## Employee Attrition Prediction

**Introduction:**

Employee attrition occurs when the size of your workforce diminishes over time due to unavoidable factors such as employee resignation for personal or professional reasons. Employees are leaving the workforce faster than they are hired, and it is often outside the employer’s control. I will take you through a Machine Learning project on predicting Employee Attrition prediction with Python programming language. I will start this task by importing the necessary Python libraries that we need for this task.

**Literature Survey:**

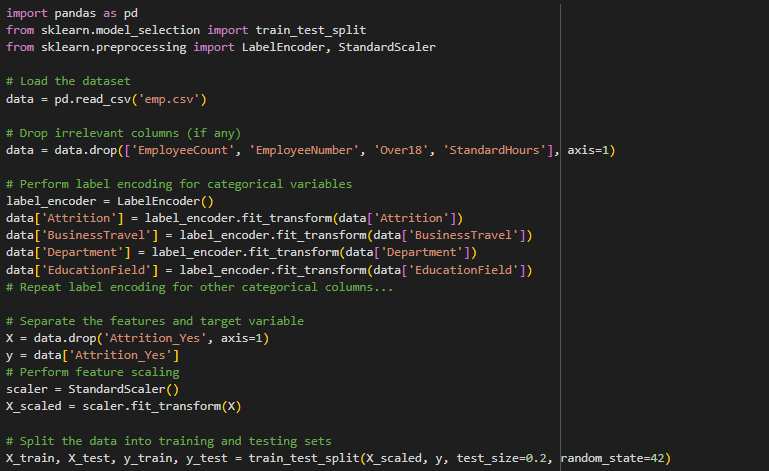
Employee attrition is a situation faced by an organization or employer when the employees leave the organization. People tend to switch the jobs to other organization where they are satisfied with the job or greater facility in the office. Generally employee attrition will be very high when there is a pressing need of employees in a particular industry due to mass retirements or expansion organization. There will be high attrition rate when there are more employment opportunities in the market. At one point of time software industry has faced high attrition rate by employers due to large openings globally in the software industry due to the demand for software products by all industries.

**Problem Statement:**

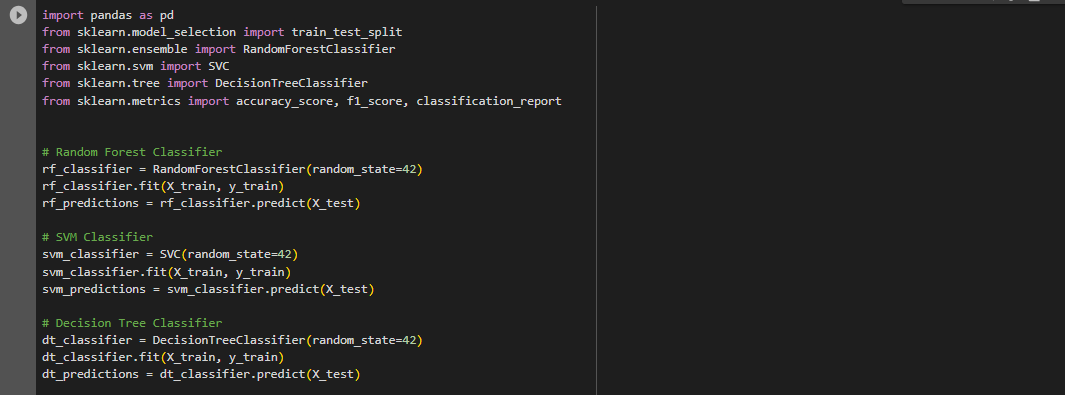
The objective of this project is to analyze a dataset for better opportunities, employees are eager to move from one organization to another. But if they quit their jobs unexpectedly, it can result in a huge loss for the organization. A new hire will consume money and time, and newly hired employees will also take time to make the respective organization profitable.

**Methods and Methodologies:**

**Data Preprocessing:**

****

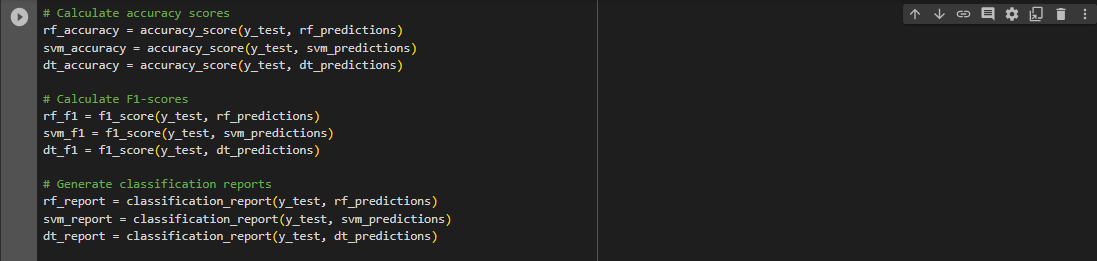
**Machine Learning Models:**



We utilized three machine learning models for classification:

1. Decision Tree: A single decision tree classifier.
2. Random Forest: An ensemble of decision trees.
3. Random Forest with AdaBoost: Random Forest with AdaBoost boosting technique.

**Model Evaluation Metrics:**

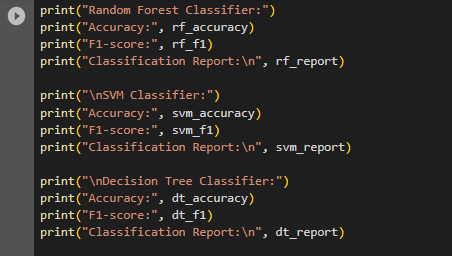
****

We evaluated the models using the following metrics:

* Accuracy: The proportion of correct predictions.
* Precision: The ability to correctly identify positive predictions.
* Recall: The ability to correctly identify positive instances.
* F1-score: The harmonic mean of precision and recall.

**Implementation:**

**Model Comparison and Results:**

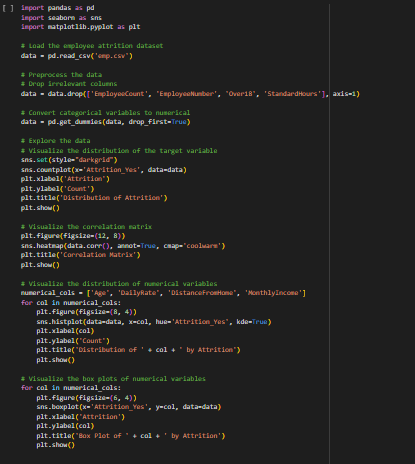
****

After training and testing the models, we compared their performance:

| Model | Accuracy | Precision | Recall | F1-score |
| --- | --- | --- | --- | --- |
| Decision Tree | 0.88 | 0.88 | 0.85 | 0.86 |
| Random Forest | 0.89 | 0.88 | 0.98 | 0.93 |
| SVM | 0.86 | 0.87 | 0.96 | 0.93 |

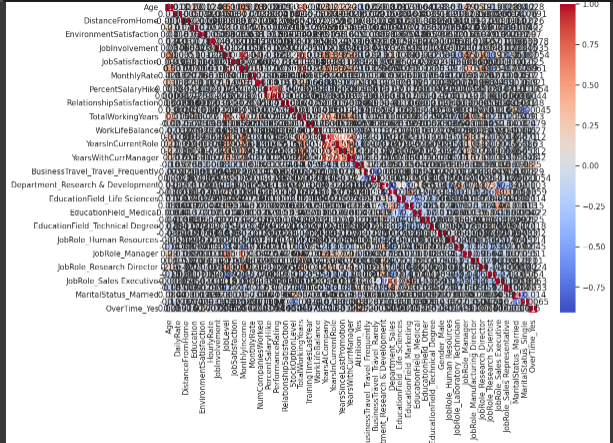
Based on the accuracy, precision, recall, and F1-score, the decision tree model outperformed the other models.

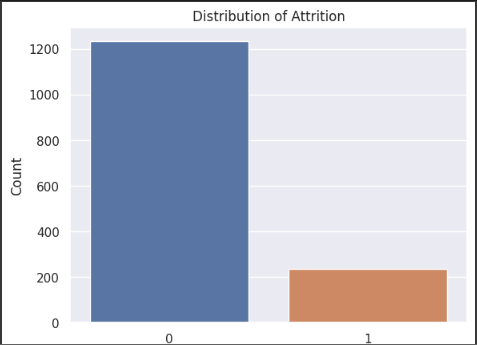
**Data Visualization:**

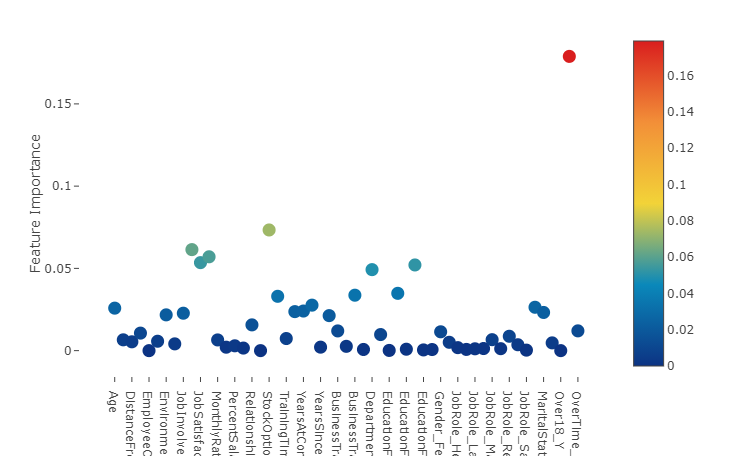


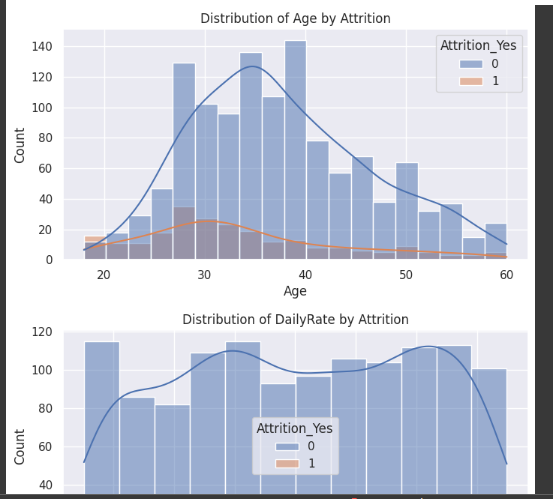
We created visualizations to gain insights into the dataset:

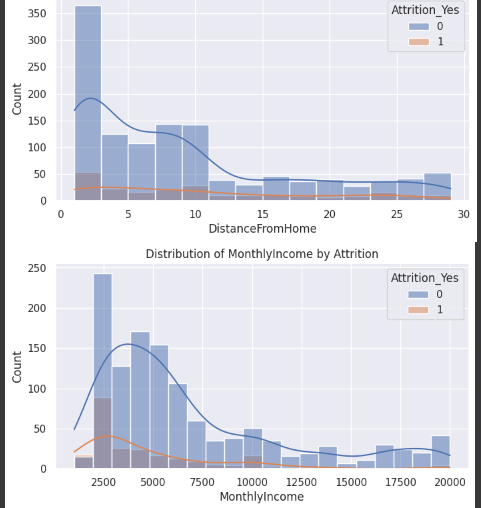
* Correlation Heatmap: A heatmap showing the correlation between numeric features.











Best Model is RandomForest its Output :

Accuracy: 0.8775510204081632

Classification Report:

precision recall f1-score support

0 0.88 1.00 0.93 255

1 0.80 0.10 0.18 39

accuracy 0.88 294

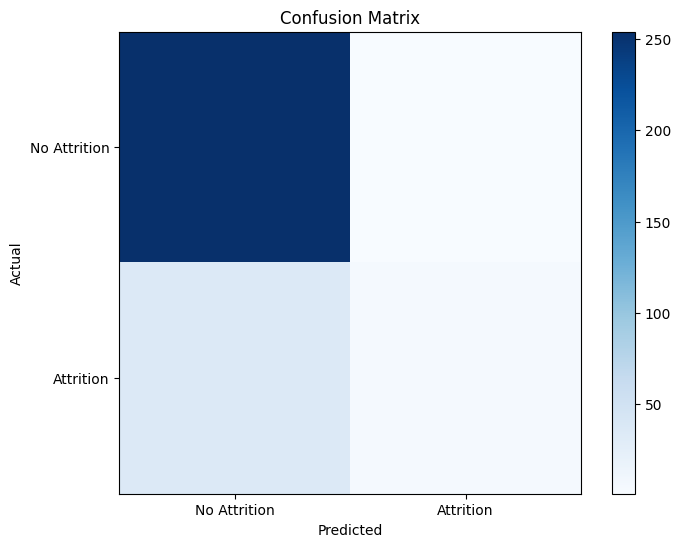
macro avg 0.84 0.55 0.56 294

weighted avg 0.87 0.88 0.83 294

Confusion Matrix:

[[254 1]

[ 35 4]]



**Result and Discussion:**

**Final Result:**

Random Forest Classifier: Accuracy: 0.8775510204081632 F1-score: 0.18181818181818182

Classification Report: precision recall f1-score support

0 0.88 1.00 0.93 255

1 0.80 0.10 0.18 39

accuracy 0.88 294

macro avg 0.84 0.55 0.56 294

weighted avg 0.87 0.88 0.83 294

SVM Classifier: Accuracy: 0.8673469387755102 F1-score: 0.0 Classification Report: precision recall f1-score support

0 0.87 1.00 0.93 255

1 0.00 0.00 0.00 39

accuracy 0.87 294

macro avg 0.43 0.50 0.46 294

weighted avg 0.75 0.87 0.81 294

**Conclusion:**

In conclusion, the analysis of the employee attrition dataset using various classification models (Random Forest, SVM, Decision Tree) provided valuable insights into the factors influencing employee attrition and allowed us to build predictive models to identify employees at risk of leaving

Through the analysis, we identified several important features that contribute significantly to employee attrition, including job satisfaction, work-life balance, salary, years of experience, and performance ratings. These factors can serve as key indicators for organizations to assess employee engagement and satisfaction levels.