National College of Ireland

MSc Research Project
Data Analytics

Oluwaseun Ayokunbi Ogunbowale
Student ID: X21193355

School of Computing
National College of Ireland

Supervisor: Vitor Horta.

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	OLUWASEUN AYOKUNBI OGUNBOWALE ………………………………………………………………………………………………
Student ID:	X21193355 …………………………………………………………………………….………………………..……
Programme:	DATA ANALYTICS……………………………… **Year:** ……2022/2023…..….
Module:	Research Project….……………………………………………………………………….………
Supervisor:	Vitor Horta……..………………………….……………………………………….………
Submission Due Date:	18/09/2023……………….………………………………………………………………..
Project Title:	Explaining Neural Networks and Random Forest for Employee Retention………………………………….……………………………………………..….
Word Count:	……9254…………………… **Page Count**………20….…………………………………….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ………Oluwaseun Ayokunbi Ogunbowale…………………………………………………………

Date: …18/09/2023………………………………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Explaining Neural Networks and Random Forest for Employee Retention.

Oluwaseun Ayokunbi Ogunbowale
X21193355

Abstract

Employees are the foundation of every business or organization, and for any company or organization to succeed or flourish in business, the employer should cherish and respect its most important resource or workforce, which is staff, so there is a need to study employee retention. Employees have a propensity to quit an organization when treated in a bad way by the management team, which can also have a negative impact on the productivity of the business and lead to dissatisfaction and customer migration. The goal of this research is to identify the motivating factors that influence employee retention using Neural networks and Random Forests with Explainable AI.

This research made use of two machine learning techniques to solve the classification problem and explored factors that contributed to employee retention using two human resources (HR)datasets from Kaggle. The result was evaluated with the use of evaluation metrics and Explainable Ai tools, with an accuracy of 99 % in the Random Forest and 96% in Neural Networks in data. Two Explainable AI models such as LIME, which is a local interpretable Model-Agnostic explainer, and SHAP, which means Shapley Additive explainer was used. 'Satifaction_level', 'number_project','average_montlyhour', and 'time_spend' contributed positively to employee attrition with features positive and negative interactions that impacted the models.

Keywords: Employee retention, Lime, Shap, Explainable Ai, Machine Learning.

1 Introduction

Employees are the most important part of an organization Lesan (2022) and employee turnover as a result of not being given the proper support, jobs not being secure, and poor management whereas employee retention is established as the tactics, or procedures utilized by the business to maintain the efficient and productive employees in their system to decrease employee turnover. This is done purposely to improve their businesses and increase profit turnover. In general, many employees are resigning, while some have been compelled to resign or relieved of their duties by the management team's decisions. According to the reports, more than 19 million US workers have resigned from their position since early 2021, and this will have a negative impact on businesses all over the world as most organizations will find it difficult to address these unforeseen circumstances because they are completely equipped to deal with or control this kind of issue or scenario.

With improvements in technology, many firms or organizations are using ML Machine learning and Artificial intelligence approaches to solve different problems Lesan (2022) in their journal paper stated that evaluating large datasets containing employee data such as performance indicators, participation, data on training, and machine learning algorithms may find or exposed previously overlooked detailed through correlations, pattern, and non-linear interactions by exploiting these data sources. This enhanced awareness of attrition helps firms to identify the real causes of employee attrition by designing and implementing data-driven strategies to successfully address this problem. The ability of ML machine learning and AI to implement predictive models is one of the primary benefits of ML and AI in employee attrition evaluation. These models are capable of precisely predicting attrition issues for individuals or certain groups, allowing employers to take initiative-taking steps to retain their sought-after staff. These models may continually learn and adapt by boosting the model's accuracy

The problem of people leaving a company is a key issue or challenge for businesses in various organizations, which impacts not just production but also expenses such as hiring, educating workers, knowledge transfer, and employee welfare. Understanding the elements or factors contributing to attrition is crucial for companies to establish effective measures to prevent its negative impacts. Explainable AI algorithms have evolved as a potential solution recently for unraveling the complicated dynamics underlying employee attrition. The combination of machine learning algorithms with interpretability methodologies (AI) to give insights into how black-box models make decisions is referred to as Explainable AI. Organizations can upgrade from predictive analysis and statistical modeling to AI to develop and have a better understanding of the fundamental factors impacting attrition by using explainability methods. This knowledge enables the Human resource department to recognize patterns, major reasons, and potential intervention areas that contribute to attrition, allowing them to build and focus on retention plans.

The goal of this research is to identify the motivating factor that influences the retention and efficiency of an employee using machine learning techniques like neural networks and Radom Forest with Explainable AI (XAI) tools to predict how long a worker stays in an organization.

The key major objective of this work is to:

I. Key Factor identification: The primary objective is to identify the key features that influence employee retention positively in an organization by using classical metrics with LIME and SHAP to study the performance of each model by feature importance to bring out the insight from the model result.

II. Assessment of importance: The quantification and ranking of the attribute according to their impact on retention rates can help companies manage their efforts and resources as another objective or goal of determining the impact of various factors on employee retention with predictive analysis.

2.0 LITERATURE/RELATED WORKS

The following are related work on employee turnover and retention in the company with the use of ML machine learning techniques and methods documented by several researchers are enumerated below.

2.1 EXPLAINABLE AI

Explainable artificial intelligence is defined as the method that allows stakeholders to understand the output of machine learning approach prediction by presenting artificial intelligence models in a simpler and easy to understand. Bhatt Uet. al (2020). This is used to analyze and describe an AI model by building trust, accuracy, and transparency in the outcome of the result for AI-based decision-making, Explainable AI is being used by most organizations to avoid making a blind decision without having full knowledge of the depth of the model because the black boxes model is difficult to interpret on how it achieved its conclusion and having a good Explainable AI is an introduction to implementing a good and responsible Ai model to meet the need of stakeholders.

2.2 Model interpretability

Model interpretability is required to confirm what the model is performing is consistent with what you anticipate and to build user trust, making the switch from human to automated operations easier. The process of extracting explanations from the machine learning technique is called the Model Agnostic Method, this method is known and used because of its flexibility as it works with all machine learning classifiers which can also be term as interpretable models as well. The explanation generated by interpretable models is rich and flexible and not limited to a specific form of explanation also it is possible for the explanation system to utilize an alternative feature of representation than the model that is being explained. Building a model is simple, interesting, and misleading as business intelligence is difficult to explain with evaluation metrics, also discovering what is happening inside the model can be tedious but with the help of Explainable AI tools the purpose of the outcome of the prediction can be uncovered and explained.

I. LIME.

This stands for a local interpretable model-agnostic explanation used to explain a single prediction of a black box means that Lime can be used on any model, it specializes in training models locally to explain the output of individual predictions with the intention of knowing why machine learning makes certain decisions. Lime finds out what happened to the outcome of prediction through the input that was supplied for machine learning by generating a new set of the data samples and a new prediction for the model and an interpretable model (linear model) is trained on the new data to extract the behavior of the complex model and it generated by taking the sample from the rows or instances closeness to the relevant instance as its weight and the local predictions made from the machine learning model could be similar to the learned model.

II. SHAP

This stands for Sharpley additive explanations which is a technique for explaining machine learning prediction to understand the importance and impact of each feature present in individual prediction by finding the Shapley values. SHAP works on the approach that is related to gaming theory using multiple features to create a prediction where each feature present represents a player and works together to make a prediction (which represents a goal). The contribution of each feature is measured by Shapley values which assist in describing how the prediction is distributed among the features present.

The equation below represents a linear model explanation of the outcome of the prediction of x row or instance by finding the performance of each factor to the prediction, where M is the maximum limit, j is the minimum limit, and $z \in \{0,1\}^m$ is the simplified features that represent the presence of one feature (1) while absent is (0),

$g(x')=\Phi 0+\sum_{J=1}^{M}(\phi j x' j)$.

2.3 Neural Networks for Employee Attrition.

In the employee attrition using Deep Neural network by Salah Ai-darraji et.al (2021) , classical machine learning and deep learning techniques with cross- fold validation was applied on the employee dataset to determine the effectiveness of their models, the researcher described the neural network architecture used are layers, activation functions, and optimization techniques used to train the model with the neural network using a cross-validation methodology, which provides information on the significance and influence of various input parameters on attrition prediction as well as by highlighting particular factors that contributed greatly to attrition, The synthetic datasets, and the unbalanced dataset, with an accuracy of 94.16 and 91.16 respectively, 89.11% accuracy obtained with 10fold cross validation, a single model was applied and it performed well, and it was difficult to conclude on the uniqueness of the result because of the model, Neural networks will be used in these studies because of its strenght to recognize non-linear correlations in the data.

Ali Raza et.al 2022 used descriptive analysis of employee data to identify various factors that contributed to employee attrition with four machine learning algorithms such as LR-Logistic Regression, DT-Decision Tree, ETC-Extra Trees Classifier, and SVM-Support Vector Machine methods were used for comparison to discover the performance that best suited the evaluation technique also a comparison analysis was performed to find the best evaluation. The synthetic oversampling method was applied to balance the imbalance dataset and the k-fold validation method, and the result was evaluated by evaluation metrics such as precision and accuracy, e t c. The key attributes contributing to employee attrition were salary, hourly income, job position, and age and the output of the prediction in terms of accuracy utilizing the four advanced machine learning techniques gives S.V.M, Lr, and D.T.C 87% for the SVM technique, 72% for the LR technique, and 83% by Decision Tree and (ETC) achieved 93% accuracy and when the dataset was used with the K fold the following were obtained, the SVM with an accuracy of 88%, the LR technique 74% accuracy, the DTC had 84% accuracy the proposed model had 93% of accuracy. The researcher proposed the use of deep learning techniques with which this work will be using Explainable AI tools to improve and enhance the dataset features.

Lesan (2022) stated in their journal that the seal to change jobs and progress in one's career were factors contributing to employee attrition, Four machine learning models were used in the study Lesan (2022), which tackles a binary classification issue, using the human resource dataset, and the accuracy results are as follows support vector machine 0.97, Random forest 0.97, decision tree with an accuracy of 0.97, and logistic regression 0.79. by using recall to evaluate the model the accuracy obtained are Logistic regression 0.561, Random Forest 0.908, Decision tree DT 0.903, and support vector machine(SVM) 0.922, and the researcher suggested using an advanced algorithm, such as an artificial neural network, as Random Forest was one of the best classifier, and SVM-support vector machine give better result for the project's business decision making when examined, the proposed model that will be used is Random Forest and Neural Network

Raja R et.al (2022) utilize machine learning techniques to predict attrition and special emphasis was placed on the expenses, issues of leadership or management, and recruitment process as explained by Lesan (2022) in their work. Work-life balance issues or excessive stress brought on by overtime were also mentioned as factors causing employee dissatisfaction. According to Raja R et al. (2022), the dataset was used in a ratio of 80% to 20% on four machine learning classifiers like K-nearest neighbor, naïve baye, linear regression, and multilayer Perceptron was used and compared it was found that there was a strong correlation between salary increase and performance rating as multi-layer perceptron outperformed every other algorithm used on the dataset and best fit accuracy was obtained as evaluation was reported by the researcher and cannot provide adequate meaning or interpretation to the result based on the attributes present.

2.4 Random Forest for Employee Attrition.

In the study by Agnibho. C et.al (2022), employee behaviour, use of appraisals and assess employee satisfaction levels in the workplace were examined that attrition affected employee performance, and leaves workers with little time for proper training, Lesan (2022) said that these factors also contributed to an increase in overhead costs related to hiring, paying off the former employee, and hiring another worker with training expenses attached. The researcher used K-nearest neighbor, (SVM)Support vector machine,(NB) Naive Baye, and Random Forest classifier on the two distinct datasets, the Naive Baye classifier has an accuracy of 43.87%, and Random Forest with a higher accuracy of 86.5%, the other dataset human resource, Random Forest turned out to give a good result, for this reason, will be considering random forest as part of the classifier to be used.

Marvin G et.al (2021) explained that the efficiency of staff has an impact on progress and improvement, and provides a competitive edge in businesses and suggest the use of (AI) artificial intelligence to improve productivity and performance as also stated by Lesan (2022), nine machine learning models: Random Forest, K-nearest neighbor, Logistic Regression, Decision Tree, Naive Bayes, Gradient Boosting Algorithms, LightGBM, and XGboost, on the imbalanced dataset to predict the probability that staff will be retained in a company or organization, Random Forest showed an accuracy of 84.6%, 99.1%, and 91.8% in testing, training, and accuracy on the dataset.

Filippo Guerranti1 et.al (2022), S. Yadav et.al (2018), and Bhartiya N et.al (2019) mentioned that staff attrition is common and one of the most significant problems that organizations encounter, citing the high cost of employee a new staff, the orientation of new staff, decline in performance, an increase in the workload for current employees Lesan (2022) Principal component analysis, descriptive analysis, and models such as the k-nearest neighbor, Random Forests, and Logistic Regression were used on the dataset. descriptive analysis in terms of trends and variations was used to determine the reason employees quit, RF (Random Forest) outperformed the other models applied and stated that the adoption of ML (machine learning) and AI (artificial intelligence) methods will be an advantage.

Bhartiya N et.al (2019) used machine learning algorithms such as Support Vector Machine, Random Forest, K-Nearest Neighbour, Decision Tree, and Naive Bayes. Naive Bayes and Support Vector Machine (SVM) performed well with respect to true positive rate, whereas Random Forest, K-nearest neighbor -KNN, and Decision Tree, performed better in terms of TN (true negative) classification. Higher accuracy was achieved as follows: Decision tree scored 81%, and Random Forest scored 83.3%. Yadav S et.al (2018) used descriptive analysis of employee behaviors and features and stated that organizations might cope with the attrition of employees who lack knowledge when compared with experienced staff. Random forest with higher accuracy of 98.61%

2.5 Explainable AI tools on Machine learning classifiers.

A research paper by Navya S (2020) on employee attrition using Explainable AI on the dataset and the Shap and Lime libraries in Python to uncover patterns and contributing factors as well as the distinction between interpretability and explain-ability and according to the researcher, difficult models can be learned and understood in multiple ways, the model was built with employee dataset using six different classifiers like logistic regression, XGBoost, Naive Baye, SVM (Support vector machines), and the LightGBM model. The models were compared, Random Forest and XGBoost achieved greater accuracy ratings of 85.7% and 86% respectively, with Explainable AI SHAP, it was also found that different models assigned different levels of importance to the same set of features, which affected both the positive and negative trends seen in the attrition model. In conclusion, it was determined that working overtime is a key factor in employee attrition suggest using artificial intelligence to improve the model.

Karthik Sekaran et.al (2022) used the Shapley Additive explainer with LIME (Local interpretable model-agnostic explainer) on the employee attrition data to interpret the factors determining employee attrition. LightGBM was used as the foundation algorithm to train the model with an accuracy of 0.95% o for train, and 0.84 % for test, and explain-ability was obtained with the application of LIME and SHAP which are AI tools used to explain the features' importance. the researcher concluded by saying that the two XAI tools applied gave the same factors the limitation of this research is the use of one model for the analysis and several models can be applied.

2.6 Conclusion.

The question of what attribute affects the retention of employees and the efficiency or productivity of workers in an organization will be answered using a Neural network and Random Forest model with Lime and Shap, this is because most papers did not discuss what factors should be improved to boost the productivity of the company. The artificial neural network was chosen because it produces better output, efficient and contributions demonstrate that it is a proactive classifier that offers or provides an appropriate outcome and prediction about employee retention while Random Forest is a well-known and efficient classifier with excellent and strong accuracy because it is adaptable and produces reliable results without hyperparameter tunning to the model.

These two proposed models with AI tools LIME and SHAP will be used for this project to evaluate predictions and contribution of each feature as ML lacks a perfect way of doing so. LIME is an effective method for achieving local interpretability for black-box models since it approximates their behavior in a clear and comprehensible manner. It enables stakeholders to develop an understanding and confidence in the model's decision by concentrating on providing local explanations for each prediction. SHAP is an integrated system for interpreting ML predictions because it offers a complete way to identify the different feature's contributions to prediction outcomes.

3. Research Methodology

A detailed workflow for the research was created with the human resources analytical dataset and employee attrition for the healthcare dataset. The main goal is to understand what are the important features that contribute to employee retention when classifying a black box model with the procedure listed below.

3.1 Dataset collection and preprocessing.

The datasets related to the above subject were obtained from the Kaggle site which is an open site for data that gives detailed information about an employee in an organization, for better understanding and insight, two of these types of datasets were obtained and preprocessing was done on the dataset to make it clean and fit for use.

- Exploratory data analysis EDA was used on the dataset to discover the effect of each factor on the attrition was noted after which feature selection was then utilized to determine the features to add to the model with correlation on the variables before model construction and prediction purposes.
- Features selection, label encoding, and scaling were done, and it was discovered that the two datasets were imbalanced, the class was not well represented through analysis, and the data was resampled used to balance the dataset and scaling was also done to normalize the variable to have consistence result.

3.2 Training of Classifiers and Performance Metrics.

I. The preprocessed dataset was prepared for modeling and the dataset was split in an 80:20 ratio as this section of the research describes various methods and systematic approaches used in processing data or extracting useful patterns and insight from the dataset by making use of knowledge discovery in a database with 80% use for training and 20% of the dataset is for testing and validation of the experiment.

II. The two models that were selected are the Neural Networks and the Random Forest model.

Random forest architecture: RandomForestClassifier(n_estimators=10, random_state=42)

This is a method used by Random Forest for classification issues and this consists of multiple decision trees (which consist of the branch as an outcome and leave node as the class label) trained on a part of features or train dataset. The selection of the subset is done by random sampling with replacement to create a unique tree to prevent features from dominating the decision-making as the predicted class has the higher votes across the tree. The random state is the hyperparameter used in the feature selection which is set to 42 to ensure reproducibility to yield the same result as the model formula when the same model is run severally while the n-estimator = 10 is the number of trees used that is hyperparameter tuned to 10. The neural network architecture is described in the implementation section.

III. Classical evaluation metric was applied to the model to measure the accuracy and the outcome of the Random Forest and the Neural Networks model for modeling and to know the classes that the outcome favored, and the results were explained in the implementation section of the project.

3.3 Explaining the Classifiers to discover important features.

I. Explainable Ai (LIME and SHAP) was applied to the already formed model to interpret the performance of each attribute present in the dataset using visualization to interpret the effect of each of the factors on the model prediction.

II. Analysis and interpretation of the result based on the effect of each of the factors as shown by the Lime and SHAP model on each of the datasets used in the experiment and its contributions to the organization.

4.0 Design Specification and Implementation

Software and Hardware specification

The research was implemented with Jupyter Notebook which has Python and its library embedded or installed on it, which has 16 GB ram, with a 64-bit operating system, and the processor is 11th Gen Intel (R)Core (TM) where preprocessing, analysis, and modeling were performed, and explainable AI tools were also used from Python libraries.

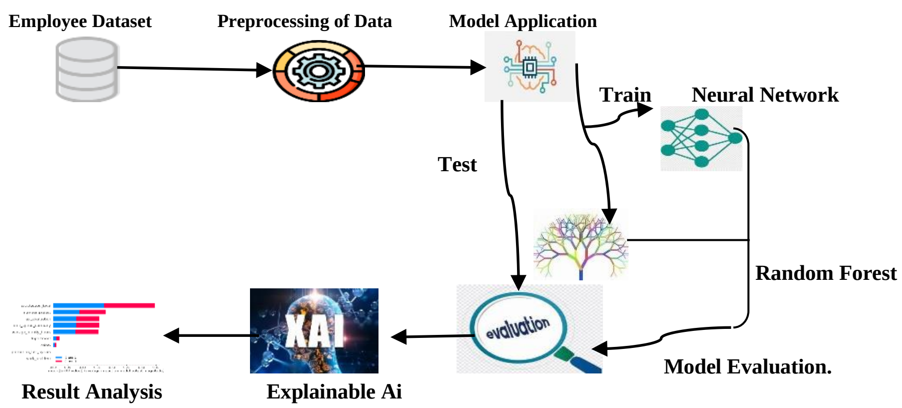


Figure 1. Methodology Chart.

4.1 Dataset collection

The two datasets that were used in this research for employee retention were sourced from Kaggle and the attributes of each dataset will be explained one after the other. The first dataset is the human resources analytical dataset[1] which consists of 14999 rows and 10 columns with integer, object, and float as the data types of the variable. The dataset consists of ten features such as 'satisfactory_level', 'last_evaluation', 'number_projects, etc. with their attributes as illustrated in **Table 1** for a clearer view and understanding.

[1] https://www.kaggle.com/datasets/giripujar/hr-analytics

Table 1: Human Resources Analytical Dataset

S/No		Data Type
1	Satisfactory level	Float
2	Last Evaluation	Float
3	Number of Projects	Integer
4	Average Monthly Hour	Integer
5	Time spent company	Integer
6	Work Accident	Integer
7	Left	Integer
8	Promotion last 5years	Integer
9	Department	Object
10	Salary	Object

The second dataset is the employee attrition for healthcare[2] which consists of 1676 rows and 35 columns with integer and object as the attributes of the dataset and these two were sourced from Kaggle which is an open dataset repository.

Table 2: Represents selected rows from the Employee attrition Dataset.

S/No	Attributes	Data type
1	Employee Id	Integer
2	Age	Integer
3	Attrition	Object
4	Daily Rate	Integer
5	Department	Object

From the table above we have the dependent variable which is also known as the target variable labeled Left which is a categorical variable that represents whether an employee is still in an organization or left the organization and the independent variables as ‘satisfactory-level’, ‘last-evaluation’, ‘number-projects’, ‘average-montly_hour’, ‘time-spent’, ‘work-accident’, ‘left’, ‘promotion-last 5years, ‘department’, and ‘salary’.

4.2 Data Processing

Data preparation is a necessary step of methodology, and it usually takes a lot of time and concentration to look out for all the errors with the dataset that occur due to oversight or errors made while typing the data during data collection also some of the related errors are missing data, an outlier in the dataset, an unstructured dataset that can be discovered with visualization. The two datasets were checked for errors with the naming consistency in the attributes and the use of the lower and upper case to ensure uniformity as this is not a good naming convention in Python, cleaning was done on this by changing all to lowercase for easy reading and computation.

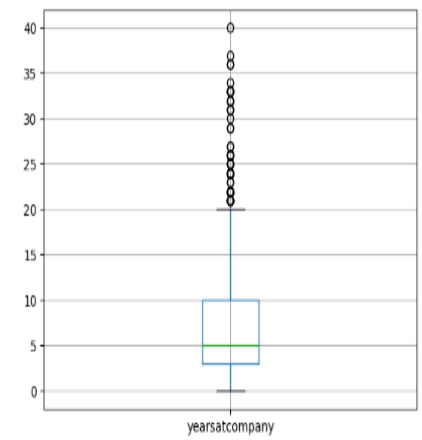
4.3 Missing Values and Duplication.

[2] .https://www.kaggle.com/datasets/jpmiller/employee-attrition-for-healthcare

It is important to verify or check if there are missing values in the dataset. This will enable clarity on what is missing and give a solution on how to replace it, if necessary, there was no missing value in the two datasets. There is a need to check the two datasets employed for duplicate values to avoid the repetition of values and figures for accurate model generation. The first dataset consists of 3008 duplicated values while the second dataset has no duplicate value, all were treated accordingly.

4.4 Data Exploratory

This is also known as exploratory data analysis EDA, this is the process of exploring and visualizing data to get a clearer knowledge of its attributes and to observe trend, patterns, and relationships in the dataset attributes, this help in finding insights and observing anomalies such as outliers, missing values and gap in the dataset that are not detectable by mere looking at the dataset but are visible by using various bars, correlation and finding the statistics of the data. The variable of the dataset consists of both numeric and those that are not numeric. Boxplot, histogram, etc. are used to visualize the attributes of the dataset and some variables that are designated as integers in the dataset with the presence of an outlier since the variables were not accurately represented but rather needed to be factored to be useful for the classification model



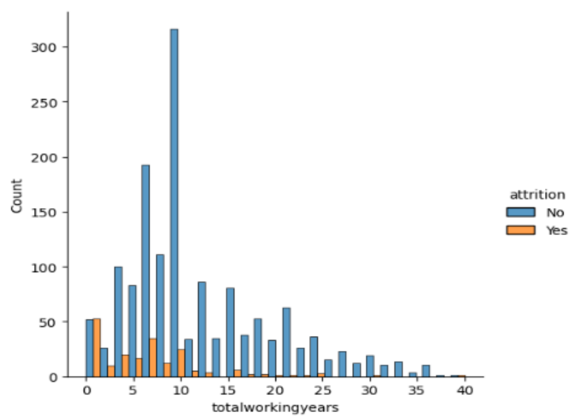


Figure 2a: Box plot for 'year at company'. **Figure 2b: Histogram for 'total working years'**

Exploratory help in showing unwanted data points in the dataset, Figure 2a shows the boxplot of 'yearatcompany' with an outlier also in some of the variables this was removed by the percentile method from statistics to get a good prediction for the model.

Observation: When an employee newly joins an organization, they decide whether to stay or not after orientation, dataset[2] in Figure 2b shows a high attrition rate which declines after 3 years and increases gain for those that set a target of working for 5 years and increases towards ten years and attrition rate decline as the number of years increases while the retention rate is showing a seasonal trend.

Distance from home: this shows that those that are living close to the organization tend to quit more than those who stay far from the office.

Year in current role: those that have spent 2 years and 6 years in their role have a higher tendency of being retained and once the staff is over ten years the retention rate also declines.

Age: Employees within the age bracket 25 years to 45 years have a higher tendency to be retained as age increases the retention rate decline.

4.5 Correlation.

In the human resource dataset, variables are not correlated with one another but in the second dataset Watson Healthcare employee dataset[2] considered, there was a positive correlation between the features as in Figure 3 below, so it is necessary to remove one out of the two variables that are correlated by using feature selection, this is done basically to select variables that have a positive impact on the model been used.

The correlation plot for the dataset

- It shows that the level of 'satisfaction_level', 'promotion_last5', and 'work accidents' have a negative correlation with those that left the organization that is once the employer is unsatisfied, not promoted, and has an accident they leave.
- Those variables that have a positive relationship with the target variable are the ones to be considered for retention in an organization and this includes 'time_spent_company', 'average_montly_hours', 'last_evaluation',and'number_project' completed by the staff.

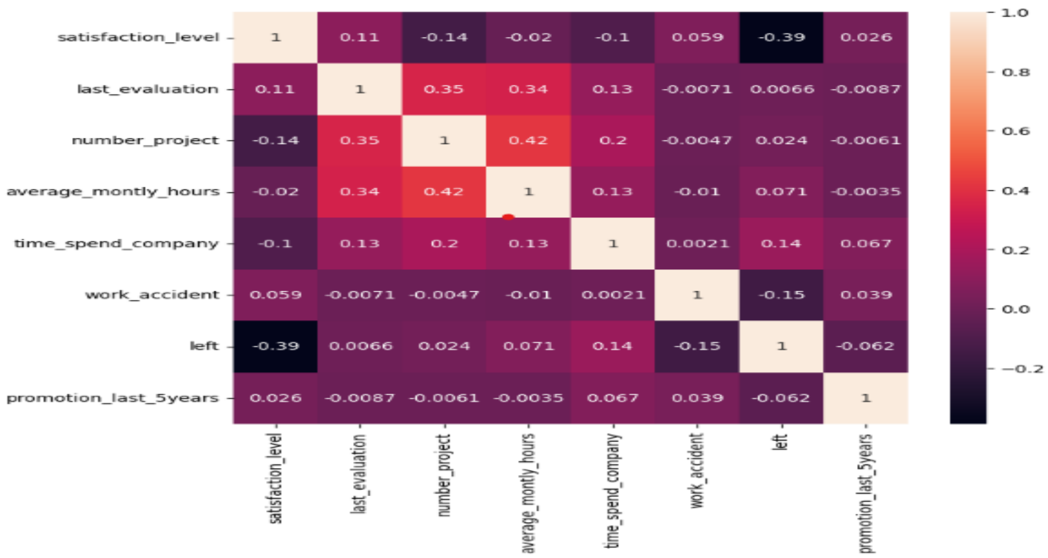


Figure 3:Heat map for correlation of Employee variables.

4.6 Categorical Encoding.

In the first dataset, some attributes needed to be transformed before the model is applied to the dataset, label encoding was done on the attributes such as Department and salary was transformed to change the properties into that which the model will be able to interpret also because the performance of the model does not only depend on the model used or the hyperparameter applied but also depend on how clean the data. Categorical encoding was done on the target variable and was converted to the class of 0 and 1 by using a label encoder from the library sklearn in Python likewise in the second dataset[2], the target 'variable' attrition' was also encoded in the class of YES or No, and other independent variables were also encoded into the format that the machine will understand.

4.7 Feature selection and scaling.

This is done on the encoded data set in this research using scaling of the dataset to standardize the features and ensure the uniformity of datasets[1] and[2] to prevent incorrect results before the model is applied and with the correlation matrix being plotted which shows the degree of relationship between some of the variables as seen in the second dataset, feature scaling is a done on the dataset to generate high-quality data by converting the numerical values of features in a dataset to a standard scale or range.

Table 3a: Summary of preprocessing done on Dataset[1]

S/NO	PREPROCESSES TO BE DONE	What was done
1	The attributes name consists of both lower and upper case	Change attribute name to lowercase
2	There is an outlier with the time spent in the company	Remove outliers
3	There are duplicated values in the dataset	Remove duplicate values and columns
4	Check correlation for variables	Correlation for numerical Datatype
5	Convert Target variable y to categorical	Orders 0 and 1
7	Imbalance class in Target variable	Up-sampling of the target variable
6	Label encoding of Department and Salary	Change from categorical to Integer
7	Scaling of the dataset for uniformity	Scale to value between 0 and 1

Table 3b Summary of preprocessing done on Dataset[2]

S/NO	PREPROCESSES TO BE DONE	What was done
1	The attribute’s name consists of both lower and upper case.	Change attribute name to lowercase
2	Check correlation for attributes	Correlation for numerical datatype
3	Unwanted column variable because of correlation	Removal of unwanted columns
4	Convert Target variable y to categorical	Orders 0 and 1
5	Imbalance class in Target variable	Up-sampling of the target variable
6	Label encoding of Department, gender, and Education field.	Change from categorical to Integer
7	Scaling of the dataset for uniformity	Scale to value between 0 and 1

4.8 Model formulation and splitting of the dataset.

The last procedure in data preprocessing is to divide the data into two sets (train and test) to guarantee that our model is evaluated correctly. A train set is a part or subset of a dataset used to train a machine learning algorithm, whereas the test data is a small portion used to validate machine learning. This can range from 80% train to 20% test depending on the volume and shape of the data being used, Neural networks and Random Forest were trained on the dataset and test accuracy was obtained in a ratio of 80% to 20% and evaluated after which Explainable Ai was used on the model.

4.8.1 Neural Network Architecture.

The structure of the neural network used consists of the input layer, the hidden layers, the output layer, model training, the number of iterations, and the random state of the model.

I. The input layer: The first dataset consists of 10 features while the second dataset used 26 features, these represent the number of features in each of the sample datasets.

II. The hidden layer: The model consists of three hidden layers that have different numbers of neurons, to capture all the features the number of neurons should be more than the count of features present in the model so also to increase the performance of the model. The first layer consists of 64 neurons and the second layer is made up of 32 neurons and the third has as features as the number of features present in the dataset and used the relu activation function the input was checked on the hidden layer for best fit. Max iteration represents where the model will stop whether it converges or not as in the case with the hidden layer 30,10 and 26 which represent the number of features present in the data and the model was stabilized with a random state. Table 4 below shows various hidden layers set up or used for the experiment and their result.

Table 4: Hidden layer selection for Neural Network

NO	No features	Three Hidden Layers	Accuracy	Observation	Reason
1	26	5, 2, 26	0.88	The number of attributes is more than the number of neurons.	All features were captured but low performance.
2	26	30,10, 26	0.967	Accuracy was obtained but the solution does not converge.	No result for the classification
3	26	64, 32, 26	0.962	The solution converges and the complexity of the input data captured	All features were captured with classes

III. Output layer: This shows the result obtained from the predicted classes in the model with the probability value of 0 and 1 which used a single neuron in this layer to produce the classes with the application of the sigmoid activation function.

F(x) = $1/(1+e^{-x})$

which is a default activation function used in multilayer perceptron classifiers with the application of the backpropagation method or reverse mode training to regularize the weight and the biased of the system based on the difference between the actual class and the predicted class.

4.8.2 Model Evaluation.

The two models will be evaluated with performance metrics which are precision, accuracy, sensitivity, Fi score, recall, and specificity to check how each model is faring.

Model interpretability.

The dataset for each of the model's train and test formats and the library for the LIME and SHAP were imported to make an explainer object.

4.8.3 Lime.

An explainer was used on the already created model which is the Random Forest and neural network model.

Step 1.

For example, in the Random Forest with the following parameters stated in the LIME explainer, which created an object of the lime to be explained.

- Features name: The names of the column or attributes in the training set.
- Mode: The type of problem to be explained
- Class: The classes in the target variable. Employee Left: (Yes or No)

- Data: This is the data being used to create an explainer, it is basically the training dataset.

Step 2.
Lime explains the contribution of each of the features by training the small perturb of the instance to give a good local interpretation of the model, this is done by finding the probability function and selecting an instance and number of features to be explained and the weight of each of the feature is used in explaining the local behavior. The result of the feature's contribution is displayed to show the contributions of each feature.

4.8.4 Shap.

This model was applied to the Random Forest and neural network model to explain the performance of the attributes in the model by using the method of gaming theory to explain and assign importance to each attribute in the model and the contribution of each feature is measured by the Shapley values where positive indicate the features that are increasing the prediction to higher value and negative values show those attributes that are pushing the prediction to the negative side, this two shows the magnitude of the Shapley values. The result of the above interpretation was plotted with summary plots and a waterfall plot which shows a significant level of features' importance and the contributions of each of the features on the model prediction.

5.0 Evaluation.

The performance of the model was measured with performance metrics as enumerated below and the research used two major evaluation metrics such as LIME and SHAP to analyze the model variables and their effects on the model performance with two experiments performed on the HR-analytics dataset and Employee attrition for healthcare

5. 1 Performance metric.

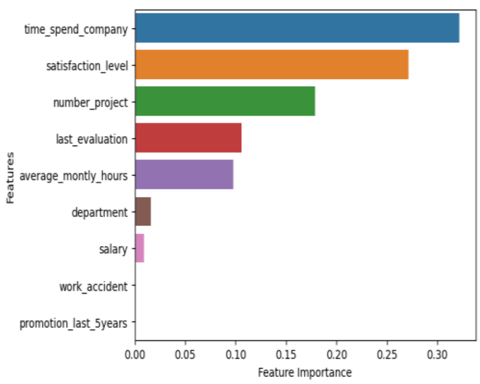
The performance of the model was measured using classical evaluation metrics such as accuracy, recall, F1 score, and precision in the models used as stated below obtained from FP: False Positive, TP: True Positive, FN: False Negative, and TN: True Negative,

Table 5: Result for HR analytics with imbalanced dataset.

Model \ Class	Accuracy	Precision	Recall	F1 Score	support	Specificity
Random Forest \ 0		0.98	1.00	0.99	1508	0.92
1	0.98	0.99	0.92	0.95	361	
Neural Networks \ 0		0.98	0.99	0.98	1508	0.91
1	0.97	0.96	0.91	0.93	361	

Table 6: Result for HR analytics with balanced dataset. (Sampling was used)

Model \ Class	Accuracy	Precision	Recall	F1 Score	support	Specificity
Random Forest \ 0		1.00	1.00	1.00	1518	0.99
1	0.99	1.00	1.00	1.00	1509	
Neural Network\0		0.97	0.95	0.96	1518	0.97
1	0.96	0.95	0.97	0.96	1508	



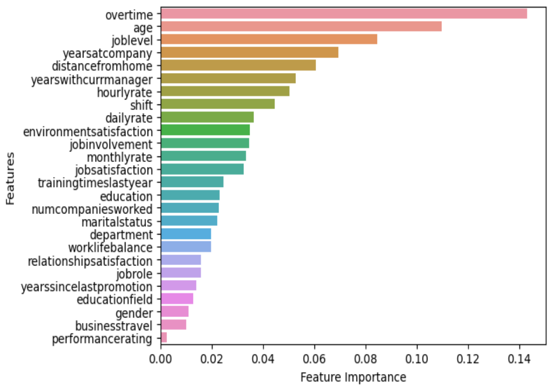


Figure 4a: Features plot in experiment 1. **Figure 4b: Features plot in experiment 2.**

Features importance in Random Forest is shown in figure 4a, the model prediction is in favor of those employees that have left the organization that is predicted left as applied to the first dataset with 99% accuracy seen in table 6 with the sampled dataset, this shows some of the features that contributed to employee turnover and an organization need to work on these factors to be able to retain its most effective and efficient worker as these are important for an employee to be retained in a firm or organization. The factors are listed in descending order from 'time_spent', 'satisfaction_level', 'number_projects', 'last_evaluation' 'average_hours', 'department', 'salary', 'work accident', and 'promotion_last'. 'Promotion_5years' does not have any effect on employee attrition or retention in an organization as seen above so an employee can decide to stay or leave whether he or she got promoted. The limitation of the Neural Network is that the contributing factors in the model are not easily interpretable as done in Random Forest, but this is possible using an explainable ai model that will be used in the second experiment below.

Table 7 Result for dataset 2 for Balanced

	Accuracy	Precision	Recall	F1 Score	Support	Specificity
Random Forest \0	0.99	1.00	0.97	0.99	312	1.00
1		0.97	1.00	0.99	279	
Neural Network \0	0.96	1.0	0.93	0.97	312	1.00
1		0.93	1.00	0.96	279	

The model predicted that employees are not in the organization, as described by the features importance plot in Random Forest in descending order, 'overtime', 'age', 'job_level', 'yearatcompany', 'distance_from_home', 'yearscurmanager', 'hourly_rate', 'shift', montly rate, shift, etc are the attributes contributing to employee turnover as an organization needs to work on and improve and check regularly to retain it most versatile employees to boost productivity and this has a significant effect on model prediction and the model has higher accuracy in recall and precision which is clearly identify the probability of employee leaving the organization and low impact by these features 'business travel', 'performance rating', gender, and 'Education' as seen in Figure 4b above.

Table 8: Model Evaluation for the Datasets

Dataset	Predicted class	Meaning
1 HR analytics	Class Yes for Left attribute	Employee has left
2 Healthcare	Class Yes for Attrition attribute	The employee has left

5.2 Experiment 1 on the HR analytics dataset

1. Lime

In this experiment, the LIME model was applied to the already created models to explain the factors that an organization needed to work on to improve employee retention in an organization. This was applied to the Random Forest model with five instances and five features selected to be explained for the correct prediction as shown in Figure 5a below.

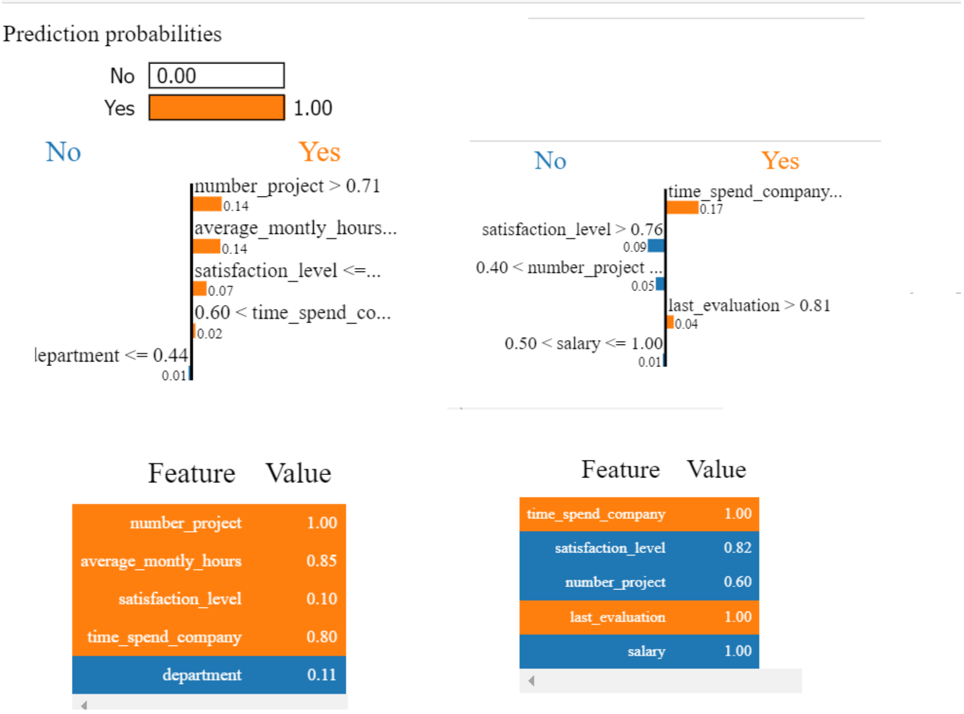


Figure 5a: Lime for Random Forest and Neural Network Model.

This shows LIME on the Random Forest model with class ‘Yes’ as the predicted class, which shows that the employees are no longer in the organization in yellow and happened to be a correct prediction for the model with five instances and features being explained from the data. The following are the factors that the organization needs to improve and work on to retain its staff,‘number_project’,‘average_montlyhour’, ‘time_spend’, and ‘satisfaction_level’ contributed positively with feature probability in descending order while the ‘department’ with a feature probability of 0.01 contributed negatively to the model prediction outcome for the instance selected. For 15 rows and five features, the two models have ‘satisfaction level as the feature that contributed to attrition for class ‘Yes’. It was found that the Random Forest model and Neural network assigned different levels of importance as ‘time-spend_comany’ is the only feature in common for the same instance explained.

2. Shap.

This is the global interpretation of the attributes that are responsible for or influencing employee retention based on the SHAP values that is the higher the SHAP value the greater the impact or positive contribution to the model prediction.

I. Summary plot for HR-analytics dataset.

In Figure 6a below, the features are ranked according to their importance as the uppermost features have a significant contribution to the model prediction, and the bottom features with less or no impact are arranged in this order: 'satisfaction_level', 'number_projects', 'time_spend', 'average_montly_hours' down to 'work_accident', as the uppermost features influenced the model prediction positively in the neural network model

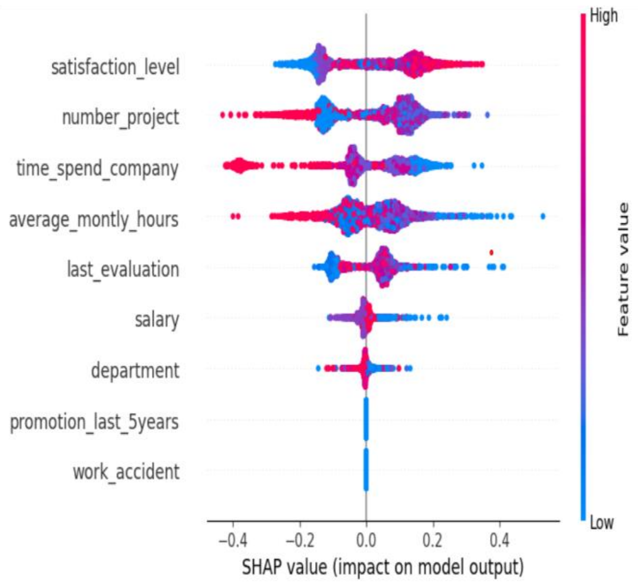


Figure 6a: Summary Plot in Neural Networks

Figure 6b: Waterfall Plot Neural Networks

The features with red color are pushing the model prediction to the positive side of predicting those that quit their jobs as also observed by the feature importance in Random Forest and the and in LIME. From Figure 6, the global effect of the factors on the model prediction depends on the instance being explained, there are times that those features are also pushing the model to the negative side could also bring about positive change. The blue color in some of the features 'satisfaction_level' reduces the model outcome by -0.3 and increases the model prediction by 0.4 while 'time_spend' increases by -0.42 and reduces by 0.39, and 'average_monthly_hours' have high SHAP values but decrease the model by 0.42 and increases by -0.4 respectively

The summary plot for the two models has three same factors in common in descending order which is 'satisfaction level', 'number_project', and 'time spend company', while the last two are interchanged ('average_monthly_hours', and 'last_evaluation') for Random forest contributed to the model prediction, the position of the next two factors is interchanged in the models while the last two variables on the plot have no impact on both models.

II. Waterfall plot for HR-analytics dataset

From Figure 6b above, 'average_montly_hour' and satisfaction_level was pushing the model prediction up without interaction with other variables while interaction is observed in 'number_projects', 'time_spend', 'department', and 'last_evaluation' as the first three boost predictions outcome that is factors with high Shap values have a positive impact on the model result while'last_evaluation' decreases the prediction. 'Salary', 'work_accident', and 'promotion_last _5years' slightly shows likely impact on the models. The waterfall plot in the Random Forest shows that three factors are pushing the prediction up and these are 'satisfaction_level', 'average_monthly_hour', and 'number _projects' while 'time_spent', 'last_evaluation', 'department', and 'salary' are pushing it to the negative side and the last two factors have low or no effect or impact on the model prediction.

5.2 Experiment 2 on Employee attrition for healthcare dataset

1. Lime

Figure 7a. below shows the predicted outcome with ten features and two instances or rows that employees have not left the organization with 99% probability and1% represents the percentages of those that quit their job as the distance from home is the reason or indicator why they left the organization as shown with Lime on Random Forest classifier. For every instance checked, some factors are contributing to the model prediction either positively or negatively. The features in blue support the model outcome that employees are still in the organization with high features probability of 0.24 as seen in overtime.

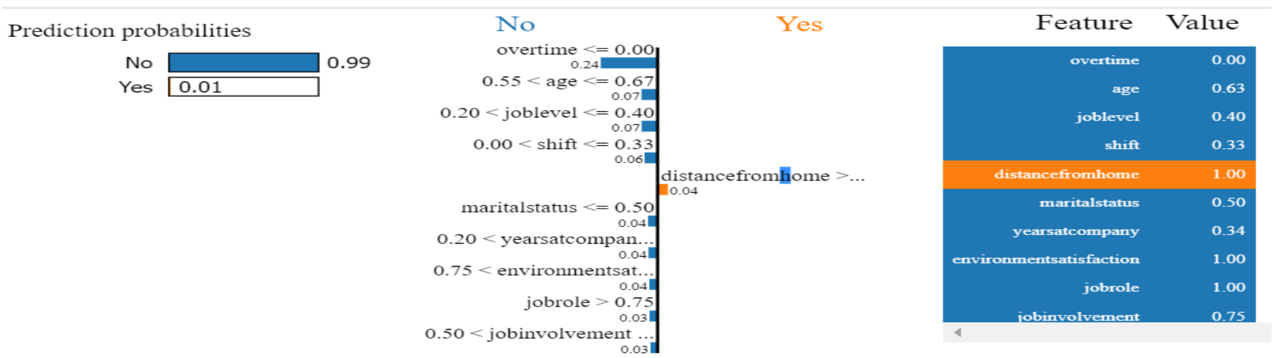


Figure 7a: Lime on Random Forest model for Employee attrition for healthcare

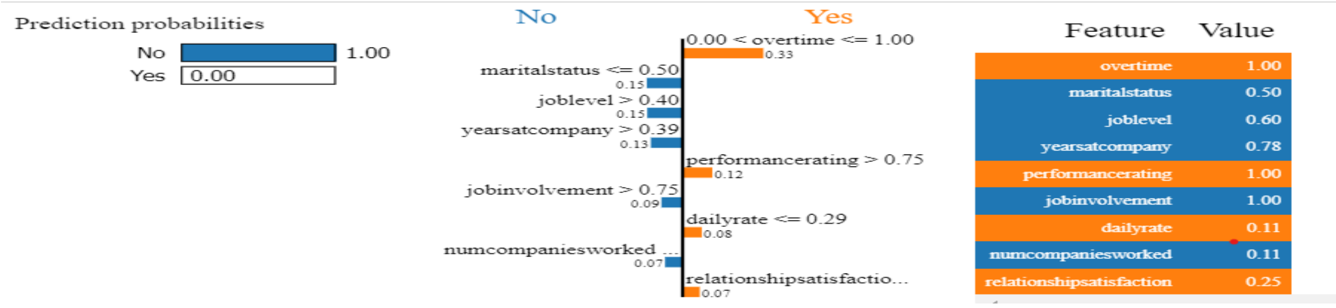


Figure 7b: Lime on Neural Network model for Employee attrition for healthcare

In the Figure 7b above, lime was used on the Neural network model with nine features, and two rows or instances from the dataset were explained, the result shows that employees are still in the organization with 100% probability with features in blue color as an indicator: 'maritalstatus' with feature probability of 0.15, 'joblevel' 0.15'yearatcompany' 0.13 are the factors that supported employee retention in the organization while 'overtime', 'performance rating','dailyrate' and 'relationship' are the reason while others quit for the instances being explained.

2. Shap on employee attrition for healthcare dataset

The second dataset[2] predicted no attrition as the outcome of the model as the significant effect of each of the factors is seen in figure 8a and the global impact of each of the factors can be seen from the waterfall plot from fig 8b as 'department', 'joblevel', 'worklifebalance', 'daily rate', 'jobinvolvement' contributed positively with interaction with those factors that are contributed negatively to the prediction, it shows that the model predicted that employees are still in the organization.

Summary plot for attrition for healthcare dataset.

According to the summary plot in the second data in Figure 8a with the Random Forest model, the six most significant factors that are contributing to employee retention are 'overtime', 'age', 'job level', 'year at company', 'shift' and 'years with current manager', depending on the instance been selected and explained. A factor may have a negative effect at some point and a positive effect at some point in the model that is contributing both positively and negatively to the model performance as seen below in the force plot in the Random Forest model with a cross-section of the effect of the factors on all X test variables as seen in Figure 9.

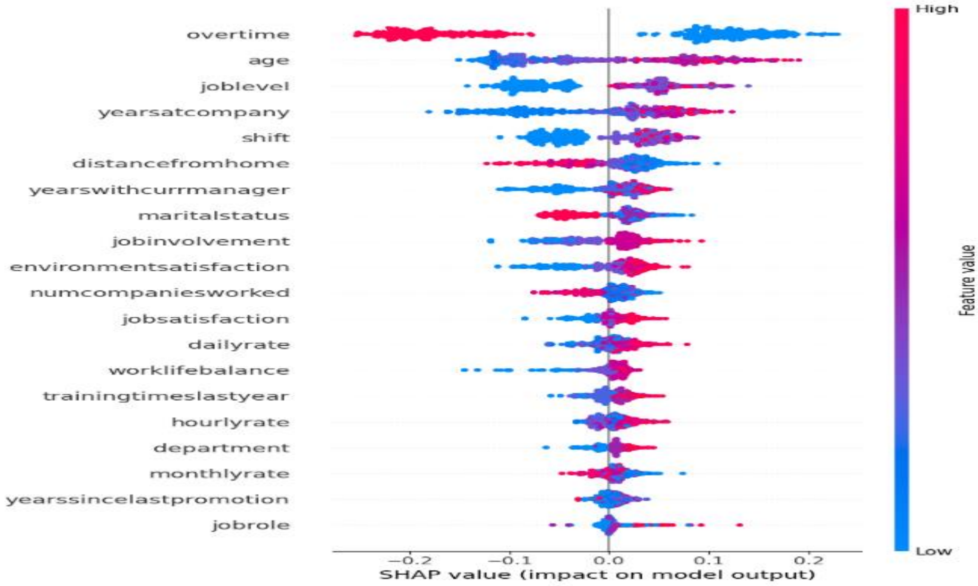


Figure 8a: Summary Plot in Neural Network

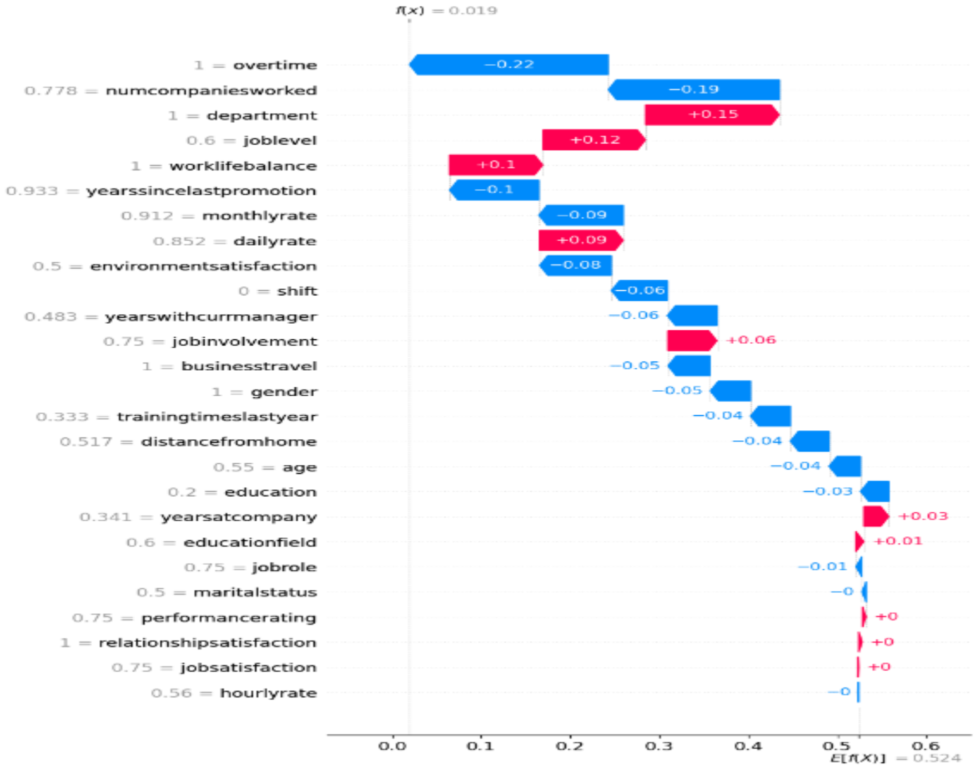


Figure 8b: Waterfall Plot Neural Network

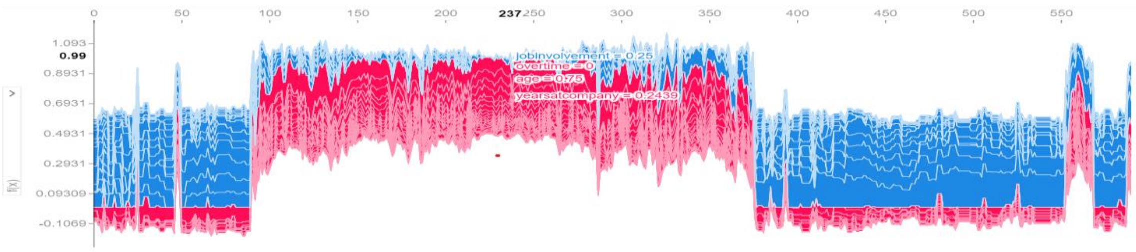


Figure 9: Force plot

The following factors ‘department’, ‘job level’, ‘work-life balance’, ‘daily rate’, and ‘job involvement’ have significant effects and positive impacts on the result by pushing the prediction of the model up with factors such as overtime, ‘numcompaniesworked’, ‘year sincelastpromotion’, ‘montlyrate’, ‘shift’ and ‘environmentsatisfaction’ contributed negatively to the prediction by lowering the outcome of the result.

5.4 Discussion

From the results obtained from the experiments above, it was discovered that the models predicted that employees are no longer in the organization in experiment one with 'number_project', 'average_montly_hours’, 'satisfaction_level', and 'time_spend' as the most significant features that contributed to this. The random forest model gives more feature contributions to employee attrition than the Neural network model this is because the random forest used features that are more relevant to the prediction in the selected instance while the Neural network captures the interaction between the features to achieve the prediction result. It was discovered that the higher the Shap values the higher the impact of features (positive contribution) and vice versa and interactions between the features resulted in changes in the result which is either positive or negative. For Employee attrition for healthcare, two models have the following features in common that contributed to retention, these are ‘joblevel’, ‘yearatcompany’,‘jobinvolvement’while‘overtime’,‘age’,’joblevel’,‘shift’,‘yearatcompany’,‘ distancefromhome’, ‘yearswithcurmanager’ are part of the first seven factors that have a significant impact on the result of prediction that is those features with higher shap values as shown in figure 8a above. Several interactions occur among the features in the waterfall plot in Figure 8b ‘number_companies_worked’ and ‘department’, ‘worklifebalance’ and ‘yearsincelastpromotion’, ‘monthlyrate’, ‘dailyrate’ and ‘environmentsatisfation’ to in influence prediction outcome and some features have high Shap values in red and those with low Shap values in blue.

Are the results obtained by LIME more informative than SHAP:

The result obtained through LIME is a simple linear model interpretation that captured a single instance being selected for explanation, that is, individual prediction for a specific instance, as this does not explain or give the broader view but a subset or an extraction of information for employee retention. The selection of the instance and the number of variables (proper selection) for explanation also affect the information or result produced by LIME, less information if selection of instance is not properly done. For an organization that is just developing with the small team or handful of employees, LIME will be the best option to explain individual model predictions for employee attrition, that is, why an employee quits from a certain department or unit to the stakeholders when the model is not a complex one, and this will enable the management to handle the attrition problem by providing suitable retention strategies as the case may be.

SHAP provided both global and local interpretation for the model output and captured feature interactions and relationship within the attribute or features used as shown by waterfall plot in figure 8b used to explain the contribution or effect of each feature by measuring or computing the SHAP values by considering all possible features combination to get the feature importance of the attributes.

Shap is more informative than LIME when applied in a big organization, but lime will be useful in a developing organization.

5.4.1 Retention Plan.

The Human resources department should adopt the following in their retention plan for healthcare and companies that take or give priority to their workforce. Regular surveys and feedback could be put into place to check employee satisfaction in companies and all issues raised should be addressed by the human capital by maintaining open communication, recognition for productive employees and team, work-life balance with convenient work arrangements and social gatherings in place, offering a competitive salary as stated by Sorn, M.K. et al.2023.The management sees that employees are not loaded with too much work with the help of project management tools in place, the head of the department should set targets and encourage teamwork to reduce stress on individual employees Sandhya K. et.al (2021). There should be designated time allotted for breaks for workers to ease out stress and regular training on tasks and time management to improve and enhance productivity, the organization should encourage employees to do a SWOT analysis of their job functions to discover their strengths, weaknesses, opportunities attached, a threat to their productivity, and ways to enhance their skills and performance by celebrating their strengths with recognition.
it was observed that overtime is one of the significant factors but contributed negatively to retention[2], proper job rotation should be encouraged, and workload shared among teams as suggested by Sandhya k, et. al (2021). Gender inclusion should be encouraged or work patterns and should not be biased regardless of gender or sex with equal pay.

6. Conclusion and Future Work

In this research, Explainable AI methods were used to discover the main factors that contribute to employee turnover when using Machine Learning models (Neural Networks and Random Forest) on the two datasets, and the two models performed well with higher accuracy, but this cannot be used to determine the contributions of the features or factors for employee retention in an organization. To address this, LIME and SHAP were used to explain factors contributing to employee retention for better understanding by stakeholders.
On human resource (hr) analytical data with Lime, it was discovered that ‘number-project’, ‘average_montly_hour’, ‘time-spend’, and ‘satisfaction_level’ in their order of importance contributed and the contribution on the features determine the model outcome to employee attrition that needed to be addressed for an organization to be productive in its businesses, Shap was used on the same model and it was discovered that ‘satisfaction_level’, ‘number_project’, ’time_spend’, and ‘average_montlyhour’ are the key contributors to employee attrition with varying interactions and interpretation in the features as LIME only provides the influence of the factors to the specific instance been explained that is a specific data point while Shap gives an average effect and interaction of the factors across the whole dataset.
The organization should work on the following to retain its employees; ‘overtime’, ‘numcompaniesworked’, ‘age’, ‘education’, ‘gender’, and ‘distancefromhome’ being negative contributors as stated in the ‘employee attrition-for-healthcare’. Priority should be

given to all these features when setting up a retention plan as salary was not the major reason why an employee quit. HR could invest their resource by designing a good retention strategy that uses the ideas listed in **Section 5.4** by working with the management to key into the goals, improve employee work-life balance and improve the organization’s productivity without harming the workforce.

The limitation of this research is that the first dataset has fewer attributes with large data and while the second data had a lot of attributes with the dataset being small and the Random Forest model gave a higher accuracy with the data without the use of hyperparameter tuning, also the class present in the target variable was not well represented as the class was imbalance.

For future work, this approach can also be explored on unstructured text data from employee feedback with an explainable ai to explain and obtain information on retention features also in Natural language processing likewise a large dataset with more features could be used to capture a broader view of the features with the application of hyperparameter tuning to improve the performance of the model applied.

References

Alshehhi, K., Zawbaa, S.B., Abonamah, A.A. and Tariq, M.U., 2021. Employee retention prediction in corporate organizations using machine learning methods. *Academy of Entrepreneurship Journal*, *27*, pp.1-23.

Bhartiya, N., Jannu, S., Shukla, P. and Chapaneri, R., 2019, March. Employee attrition prediction using classification models. In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) (pp. 1-6). IEEE.

Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., Ghosh, J., Puri, R., Moura, J.M. and Eckersley, P., 2020, January. Explainable machine learning in deployment. In *Proceedings of the 2020 Conference on fairness, accountability, and transparency* (pp. 648-657).

Chowdhury, A.H., Malakar, S., Seal, D.B. and Goswami, S., 2022. Understanding employee attrition using machine learning techniques. In *Data Management, Analytics, and Innovation: Proceedings of ICDMAI 2021, Volume 2* (pp. 101-109). Springer Singapore.

C.N. and Krishna, A., 2019. Employer branding: perceived organizational support and employee retention–the mediating role of organizational commitment. (“Employer branding: perceived organizational support and employee ...”) *Industrial and Commercial Training*, *51*(3), pp.174-183.

Dutta, S., and Bandyopadhyay, S.K., 2020. Employee attrition prediction using neural network cross validation method. *International Journal of Commerce and Management Research*, *6*(3), pp.80-85.

Guerranti, F. and Dimitri, M., 2023. A Comparison of Machine Learning Approaches for Predicting Employee Attrition. *Applied Sciences*, *13*(1), p.267.

Idit Cohen (2021)Explainable AI (XAI) with SHAP -Multi-Class Classification Problem .Available at:URL:https://towardsdatascience.com/explainable-ai-xai-with-shap-multi-class-classification-problem-64dd30f97cea [Accessed July 2023].

Joseph, R., Udupa, S., Jangale, S., Kotkar, K. and Pawar, P., 2021, May. Employee attrition using machine learning and depression analysis. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1000-1005). IEEE.

Khalifa, N., Alnasheet, M. and Kadhem, H., 2022, May. Evaluating Machine Learning Algorithms to Detect Employees' Attrition. In *2022 3rd International Conference on Artificial Intelligence, Robotics and Control (AIRC)* (pp. 93-97). IEEE.

Maria-C Lesan., 2022, April. Classical machine-learning classifiers to predict employee turnover. In *Education, Research and Business Technologies: Proceedings of 20th International Conference on Informatics in Economy (IE 2021)* (pp. 295-306). Singapore: Springer Singapore.

Marvin, G., Jackson, M., and Alam, M.G.R., 2021, August. A machine learning approach for employee retention prediction. In *2021 IEEE Region 10 Symposium (TENSYMP)* (pp. 1-8). IEEE.

Qutub, A., Al-Mehmadi, A., Al-Hssan, M., Aljohani, R. and Alghamdi, H.S., 2021. Prediction of employee attrition using machine learning and ensemble methods. *Int. J. Mach. Learn. Comput*, *11*(2), pp.110-114.

Rajeswari, G.R, Murugesan, R, Aruna, R., Jayakrishnan, B. and Nilavathy, K., 2022, October. Predicting Employee Attrition through Machine Learning. In *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 1370-1379). IEEE.

Sabbineni, N., 2020. Understanding Employee Attrition Using Explainable AI.

Sandhya, K. and Kumar, D.P., 2021. Employee retention by motivation. *Indian Journal of science and technology*, *4*(12), pp.1778-1782.

Sekaran, K. and Shanmugam, S., 2022, March. Interpreting the factors of employee attrition using Explainable AI. In *2022 International Conference on Decision Aid Sciences and Applications (DASA)* (pp. 932-936). IEEE.

Shilpi Bhattacharyya(2020) Explaining Black Box Models: Ensemble and Deep Learning Using LIME and SHAP. Available at: URL:https://www.kdnuggets.com/2020/01/explaining-black-box-models-ensemble-deep-learning-lime-shap.html.

Sorn, M.K., Fienena, A.R., Ali, Y., Rafay, M. and Fu, G., 2023. The Effectiveness of Compensation in Maintaining Employee Retention. *Open Access Library Journal*, *10*(7), pp.1-14.

Understanding Model Predictions with LIME (2018) Available at : URL: https://www.datacamp.com/tutorial/understanding-model-predictions-lime

Yadav, S., Jain, A. and Singh, D., 2018, December. Early prediction of employee attrition using data mining techniques. In *2018 IEEE 8th international advance computing conference (IACC)* (pp. 349-354). IEEE