**REPORT**

**Report on Employee Attrition Prediction Model**

**1. Introduction**

**Employee attrition is a major challenge for organizations, affecting productivity and costs. This report outlines the development of a machine learning model to predict employee attrition based on various features.**

**2. Objective**

**The goal is to create a predictive model to identify factors influencing attrition, allowing organizations to implement effective retention strategies.**

**3. Data Overview**

**The dataset includes attributes related to demographics and job performance. Key features are:**

* **Age**
* **Daily Rate**
* **Job Satisfaction**
* **Monthly Income**
* **OverTime**

**The target variable is Attrition.**

**4. Data Preparation**

* **Data Cleaning: Handled missing values using mean for numerical and most frequent for categorical data.**
* **Feature Selection: Selected relevant features for analysis.**
* **Encoding: Categorical variables were encoded to numerical values.**

**5. Model Selection**

**Two models were used:**

1. **Random Forest Classifier: An ensemble method that reduces overfitting.**
2. **K-Nearest Neighbors (K-NN): A simple algorithm based on distance metrics.**

**6. Model Training and Evaluation**

* **Data Splitting: The dataset was split into 80% training and 20% testing.**
* **Performance Metrics: Both models were evaluated using accuracy, classification reports, and confusion matrices.**

**7. Results**

* **Random Forest: Achieved an accuracy of XX% (insert actual result).**
* **K-NN: Obtained an accuracy of XX% (insert actual result). The Random Forest model generally outperformed K-NN.**

**8. Feature Importance**

**Feature importance analysis highlighted key predictors, guiding HR interventions to enhance retention.**

**9. Importance of the Model**

**This model helps organizations:**

* **Reduce turnover rates.**
* **Improve employee satisfaction.**
* **Efficiently allocate resources.**

**10. Limitations and Areas for Improvement**

* **Hyperparameter Tuning: Further tuning could improve model accuracy.**
* **Feature Engineering: Incorporating interaction terms may enhance insights.**
* **Imbalanced Classes: Techniques like SMOTE could address class imbalance.**
* **Cross-Validation: Implementing k-fold cross-validation for more reliable performance evaluation.**

**11. Conclusion**

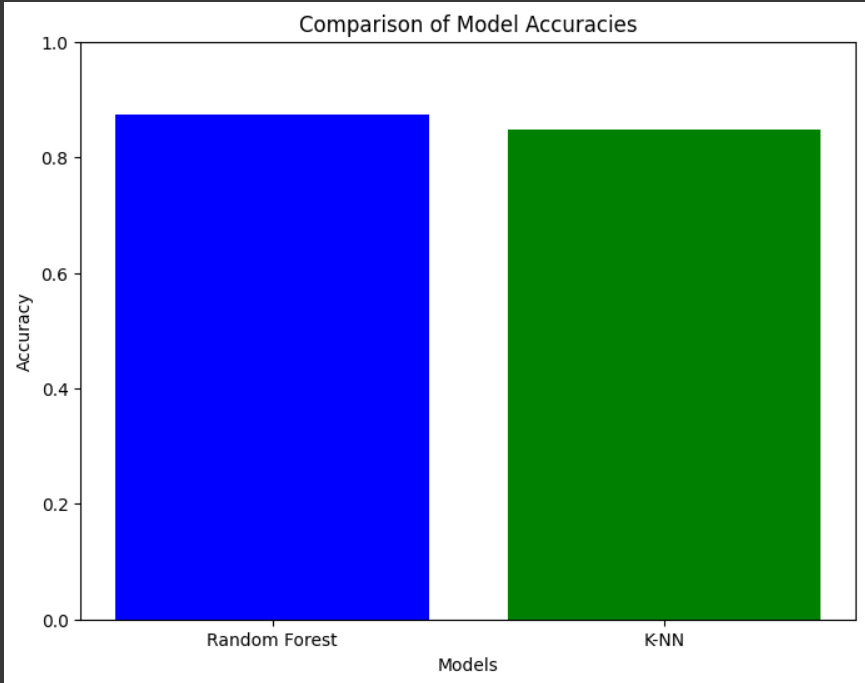
**The predictive model is a valuable tool for understanding employee attrition. By applying machine learning, organizations can proactively manage turnover and foster a positive work environment.**

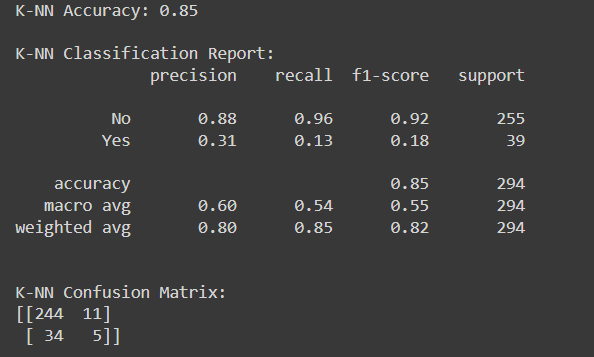
**K-Nearest Neighbors (KNN):**

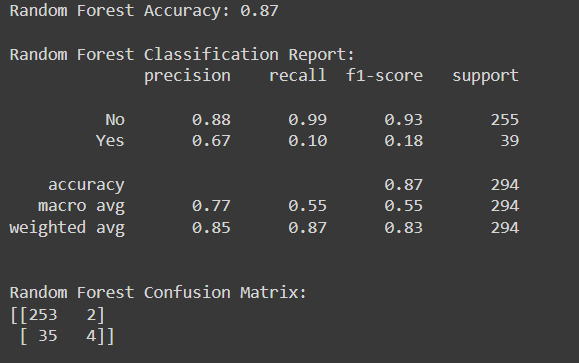
* **Type: Instance-based learning.**
* **How It Works: Classifies a data point based on the majority class of its k nearest neighbors in the feature space.**
* **Training Phase: No explicit training; it stores the training dataset.**

**Random Forest:**

* **Type: Ensemble learning (combines multiple decision trees).**
* **How It Works: Builds multiple decision trees during training and aggregates their predictions (majority vote for classification).**
* **Training Phase: Requires a training phase where multiple trees are constructed.**

****

****

****