A FIELD PROJECT REPORT

on

"Predicting Employee Attrition with Deep Learning and Ensemble Techniques: A Comprehensive Study"

Submitted

by

221FA04099

Mellachervu Chandana

221FA04142

Kakumanu Pavan Sai

221FA04105

Sanikommu Renuka

221FA04701

Shatakshi Bajpai

Under the guidance of

M Bhargavi

Assistant Professor, Department of CSE, VFSTR.

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH

Vadlamudi, Guntur.

ANDHRA PRADESH, INDIA, PIN-522213.

VIGNAN'S
FOUNDATION FOR SCIENCE, TECHNOLOGY & RESEARCH
(Deemed to be University) - Estd. u/s 3 of UGC Act 1956

CERTIFICATE

This is to certify that the Field Project entitled **"Predicting Employee Attrition with Deep Learning and Ensemble Techniques: A Comprehensive Study"** that is being submitted by 221FA04099 (Chandana), 221FA04142 (Pavan Sai) ,221FA04105 (Renuka) and 221FA04701(Shatakshi) for partial fulfilment of Field Project is a bonafide work carried out under the supervision of M Bhargavi, Assistant Professor, Department of CSE, VFSTR.

M.Bhargavi	Dr. S. V. Phani Kumar	Dr.K.V. Krishna Kishore
Assistant Professor, CSE	HOD,CSE	Dean, SoCI

VIGNAN'S
FOUNDATION FOR SCIENCE, TECHNOLOGY & RESEARCH
(Deemed to be University) - Estd. u/s 3 of UGC Act 1956

DECLARATION

We hereby declare that the Field Project entitled **“Predicting Employee Attrition with Deep Learning and Ensemble Techniques: A Comprehensive Study”** is being submitted by 221FA04099 (Chandana), 221FA04142 (Pavan Sai) ,221FA04105 (Renuka) and 221FA04701(Shatakshi) in partial fulfilment of Field Project course work. This is our original work, and this project has not formed the basis for the award of any degree. We have worked under the supervision of M Bhargavi, Associate Professor, Department of CSE, VFSTR.

By
221FA04099 (Chandana),
221FA04142 (Pavan Sai),
221FA04105 (Renuka),
221FA04701(Shatakshi).

Date:

ABSTRACT

Employee attrition is a substantial problem for many organizations, which leads to disruption and high costs. The capability of understand- ing the reason behind employee departures is necessary for creating an optimistic work environment and enhancing recruitment strategies.This work aims to undertook a comprehensive analysis for predicting employee attrition using a wide range of machine learning models, from standard to advanced levels, which comprise both deep learning and ensemble models. To analyze the features of employee attrition, we initiated our research by evaluating conventional machine learning models, followed by deep learning models, and ultimately exploring ensemble models. Our study revealed Random Forest as the highest achiever with an accuracy of 98.3%, followed by Gradient Boost and XGBoost with an accuracy of 97.9%.

Keywords: Machine learning, Deep learning, Ensemble models, Recruitment strategies, Random Forest, Gradient Boost, XGBoost

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER-1

INTRODUCTION

1. INTRODUCTION

1.1What is Employee Attrition?

Employee attrition refers to the gradual reduction of a workforce due to employees leaving an organization for various reasons, including voluntary resignations, retirements, or terminations. This phenomenon can significantly impact organizational dynamics, leading to operational disruptions, increased recruitment costs, and the potential loss of valuable institutional knowledge and experience. Understanding the reasons behind employee exits is crucial for organizations aiming to maintain a stable and productive workforce.

1.2 The Importance of Predicting Attrition

Predicting employee attrition is essential for several reasons. Firstly, high attrition rates can lead to increased costs associated with recruiting and training new employees. Secondly, understanding the factors contributing to attrition enables organizations to implement effective retention strategies, fostering a more engaged and committed workforce. By identifying potential turnover risks early, companies can proactively address underlying issues, ultimately enhancing employee satisfaction and organizational performance. In a competitive labor market, the ability to predict and mitigate attrition becomes a vital component of a successful human resource strategy.

1.3 Role of Machine Learning in Predicting Attrition

Machine learning plays a pivotal role in enhancing the accuracy and effectiveness of attrition prediction. Traditional human resource practices often struggle to identify complex patterns and relationships within employee data. However, machine learning algorithms can analyze large datasets, uncovering hidden trends that may not be immediately apparent. By leveraging a range of techniques—from traditional algorithms like K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) to advanced methods such as deep learning and ensemble techniques—organizations can develop robust models that predict attrition with greater precision. These insights empower HR departments to make data-driven decisions, thereby equipping organizations to maintain an agile and resilient workforce in an ever-evolving business landscape.

CHAPTER-2

LITERATURE SURVEY

2. LITERATURE SURVEY

2.1 Literature review

We have carried out a literature survey to include all related works with our study on Employee Attrition. The crux of ideas from these papers has been summed up below:

N. Bhartiya et al.
The models like Decision Trees, SVM and Random Forest are developed more accurate and efficient Attrition predicting tools, leading to earlier interventions and improved outcomes for individuals into consideration where maximum accuracy has reached as high as 85.6% (Random Forest).
Limitation: Limited generalizability due to small dataset.

No	Author(s)	Model/Approach	Accuracy/Results	Limitation
1	N. Bhartiya et al.	Decision Trees, SVM, Random Forest	85.6% (Random Forest)	Limited generalizability due to small dataset
2	S. Gupta et al.	Logistic Regression, XGBoost	87.5% (XGBoost)	Lack of interpretability in complex models
3	Doohee Chung et al.	Stacking Ensemble Learning	89.8%	High computational cost
4	P. K. Jain et al.	Random Forest, Gradient Boosting	86% (Random Forest)	Overfitting on small datasets
5	Yedida et al.	Neural Networks	80%	High training time
6	A. Raza et al.	Random Forest, Gradient Boosting	88.1% (Gradient Boosting)	Lack of deep learning comparison
7	N. Mansor et al.	Logistic Regression, KNN	75.4% (KNN)	Low accuracy on imbalanced data
8	Raj M. et al.	Random Forest, Decision Trees	84.3%	Limited to IT sector data
9	M. Lazzari et al.	Logistic Regression, SVM	82%	Does not account for time-based attrition factors
10	M. Nandal et al.	Decision Trees, XGBoost	87% (XGBoost)	Limited feature engineering
11	M. Subhashini et al.	Logistic Regression, SVM	80% (SVM)	Poor performance on large datasets

12	Mohammad R. Shafie et al.	Artificial Neural Networks (ANN), Data Augmentation	91%	Requires high computational resources
13	L. Zhang et al.	Decision Trees, Random Forest	85.7% (Random Forest)	Limited feature diversity
14	S. Singh et al.	Hybrid Approach (SVM + Random Forest)	88%	Limited scalability for large data
15	A. S. Mohammed et al.	SVM, KNN, Decision Trees	83% (SVM)	High sensitivity to parameter tuning
16	H. Nguyen et al.	Gradient Boosting, Deep Learning	92% (Deep Learning)	Requires substantial data preprocessing
17	S. K. Singh	Ensemble Learning	90%	Complex models are less interpretable
18	K. P. Jadhav et al.	Logistic Regression, Random Forest	86.5% (Random Forest)	Model interpretability issues
19	M. A. Sharif et al.	Deep Learning	88%	High training time
20	B. S. Ayyub	Logistic Regression, SVM	82% (SVM)	Low performance on imbalanced data
21	Y. Han and D. Wu	Convolutional Neural Networks (CNN)	89.5%	Data preparation and computation-heavy
22	P. Saxena and R. Bhargava	Neural Networks, Time Series Analysis	90% (Neural Networks)	Limited by time series anomalies
23	A. Sarkar and A. K. Saha	Deep Learning Framework	87.2%	Requires large training data
24	S. Z. Qureshi	Predictive Analytics using Random Forest	85.4%	Limited applicability beyond specific industry
25	L. Liu and Y. Zhang	Decision Trees, Random Forest	86% (Random Forest)	Limited to banking sector data

2.2 Motivation

Attrition of employees is a serious concern for any organization in view of the considerable costs incurred through recruiting, training, and lowered productivity due to loss of employees. It would be better if a company could predict the moment when there could be an attrition of employees so that the factors leading to it could be addressed in advance, thereby enhancing the retention plans and the happiness of workers at the workplace. Big data has become a catalyst for machine learning techniques that analyze patterns in the behavior of employees and the dynamics of organizations with better accuracy. All these techniques will be consolidated to build accurate attrition-prediction models, which can help human resources operate on data-driven decision support. Turnover can be minimized so that it may promote long-term growth and employee engagement in a business.

CHAPTER-3

PROPOSED SYSTEM

3. PROPOSED SYSTEM

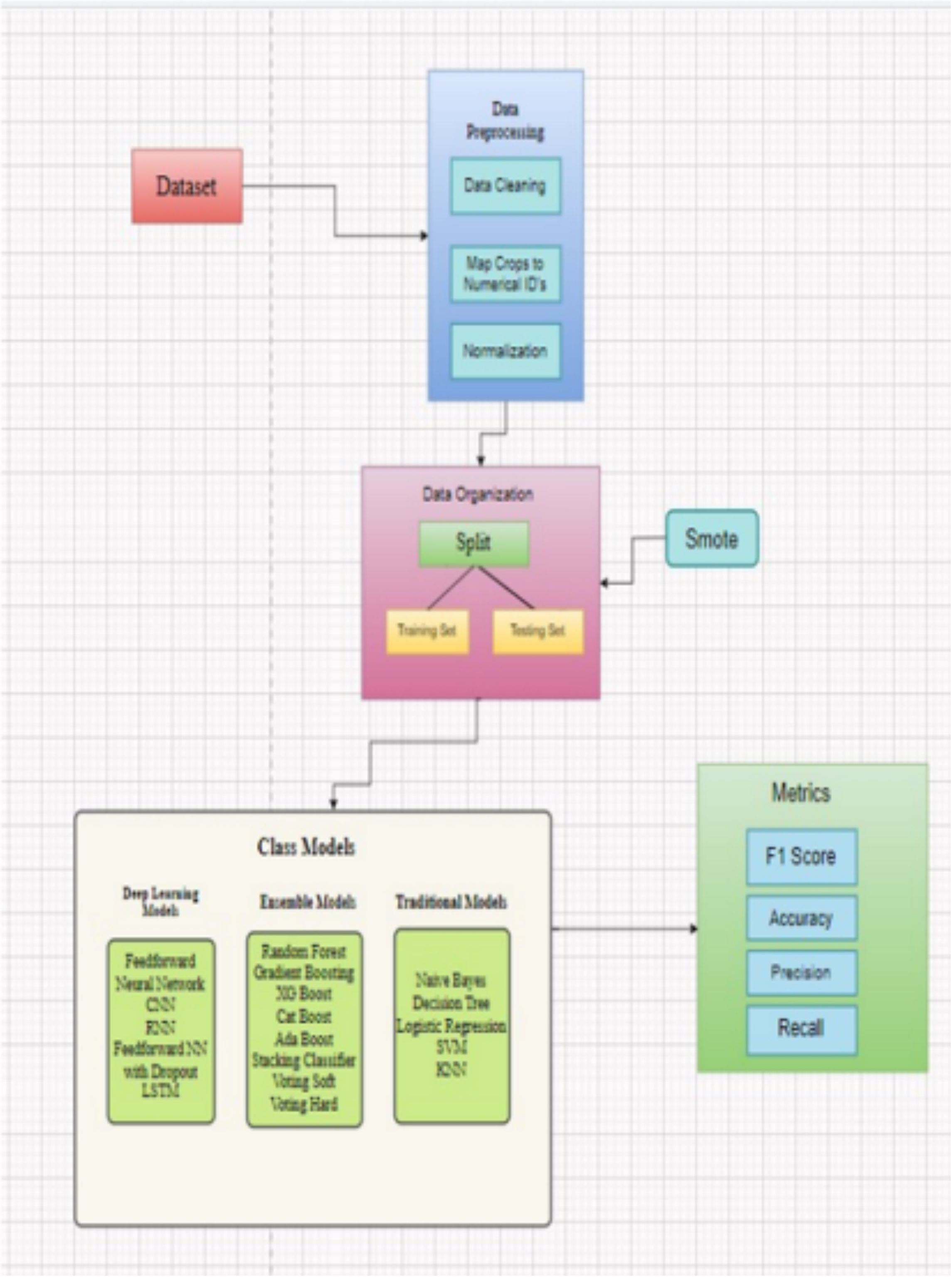
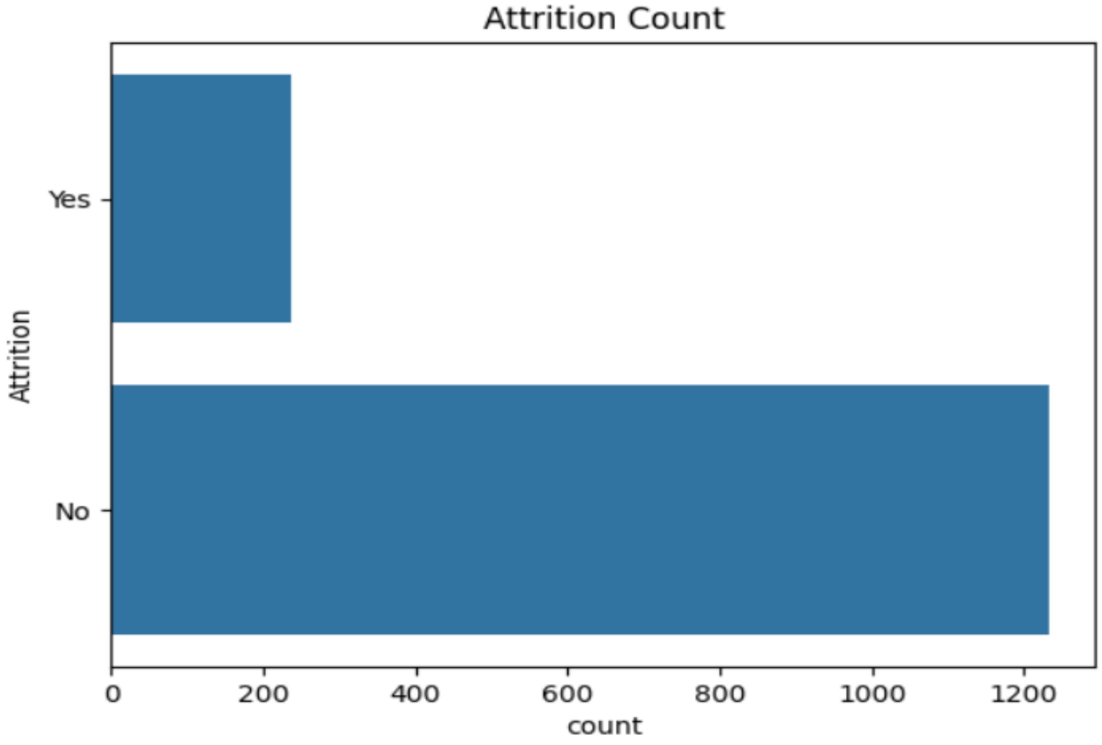


Figure 1: Proposed Architecture

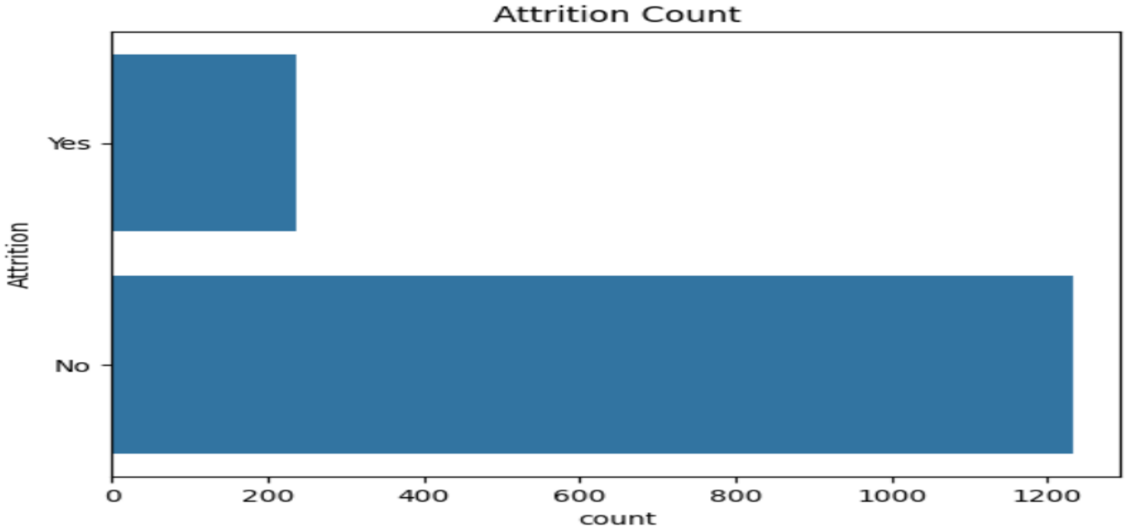
3.1 Input dataset

The dataset utilized is HR Analytics of IBM in Kaggle. This dataset comprises demographics, job attributes, and performance measures. Key attributes in the dataset include Age, Attrition, Job Satisfaction, and Monthly Income.

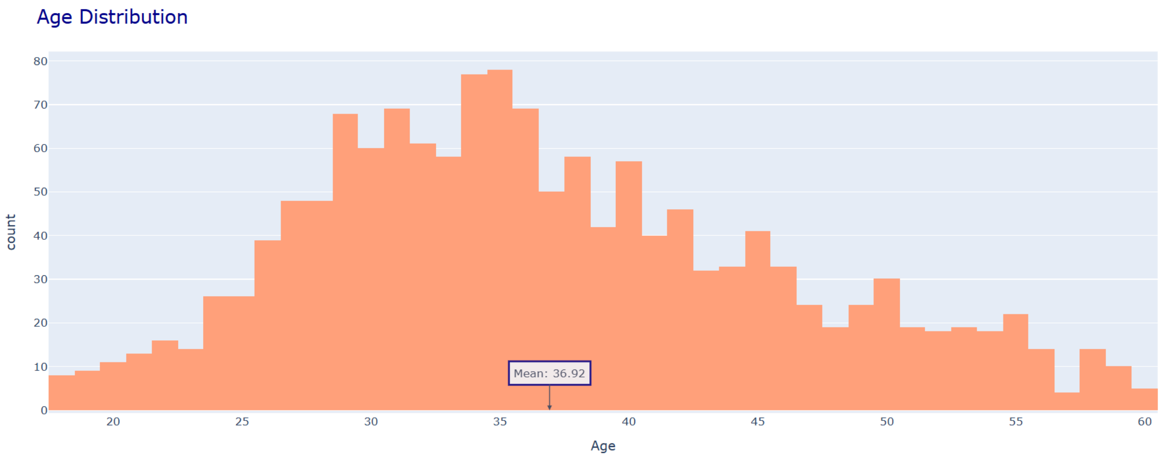
3.1.1 Detailed Features of the Dataset



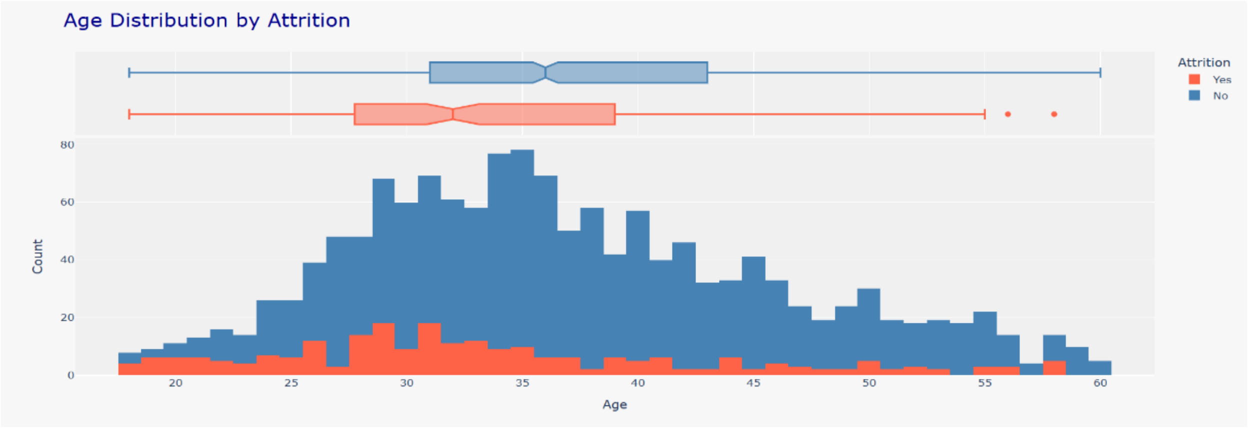
a. Count Plot



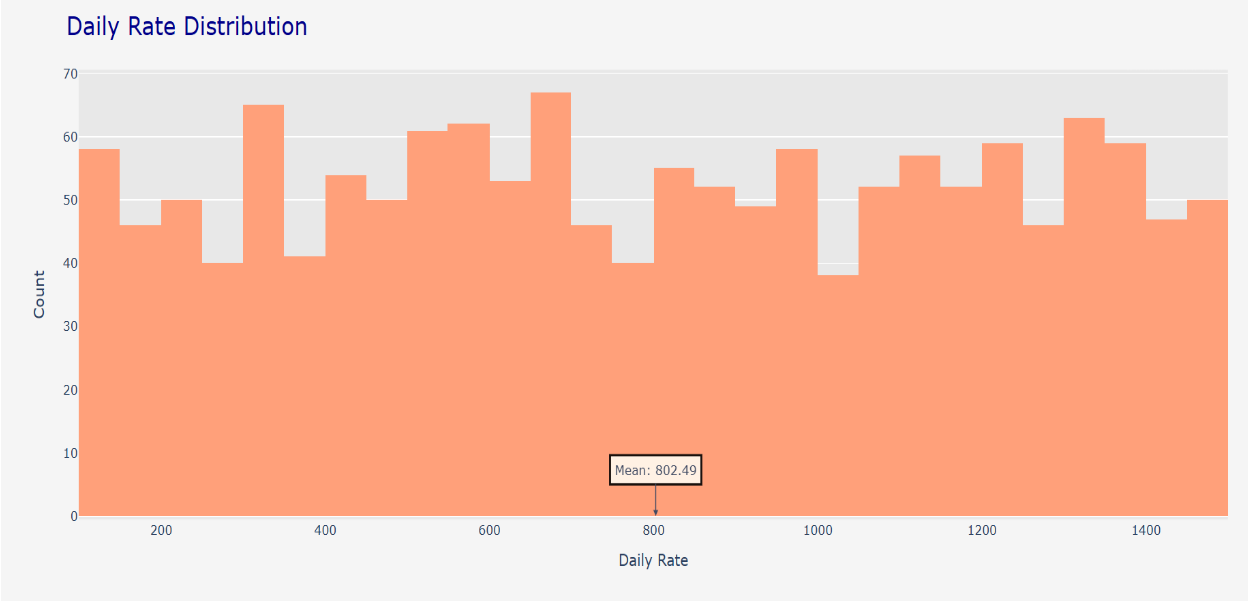
b. Attrition Count



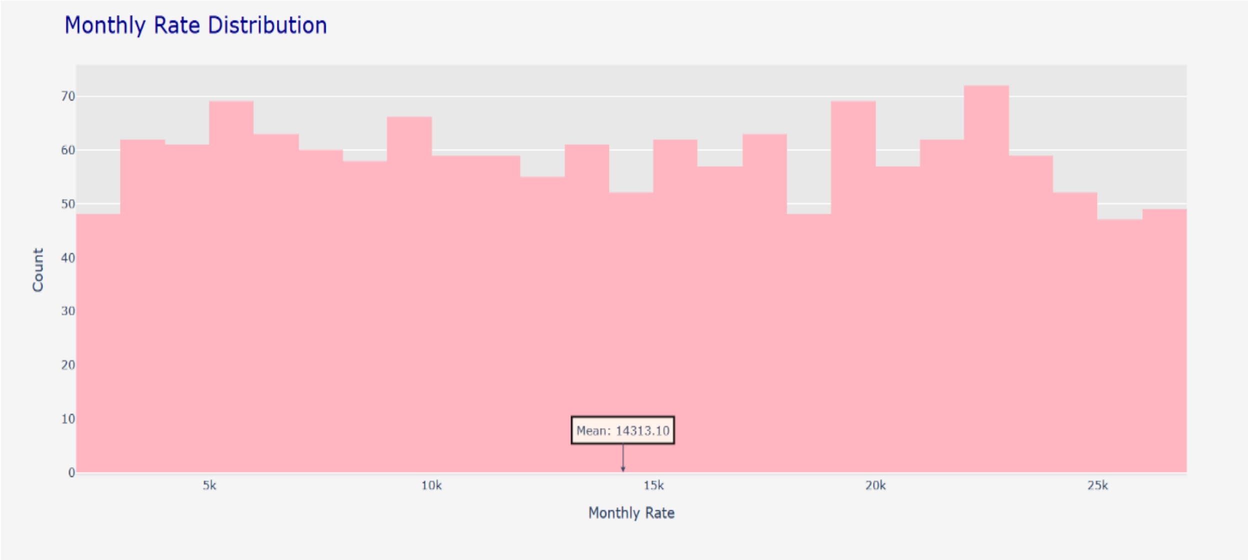
c. **Age Distribution**



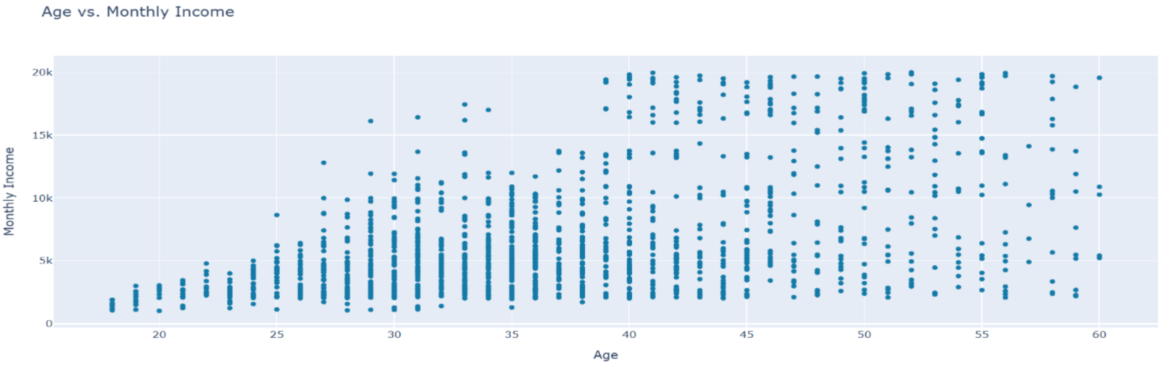
d. **Age Distribution by Attrition**



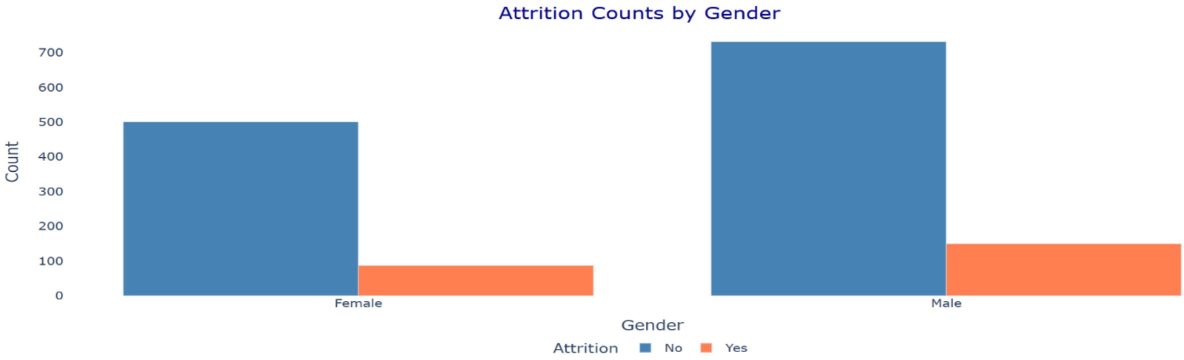
e. **Daily Rate Distribution**



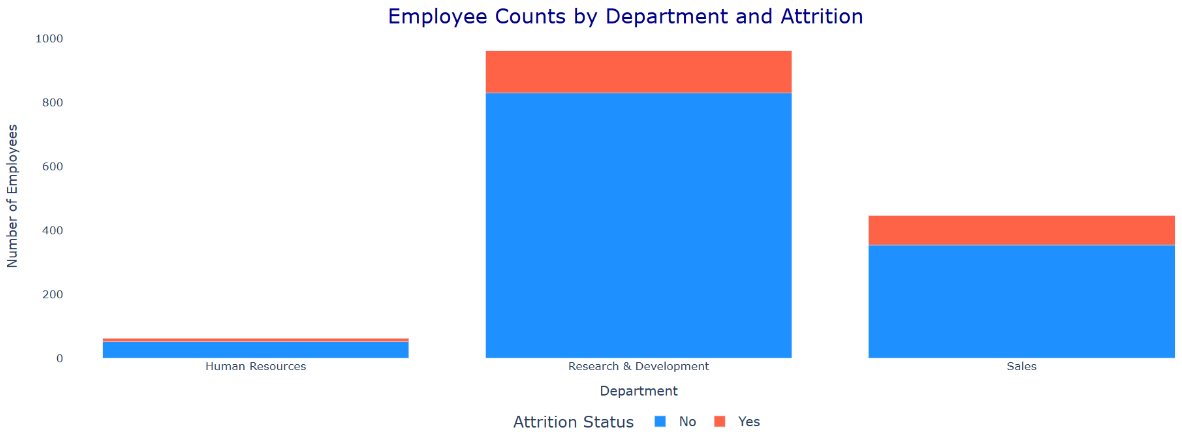
f. **Monthly Rate Distribution**



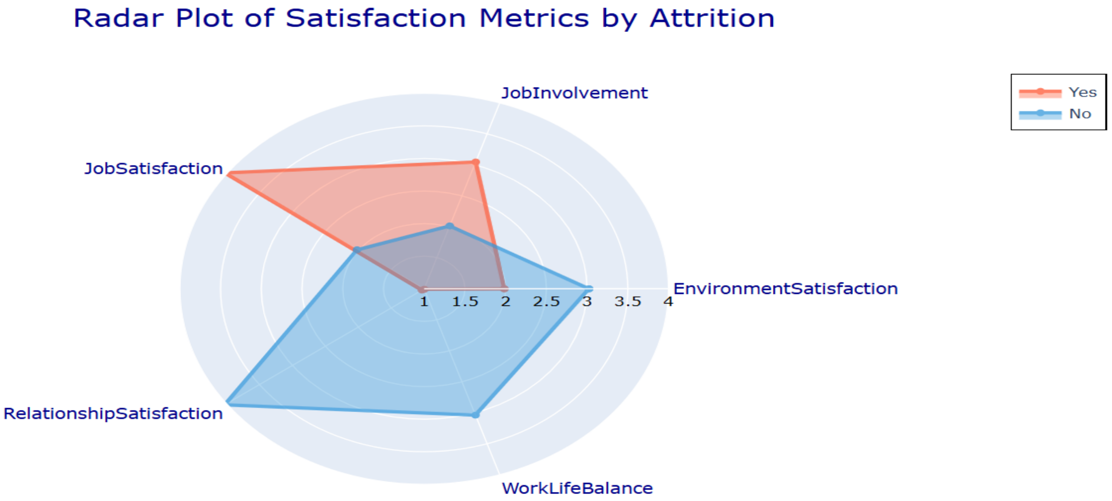
g. **Age vs. Monthly Income**



h. **Attrition Counts by Gender**



i. Employee Counts by Department and Attrition



j. Radar Plot of Satisfaction Metrics by Attrition

EDA is performed for exploring issues with attrition of employees and factors that influence it.

(a) Count Plot : Displays the count of the employee by their status on attrition. Since most employees stayed in instead of leaving, it might indicate class imbalance and 4 this would be a challenge to modeling.

(b) Age: Histograms of Attrition Status Age distributions for those leaving with those staying, so it would seem that the younger employee is more likely to leave, suggesting different expectations in the Histogram

(c) Daily Rate Distributions :This would give an idea about the variation in salaries for the employees. It would show that low daily rates may have a good deal with high attrition and, therefore, competitive pay policies might be required.

(d) Monthly Rate Distributions: As with a daily rate, so too does it indicate that the lower rate of salary in monthly may be associated with higher attrition rates, and fair compensation becomes all the more crucial.

(e) Scatter Plot between Age and Monthly Income :Scatter plot will plot an increasing trend of income with age, that is, older-aged workers tend to stay longer perhaps because they have more experience, thus making career development opportunity vital for them.

(f) Bar Chart of Counts by Gender :Gender count of Bar chart indicates that men have a higher rate of attrition compared to women, hence, this might suggest retention issues could be gender specific to some employees

(g) Bar Charts of Counts by Department :Department counts of Bar chart focuses on the departments that show the highest rates of attrition hence giving a few pinpointed areas that need focus on departmental areas that would need retention strategies and better working conditions.

(h) Radar Plot: It uses the leavers vs. stayers metrics of visual satisfaction thus implies a lower satisfaction level among the leavers, which means it is one of the employee engagement areas where it needs improvement.

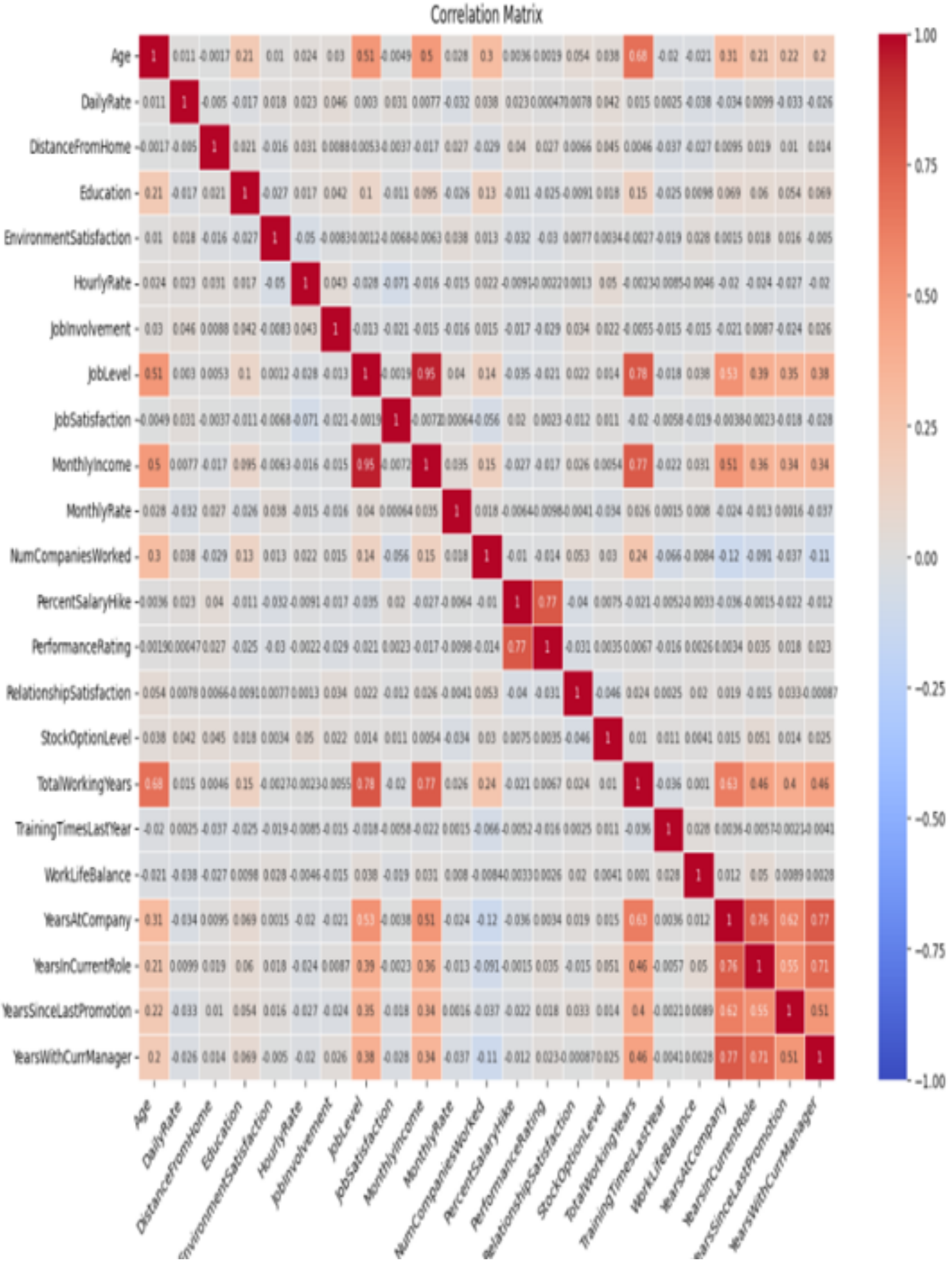


Figure 2: Matrix of Correlation

3.2 Data Pre-processing

3.2.1 Handling Missing Values

Exploratory analysis was conducted prior to this step, including data type inspection and identification of missing values. This provided an overview of the dataset's structure and highlighted columns requiring further processing.

3.2.2 Removal of Irrelevant Columns

Several columns were dropped from the dataset due to their lack of meaningful contribution to predictive modeling:

Over18: This attribute is always true and, therefore, not useful as a predictor.

EmployeeNumber: This attribute does not contribute to data interpretation.

EmployeeCount: This attribute remains constant across all records.

StandardHours: This attribute exhibits no variation.

The removal of these irrelevant columns effectively reduced noise in the data, thereby improving model performance.

3.2.3 Encoding Categorical Variables

To prepare categorical features for modeling, we applied two encoding methods based on the number of unique values:

For binary categorical variables such as Attrition, OverTime, and Gender, we utilized LabelEncoder to convert categories into a numerical format

(e.g., converting ['Yes', 'No'] to [1, 0]).

For categorical variables with multiple unique levels (e.g., BusinessTravel, Department, JobRole), we employed one-hot encoding using pd.get dummies.This approach creates binary columns for each category, avoiding the creation

of misleading ordinal relationships among categorical variables.

3.2.4 Scaling Numeric Features

To ensure equitable contributions from all numeric features during training, we

applied Min-Max scaling using MinMaxScaler,to scale down with a range of [0, 1]. This normalization technique is commonly used in machine learning, especially when feature attributes exhibit widely varying magnitudes, to enhance model convergence and performance.

3.2.5 Addressing Class Imbalance with SMOTE and implementing RFE

The attrition have associated with class imbalance, we applied (SMOTE). SMOTE creates synthetic samples for the minority class (e.g., employees who left) to balance the dataset This balancing improves ability of model by learning from classes like underpresent, so that it improves predictive performance overall.The dataset is splitted into testing and training applying an 20:80 ratio and there by applying feature recursive elimination.

3.3 Model Building

3.3.1 Logistic Regression

Logistic regression is applied to predict binary outputs by leveraging the

logistic function to model the probability of the target variable.

$$P(y = 1|Z) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 Z_1 + \cdots + \beta_n Z_n)}}$$

Here model parameters are β0, β1, . . . , βn .

3.3.2 K-Nearest Neighbors (KNN)

KNN is a label for a data point which depends on the majority class of itsclosest ’k’ neighbors, calculated using distance metrics .

$$d(q, q') = \sqrt{\sum_{i=1}^{n} (q_i - q_i')^2}$$

Where the Euclidean distance between points q and q′ is d(q, q′).

3.3.3 Support Vector Machine (SVM)

SVM separates classes by finding the hyperplane with the largest possible margin between the data points of each class.

$$f(x) = b + w^T y$$

Where b is bias, y is input vector, and w is weight vector.

3.3.4 Naive Bayes

Naive Bayes applies Bayes’ theorem with an independence assumption between predictors to classify data.

$$P(C|Q) = \frac{P(Q|C)P(C)}{P(Q)}$$

Where the feature is Q, and the class is C.

3.3.5 Decision Tree

Decision Trees split data at decision nodes to categorize it, using measures like Gini impurity to determine splits.

$$Gini = 1 - \sum_{i=1}^{n} q_i^2$$

Where qi is the probability for class i.

3.3.6 Long Short-Term Memory (LSTM)

LSTMs represent a RNN utilizing memory cells to learn long-term dependencies, overcoming the problem of vanishing gradients in standard RNNs.

$$h_t^{\alpha} = o_t^{\beta} \cdot \tanh\left(C_t^{\gamma}\right)$$

$$f_t^{\delta} = \sigma\left(b_f^{\epsilon} + W_f^{\zeta} \cdot \left[h_{t-1}^{\eta}, x_t^{\theta}\right]\right)$$

$$C_t^{\iota} = C_{t-1}^{\kappa} \cdot f_t^{\lambda} + i_t^{\mu} \cdot \widetilde{C_t^{\nu}}$$

$$i_t^{\xi} = \sigma\left(\left[h_{t-1}^{\pi}, x_t^{\rho}\right] \cdot W_i^{\sigma} + b_i^{\tau}\right)$$

$$o_t^{\upsilon} = \sigma\left(W_o^{\chi} \cdot \left[h_{t-1}^{\psi}, x_t^{\omega}\right] + b_o^{\varphi}\right)\mathrm{h}$$

3.3.7 Convolutional Neural Network (CNN)

CNNs are applied to process and analyze images. CNNs employ convolutional layers that automatically learn spatial relationships within input images.

$$S(p,q) = (I * K)(p,q) = \sum_{a} \sum_{b} I(a,b)K(p-a, q-b)$$

Here I is the image input, S is the feature output map, and K is the kernel/filter.

3.3.8 Recurrent Neural Network (RNN)

RNNs are tailored for predicting sequences, utilizing an internal hidden state to retain information from earlier inputs, enabling them to forecast outcomes based on sequential data.

$$h_t = \sigma(b + W_h \cdot h_{t-1} + W_r \cdot r_t)$$

Where b is the bias, ht is the hidden state at time t, input weight matrix is Wr, Wh is the weight matrix for the hidden state, and rt is the input at time t.

3.3.9 Feedforward Neural Network (FNN)

A FNN is a kind of neural network [10]characterized by one-way connections without cycles. Data flows from input nodes, passes through hidden nodes, finally it reaches the output nodes.

$$a^{(l)} = f\left(z^{(l)}\right) = f\left(b^{(l)} + W^{(l)}a^{(l-1)}\right)$$

Here,f is the activation function (e.g., ReLU, sigmoid), activation of layer l is a (l), layer l weight is W(l), and layer l bias is b(l).

3.3.10 Feedforward Neural Network with Dropout

Dropout is a regularizer that prevents overfitting by randomly setting some neurons to zero during training, allowing the network to learn robust features.

$$a^{(l)} = f\left(z^{(l)}\right) = f\left(b^{(l)} + W^{(l)}a^{(l-1)}\right) \cdot r$$

Here, r is a vector of random variables drawn from a Bernoulli distribution,indicating whether to keep or drop a neuron during training.

3.3.11 Random Forest

Random Forest creates numerous decision trees[11] while training process and gives the most (common class) of classification or average prediction of (regression) as its output.

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} f_i(q)$$

Here prediction from the i-th tree is fi(q) , and N is the number of trees.

3.3.12 Gradient Boosting

Gradient Boosting constructs a series of models where each new model addresses the mistakes made by the preceding one, using gradient descent to minimize the loss function.

$$\widehat{y_t} = \widehat{y_{t-1}} + \eta f_t(x)$$

Here ˆyt is the prediction of iteration t, ˆyt−1 is prediction of iteration t − 1, η is the learning rate, and ft(x) is the new weak learner.

3.3.13 Extreme Gradient Boosting(XGBoost)

XGBoost is gradient boosting but advanced version in terms of speed and performance, incorporating regularization to prevent overfitting.

$$\hat{y} = \sum_{t=1}^{T} \eta f_t(x)$$

Here the learning rate is η , ft(x) is prediction from t-th tree, and T is the number of trees.

3.3.14 AdaBoost (Adaptive Boosting)

AdaBoost combines several weak classifiers into one strong classifier, assigning more weight to misclassified instances in subsequent classifiers.

$$\hat{y} = \sum_{t=1}^{T} \alpha_t f_t(x)$$

Here t-th classifier weight is αt , and prediction is ft(x) from t-th classifier.

3.3.15 CatBoost (Categorical Boosting)

CatBoost handles categorical features natively, preventing overfitting with ordered boosting.

$$\hat{y} = \sum_{t=1}^{T} \eta f_t(x)$$

Similar to XGBoost, with adjustments for categorical features.

3.3.16 Stacking

Stacking combines multiple models (base learners) by training a meta-learner on the predictions of the base models.

yˆ = q(p1(x), p2(x), . . . , pn(x))

Where q is the meta-learner, and pi(x) is the prediction from i-th base learner.

3.3.17 Voting

Voting combines multiple classifiers to make a single prediction, using either soft voting on predicted probabilities or hard voting on class labels. Soft Voting on predicted probabilities or hard voting on class labels.

$$\hat{y} = \arg\max\left(\frac{1}{N}\sum_{i=1}^{N} P_i(y|x)\right)$$

Where Pi(y|x) is the predicted class probability y from the i-th classifier

CHATPER – 4

Interpretation and test Results

A comparison of several models metrics are discussed among that best algorithm for employee attrition prediction is the Random Forest classifier with an accuracy of 98.3%. Such robust performance can definitely be attributed to the ensemble learning technique used by Random Forest, where it is possible to aggregate the predictions from multiple decision trees in order to enhance the overall accuracy and avoid overfitting.[12]

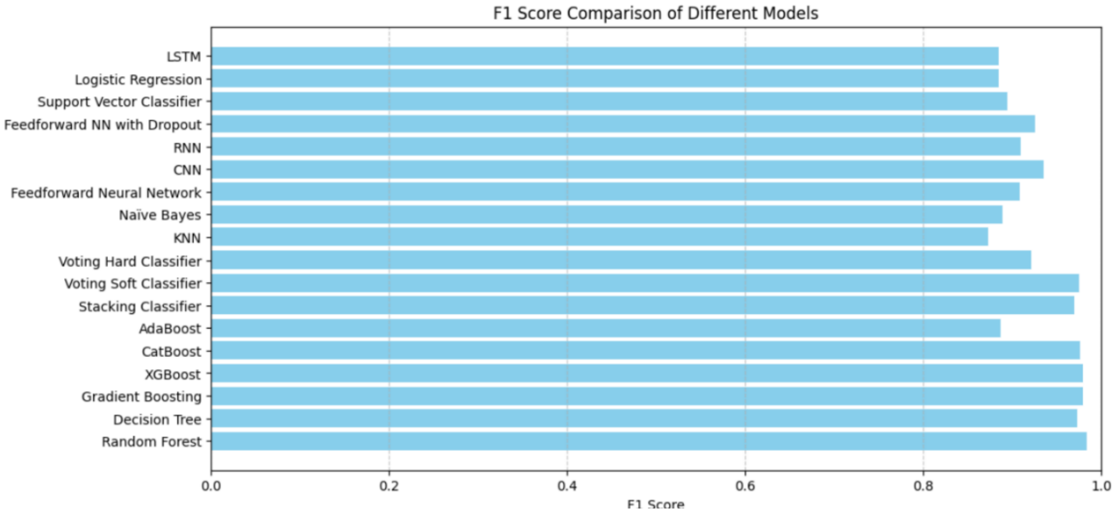


Figure 12: F1 Score Comparison of Different Models

Figure 12: Comparision of Different models F1 Score for balancing precision,recall to better uderstand classification of employees that are at risk There are numerous reasons why the model was able to represent high-dimensional data and capture complex interactions among features in an efficient way, which led to its successful prediction of employee attrition in a significant way. Here are metrics of evaluation for different model.

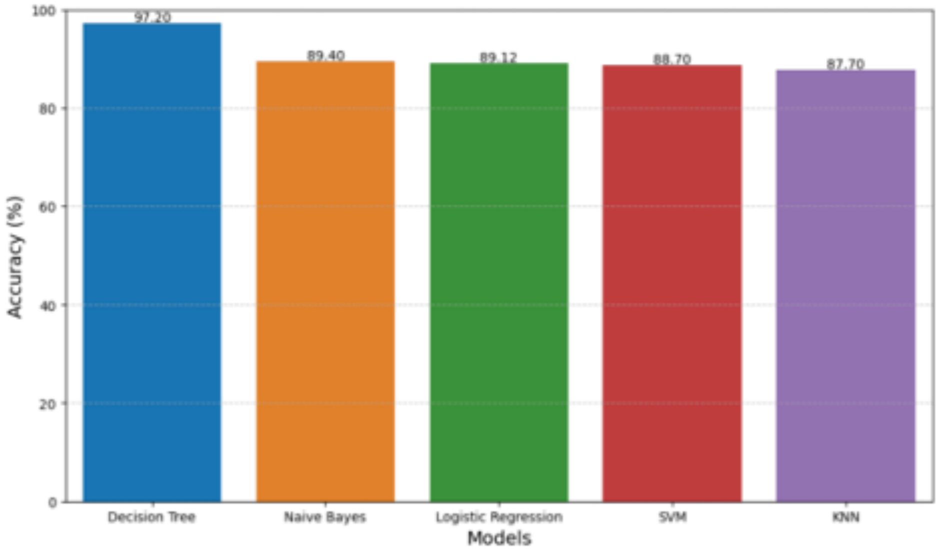


Figure 13: Conventional Models Accuracies

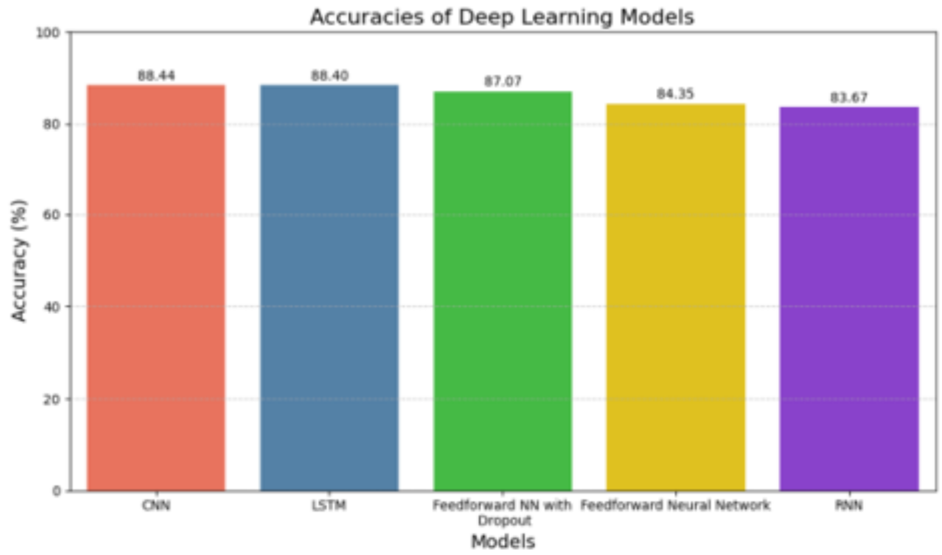


Figure 14: Deep Learning Models Accuracies

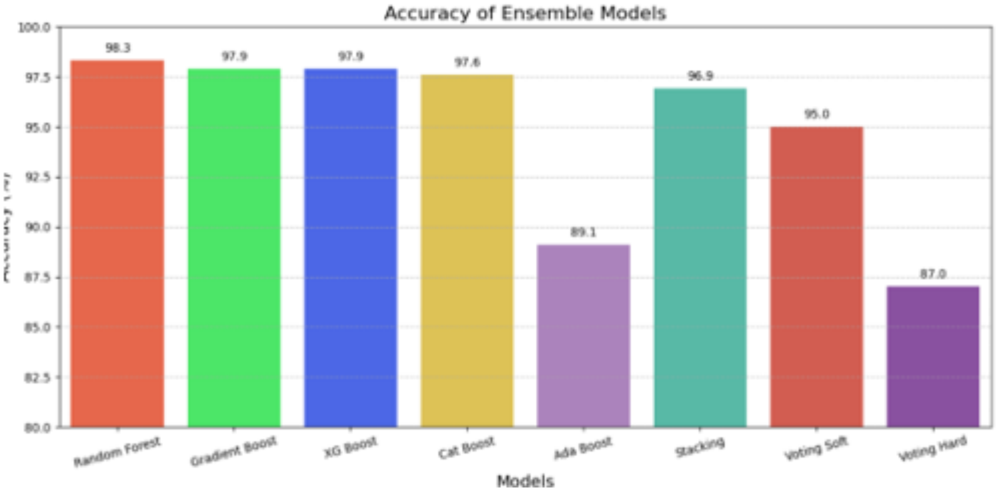
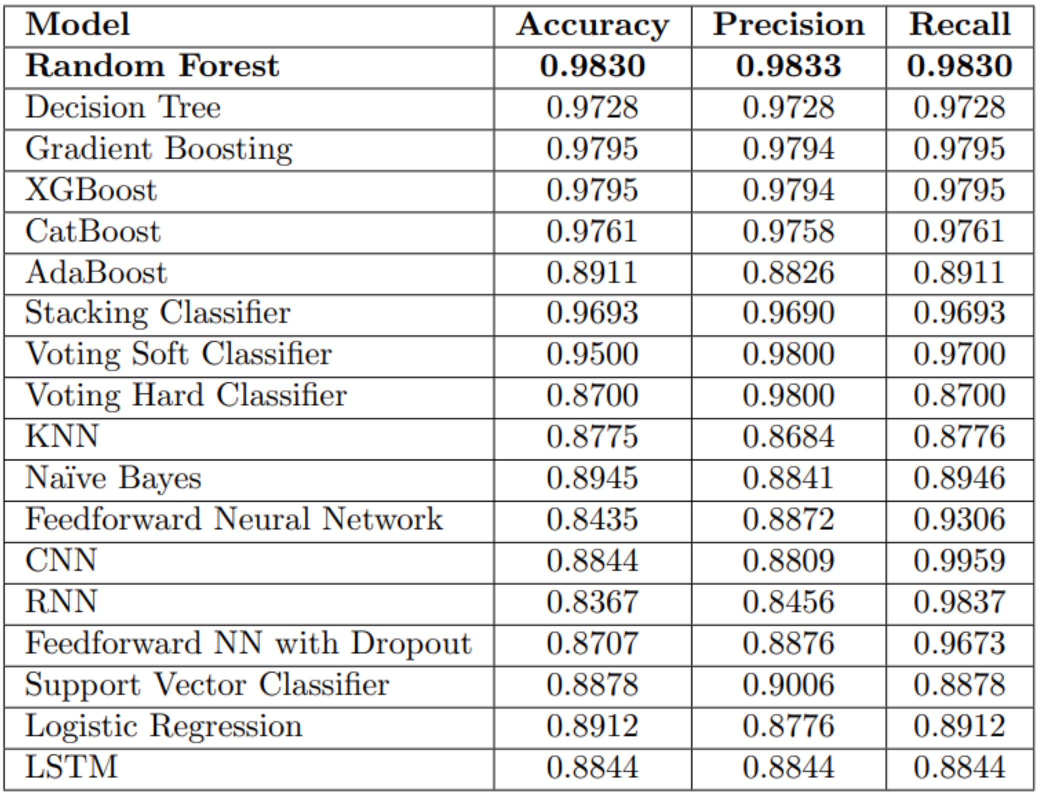


Figure 15: Ensemble Models Accuracies

Table 1: Derived Model Accuracy, Precision, and Recall



Model	**Accuracy**	**Precision**	**Recall**
Random Forest	**0.9830**	**0.9833**	**0.9830**
Decision Tree	0.9728	0.9728	0.9728
Gradient Boosting	0.9795	0.9794	0.9795
XGBoost	0.9795	0.9794	0.9795
CatBoost	0.9761	0.9758	0.9761
AdaBoost	0.8911	0.8826	0.8911
Stacking Classifier	0.9693	0.9690	0.9693
Voting Soft Classifier	0.9500	0.9800	0.9700
Voting Hard Classifier	0.8700	0.9800	0.8700
KNN	0.8775	0.8684	0.8776
Naïve Bayes	0.8945	0.8841	0.8946
Feedforward Neural Network	0.8435	0.8872	0.9306
CNN	0.8844	0.8809	0.9959
RNN	0.8367	0.8456	0.9837
Feedforward NN with Dropout	0.8707	0.8876	0.9673
Support Vector Classifier	0.8878	0.9006	0.8878
Logistic Regression	0.8912	0.8776	0.8912
LSTM	0.8844	0.8844	0.8844

The following confusion matrix clarifies further on performance by representing Correctly classified as positive,Correctly classified as negative,Incorrectly classified as positive,Incorrectly classified as negative for top accuracy achiever RF model

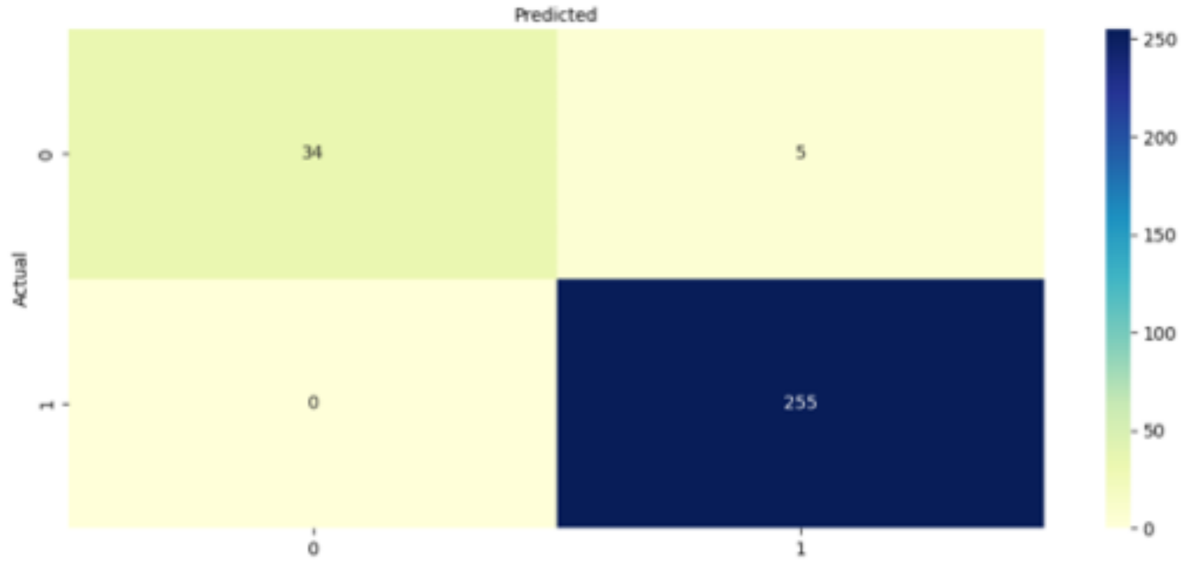


Figure 16: Confusion Matrix of Random Forest

CHAPTER – 5

Conclusion

In conclusion this study underlines the important role of accurately predicting employee attrition in an organization to improve organizational sustainability and performance. Using a variety of classification algorithms, this study displayed that the Random Forest model got an impressive accuracy of 98.3% 15 with regard to employee attrition prediction. Such high accuracy underscores the capability to identify complex patterns in the data even amidst inherent noise. Indeed, through this study, it becomes evident that implementation of the Random Forest classifier will help organizations to provide proactive identification of at-risk employees, thus enabling opportunity for timely intervention and retaining employees. Ultimately, this contributes valuable insights into human resource management by focusing on how advanced analytics has the potential to provide a committed and stable workforce.

REFERENCES

[1] N. Bhartiya, S. Jannu, P. Shukla, and R. Chapaneri, ”Employee Attrition Prediction Using Classification Models,” 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 2019, pp. 1-6. Available: https://doi.org/10.1109/I2CT45611.2019.9033784.

[2] S. Gupta, G. Bhardwaj, M. Arora, R. Rani, P. Bansal, and R. Kumar, ”Employee Attrition Prediction in Industries using Machine Learning Algorithms,” in 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 945-950.

[3] Doohee Chung, Jinseop Yun, Jeha Lee, Yeram Jeon, ”Predictive model of employee attrition based on stacking ensemble learning,” Expert Systems with Applications, vol. 215, 2023, 119364. ISSN 0957-4174. Available: https://doi.org/10.1016/j.eswa.2022.119364.

[4] Praphula Kumar Jain, Madhur Jain, and Rajendra Pamula, ”Explaining and predicting employees’ attrition: a machine learning approach,” SN Appl. Sci., vol. 2, no. 757, 2020. Available: https://doi.org/10.1007/s42452- 020-2519-4.

[5] Yedida, Rahul, Reddy, Rahul, Vahi, Rakshit, Jana, Rahul, Gv, Abhilash, and Kulkarni, Deepti. (2018). Employee Attrition Prediction. 10.48550/arXiv.1806.10480.

[6] A. Raza, K. Munir, M. Almutairi, F. Younas, and M. M. S. Fareed, ”Predicting Employee Attrition Using Machine Learning Approaches,” Applied Sciences, vol. 12, no. 6424, 2022. Available: https://doi.org/10.3390/app12136424.

[7] N. Mansor, N. S. Sani, and M. Aliff, ”Machine Learning for Predicting Employee Attrition,” International Journal of Advanced Computer Science and Applications (IJACSA),vol. 12, no. 11, 2021. Available: http://dx.doi.org/10.14569/IJACSA.2021.0121149.

[8] Raj M., Arjun, Mishra, Arjyalopa, and Forest, George. (2024). ”Predictive Analytics for Employee Attrition: Leveraging Machine Learning for Strategic Human Resources Management,” pp. 13-29. Available: 16 https://www.researchgate.net/publication/379861220 Predictive Analytics for Employee Attrition Leveraging Machine Learning for Strategic Human Resources Management.

[9] M. Lazzari, J. M. Alvarez, and S. Ruggieri, ”Predicting and explaining employee turnover intention,” International Journal of Data Science and Analysis, vol. 14, pp. 279–292, 2022. Available: https://doi.org/10.1007/s41060- 022-00329-w.

[10] M. Nandal, V. Grover, D. Sahu, and M. Dogra, ”Employee Attrition: Analysis of Data Driven Models,” EAI Endorsed Transactions on Internet of Things, vol. 10, Jan. 2024.

[11] M. Subhashini and R. Gopinath, ”Employee Attrition Prediction in Industry Using Machine Learning Techniques,” International Journal of Advanced Research in Engineering and Technology, vol. 11, no. 12, 2020, pp. 3329-3341. Available: https://doi.org/10.17605/OSF.IO/9XDWE.

[12] Mohammad Reza Shafie, Hamed Khosravi, Sarah Farhadpour, Srinjoy Das, and Imtiaz Ahmed, ”A cluster-based human resources analytics for predicting employee turnover using optimized Artificial Neural Networks and data augmentation,” Decision Analytics Journal, vol. 11, 2024, 100461. ISSN 2772-6622. Available: https://doi.org/10.1016/j.dajour.2024.100461.