| RESEARCH ARTICLE

Employee Attrition Prediction in the USA: A Machine Learning Approach for HR Analytics and Talent Retention Strategies

Md Sumon Gazi[1] ✉ **Md Nasiruddin[2], Shuvo Dutta[3], Rajesh Sikder[4], Chowdhury Badrul Huda[5]** and **Md Zahidul Islam[6]**

[16]*MBA Business Analytics, Gannon University, USA*

[2]*Department of Management Science and Quantitative Methods, Gannon University, USA*

[3]*Master of Arts in Physics, Western Michigan University, USA*

[4]*PhD Student in Information Technology, University of the Cumberlands, KY, USA*

[5]*Masters of Science in Management, ST. Francis College*

Corresponding Author: Md Sumon Gazi, **E-mail**: sumonmkt87@gmail.com

| ABSTRACT

In the dynamic business domain in the USA, human capital is one of the most instrumental assets for companies. Maintaining high performance and reducing employee attrition has become an utmost priority in the USA since the costs related to employee attrition can be significant. The chief objective of this study was to explore the application of machine learning in terms of forecasting employee attrition and its ramifications for HR analytics and talent retention strategies. In this study, the investigator used Jupyter Notebook, an interactive platform for Python users, to design machine learning algorithms. The dataset utilized in this research was attained from the IBM Human Resource workforce attrition survey dataset. In the current research, the investigator proposed an array of machine learning models, most notably, Decision Tree, Ada-boost classifier, Random Forest, and gradient-boosted classifier. By referring to the model's performance evaluation, it was apparent that the Random Forest algorithm had the highest accuracy, followed by Gradient Boosting and Decision Tree respectively. AdaBoost had the lowest accuracy. Concerning precision, the Random Forest algorithm again had the highest precision followed by Gradient Boosting and AdaBoost accordingly. By implementing the proposed models' organizations in the USA can identify high-performing employees at risk of quitting, and subsequently take proactive steps to retain them, saving significant organizational resources. Ultimately, the proposed machine learning techniques can assist the government in maintaining high-performing employees, reducing the impact of labor shortages, and maintaining business continuity.

| KEYWORDS

| ARTICLE INFORMATION

1. Introduction

In the recent past, employee attrition or turnover rate has become a substantial concern for companies in the USA, because it results in loss of productivity, human capital, increased organizational costs, and the demand for talent acquisition. Employee attrition refers to the involuntary or voluntary departure of personnel from a company. Global surveys by the World Bank demonstrate that 15-25% of staff voluntarily quit their jobs annually (Naik, 2023). Therefore, retaining top performers is paramount for organizational success. In that respect, companies require strategies and tools to strategically pinpoint at-risk employees and resolve their needs respectively. With the large volume of employee data now available on the company database, machine learning can be utilized to build predictive models for employee attrition (Musanga, 2023). This study aims to delve into the

application of machine learning in terms of forecasting employee attrition and its ramifications for HR analytics and talent retention strategies.

Marvin and Jackson (2021), contend that in the versatile business spectrum in the USA, human capital is one of the most instrumental assets for companies. Retaining top performing and diminishing employee attrition has become an utmost priority in the USA since the costs related to employee attrition can be significant. These organizational costs comprise the disruptions in productivity, loss of institutional knowledge, and the high cost involved in training and recruiting new employees. As such, efficient talent retention tactics are instrumental for retaining a skilled and productive workforce, affirming business continuity, and promoting a competitive advantage.

IJRASET (2021), indicates that conventional methods for employee turnover rates frequently depend on reactive tools, such as the employee departure survey questionnaire, which gives Human Resource managers valuable information after the staff has already decided to quit their employment. Nevertheless, by deploying advanced machine learning and data analytics methods, organizations can take a comprehensive method to detect and reduce the risk of employee attrition. Machine learning, a sub-category of artificial intelligence, allows the crating of predictive models that can evaluate large volumes of datasets and expose relationships and patterns that may not be instantly visible. By deploying machine learning models to historical employee data, companies can build algorithms capable of forecasting the probability of employee churning grounded on various factors, such as career development opportunities, compensation, job satisfaction, and work-life balance.

1.1 Problem Statement
In retrospect, employee attrition presents a myriad of challenges for companies across industries in the USA, particularly, high attrition rates result in the loss of valuable talent, and human capital, disrupt team dynamics, as well as increase recruitment and training costs. The departure of well-trained and high-performing employees from an organization frequently results in a major organizational loss. In that light, this research paper aims to explore the root causes that lead to employee attrition, such as salary changes, promotional opportunities, the nature of the work environment, business travel opportunities, the nature of relations between superiors and subordinates, recognition and appreciation, and the time gap since the last promotion. Machine learning techniques will be the basis on which organizations in the USA can predict employee attrition. Such predictions can be made for the retention of valuable employees through various AI and machine learning methodologies.

2. Related Works
A myriad of groundbreaking studies have been conducted to explore the efficacy of machine learning algorithms in predicting employee attrition. For instance, Qutub et al. (202), investigated employee attrition by applying distinct machine learning techniques, most notably, Logistic Regression (LR), K-Nearest Neighbors (KNN), Naive Bayes (NB), Support Vector Machines (SVM), Decision Trees (DT), and random forests (RF). The investigators performed two different experiments – one integrating feature selection and the other without. As a result, across both experimental instances, the support vector machine (SVM) algorithm proved to be the most effective and accurate technique for predicting employee attrition.

By contrast, Sani (2023), in his in-depth study, discovered they found out that Naive Bayes had a relatively higher recall rate of 0.541. Nonetheless, they also discovered that deploying Naive Bayes on real-world turnover data presents challenges because of the assumption of predictor autonomy, which may not hold. This, combined with its lesser performance, results in relatively lower accurate predictions of employee attrition. In their study, they also established that the Decision tree algorithm performed second best, with logistic regression following closely in performance.

Raza et al. (2022), in their experimentation, ascertained that the XG-Boost (XGB) model outshined Logistic Regression, Random Forest, and Naive Bayes in terms of accuracy. They also outlined that XGB's innate regularization assists in terms of overcoming overfitting issues. The researchers equally pinpointed that the data used forecasting employee churning contains noise, and XGB can efficiently handle this noise, preventing it from being misinterpreted as relevant data.

Ajit (2022), designed an algorithm deploying a gradient-boosting classifier. After evaluating the algorithm's performance through precision, recall, and accuracy measurements, the gradient boosting tree model showcased superior outcomes compared to other techniques, attaining an accuracy of 97%. The experiment findings indicated that workers who are dissatisfied are more likely to quit and leave the company, while those with longer tenure and higher work involvement were less likely to resign.

IJRASET (2021), in their research, evaluated the efficiency of machine learning models in forecasting employee attrition. The goal of the investigators was to compare and contrast the accuracy of employee attrition forecasting attained via distinct machine learning algorithms, including Naive Bayes, Support Vector Machines, and Decision Trees. The result suggested that Random Forest displayed a relatively solid forecasting capability. The assessment indicated that employing more complicated machine learning

methods and validating feature importance through multi-criteria models could have potentially led to even higher levels of accuracy.

3. Methodology

In this research paper, the investigator uses Jupyter Notebook, an interactive platform for Python users, to design machine learning algorithms. According to Hasan et al (2024), Python is an open-source program prevalently utilized for data science and machine learning tasks. Besides, R programming was also used, which is also an open-source language for developing machine learning algorithms. R-programming affords various graphical and statistical approaches for evaluating data and crafting predictive analytics resolutions. Both Python and R-system were considered appropriate programming languages for the objective of this study, as they provide robust libraries and functionality for employing machine learning techniques and constructing predictive models from data.

3.1 Dataset

The dataset utilized in this research was attained from the IBM Human Resource workforce attrition survey dataset. Initially, the dataset comprised 35 features or variables. Nevertheless, 14 of these attributes were pinpointed as redundant and eliminated via data cleaning. This culminated in a final dataset with 21 attributes. Table 1 highlights and portrays these 21 attributes, entailing their distinguished data types as either numeric or non-numeric categorical variables (Pro-AI-Rokibul, 2024). Ultimately, these attributes capture distinct components of employees' job satisfaction, demographic characteristics, and work environment, to assist in evaluating and predicting the likelihood of attrition.

3.2 Pre-Processing

Pre-processing entailed data reduction and cleaning was performed, comprising tasks such as the transformation of feature type from numerical to nominal. Based on the pre-processing, four (4) features were pinpointed and subsequently eliminated, resulting in a remaining set of 21 attributes (Pro-AI-Rokibul, 2024). After the production of the interquartile filter, outliers were detected and eliminated in the dataset.

| S/No	Attribute	Data Type
1	Age bracket	Numeric type
2	Daily Rate	Numeric
3	Department	Categorical
4	Business Travel	Categorical
5	Education	Categorical
6	Gender	Categorical
7	Distance from home	Numeric
8	Job involvement	Categorical
9	Hourly Rate	Numeric
10	Job Satisfaction	Categorical
11	Job level	Categorical
12	Job role	Categorical
13	Monthly rate	Numeric
14	Monthly income	Numeric
15	Number of organizations worked	Numeric
17	Performance rating	Categorical
18	Total Working Years	Numeric
19	Years at the Organization	Numeric
20	Years at the Present Role	Numeric
21	Work-Life Balance	Categorical

3.3 Feature Engineering Selection

Feature selection in this research entailed the careful selection of relevant features while discarding redundant and irrelevant information from the dataset. The objective was to diminish the dimensionality of the dataset, therefore enhancing accuracy, combating overfitting, minimizing training time, and pinpointing the most subjective and predictive domains for the evaluation (Pro-AI-Rokibul, 2024). Several feature selection techniques, such as Symmetrical Uncertainty Attributes, Gain Ratio Attributes, and Correlation Attributes, were adopted to choose the top twenty-one features out of the total of 30 features.

4. Proposed Models and Metrics

In the current research, the investigator proposed an array of machine learning models, most notably, a Decision Tree, Ada-boost classifier, Random Forest, and Gradient Boosted Classifier. The assessment and evaluation were performed using Jupyter Notebook, an engaging algorithm development forum for Python, an open-source programming language predominantly adopted in machine learning algorithm development (Hasan et al. 2024). Besides, R-programming was also utilized in building the models.

4.1 Random Forest
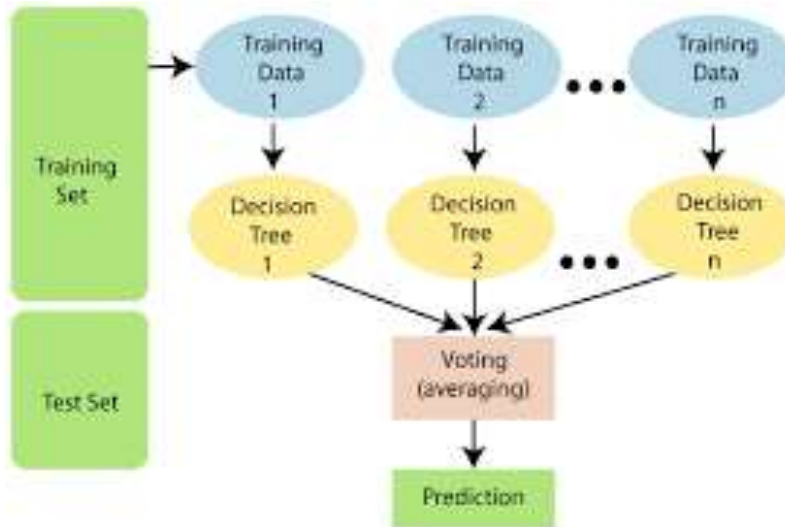
Random Forest is a form of ensemble classifier that incorporates multiple algorithms to categorize items by making a collective decision. For example, it can consolidate algorithms such as Decision Tree, Support Vector Machines (SVM), and Naive Bayes. After consolidating these systems, it makes the final votes from each of the algorithms to develop the model's classification. In the Random Forest classifier, a category of decision trees is initially established utilizing a sub-category of the training data (Hasan et al. 2024). Afterward, the votes from the different trees are combined to make the ultimate result of the algorithm's class. The performance of the Random forest is demonstrated in Figure 1 below, portraying the evaluation of its effectiveness.

4.2 Gradient Boosting
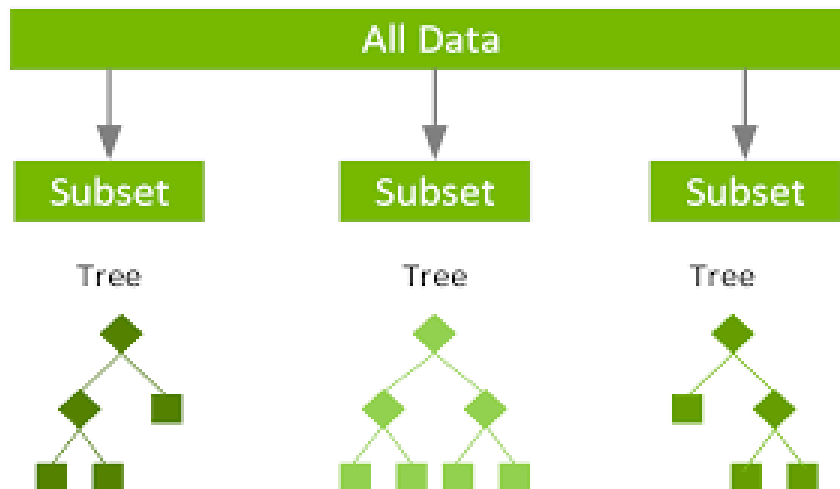
Gradient boosting is a dynamic algorithm employed for both classification and regression tasks, facilitating the development of prediction algorithms. What distinguishes Gradient Boosting from the rest is its scalability and remarkable speed during deployment. In the setting of classification, gradient boosting functions sequentially, with every novel predictor learning from the errors made by preceding predictors. This repeated learning process enables swift classification and faster production of outcomes because of the algorithm's capability to learn from previous mistakes (Jain & Nayyar, 2018). The performance measure of the Gradient Boosting classifier is demonstrated in Figure 3 below, illustrating the evaluation of its effectiveness.

4.3 Ada-Boost Algorithm
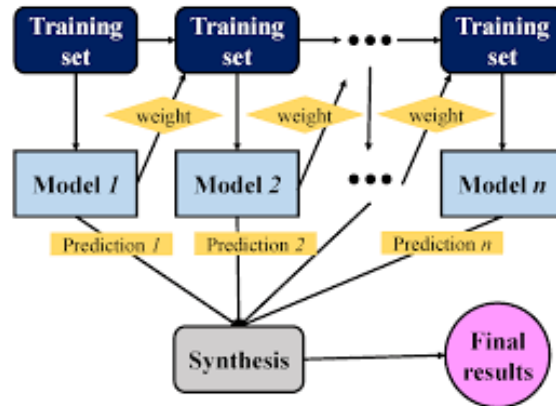
Hasan et al. (2024), indicate that the AdaBoost algorithm is a classifier that integrates weak classifier models to establish a solid classifier. Independently, a single model may result in poor classification performance. Nevertheless, by incorporating multiple classifiers and cautiously choosing the training set at every experiment, along with strategically designating weights during the last voting process, it becomes possible to attain a high-level accuracy score for the final classifier. This classifier employs the collaborative strength of multiple classifiers to elevate the effectiveness and accuracy of the classification process.

4.4 Decision Tree Algorithm
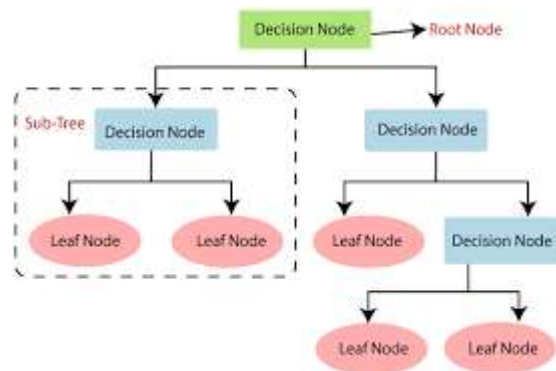
Decision Trees are prominent classifier models due to their simplicity in implementation and interpretation. The model develops a tree architecture from the training dataset, where every leaf node stands for a feature, and the branches represent the respective feature values (Hasan et al. 2024). This hierarchical portrayal enables straightforward visualization and representation of the decision-making process within the algorithm.

4.5 Importing Libraries
Initially, the analyst loaded the Human Resource Employee Attrition dataset into Jupyter Notebook which is an IDE of Python on which the analyst performed all the analysis, the figure below displays the loading of the dataset and dataset attributes.

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv("HR-Employee-Attrition.csv")
data
```

Output:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Employ
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	
...	...	...	...	...	...	...	...	...	...	
1465	36	No	Travel_Frequently	884	Research & Development	23	2	Medical	1	
1466	39	No	Travel_Rarely	613	Research & Development	6	1	Medical	1	
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sciences	1	
1468	49	No	Travel_Frequently	1023	Sales	2	3	Medical	1	
1469	34	No	Travel_Rarely	628	Research & Development	8	3	Medical	1	

After the dataset was loaded, the researcher progressed with data preprocessing steps, which entailed eliminating records that comprised null values and converting the data to satisfy the protocol of the model being developed.

```
In [8]:  data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Age                       1470 non-null   int64
 1   Attrition                 1470 non-null   object
 2   BusinessTravel            1470 non-null   object
 3   DailyRate                 1470 non-null   int64
 4   Department                1470 non-null   object
 5   DistanceFromHome          1470 non-null   int64
 6   Education                 1470 non-null   int64
 7   EducationField            1470 non-null   object
 8   EmployeeCount             1470 non-null   int64
 9   EmployeeNumber            1470 non-null   int64
 10  EnvironmentSatisfaction   1470 non-null   int64
 11  Gender                    1470 non-null   object
 12  HourlyRate                1470 non-null   int64
 13  JobInvolvement            1470 non-null   int64
 14  JobLevel                  1470 non-null   int64
 15  JobRole                   1470 non-null   object
 16  JobSatisfaction           1470 non-null   int64
 17  MaritalStatus             1470 non-null   object
 18  MonthlyIncome             1470 non-null   int64
 19  MonthlyRate               1470 non-null   int64
 20  NumCompaniesWorked        1470 non-null   int64
 21  Over18                    1470 non-null   object
 22  OverTime                  1470 non-null   object
 23  PercentSalaryHike         1470 non-null   int64
 24  PerformanceRating         1470 non-null   int64
 25  RelationshipSatisfaction  1470 non-null   int64
 26  StandardHours             1470 non-null   int64
 27  StockOptionLevel          1470 non-null   int64
 28  TotalWorkingYears         1470 non-null   int64
 29  TrainingTimesLastYear     1470 non-null   int64
 30  WorkLifeBalance           1470 non-null   int64
 31  YearsAtCompany            1470 non-null   int64
 32  YearsInCurrentRole        1470 non-null   int64
 33  YearsSinceLastPromotion   1470 non-null   int64
 34  YearsWithCurrManager      1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
```

After the preprocessing steps, the investigator explored the dataset, particularly exploring the correlations between different variables.

```python
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report
import xgboost as xgb
```

```python
# Encode categorical variables
label_encoder = LabelEncoder()
categorical_cols = ['Attrition', 'BusinessTravel', 'Department', 'EducationField', 'Gender', 'JobRol
for col in categorical_cols:
    data[col] = label_encoder.fit_transform(data[col])
```

```python
# Split features and target variable
X = data.drop('Attrition', axis=1)
y = data['Attrition']
```

```python
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

To ascertain the age distribution in the dataset, the analyst applied a code snippet to generate a suitable histogram respectively, the results can be showcased as follows:
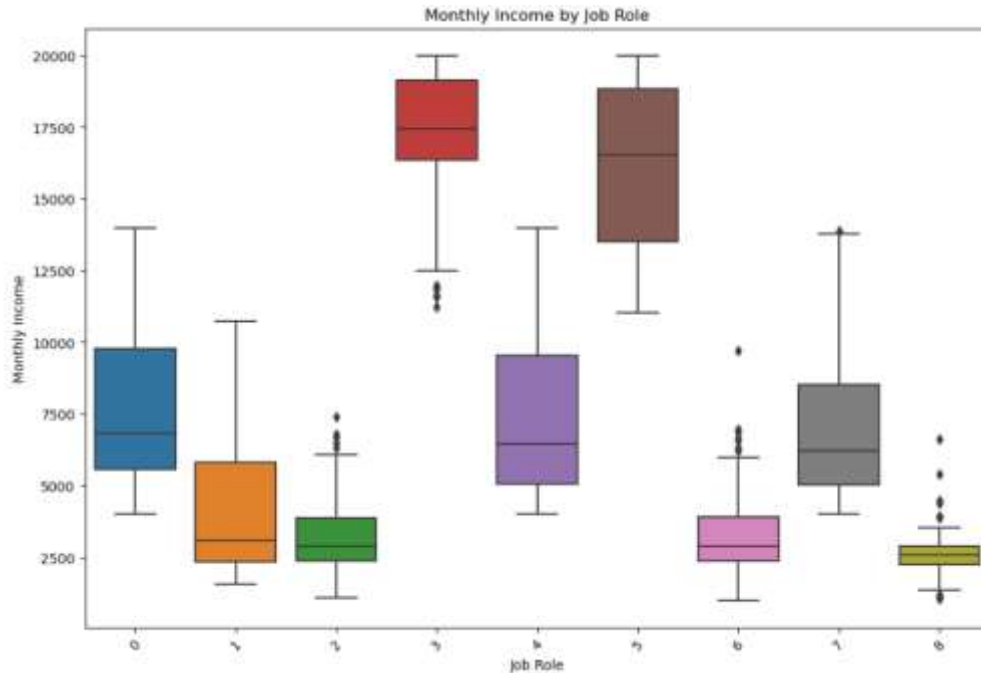
```python
# EDA
# 1. Distribution of Age
plt.figure(figsize=(10, 6))
sns.histplot(data['Age'], bins=30, kde=True, color='skyblue')
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()
```

Outcomes:

Distribution of Age
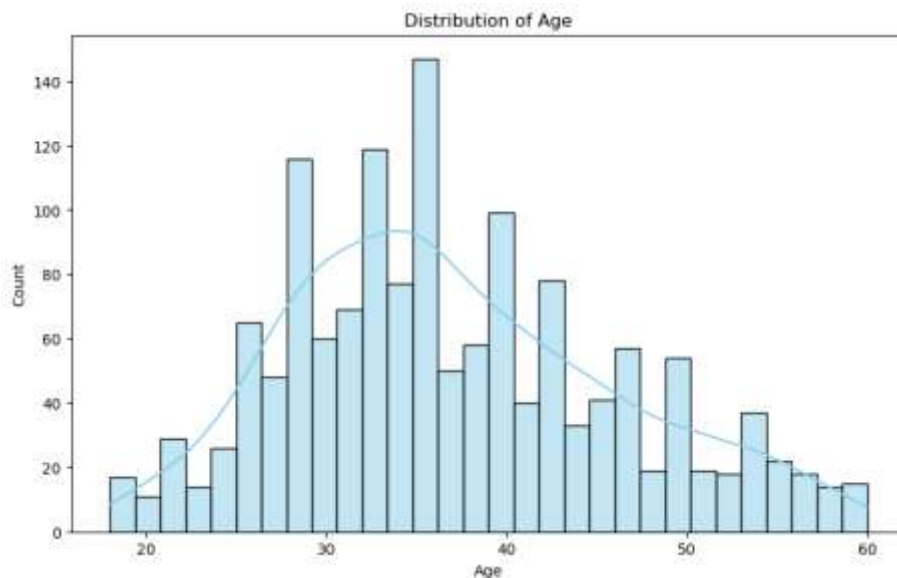
To determine the correlation between job role and monthly income, an ideal code snippet was imposed to generate a scatter plot as showcased below:

Output:

To establish the correlation between distance from home and the years at the organization, the analyst equally employed a suitable code snippet to generate an appropriate scatter plot which can be exhibited below:

```
# 10. Distance from Home vs. Years at Company
plt.figure(figsize=(10, 8))
sns.scatterplot(x='YearsAtCompany', y='DistanceFromHome', data=data, hue='Attrition', palette='Set1'
plt.title('Distance from Home vs. Years at Company')
plt.xlabel('Years at Company')
plt.ylabel('Distance from Home')
plt.legend(title='Attrition', loc='upper right')
plt.show()
```

Output:

```
# 3. Monthly Income vs. Job Role
plt.figure(figsize=(12, 8))
sns.boxplot(x='JobRole', y='MonthlyIncome', data=data)
plt.title('Monthly Income by Job Role')
plt.xlabel('Job Role')
plt.ylabel('Monthly Income')
plt.xticks(rotation=45)
plt.show()
```

Distance from Home vs. Years at Company
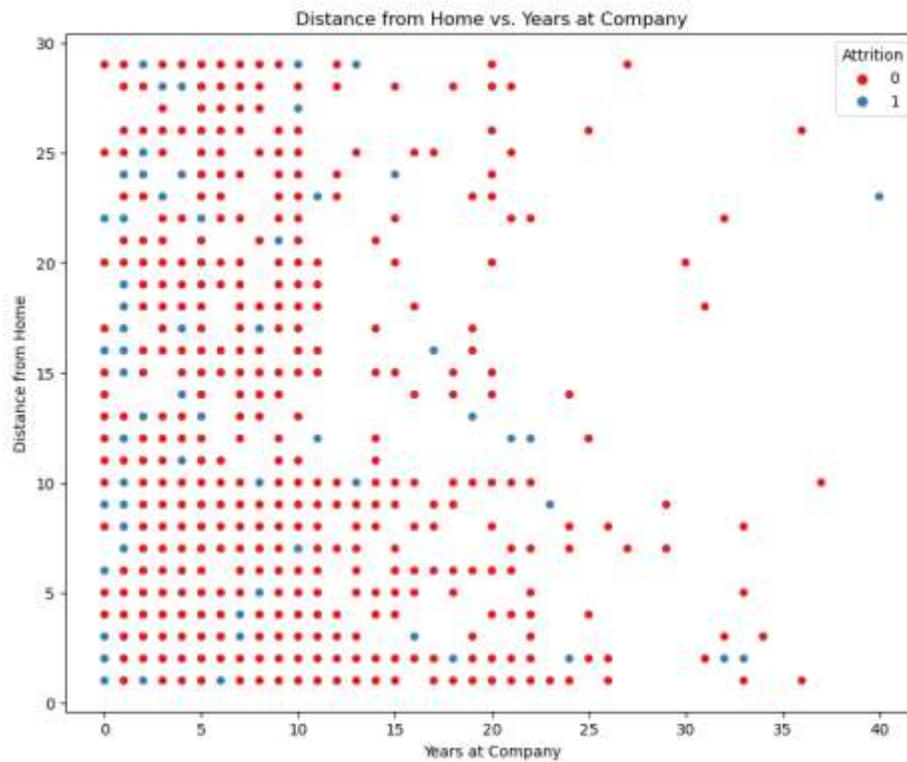
To substantiate the association between monthly income by attrition and department, the following code snippet was applied to generate the appropriate scatter plot:

```
# 9. Monthly Income Distribution by Attrition and Department
plt.figure(figsize=(12, 8))
sns.boxplot(x='Department', y='MonthlyIncome', hue='Attrition', data=data, palette='viridis')
plt.title('Monthly Income Distribution by Attrition and Department')
plt.xlabel('Department')
plt.ylabel('Monthly Income')
plt.legend(title='Attrition', loc='upper right')
plt.show()
```

Output:
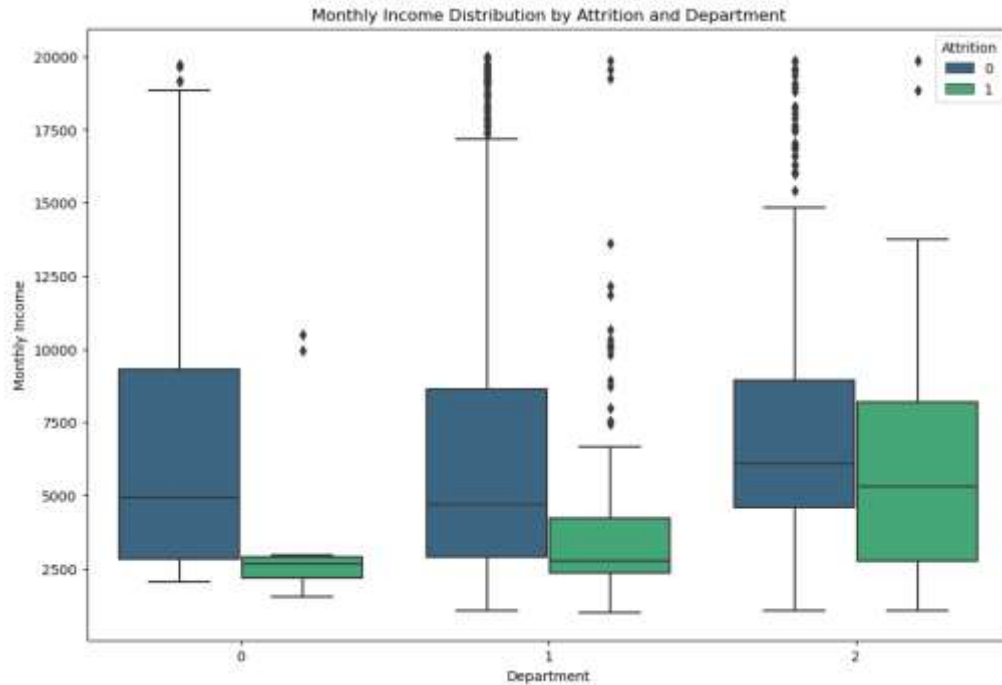
To ascertain the connection between education level and job satisfaction, the following code snippet was imposed:

```
# 7. Education Level and Job Satisfaction
plt.figure(figsize=(8, 6))
sns.boxplot(x='Education', y='JobSatisfaction', data=data, palette='Set3')
plt.title('Education Level and Job Satisfaction')
plt.xlabel('Education Level')
plt.ylabel('Job Satisfaction')
plt.show()
```

Output:
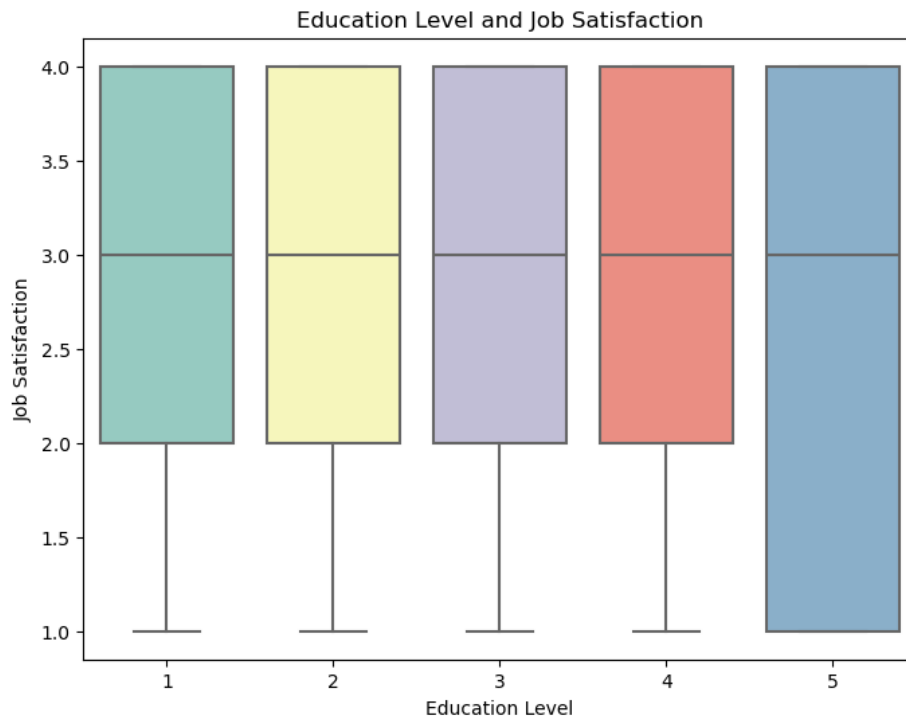
5. Models Performance Evaluation

Confusion matrix was applied to assess the performance of the models in terms of accuracy, precision, and recall.

Actual Class		Predicted Class	
		Class=No	Class=Yes
	Class=No	False Positive	True Negative
	Class=Yes	True Positive	False Negative

Where

Accuracy = TP+TN/TP+FP+FN+TN

Precision = TP/TP+FP

Recall = TP/TP+FN

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

5.1 Algorithm Performance Evaluation

Models	Precision	Accuracy	F1 Score	Recall
Random forest	83.33%	86.05%	32.78%	20.40%
Decision Tree	52.38%	83.67%	31.42%	22.44%
Gradient Boosting	62.96%	85.71%	44.73%	34.69%
AdaBoost	56.52%	84.35%	36.11%	26.53%

	Decision Tree	Random Forest	AdaBoost	Gradient Boosting
Accuracy	83.67	86.05	84.35	85.71
Precision	52.38	83.33	56.52	62.96
Recall	22.44	20.4	26.53	34.69
F1 Score	31.42	32.78	36.11	44.73
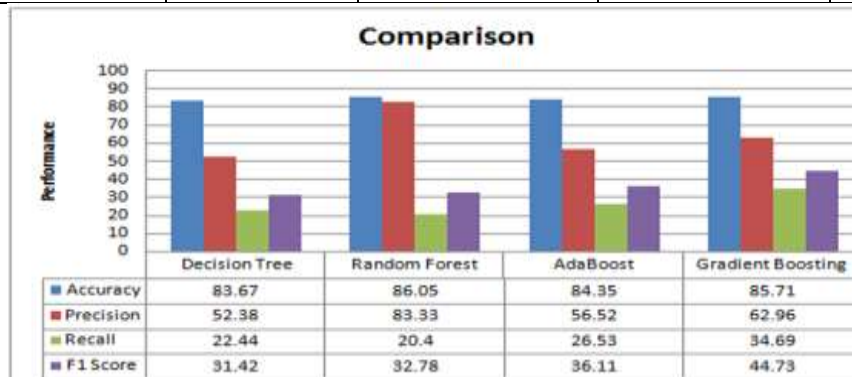
By referring to the table above, it was evident that the Random Forest algorithm had the highest accuracy (86.05%), followed by Gradient Boosting (85.71%) and Decision Tree (83.67%) respectively. AdaBoost has the lowest accuracy (84.35%). Concerning precision, the Random Forest algorithm again had the highest precision (83.33%), followed by Gradient Boosting (62.96%) and AdaBoost (56.52%) accordingly. The Decision Tree algorithm had a relatively lower precision (52.38%).

5.2 Business Impact
5.2.1 Benefits of Implementing the Models on the Organizations in the USA
Diminishing turnover costs: By deploying the proposed models' organizations in the USA can identify high-performing employees at risk of quitting, and subsequently take proactive steps to retain them, saving significant organizational resources.

Enhanced Talent Management: Implementing the proposed models can assist businesses in the USA understand the root causes of employee attrition. Organizations can target retention efforts toward specific areas by evaluating aspects such as compensation, workload, and performance rating data, creating a more engaging work environment.

Improved Employer Brand: The recommended models can assist organizations in proactively resolving workers' issues and promoting a positive work atmosphere. As a result, the organization can enhance its employer brand. Consequently, this attracts top talent and minimizes the need for expensive and time-consuming recruitment efforts.

5.2.2 Benefits to the USA Economy

Combating Labor Shortages: The US economy is presently confronting a tight labor market. As a result, Machine learning can assist the government in maintaining high-performing employees, reducing the impact of labor shortages, and maintaining business continuity.

Elevated productivity: The proposed can automate repetitive tasks across different sectors, liberating human workers to concentrate on higher-value activities. As a result, this leads to improved efficiency and overall productivity in the American economy.

Competitive Advantage: By employing the proposed models, consequently the government can affirm a significant competitive advantage, by retaining valuable and high-performing employees. As a result, this can lead to increased exports and a better overall position for the US economy in the global market.

Enhanced public service: The recommended algorithms can be employed to optimize government operations, predict attrition patterns and customize social services. This can result in a more equitable allocation of resources and better outcomes for US citizens.

5.3 How to Implement the Model

Step 1: Define Business Problem and Data Needs-Business in the USA should first consider the specific business challenges confronting their respective organization and determine the data required to solve those challenges.

Step 2: Select the Right Model- The random Forest algorithm might be suitable for general classification with low error margins, while the Gradient Boosting algorithm could be appropriate for detecting positive case identification.

Step 3: Data Pre-processing and Preparation- Subsequently, the business analyst should resolve the outliers, inconsistencies, and missing values. Besides, the analyst should standardize data formats for algorithm compatibility.

Step 4: Model Training & Evaluation-Afterwards the business analyst should split the data into training, validation, and testing. As a result, the analyst should utilize the testing data to evaluate metrics such as precision, accuracy, and recall.

Step 5: Model Implementation and Monitoring: The business should track the metrics constantly and update the model with new data to avoid degradation.

6. Conclusion

The focal aim of this research paper was to delve into the application of machine learning in terms of forecasting employee attrition and its ramifications for HR analytics and talent retention strategies. In this study, the investigator used Jupyter Notebook, an interactive platform for Python users, to design machine learning algorithms. The dataset utilized in this research was attained from the IBM Human Resource workforce attrition survey dataset. In the current research, the investigator proposed an array of machine learning models, most notably, a Decision Tree, Ada-boost classifier, Random Forest, and Gradient Boosted Classifier. By referring to the performance evaluation, it was evident that the Random Forest algorithm had the highest accuracy, followed by Gradient Boosting and Decision Tree respectively. AdaBoost had the lowest accuracy. Concerning precision, the Random Forest algorithm again had the highest precision followed by Gradient Boosting and AdaBoost accordingly. By implementing the proposed models' organizations in the USA can identify high-performing employees at risk of quitting, and subsequently take proactive steps to retain them, saving significant organizational resources. Ultimately, the proposed machine learning techniques can assist the government in maintaining high-performing employees, reducing the impact of labor shortages, and maintaining business continuity.

Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Ajit, P. (2019). Prediction of employee turnover in organizations using machine learning algorithms. *algorithms, 4*(5), C5.

[2] Ahmad, M., Ali, M. A., Hasan, M. R., Mobo, F. D., & Rai, S. I. (2024). Geospatial Machine Learning and the Power of Python Programming: Libraries, Tools, Applications, and Plugins. In *Ethics, Machine Learning, and Python in Geospatial Analysis* (pp. 223-253). IGI Global.

[3] Gurung, N., Gazi, M. S., & Islam, M. Z. (2024). Strategic Employee Performance Analysis in the USA: Deploying Machine Learning Algorithms Intelligently. *Journal of Business and Management Studies*, *6*(3), 01-14.

[4] IJRASET (2021). Prediction of employee attrition using machine learning approach. *www.academia.edu*. https://www.academia.edu/52266979/Prediction_of_Employee_Attrition_Using_Machine_Learning_Approach?sm=b

[5] Jain, R., & Nayyar, A. (2018, November). Predicting employee attrition using the xgboost machine learning approach. In *2018 international conference on system modeling & advancement in research, trends (smart)* (pp. 113-120). IEEE.

[6] Marvin, G., & Jackson, M. (2021). A machine learning approach for employee retention prediction. *Mak*. https://www.academia.edu/56415761/A_Machine_Learning_Approach_for_Employee_Retention_Prediction?sm=b

[7] Musanga, V. (2023). A supervised machine learning model to optimize human resources analytics for employee churn prediction. *www.academia.edu*. https://www.academia.edu/99465169/A_Supervised_Machine_Learning_Model_to_Optimize_Human_Resources_Analytics_for_Employee_Churn_Prediction?sm=b

[8] Naik, P. (2023). Machine Learning approach for employee attrition analysis. *www.academia.edu*. https://www.academia.edu/75198011/Machine_Learning_Approach_for_Employee_Attrition_Analysis?sm=b

[9] Pro-AI-Rokibul. (2024). *Employee-Attrition-Prediction/Model/main.ipynb at main · proAIrokibul/Employee-Attrition-Prediction*. GitHub. https://github.com/proAIrokibul/Employee-Attrition-Prediction/blob/main/Model/main.ipynb

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

[11] Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. M. S. (2022). Predicting employee attrition using machine learning approaches. *Applied Sciences*, *12*(13), 6424.

[12] Sani, N. S. (2023). Machine learning for predicting employee attrition. *www.academia.edu*. https://www.academia.edu/98899513/Machine_Learning_for_Predicting_Employee_Attrition?sm=b

[13] Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., & Zhu, X. (2019). Employee turnover prediction with machine learning: A reliable approach. In *Intelligent Systems and Applications: Proceedings of the 2018 Intelligent Systems Conference (IntelliSys) Volume 2* (737-758). Springer International Publishing.