Employee Attrition Prediction using an Explainable FT-Transformer Deep Learning Model

K.Kanthimathi*[1], Aarathy T.S[1]
[1]School of Management and Commerce, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamilnadu, India-600073
kkanthimathi.karanam@gmail.com

Abstract. As the attrition rate increases, it is difficult for Human Resource professionals to predict the influencing factors. Taking necessary actions at the right time to retain employee will reduces employee recruitment costs become difficult for Human Resource Personnel Therefore there is a need for efficient decision making system for Human Resource Management. In this study, to predict Employee Attrition, one of the advanced Deep learning techniques called Feature Tokenizer Transformer (FT-Transformer) architecture model was implemented which works efficiently for tabular type of datasets. The dataset used in this study was the IBM HR Analytics Employee Attrition obtained from Kaggle Repository. Performance of the model was evaluated using standard metrics of accuracy, precision, recall, F1-score, and Area Under the ROC curve (AUC) and along with confusion matrix. An Explainable Artificial Intelligence (XAI) method called SHAP (SHapley Additive exPlanations) was applied to the FT-Transformer predictions to know the strong attributes that influences an employee to leave an organization. The Proposed model achieved an accuracy of 87.07% and an AUC of 0.763 with highest SHAP value 0.041003 for the attribute age.

Keywords: Employee Attrition Prediction, Deep Learning, FT-Transformer, Explainable Artificial Intelligence (XAI), SHAP, IBM HR Analytics Employee Attrition, HR Professional, Decision-Support systems.

1. Introduction

Attrition is the gradual reduction of employees in an organization due to resignations and retirements. Causes for employee attrition include poor work environment, work-life balance, job satisfaction and compensation that impact the performance of an organization. HR professional monitors to understand the causes for leaving an organization and manages attrition in order to reduce job recruitment and training cost and to increase productivity [1]. As modern organizations increasingly depend on data-driven decision-making, the ability to accurately predict employee attrition has become an important objective in human resource (HR) analytics. Early identification of employees at risk of leaving organization enables to implement retention strategies and optimize workforce planning [8]. Traditionally, statistical and rule based methods were used for employee attrition analysis but it lacks to predict the complex and nonlinear relationships between the attributes. With the growing availability of public datasets and emerging Machine Learning (ML) techniques [2] such as Decision Support Trees, Logistic Regression, Support Vector Machines, and XGBoost have

S. P. Vijayaragavan et al. (eds.), *Proceedings of the Global Conference on Sustainable Energy Systems, Smart Electronics and Intelligent Computing (GCSESEIC 2025)*, Advances in Engineering Research 297,
https://doi.org/10.2991/978-94-6239-654-8_50

been widely implemented to improve predictive performance. Though these techniques increases performance but lacks in handling class imbalances and high dimensional feature interactions [9]. Recent advancements in deep learning [3] have the capability to learn more complex patterns such as text, image and audio but limits its applications in processing tabular dataset. Due to its black-box nature, it lacks its adoption in sensitive domains such as human resource management, where transparency, fairness, and trust are essential [10].Transformer-based architectures have recently gained attention for their ability to model complex feature interactions and long-range dependencies. Although transformers were developed for Natural Language Processing, recent applications have proved their effectiveness on tabular data [11]. The Feature Tokenizer Transformer (FT-Transformer) is a specialized transformer architecture designed for processing tabular datasets such as IBM HR analytics, where each feature is treated as a token and processed through self-attention mechanisms [12]. Though the model proves its efficiency but fails in interpreting its results. In this study, FT-Transformer model was implemented to predict attrition using IBM HR Analytics Employee Attrition dataset and an Explainable Artificial intelligence (XAI) [4] method SHAP was implemented to interpret the results to increase decision making capability of HR professionals to understand what strong attributes influence employee attrition. . Model performance was assessed using metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC), along graphical representations such as confusion matrices, ROC curves, and precision–recall curves.This is paper is divided into 5 sections where section 1 describes about employee attrition and its prediction models. In section 2, Dataset collections and its description are given. Methodology used in this study is given in section 3. Experiments and results are discussed in section 4. Finally conclusions are given section 5.

2. Dataset

The dataset used in this study was IBM HR Analytics Employee Attrition collected from the Kaggle repository [5]. The dataset contains 1,470 employee records with 35 attributes. These attributes include demographic, education, work-life balance and professional details. Among these 35 attributes, 4 attributes having constants were removed, and finally, 30 attributes were taken as input and an attribute Attrition as a target attribute. The updated dataset attributes along with their descriptions and data types were presented in Table 1.

Table 1: IBM HR Analytics Employee Attrition Dataset Attributes, Description and its Data Types

S.No	Attribute Name	Description	Data Type
1	Age	Employee age (years)	Numerical

2	BusinessTravel	Business travel frequency	Categorical
3	DailyRate	Rate of Daily wages	Numerical
4	Department	Employee Department	Categorical
5	DistanceFromHome	Distance from home to workplace (km)	Numerical
6	Education	Level of education (1–5)	Ordinal
7	EducationField	Education Domain	Categorical
8	EnvironmentSatisfaction	Work environment satisfaction (1–4)	Ordinal
9	Gender	Employee Gender(Male/Female)	Categorical
10	HourlyRate	Rate of Hourly wages	Numerical
11	JobInvolvement	Type of job (1–4)	Ordinal
12	JobLevel	Job level of the employee	Ordinal
13	JobRole	Role/designation of the employee	Categorical
14	JobSatisfaction	Job satisfaction (1–4)	Ordinal
15	MaritalStatus	Employee Marital status	Categorical
16	MonthlyIncome	Monthly income of employee	Numerical
17	MonthlyRate	Monthly pay of employee	Numerical
18	NumCompaniesWorked	Number of companies worked previously	Numerical
19	OverTime	Status of employee overtime working	Binary (Yes/No)
20	PercentSalaryHike	Hike percentage salary	Numerical
21	PerformanceRating	Performance rating (1–4)	Ordinal
22	RelationshipSatisfaction	Satisfaction with work relationships (1–4)	Ordinal
23	StockOptionLevel	Level of Stock Option	Ordinal
24	TotalWorkingYears	Total professional experience(years)	Numerical
25	TrainingTimesLastYear	Trainings attended in last year	Numerical
26	WorkLifeBalance	Work-life balance rating (1–4)	Ordinal
27	YearsAtCompany	No. of Years worked in the current company	Numerical
28	YearsInCurrentRole	No. of Years in current job role	Numerical
29	YearsSinceLastPromotion	No. of Years since last promotion	Numerical
30	YearsWithCurrManager	No. of Years with current manager	Numerical
31	Attrition	Whether the employee left the organization	Binary (Yes/No)

3. Methodology

3.1 Data Pre-processing

The collected IBM HR Analytics Employee Attrition dataset has no missing values therefore no imputation is required. All the categorical attributes were encoded with label encoding, ordinal attributes were encoded using ranking relationships and Continuous numerical attributes were normalized using z-score standardization to ensure uniform feature scales. The target attribute attrition was encoded using binary label encoding. Due to the imbalance class data cost-sensitive learning was employed through class weight rather than resampling technique and finally dataset was divided into training and testing subsets using a stratified train–test split to maintain the original class distribution across both subsets.

3.2 FT-Transformer Model Architecture

The below Fig 1 FT-Transformer (Feature Tokenizer Transformer) is a transformer-based deep learning architecture [6] designed to process tabular data. In general tabular data contains attributes and records where there exists relationship between these attributes. The FT-Transformer architecture learns these relationships and performs classification efficiently. FT-Transformer architecture contains the components of input embeddings, multiple transformer encoder self-attention and multi-head attention layers, Feedforward Neural Network and a decoder or output layer. Initially the all categorical and numerical attributes are embedded with vector representation. Then the multi-head and Self-attention encoder layer finds the relationship between the attributes. The Feedforward Network calculates the ouput of each attribute transformation. Finally the decoder or output layer contains the aggregated transformed features[7]. A simple transformer architecture and its work flow is

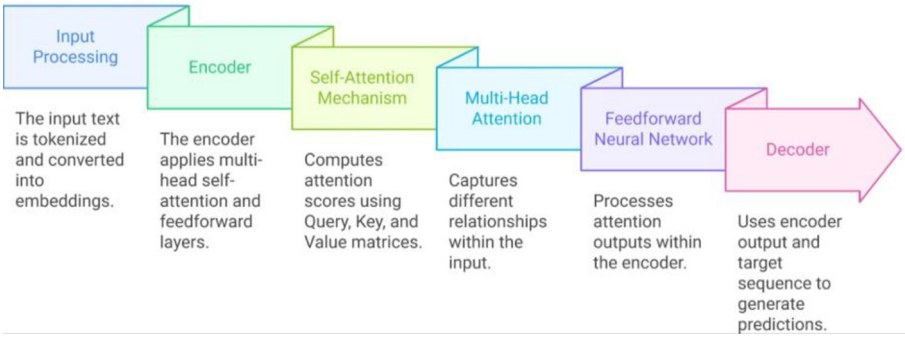

Fig 1: Architecture of the Explainable FT-Transformer Model for Employee Attrition Prediction 1

The FT-Transformer model was trained using a supervised learning approach. Binary cross-entropy loss was employed as the objective function Fig. 1. Optimization was performed using the Adam optimizer, and early stopping was applied to prevent overfitting. Hyperparameters such as embedding dimension, number of attention heads, and learning rate were selected empirically based on validation performance.

3.3 Explainable AI (XAI) Framework

To enhance model interpretability, Explainable Artificial Intelligence techniques were applied to the proposed framework. SHAP (SHapley Additive exPlanations) was used to provide explanations of the FT-Transformer predictions and were useful in identifying the most influential attributes contributing to attrition prediction.

3.4 Evaluation of Performance Metrics

The standard performance metrics including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC), and graphical representations of ROC curve, precision–recall curve and confusion matrix were evaluated to predict the proposed model's performance.

These metrics provide a comprehensive assessment of predictive performance, particularly in the presence of class imbalance. In addition, visual evaluation techniques such as the confusion matrix, ROC curve, and precision–recall curve were used to examine classification behavior under class imbalance.

4. Experiments and Results

For employee attrition prediction, proposed FT-Transformer model was implemented using PyTorch framework. The dataset used for training and testing are divided based on the stratified train–test split. There was a class imbalance in the collected dataset as shown in Fig 2. To address the problem, a cost-sensitive learning strategy was implemented during model training and AdamW optimizer was used for faster convergence Fig. 2.

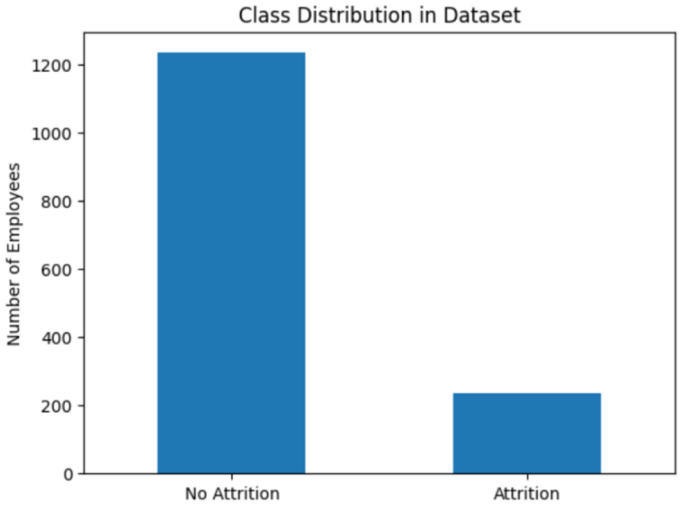

Fig 2: Class Distribution IBM HR Analytics Employee Attrition dataset

The performance metrics of the proposed FT-Transformer model on the employee attrition prediction task was summarized in Table 2.

Table 2: Performance Metrics of the Proposed FT-Transformer Model on IBM HR Attrition dataset

Metric	Value
Accuracy	**87.07%**
Precision	61.54%
Recall	51.06%
F1-score	55.81%
AUC	0.763

The model achieved attrition prediction accuracy of indicating strong predictive performance on the imbalanced dataset and an 87.07%, AUC of 0.763, which confirms its ability to distinguish between employees who are likely to leave and those who are not across different decision thresholds. Additional evaluation plots of ROC curve shown in Fig 3 which demonstrates that the model consistently performs above the random baseline and Precision-recall curve shown in Fig 4 provides the model's effectiveness on the minority attrition class.

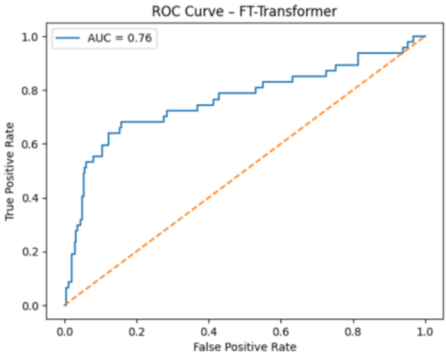

Figure 3: ROC curve

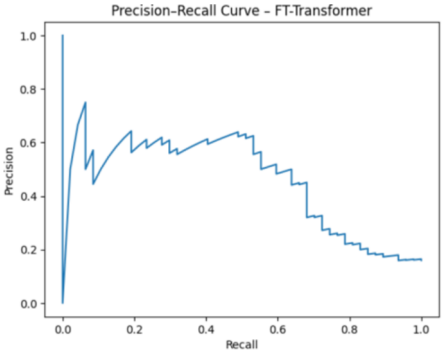

Figure 4: precision–recall curve of

A total of 294 records were tested using the proposed model and its results were represented in the form of Confusion Matrix shown in Fig 5. Among 294 cases, model correctly classified 24 attrition records and 232 No-attrition records while 15 No Attrition and 23 Attrition records were misclassified. Although some attrition cases were misclassified as non-attrition, this behavior is expected in highly imbalanced datasets.

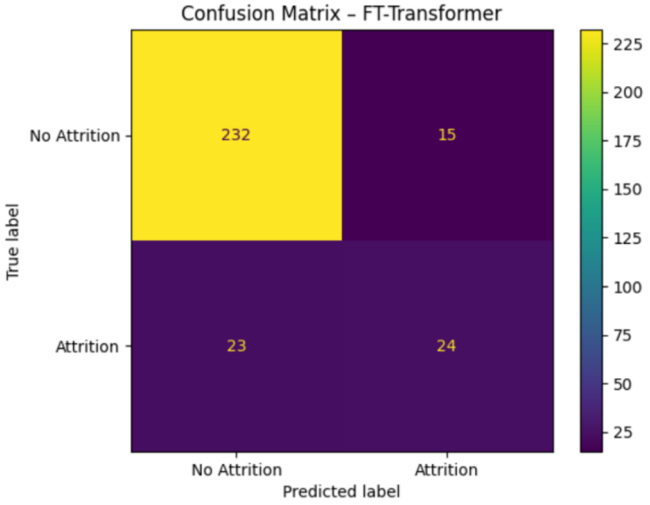

Fig 5: Confusion Matrix of IBM HR Employee Attrition Prediction using FT-Transformer

Interpretation of model's accuracy using XAI was represented in Table 3. From the table it shows that Age, OverTime, YearsAtCompany, Department, and YearsInCurrentRole are among the most influential attributes affecting attrition predictions. Most 10 attributes that influences employee attrition are listed in the table. This makes the model particularly suitable for deployment in real-world HR decision-support systems where transparency, reliability, and trust are essential.

Table 3: Explainable AI Analysis Results of FT-Transformer Model On IBM HR Attrition dataset

Rank	Attribute	Importance Score
1	Age	0.041003
2	OverTime	0.039021
3	YearsAtCompany	0.034111
4	Department	0.029977
5	YearsInCurrentRole	0.023947
6	TrainingTimesLastYear	0.019382
7	Education	0.017831
8	StockOptionLevel	0.017831
9	JobInvolvement	0.017745
10	RelationshipSatisfaction	0.016367

5. Conclusion

In this study, Explainable FT-Transformer architecture was implemented for employee attrition using IBM HR Analytics attrition dataset. Experimental results showed that the proposed model achieved an accuracy of 87.07% and an AUC of

0.763. Interpretation of results using XAI method showed that the Age, OverTime, YearsAtCompany, Department, and YearsInCurrentRole were among the most influential attributes affecting attrition predictions which are useful in decision making of HR professional to retain employees.

6. Future Work

Advanced imbalance handling techniques are needed to improve the performance of the model where the dataset has more imbalanced data. This has to be taken as a future work.

References

1. V.M Tharaka Rani, Madhavan, Ravija and Saran. S, “A Study on Job Attrition among Employees in IT Companies With the Reference to Coimbatore City”, International Journal for Research Trends and Innovation, Vol. 9, No. 4,pp:1034-1039, ISSN: 2456-3315(2024).
2. Ali Raza, Kashif Munir , Mubarak Almutairi, Faizan Younas and Mian Muhammad Sadiq Fareed,”Predicting Employee Attrition Using Machine Learning Approaches”, Appl. Sci, 12, 6424(1-17). 2022
3. D. M. Quinteros, “Predictive modelling of employee attrition using deep learning,” Acadlore Trans. Mach.Learn., vol. 2, no. 4, pp. 212–225. https://doi.org/10.56578/ataiml020404. , 2023
4. Irem Tanyıldızı Baydili and Burak Tasci Systems,” Predicting Employee Attrition: XAI-Powered Models for Managerial Decision-Making”,Systems, 13, 583(1-25). 2025
5. M. E. Hossen *et al*., "Boosting Cervical Cancer Prediction Leveraging a Hybrid FT-Transformer Model," in *IEEE Access*, vol. 13, pp. 26876-26896, 2025, doi: 10.1109/ACCESS.2025.3538566.
6. N. T. N, A. S, A. M. V, A. S. Nath, V. J. Aswin Chandran and B. Moozhippurath, "AI - Powered Student Dropout Prediction Using AutoGluon FT-Transformer," *2025 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)*, Ernakulam, India, 2025, pp. 1-6, doi: 10.1109/ACCTHPA65749.2025.11168572.
7. A. Aggarwal, P. Taneja, K. Seth and H. Singh Pannu, "Predicting Telecom Customer Churn Utilising Machine Learning Models," *2025 World Skills Conference on Universal Data Analytics and Sciences (WorldSUAS)*, Indore, India, 2025, pp. 1-6, doi: 10.1109/WorldSUAS66815.2025.11198977.
8. P. S. S. S. Sundaram, "Machine Learning for Predictive Customer Retention in Competitive Markets," *2025 IEEE 5th International Conference on ICT in*

Business Industry & Government (ICTBIG), Indore, Madhya Pradesh, India, India, 2025, pp. 1-6, doi: 10.1109/ICTBIG68706.2025.11323864.

9. Z. Liu, S. Orr and J. Grau-Bove, "Htgtm: Hybrid Temporal-Graph Tabular Model For Complex Multimodal Tabular Data Processing," *2023 IEEE 33rd International Workshop on Machine Learning for Signal Processing (MLSP)*, Rome, Italy, 2023, pp. 1-6, doi: 10.1109/MLSP55844.2023.10285918.
10. Guduri, M., Polavarapu, A., El Yabroudi, M., Thota, S.K., Gummadi, H.S.B., Chennupati, N. (2026). A Novel Framework on Cardiovascular Disease Prediction Using Transfer Learning Technique. In: Lanka, S., Cabezuelo, A.S., Tugui, A. (eds) Trends in Sustainable Computing and Machine Intelligence. ICTSM 2025. Lecture Notes in Networks and Systems, vol 1755. Springer, Cham. https://doi.org/10.1007/978-3-032-13177-5_7
11. A. M. Ayaz and S. Manoharan, "A Comparative Study of Advanced Deep Learning Models for Real Estate Price Prediction in Dubai," *2025 Second International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS)*, Coimbatore, India, 2025, pp. 282-289, doi: 10.1109/ICC-ROBINS64345.2025.11086212.
12. Wang, Y., Liang, Z., He, Y., Wu, J., Tian, P., & Ling, Z. (2025). AMFormer-based framework for accident responsibility attribution: Interpretable analysis with traffic accident features. *PLOS ONE*, *20*(7), e0329107. https://doi.org/10.1371/journal.pone.0329107