data, such as survey responses or exit interviews, alongside contextual business information would support a more accurate understanding of why employees leave. Combining quantitative analysis with qualitative insights is often recommended, as it leads to a more holistic view of employee behaviour (Shaw, 2011).

1.3 Data Preparation

Ghouzam’s report demonstrates only a basic level of data preparation, overlooking several essential steps required for reliable modelling. Although no missing values were reported, the dataset contains a large number of duplicate rows, an important issue that was not identified or addressed. Duplicate entries can inflate model confidence and reduce generalisability. This is especially problematic when models are evaluated on datasets containing similar patterns, which may not reflect real-world variability. Recent research highlights that repetitive patterns in training data can contribute to overfitting and compromise a model’s ability to perform across varied contexts (Kang et al., 2025). Addressing these data quality issues is essential for building trustworthy predictive models.

Outlier detection was also missing from the analysis. While tree-based models such as decision trees and random forests are generally strong, they can still be affected by extreme values, particularly when those values represent non-representative or duplicated behaviour. For example, rare patterns in *average_monthly_hours* or *number_project* may cause unstable splits or introduce misleading decision thresholds. Studies have shown that handling such outliers improves model reliability and decision boundary consistency (Marín Díaz et al., 2023).

Feature scaling was correctly applied using standardisation, which is essential for algorithms like PCA and KMeans. However, feature selection lacked justification. All variables were included without assessing their individual predictive value, potential redundancy, or multicollinearity. Additionally, skewness in the feature distributions was not evaluated, and no transformations were applied to correct non-normal distributions.

Interaction effects were also not investigated. These interactions could provide more insight than individual variables alone. In response, three interaction effect variables were created. An overall happiness index (*happy_idx*) combining *satisfaction_level* and *last_evaluation*. A burnout index (*burn_idx*), involving *number_project*, *time_spent_company* and *promotion_last_5years*. Finally, an overworked index (*over_idx*) showing interaction effects between *average_monthly_hours*, *time_spent_company* and *average_monthly_hours* was included.

1.4 Modelling

Ghouzam’s report relies on decision tree and random forest classifiers, both appropriate for binary classification and useful when model interpretability is important. However, the choice of models is not justified, and no alternatives are explored. Including a logistic regression model as a baseline would have provided a valuable point of comparison to assess whether the tree-based models offered a genuine improvement in performance.

A more important concern is the lack of clarity around the evaluation strategy. The report does not specify how the data was split as there is no mention of a training/test ratio or whether stratified sampling was used. While 10-fold cross-validation is used, the absence of clear data partitioning raises the risk of data leakage and inflated performance estimates. Proper splitting is essential to ensure models generalise reliably beyond the training data (Shafie et al., 2024).

The random forest model is reported to achieve 99% accuracy, which strongly indicates overfitting. Without context from precision, recall, or F1-score, especially in an imbalanced classification context, this figure is potentially misleading. High accuracy alone reflects the model’s bias toward predicting the majority class. Including a confusion matrix or class-wise performance metrics would have clarified how effectively the model captured actual leaving cases.

The decision tree visualisation is a valuable inclusion, offering clear insight into how splits occur and highlighting key features such as *satisfaction_level* and *time_spent_company*.

However, no comparable explanation is provided for the random forest. Given the complexity of ensemble models, incorporating interpretability tools such as Local Interpretable Model-Agnostic Explanations (LIME) or Shapley Additive Explanations (SHAP) would have improved transparency and supported more informed decision-making (Thi Hang et al., 2024).

1.5 Evaluation

Currently, the project is not suitable for deployment. While the initial analysis offers useful insights into key drivers of employee turnover, several core limitations prevent it from being implemented in real-world HR settings. A key limitation is that the model has not been validated on unseen data. Without testing on external scenarios or real-world organisational settings, there is a high risk that the conclusions will not generalise, particularly in dynamic or department-specific contexts.

While informative, the dataset has a limited scope. It excludes important features such as external labour market indicators, voluntary resignation reasons, and employee sentiment, which are often critical for understanding workforce dynamics and employee decision-making. However, without qualitative inputs like exit interviews or engagement surveys, the model lacks the depth to provide actionable insight beyond correlation. This limits its value for long-term strategic planning.

The project correctly identifies key predictive features, but the model would benefit from clearer validation. Key metrics like precision, recall, and F1-score are missing, making it hard to evaluate performance on the minority class (employees who leave). There is no evidence of hyperparameter tuning or regularisation, both of which are essential for preventing overfitting. Fairness is also not addressed as no subgroup analysis was done to check for bias across roles or pay levels. As Köchling & Wehner (2020) emphasise, failing to assess fairness can reinforce existing organisational biases and lead to discriminatory outcomes.

The visualisations were helpful in showing feature relevance and model logic but remained limited. They lacked temporal or comparative elements, such as time-series trends or benchmarks against industry norms, due to dataset constraints. These would be essential for contextualising insights and tracking changes over time. Model interpretability was also limited. While feature importance was shown, tools like LIME or SHAP were not used to explain individual predictions. Without clearer explanations, stakeholders may not be able to trust or act on the model's outputs, reducing the likelihood of adoption.

1.6 Deployment

While the model is not ready for application, there are several clear and achievable steps that could move it closer to operational use. The most immediate priority is planning its integration within existing HR systems. Currently, there is no deployment framework in place, with no

API, dashboard, or defined output format tailored for HR teams. For the model to be impactful in practice, it must deliver insights in a format that is both accessible and actionable. A basic interactive dashboard, developed using tools like Streamlit or Power BI, could act as a front-end interface, highlighting individual employee exit risk alongside feature-level explanations. Risk scores could then be linked to suggested HR responses, such as initiating check-ins, offering internal mobility options, or reviewing workload levels for high-risk employees.

A second major gap is the lack of a monitoring and retraining strategy. HR data evolves over time due to seasonal trends, policy changes, and broader economic conditions. Without a mechanism to track changes in data patterns and declining model performance, the model risks becoming outdated. Implementing a structured retraining process, supported by a clear Extract, Transform, Load (ETL) framework, would allow the model to adapt as new employee data becomes available. Key evaluation metrics such as F1-score, recall, and precision should be continuously tracked through a live dashboard to monitor ongoing performance. Retraining schedules could align with quarterly HR reviews, ensuring the model evolves with the organisation's workforce and operational shifts.

To support continued use, shifting toward a more prescriptive analytics approach is recommended. This would allow the system to not only identify employees at high risk of leaving but also suggest targeted interventions such as performance coaching, role redesign, flexible work arrangements, or proactive retention bonuses. Cluster analysis revealed distinct groups of leavers, suggesting that tailored strategies will be more effective than a one-size-fits-all solution. For example, overworked high performers may benefit from workload redistribution or project prioritisation, while disengaged underperformers could be offered clearer career development paths or mentoring opportunities, as highlighted in recent work on AI-supported HR practices (Tambe et al., 2019). However, the effectiveness of such interventions depends on interpretability. Currently, the model does not include explanation tools like LIME or SHAP, which makes it difficult for HR teams to understand or act on individual predictions.

Finally, ethical and regulatory considerations need to be addressed. The model does not incorporate fairness assessments or bias audits, and there is no reference to data privacy safeguards or General Data Protection Regulation (GDPR) compliance. These are particularly important when working with personal employee data in predictive systems (Leicht-Deobald et al., 2019). A responsible deployment strategy should include bias reduction measures, subgroup performance analysis, and a governance framework to ensure transparency and compliance. HR leaders should also consider establishing cross-functional review teams to evaluate predictive outputs before any action is taken, ensuring that interventions are fair, context-aware, and aligned with organisational values.

2. Improved Analysis - XGBoost Model

When developing my own model, I focused on addressing the main limitations of the previously mentioned study using an XGBoost, an accurate ensemble model that has demonstrated strong performance in employee turnover prediction tasks (Gomes et al., 2025). It offers a strong balance of accuracy and interpretability, especially when paired with LIME to visualise local feature contributions and F-scores to show global importance.

For the data preparation, I completed all steps within Ghouzam's analysis. Also, duplicate observations were removed from the dataset so they have no impact on the model and the data was split into an 80-20 train/test split to reduce overfitting and allow evaluation on unseen data. Within my model, regularisation techniques were used to reduce overfitting and address class imbalance. Hyperparameter tuning is done using gridsearch to improve the performance of the model. Furthermore, I included more performance metrics; precision, recall, F1-score and area under the curve (AUC) to give a better evaluation of the model. Without these steps, it's difficult to trust any claims about model performance or business impact.

Table 1: Performance Metrics for XGBoost Model

Metric	Scores
Accuracy	0.982
Precision	0.99
Recall	0.904
F1-Score	0.945

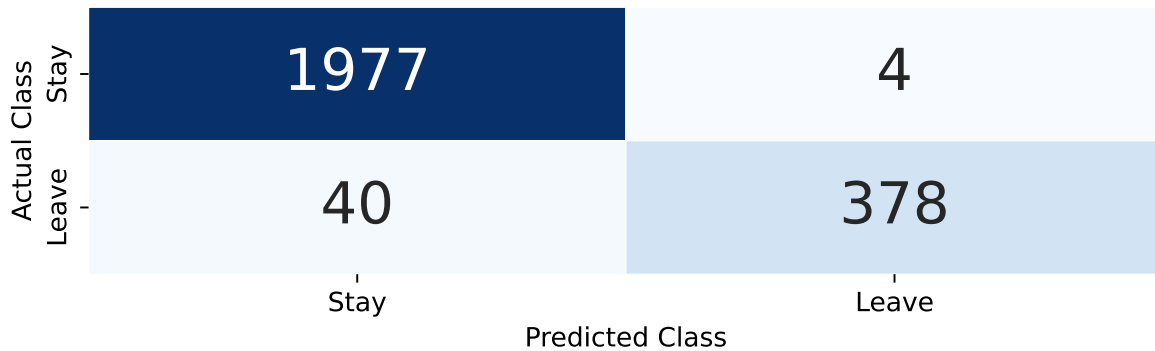


Figure 1: Confusion Matrix for XGBoost Model

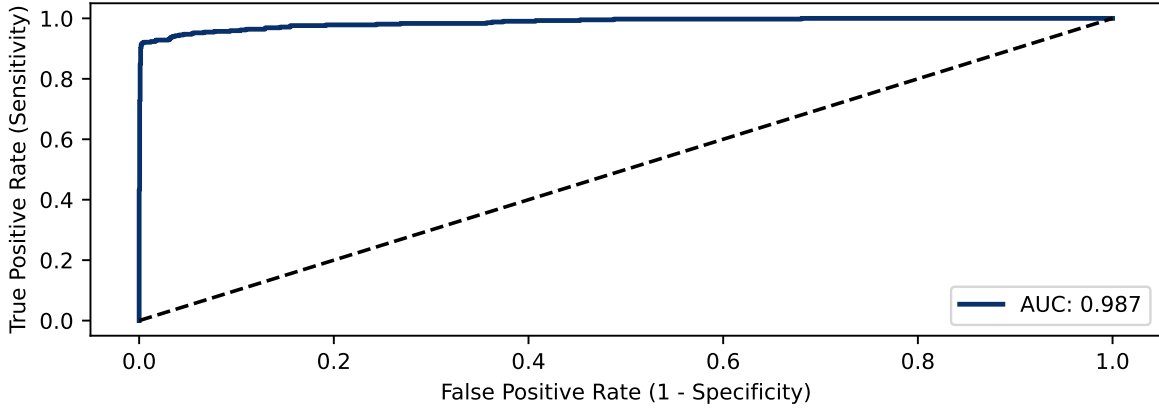


Figure 2: ROC Curve for XGBoost Model

The XGBoost model showed strong performance across all key evaluation metrics. On the test set, it achieved an accuracy of 98.2, which already indicates that the majority of predictions were correct. However, looking beyond accuracy, the precision score of 0.99 is very high. This means that when the XGBoost predicts a positive case, it is 99.0% likely to be correct. This is important in scenarios where false positives can waste resources.

Recall was also high at 0.904, which suggests the XGBoost is successfully identifying most of the actual positive cases. That being said, there were still 40 false negatives, as seen in the confusion matrix (Figure 1). While this number is relatively small, it could still be a concern as leaving employees can have a financial impact on the company due to recruitment costs.

Figure 2, the ROC curve, reiterates the XGBoost's exceptional classification ability. The AUC score of 0.987 shows that the XGBoost is excellent at distinguishing between the two classes, even under different thresholds. This, combined with the F1 score of 0.945, suggests that the XGBoost handles class balance well and is not overly biased toward the majority class. Despite the high performance metrics which traditionally could indicate potential overfitting, due to the implementation of L1 + L2 regularisation, a train/test split, and 10 cross-fold validation, these metrics can be trusted allowing this model to be deployed within a HR environment for employee exit decisions.

Figure 3 shows how LIME can enhance transparency by breaking down the individual contribution of features in a specific observation. Within this observation, the XGBoost predicted with 95% certainty that the employee would leave, driven by the overworked index being above 0.43, not having a work accident, having a high burnout index above 0.66, and a high time spent at the company (>4 years). Conversely, salary being high or medium, and average monthly hours being between 199 and 234 are important factors when predicting someone to stay at the company.

The significant role of interaction effects highlights the importance of including them in Ghouzam's report to ensure a more accurate and reliable analysis. However, LIME explains

only one observation which is why feature importance is essential. While Ghouzam’s project correctly incorporates feature importance, the analysis could be strengthened by including the F-score to quantify importance, rather than relying solely on relative values.

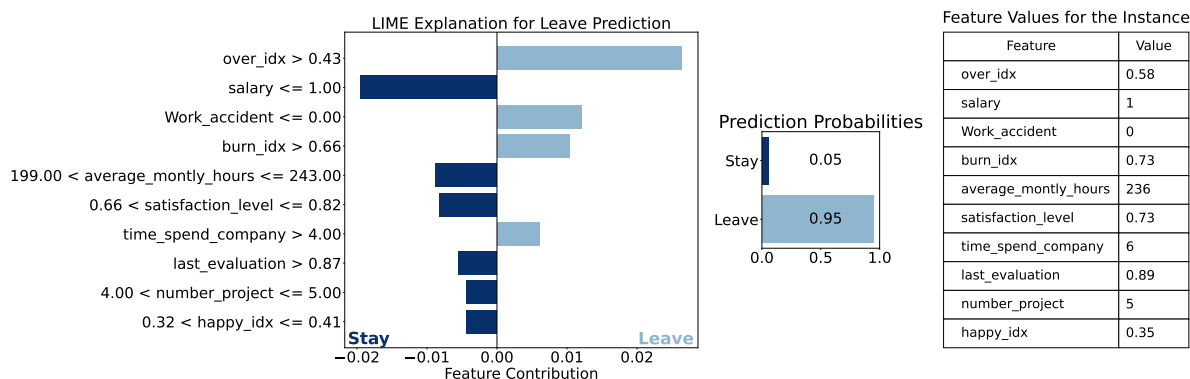


Figure 3: LIME for XGBoost Model

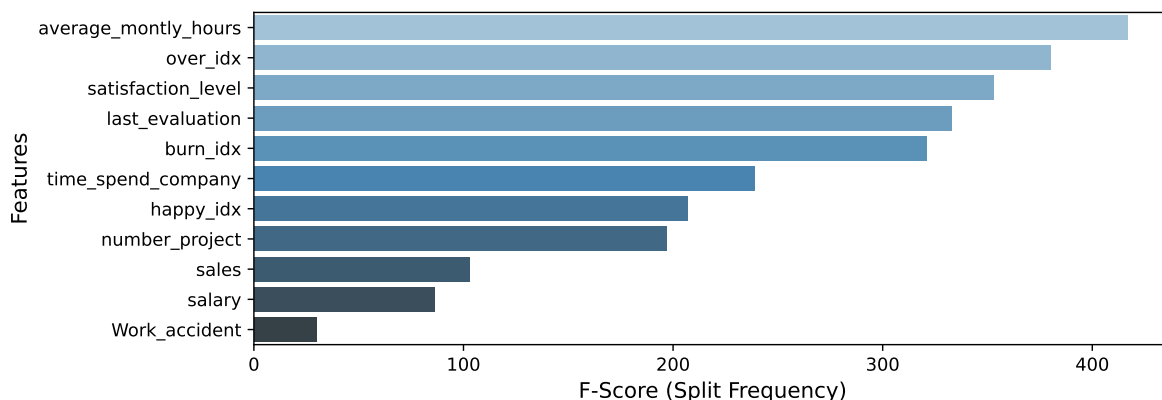


Figure 4: Feature Importance for XGBoost Model

Figure 4 shows global feature importance based on XGBoost’s F-score, which counts how often each feature is used to split the data across all trees in the XGBoost. High F-scores for *satisfaction_level*, *average_monthly_hours*, and *last_evaluation* indicate these features are frequently used in decision rules, suggesting strong individual predictive power. The engineered features *over_idx* and *burn_idx* rank highly, confirming their added value as a standalone inputs. Interestingly, *Work_accident* ranks as the least important feature in Figure 4 whereas it is shown to be the third most influential predictor in Figure 3. This confirms that people will leave the company for different reasons and highlights the need for global feature importance.

This analysis could be used by the company to proactively identify at-risk employees who exhibit similar patterns. For example, those who are highly tenured and under-recognised (e.g. no recent promotion), and showing signs of burnout despite strong performance. HR

interventions such as role reassessment, workload balancing, or wellbeing initiatives could be targeted based on the specific drivers of turnover uncovered in individual cases. While the model performs well overall, future analysis should explore subgroup performance to ensure fairness. For instance, analysing precision and recall across salary bands or tenure levels may reveal disparities that impact HR decision-making.

[Link to Github Repository = BEM2031 Introduction To Business Analytics](#)

3. References

- Ayodele, O. A., Chang-Richards, A., & González, V. (2020). Factors affecting workforce turnover in the construction sector: A systematic review. *J. Constr. Eng. Manag.*, 146(2), 03119010. [https://doi.org/https://doi.org/10.1061/\(asce\)co.1943-7862.0001725](https://doi.org/https://doi.org/10.1061/(asce)co.1943-7862.0001725)
- Gomes, A., Silva, L. M., & Cruz, J. P. (2025). Predicting employee attrition in a multi-company setting. In *Lecture notes in computer science* (pp. 337–348). Springer Nature Switzerland.
- Hancock, J. I., Allen, D. G., Bosco, F. A., McDaniel, K. R., & Pierce, C. A. (2013). Meta-analytic review of employee turnover as a predictor of firm performance. *J. Manage.*, 39(3), 573–603. <https://doi.org/https://doi.org/10.1177/0149206311424941>
- Indarapu, S. R. K., Vodithala, S., Kumar, N., Kiran, S., Reddy, S. N., & Dorthi, K. (2023). Exploring human resource management intelligence practices using machine learning models. *J. High Technol. Manag. Res.*, 34(2), 100466. <https://doi.org/https://doi.org/10.1016/j.hitech.2022.100466>
- Kang, X., Cheng, J., Yang, Y., & Liu, F. (2025). Repetitive transient impact detection and its application in cross-machine fault detection of rolling bearings. *Mech. Syst. Signal Process.*, 228(112422), 112422. <https://doi.org/https://doi.org/10.1016/j.ymssp.2019.107186>
- Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Bus. Res.*, 13(3), 795–848. <https://doi.org/https://doi.org/10.1007/s40685-020-00134-w>
- Kwon, K., & Rupp, D. E. (2013). High-performer turnover and firm performance: The moderating role of human capital investment and firm reputation. *J. Organ. Behav.*, 34(1), 129–150.
- Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., & Kasper, G. (2019). The challenges of algorithm-based HR decision-making for personal integrity. *J. Bus. Ethics*, 160(2), 377–392. <https://doi.org/https://doi.org/10.1007/s10551-019-04204-w>
- Maertz, C. P., Jr, Keith, M. G., Raghuram, S., Porter, C. M., & Dalton, G. L. (2022). Advancing theory and practice on managing dysfunctional turnover: Developing an improved measure of turnover reasons. *Group Organ. Manag.*, 105960112110658. <https://doi.org/https://doi.org/10.1177/10596011211065807>

- Marín Díaz, G., Galán Hernández, J. J., & Galdón Salvador, J. L. (2023). Analyzing employee attrition using explainable AI for strategic HR decision-making. *Mathematics*, 11(22), 4677. <https://doi.org/https://doi.org/10.3390/math11224677>
- Sajjadi, S., Kammeyer-Mueller, J., & Benson, A. (2023). Who is leaving and why? The dynamics of high-quality human capital outflows. *Acad. Manage. J.*, 66(6), 1929–1953. <https://doi.org/https://doi.org/10.5465/amj.2021.1327>
- Shafie, M. R., Khosravi, H., Farhadpour, S., Das, S., & Ahmed, I. (2024). A cluster-based human resources analytics for predicting employee turnover using optimized artificial neural networks and data augmentation. *Decision Analytics Journal*, 11(100461), 100461. <https://doi.org/https://doi.org/10.1016/j.eti.2024.103655>
- Shaw, J. D. (2011). Turnover rates and organizational performance: Review, critique, and research agenda. *Organ. Psychol. Rev.*, 1(3), 187–213. <https://doi.org/https://doi.org/10.1177/2041386610382152>
- Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *Calif. Manage. Rev.*, 61(4), 15–42. <https://doi.org/https://doi.org/10.1177/0008125619878580>
- Thi Hang, H., Mallick, J., Alqadhi, S., Bindajam, A. A., & Abdo, H. G. (2024). Exploring forest fire susceptibility and management strategies in western himalaya: Integrating ensemble machine learning and explainable AI for accurate prediction and comprehensive analysis. *Environ. Technol. Innov.*, 35(103655), 103655. <https://doi.org/https://doi.org/10.1016/j.eti.2022.103655>
- Trevor, C. O. (2001). Interactions among actual ease-of-movement determinants and job satisfaction in the prediction of voluntary turnover. *Acad. Manage. J.*, 44(4), 621–638. <https://doi.org/https://doi.org/10.5465/3069407>