

TABLE OF CONTENTS

INTRODUCTION	3
PREDICTIVE ANALYTICS AND MACHINE LEARNING IN HR.....	3
BUSINESS SIGNIFICANCE OF EMPLOYEE ATTRITION.....	3
CHAPTER ONE - DATA UNDERSTANDING, PREPARATION AND EDA.....	4
DATASET DESCRIPTION.....	4
PROBLEM DEFINITION	4
CHAPTER TWO - DATASET CHARACTERISTICS AND DATA TYPES.....	5
FEATURE OVERVIEW	5
DATA TYPE IDENTIFICATION	5
DATA CLEANING AND PREPROCESSING	5
MISSING VALUE ANALYSIS	5
DUPLICATE RECORDS ANALYSIS.....	6
ENCODING OF CATEGORICAL VARIABLES	6
EXPLORATORY DATA ANALYSIS	7
DESCRIPTIVE STATISTICS	7
ATTRITION DISTRIBUTION ANALYSIS	8
CORRELATION ANALYSIS.....	8
KEY FACTORS INFLUENCING ATTRITION	9
CHAPTER THREE - SUPPORT VECTOR MACHINE CLASSIFICATION	11
SVM MODEL DEVELOPMENT	11
SVM THEORY AND RATIONALE.....	11
DATA SPLITTING AND FEATURE SELECTION	11
MODEL TRAINING	11
SVM PERFORMANCE EVALUATION	11
EVALUATION METRICS	11

RESULTS INTERPRETATION	12
<u>3D VISUALIZATION OF SVM CLASSIFICATION BOUNDARY</u>	<u>12</u>
<u>CHAPTER TWO - XGBOOST OPTIMIZATION MODEL</u>	<u>14</u>
<u>XGBOOST MODEL IMPLEMENTATION.....</u>	<u>14</u>
XGBOOST ALGORITHM OVERVIEW	14
MODEL TRAINING	14
<u>HYPERPARAMETER OPTIMIZATION</u>	<u>14</u>
LEARNING RATE OPTIMIZATION	14
NUMBER OF ESTIMATORS OPTIMIZATION	15
<u>XGBOOST PERFORMANCE ANALYSIS</u>	<u>16</u>
<u>CONCLUDING REMARKS</u>	<u>18</u>
<u>BIBLIOGRAPHY.....</u>	<u>19</u>
<u>APPENDIX (IF NECESSARY)</u>	<u>21</u>

INTRODUCTION

Predictive Analytics and Machine Learning in HR

Predictive analytics is a major contributor in the modern organizations since it has the potential of transforming the past data into actionable information that assists in making decisions in the future. Statistical and machine learning methods give organizations an opportunity to identify the trends, future outcomes, and improve the performance of their activities. The use of predictive analytics in Human Resource (HR) management helps organisations understand the behaviour of their workforce, enhances the engagement of the employees and reduces the threat of turnover. Machine learning also enhances HR analytics by automatically uncovering concealed connections in employee data. Therefore, these technologies are becoming a prediction tool that organizations are utilizing in employee attrition, talent management, and proactive and data-driven workforce planning.

Business Significance of Employee Attrition

Employee attrition is one of the most crucial issues that organizations nowadays must address. The high staff turnover can result in a huge financial loss, which can be associated with recruitment, onboarding, training and loss of productivity. When the experienced employees are lost to an organization, it is accompanied by a loss of valuable institution knowledge and expertise, which in most instances affect the performance of a team and quality of service delivery.

The IBM HR Analytics Employee Attrition dataset utilized in this paper has the data on 1,470 employees and 35 variables related to the workforce that can be used to explore the variables affecting employee turnover. Analyzing the data, it can be seen that most of the workforce remains in the company and the number of people who quit is lower, this is why it is important to evaluate properly employees who may be inclined to leave the company. Organizations can potentially predict potential patterns of attrition in the early stages and design retention-specific strategies with the help of machine learning algorithms such as Support Vector Machines (SVM) and Extreme Gradient Boost (XGBoost). Early forecasting will enable the HR managers act in time, improve on employee satisfaction and reduce massive costs incurred in employee turnover.

CHAPTER ONE - DATA UNDERSTANDING, PREPARATION AND EDA

Dataset Description

The dataset to be used in this study is the IBM HR Analytics Employee Attrition and Performance dataset that is accessible on Kaggle. It has 1,470 records of employees and 35 variables including demographic, job related, compensation and satisfaction variables. The most significant ones are Age, Department, JobRole, MonthlyIncome, Education, JobSatisfaction and OverTime. Both numerical and categorical types of data are included in the dataset, which is appropriate in terms of data pre-processing, Exploratory Data Analysis (EDA), and machine learning (Sun et al., 2023). It is multidimensional, and thus offers useful insights on the behaviour of employees, and the dynamics of the workforce, so it is a suitable dataset in considering employee attrition via predictive analytics and classification modelling.

Problem Definition

The employee attrition prediction problem is a machine learning classification problem that is supervised and attempts to predict those employees who will most probably leave an organisation. Attrition is the target variable and has two categories; Yes and No. Predictive models could be used to identify patterns that relate to turnover, by analysing the characteristics of the employees and the workplace features. The business aim of it is to accurately forecast high-risk employees who may turn over, in order that organizations should use retention strategies to retain their employees, reduce the expense of recruitment, increase workforce stability and to foster data-driven human resource management.

CHAPTER TWO - DATASET CHARACTERISTICS AND DATA TYPES

Feature Overview

The IBM HR Analytics Employee Attrition data contains demographic, organizational, financial and behavioural variables which are relevant in the analysis of employee turnover. Some of them include Age, MonthlyIncome, Department, JobRole, YearsAtCompany and JobSatisfaction. These variables provide information on worker characteristics, working environment, pay and experience. Both machine learning classification and exploratory data analysis are enabled by the dataset as it has both numerical and categorical attributes in it. Its multidimensional design allows identifying factors which have a role in employee attrition and makes organizations make informed workforce management decisions.

Data Type Identification

There are both numerical and categorical variables in the dataset. The interpretation of these types of data is crucial since various preprocessing and machine learning methods have to be used based on the nature of each variable.

Feature	Type
Age	Numerical
MonthlyIncome	Numerical
Attrition	Categorical
Department	Nominal
Education	Ordinal
JobLevel	Ordinal

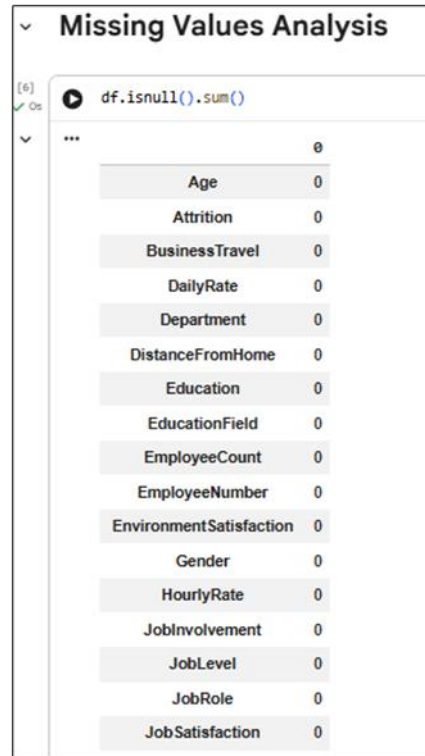
Numerical variables are variables that can take numbers as their measures such as Age and MonthlyIncome. These variables are employed to make calculations and statistical analysis. Categorical variables are categorical or groups of variables instead of continuous values (Huang, Algahtani and Kaewunruen, 2022). Attrition is categorical variable since employees fall under the category of Yes or No. Nominal variables like Department have categories that have no natural order. Ordinal variables, on the other hand, have order or hierarchy.

Data Cleaning and Preprocessing

Missing Value Analysis

The quality of the data was tested before the model was developed to determine the reliability of the analysis findings. Missing value analysis revealed the absence of a null or missing value in all the data and complete information in all the variables. The outcomes of the preprocessing stage revealed that

there were no missing values of any feature (Oukhouya and El Himdi, 2023). This implies that it is a good quality dataset that does not require imputation such as mean substitution, median replacement and predictive filling techniques. The fullness of the dataset minimizes possible bias and enhances the efficiency of the other machine learning models.



The screenshot shows a window titled "Missing Values Analysis". It displays the result of the code `df.isnull().sum()`. The table below lists 18 features, all of which have a count of 0, indicating no missing values.

Feature	Count
Age	0
Attrition	0
BusinessTravel	0
DailyRate	0
Department	0
DistanceFromHome	0
Education	0
EducationField	0
EmployeeCount	0
EmployeeNumber	0
EnvironmentSatisfaction	0
Gender	0
HourlyRate	0
JobInvolvement	0
JobLevel	0
JobRole	0
JobSatisfaction	0

Figure 1: Missing Values Analysis

(Source: Google - Colab)

Duplicate Records Analysis

Duplicate records were made to identify the repeated observations of the employees that could confound the model training and assessment. The duplicate detection procedure made sure that the number of duplicates in the dataset was zero. As a result, there were no rows to be removed as part of preprocessing (Kim et al., 2023). The lack of duplicate entries will make every employee observation unique to the analysis and help in the validity of the predictive modelling process.

Encoding of Categorical Variables

Machine learning algorithms take in numerical data; hence, categorical variables were encoded into numerical data using encoding methods. The target variable was Attrition where Label Encoding converted the responses to the yes and no to binary numbers. Other categorical variables, such as Gender, Department, and JobRole were also converted into preprocessing. Label encoding is used whereby the categories can be encoded with integer values in an efficient manner. In more complex

settings, the One-Hot Encoding could be used to have a single binary column reflecting each category, thereby preventing undesired ordinal relationships between nominal categories. Encoding assists machine learning models, such as Support Vector Machines and XGBoost to encode categorical data without degrading prediction.

Exploratory Data Analysis

Descriptive Statistics

The descriptive statistics were generated to present the most important characteristics of the data. The analysis revealed that the average age of the employees is approximately 36.9 years, which implies that the employees are very mature. The monthly earnings are around 6,503, but there is significant disparity among workers because of the job description and level of the organization (Shao, Ahmad and Javed, 2024). The dataset also demonstrates high diversity of the work experience and tenure related variables such as TotalWorkingYears and YearsAtCompany. These findings provide an overview of the labor market and creates a context of the turnover trends of the employees. The summary statistics also show that there is a considerable variance in the measures of salary, experience and satisfaction related variables implying that these variables may also be a cause of employee turnover behaviour.

Descriptive Statistics

```
df.describe().T
```

	count	mean	std	min	25%	50%	75%	max
Age	1470.0	36.923810	9.135373	18.0	30.00	36.0	43.00	60.0
Attrition	1470.0	0.161224	0.367863	0.0	0.00	0.0	0.00	1.0
BusinessTravel	1470.0	1.607483	0.665455	0.0	1.00	2.0	2.00	2.0
DailyRate	1470.0	802.485714	403.509100	102.0	465.00	802.0	1157.00	1499.0
Department	1470.0	1.260544	0.527792	0.0	1.00	1.0	2.00	2.0
DistanceFromHome	1470.0	9.192517	8.106864	1.0	2.00	7.0	14.00	29.0
Education	1470.0	2.912925	1.024165	1.0	2.00	3.0	4.00	5.0
EducationField	1470.0	2.247619	1.331369	0.0	1.00	2.0	3.00	5.0
EmployeeCount	1470.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
EmployeeNumber	1470.0	1024.865306	602.024335	1.0	491.25	1020.5	1555.75	2068.0
EnvironmentSatisfaction	1470.0	2.721769	1.093082	1.0	2.00	3.0	4.00	4.0
Gender	1470.0	0.600000	0.490065	0.0	0.00	1.0	1.00	1.0
HourlyRate	1470.0	65.891156	20.329428	30.0	48.00	66.0	83.75	100.0
JobInvolvement	1470.0	2.729932	0.711561	1.0	2.00	3.0	3.00	4.0

Figure 2: Descriptive Statistics Part

(Source: Google - Colab)

Attrition Distribution Analysis

The employee turnover was also looked at to obtain the graph of the percentage of employees who remained in the organization and those who left. The count plot indicates that majority of the employees are either in the category of No Attrition or in the relatively lower percentage in the category of Attrition. This disequilibrium means that employee turnover rate is lower as compared to the rate of employee retention. This type of class imbalance is typical in HR analytics and can affect the performance of models by making them tend to favour the majority class when predicting. The right evaluation measurements are thus needed when assessing the models.

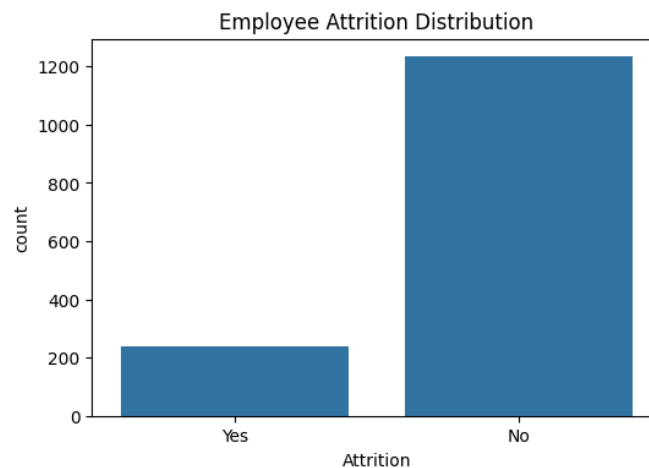


Figure 3: Employee Attrition Distribution

(Source: Google - Colab)

Correlation Analysis

The correlation analysis was to determine the relationship between the numerical variables and what might have been anticipated to be the predictor of employee attrition. It is possible to conclude that higher-organizational workers are likely to earn more since the correlation heatmap shows that MonthlyIncome and JobLevel are positively related, and there are strong correlations between these variables. On the same note, TotalWorkingYears has positive correlations with YearsAtCompany and JobLevel, which indicate that older workers are likely to be in higher positions (Singh, Biswas and Pal, 2022). YearsAtCompany is also significantly associated with variables managerial experience, career development in the company. The heatmap gives useful information on the structure of the workforce and shows the variables that could contribute to the prediction of attrition. The awareness of such relations will be useful in the selection of features and interpretation of the outputs of machine learning models.

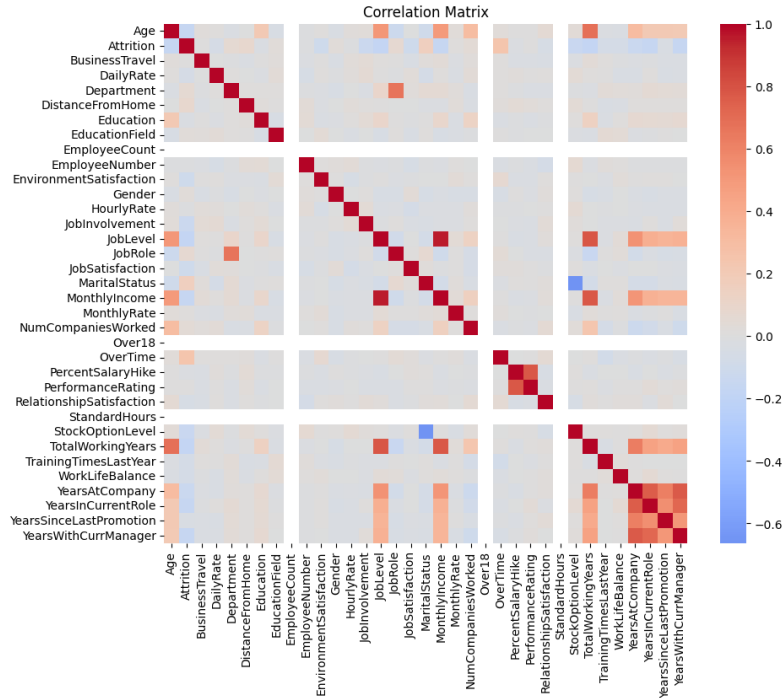


Figure 4: Correlation Heatmap

(Source: Google - Colab)

Key Factors Influencing Attrition

A few exploratory visualizations were employed to determine variables that were related to employee attrition. Analysis of the trends of overtime shows that employees who work overtime tend to have higher attrition rates than those who do not work overtime suggesting the existence of a work pressure-turnover relationship.

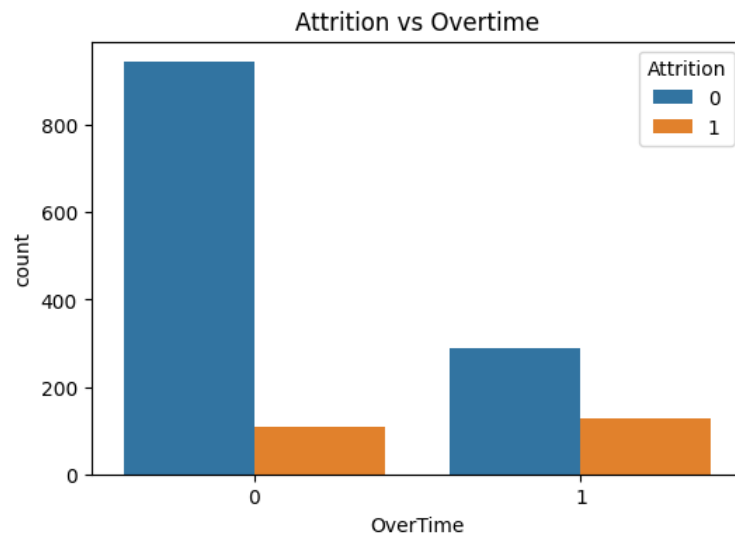


Figure 5: Attrition by Overtime

(Source: Google - Colab)

Job role analysis also reveals that the rate of attrition has been higher in some occupational groups, which suggests that some jobs can be more associated with dissatisfaction of employees or external hiring (Imani, Beikmohammadi and Arabnia, 2025).

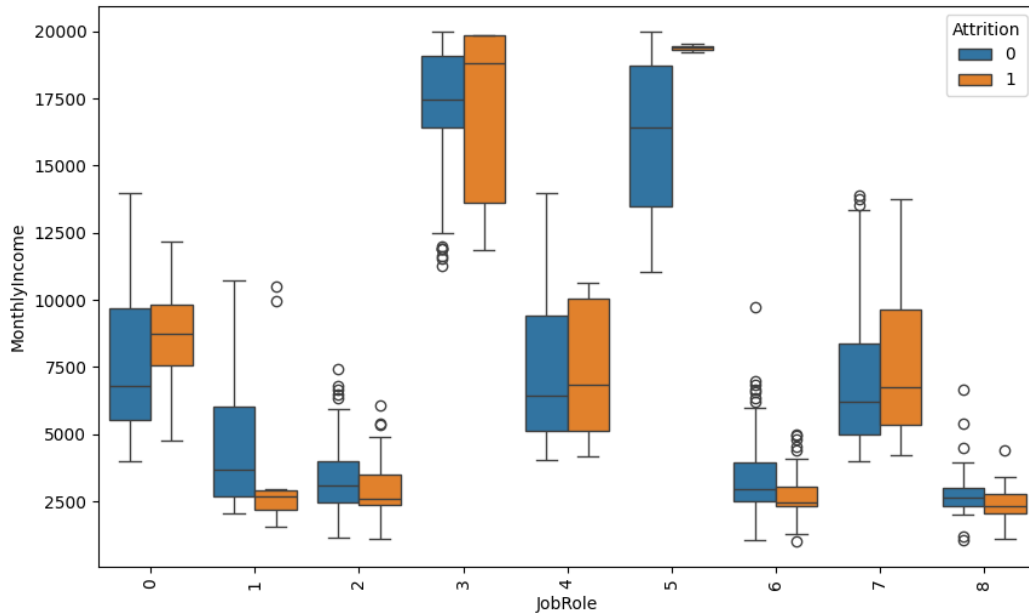


Figure 6: Attrition by Job Role

(Source: Google - Colab)

Comparisons of salary allocations done according to income across the occupation show that there is a big difference in income disbursement between the leavers and those who remain. The department level analysis also implies that there is a difference in turnover behaviour among the organizational units.

CHAPTER Three - SUPPORT VECTOR MACHINE CLASSIFICATION

SVM Model Development

SVM Theory and Rationale

The Support Vector Machine (SVM) is a supervised machine learning algorithm which is widely used in classifications due to the ability of generating the best decision boundaries between classes. The basic idea of SVM is the hyperplane which will serve as a barrier dividing the observations of two or more classes. The algorithm aims to maximize the distance between the hyperplane and the closest data points, or support vectors, thus enhancing the classification accuracy and generalization ability (Ileri, 2025). SVM is specifically useful in situations of high dimensionality data and complex relationships among variables. It is powerful, predicts well, and is capable of predicting linear and non-linear classification issues, making it suitable in employee attrition predictions.

Data Splitting and Feature Selection

Before the models were developed the data was split in 80:20 train test split. In so doing, the model will be able to extract trends using the available employee data and leave the observations that could not be seen as a part of the performance evaluation. Attrition was the target variable and the rest of the variables were utilized as predictive features (Villar and de Andrade, 2024). The StandardScaler technique then used to standardize all the numerical variables so that all variables were placed on a similar scale, thus enhancing the performance of the SVM and minimizing bias due to differences in the magnitude of the features.

Model Training

The Support Vector Classifier (SVC) found in the Scikit-Learn library was used to implement the SVM model. Linear kernel has been selected in order to come up with a plain decision-boundary between those employees that can leave and those that can remain. The model has been trained on the standardized training data and the relationship between the employee traits and the attrition results have been learnt. The test dataset was used to make predictions after training, which allowed assessing the results of the classification in terms of accuracy, precision, recall, F1-score, and confusion matrix analysis.

SVM Performance Evaluation

Evaluation Metrics

Measuring the performance of the Support Vector Machine model as an employee attrition predictor was done by measuring its performance using several measures of classification. The model managed to make an approximation of 86.4 meaning that most of the records of the employees were well-classified.

It was approximately 76.9 which demonstrates that a considerable portion of those classified as the cases of attrition were correctly identified among the employees (Imani and Arabnia, 2023). The recall was approximately 21.3 per cent which means that the model detected a low percentage of actual cases of attrition. F1- score of approximately 0.33 suggests the trade off between the precision and recall and is further demonstration that the predictors could still be enhanced to help identify the employees at the risk of leaving. Furthermore, the results of classification were analyzed in a confusion matrix to examine results in greater detail. The matrix showed 244 true Negatives and 10 true Positives and 37 attrition cases were not properly classified as non-attrition and 3 retention cases were not properly classified as attrition. The results are an overall evaluation of the model performance and classification behaviour.

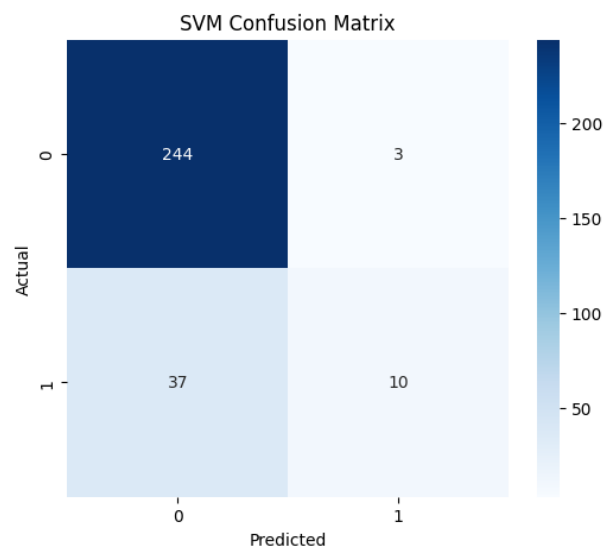


Figure 7: SVM Confusion Matrix

(Source: Google - Colab)

Results Interpretation

The SVM model showed good predictive results in terms of the relatively good accuracy and precision scores. The model successfully predicted those employees that would stay in the organization and had few false-positive predictions. Nevertheless, the recall value is rather low, which means that it is challenging to pick up all employees who left the organization actually. It is likely because the classes are not varied, and it can also be seen in the dataset, as the cases of non-attrition are much more common than attrition cases. Despite this weakness, the SVM model provides good predictive hints and can be adopted as a good benchmark to compare with the XGBoost model.

3D Visualization of SVM Classification Boundary

Three-dimensional visualization was plotted using the Age, MonthlyIncome, YearsAtCompany and it was used to understand how employees were leaving the company based on the data. The scatter plot

assists to visualize the clusters of the employees and determine the differences among the observations of attrition and non-attrition (Omar et al., 2024). The feature space is densely populated by most of the employees who are clustered in moderate age groups, average income group and less organizational tenure. Although there is some overlap in the two classes, a certain division is visible in some places in the plot. The visualization will help make the classification process much easier because it will demonstrate how the interaction of different characteristics of employees works.

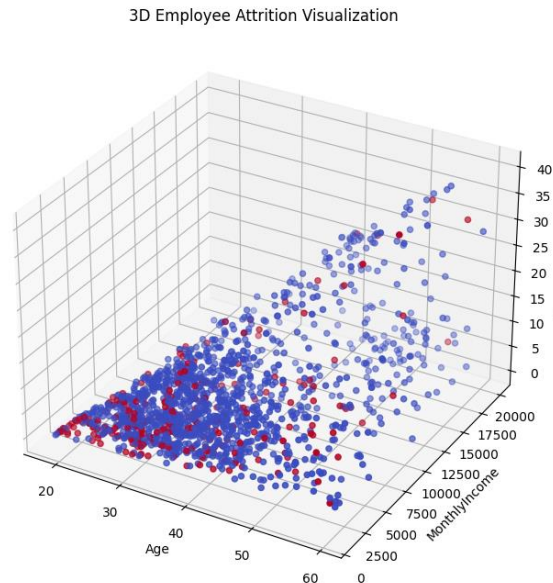


Figure 8: 3D Employee Attrition Visualization

(Source: Google - Colab)

CHAPTER TWO - XGBOOST OPTIMIZATION MODEL

XGBoost Model Implementation

XGBoost Algorithm Overview

Extreme Gradient Boosting (XGBoost) is a highly developed machine learning algorithm which is founded on the gradient boosting framework. The algorithm constructs a series of decision trees each subsequent tree tries to rectify the errors of the previous trees in prediction. This process of continuous learning enables XGBoost to be increasingly efficient at making predictions and minimizes classification errors (Nutakki, Koduru and Mandava, 2024). XGBoost being a tree ensemble learning methodology is a combination of the work of many weak learners to create a strong predictive model.

Model Training

The preprocessed employee attrition data have been coded and train-test split to train the XGBoost classifier. Multiple hyperparameters were set in the model training to control the learning process like the learning rate and the number of estimators. The learning rate indicates the worth of each tree to the entire model and the amount of estimators determines the total amount of boosting steps. The model was repeatedly trained to learn the relations between the attributes of employees and the attrition outcomes.

Hyperparameter Optimization

Learning Rate Optimization

The hyperparameters were optimized to make the model more effective and define the most efficient learning rate which would be beneficial in predicting the employee turnover. Several learning rate values were tested, including 0.01, 0.05, 0.10, 0.20, and 0.30. The results of the optimization indicated that there were variations in prediction accuracy across configurations. Lower learning rates led to less sharp learning behaviour, and an increase in learning rates increased the responsiveness of the model, but increased the risk of overfitting (Mishra et al., 2024). The resultant graph obtained on experimenting depicts the relationship between the values of the learning rates and prediction precision.

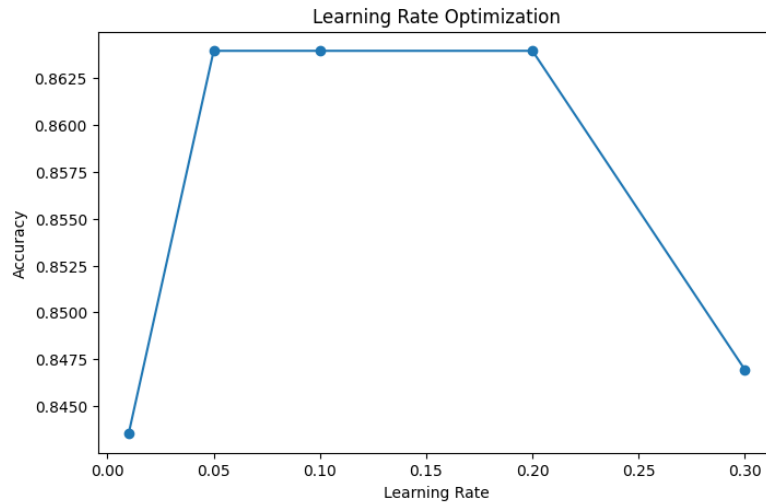


Figure 9: Learning Rate Optimization Results

(Source: Google - Colab)

Number of Estimators Optimization

The second optimization test was implemented to determine the best number of boosting trees in the ensemble model. The estimator values of 50, 100, 150, 200 and 300 were compared and assessed on the basis of the accuracy of the prediction. The higher the number of estimators, the higher the model learning capacity is likely to be since additional trees will counteract the residual errors. However, when the level is reached, the ensemble size can be bigger, making it costly and leading to a decrease in performance gain (Amjad et al., 2022). The optimization plot illustrates how the accuracy of estimating the estimates changed depending on the settings of the estimators. The selected estimator value was most favourable predictive accuracy, model generalization and computational efficiency.

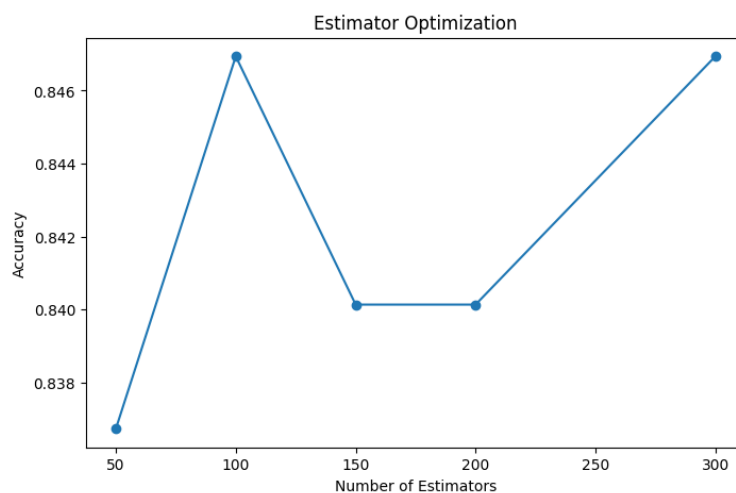


Figure 10: Number of Estimators Optimization Results

(Source: Google - Colab)

XGBoost Performance Analysis

The XGBoost model was optimized to produce good classification results, and provided a meaningful comparison to Support Vector Machine model.

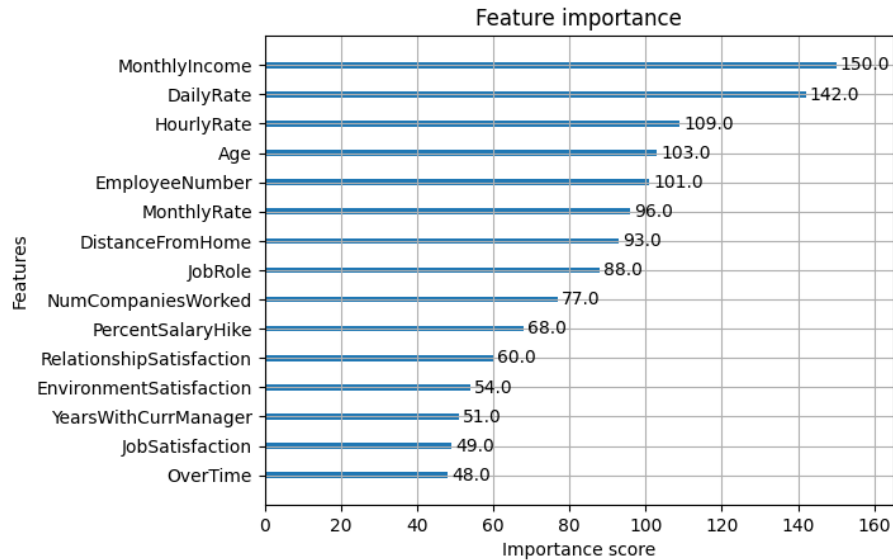


Figure 11: XGBoost Feature Importance Plot

(Source: Google - Colab)

As the analysis has revealed, XGBoost was more successful in the recall aspect, enabling the model to forecast more employees, who are susceptible to leave the company (Gayathri et al., 2022). Even though the general accuracy of the SVM model was good, the recall of the model was low, thus limiting its ability to detect cases of attrition.

	Model	Accuracy	Precision	Recall	F1 Score
0	SVM	0.863946	0.769231	0.212766	0.333333
1	XGBoost	0.846939	0.550000	0.234043	0.328358

Figure 12: Performance Comparison Between SVM and XGBoost

(Source: Google - Colab)

The SVM model based on the performance comparison was more accurate (86.39) and more precise (76.92) as compared to the XGBoost model, which implies that the SVM model performs better with respect to its performance overall. However, the XGBoost had a slightly higher recall rate (23.40% compared to 21.28%), which means that it was a little more effective in forecasting who among the employees would have left the organization. On the other hand XGBoost had lower bias trade-off

between accuracy, precision and recall and therefore it is particularly useful in employee retention related analysis where it is important to identify the potential instances of attrition.

CONCLUDING REMARKS

As demonstrated in this paper, predictive analytics, and machine learning can be utilized to address the employee attrition issue using the IBM HR Analytics Employee Attrition dataset. The Exploratory data analysis revealed important workforce variables, including the disproportion of the classes in the attrition variable and the high relationships between the salary, experience, and the organizational position. Analysis revealed that the overtime, monthly earnings, job position, age and years of working with the company are among the factors involved in employee turnover. Both Support Vector Machine and XGBoost models provided valuable predictive insights, with XGBoost delivering more balanced classification performance.

BIBLIOGRAPHY

- Amjad, M., Ahmad, I., Ahmad, M., Wróblewski, P., Kamiński, P. and Amjad, U., 2022. Prediction of pile bearing capacity using XGBoost algorithm: modeling and performance evaluation. *Applied Sciences*, 12(4), p.2126.
- Gayathri, R., Rani, S.U., Čepová, L., Rajesh, M. and Kalita, K., 2022. A comparative analysis of machine learning models in prediction of mortar compressive strength. *Processes*, 10(7), p.1387.
- Huang, J., Algahtani, M. and Kaewunruen, S., 2022. Energy forecasting in a public building: A benchmarking analysis on long short-term memory (LSTM), support vector regression (SVR), and extreme gradient boosting (XGBoost) networks. *Applied Sciences*, 12(19), p.9788.
- Ileri, K., 2025. Comparative analysis of CatBoost, LightGBM, XGBoost, RF, and DT methods optimised with PSO to estimate the number of k-barriers for intrusion detection in wireless sensor networks. *International Journal of Machine Learning and Cybernetics*, 16(9), pp.6937-6956.
- Imani, M. and Arabnia, H.R., 2023. Hyperparameter optimization and combined data sampling techniques in machine learning for customer churn prediction: a comparative analysis. *Technologies*, 11(6), p.167.
- Imani, M., Beikmohammadi, A. and Arabnia, H.R., 2025. Comprehensive analysis of random forest and XGBoost performance with SMOTE, ADASYN, and GNUS under varying imbalance levels. *Technologies*, 13(3), p.88.
- Kim, M.C., Lee, J.H., Wang, D.H. and Lee, I.S., 2023. Induction motor fault diagnosis using support vector machine, neural networks, and boosting methods. *Sensors*, 23(5), p.2585.
- Mishra, P., Al Khatib, A.M.G., Yadav, S., Ray, S., Lama, A., Kumari, B., Sharma, D. and Yadav, R., 2024. Modeling and forecasting rainfall patterns in India: a time series analysis with XGBoost algorithm. *Environmental Earth Sciences*, 83(6), p.163.
- Nutakki, M., Koduru, S. and Mandava, S., 2024. Quantum support vector machine for forecasting house energy consumption: a comparative study with deep learning models. *Journal of Cloud Computing*, 13(1), pp.1-12.
- Omar, E.D., Mat, H., Abd Karim, A.Z., Sanaudi, R., Ibrahim, F.H., Omar, M.A., Ismail, M.Z.H., Jayaraj, V.J. and Goh, B.L., 2024. Comparative analysis of logistic regression, gradient boosted trees, SVM, and random forest algorithms for prediction of acute kidney injury requiring dialysis after cardiac surgery. *International journal of nephrology and renovascular disease*, pp.197-204.
- Oukhouya, H. and El Himdi, K., 2023, April. Comparing machine learning methods—svr, xgboost, lstm, and mlp—for forecasting the moroccan stock market. In *Computer Sciences & Mathematics Forum* (Vol. 7, No. 1, p. 39). MDPI.

Shao, Z., Ahmad, M.N. and Javed, A., 2024. Comparison of random forest and XGBoost classifiers using integrated optical and SAR features for mapping urban impervious surface. *Remote Sensing*, 16(4), p.665.

Singh, R., Biswas, M. and Pal, M., 2022. Cloud detection using sentinel 2 imageries: a comparison of XGBoost, RF, SVM, and CNN algorithms. *Geocarto International*, pp.1-32.

Sun, J., Dang, W., Wang, F., Nie, H., Wei, X., Li, P., Zhang, S., Feng, Y. and Li, F., 2023. Prediction of TOC content in organic-rich shale using machine learning algorithms: Comparative study of random forest, support vector machine, and XGBoost. *Energies*, 16(10), p.4159.

Villar, A. and de Andrade, C.R.V., 2024. Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study. *Discover Artificial Intelligence*, 4(1), p.2.

APPENDIX (if necessary)



Dataset Link:

<https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-datase>

Google – Colab Link:

https://colab.research.google.com/drive/1q4i2d2lUVjGFPwHjo_u-aqPri3YIOz_f?usp=sharing