**Project Report: Employee Attrition Prediction**

**Name: Aiyede Sunday Victor**

**Date: 17-08-2025**

**1. Introduction**

Employee attrition poses a critical challenge for organizations, leading to increased hiring costs, reduced productivity, and potential loss of institutional knowledge. Understanding the key drivers of attrition enables management to take proactive measures.  
This project leverages data analytics and machine learning to predict employee attrition and identify influential factors contributing to turnover.

**2. Objective**

The main objectives of the project are:

* To analyze employee data and identify patterns related to attrition.
* To build predictive models that classify employees as "likely to leave" or "likely to stay."
* To provide insights and recommendations for HR strategies to reduce attrition rates.

**3. Dataset Description**

The dataset used in this project contains multiple features such as:

* **Demographic Information**: Age, Gender, Marital Status.
* **Job-Related Features**: Department, Job Role, Years at Company, Distance from Home, Work-Life Balance, Job Satisfaction.
* **Compensation and Performance**: Monthly Income, OverTime, Performance Rating.
* **Target Variable**: *Attrition* (Yes/No).

**4. Methodology**

**4.1 Data Preprocessing**

* Handled missing values and outliers.
* Encoded categorical variables.
* Normalized numerical features for modeling.

**4.2 Exploratory Data Analysis (EDA)**

* Distribution of attrition across age groups, job roles, and departments.
* Higher attrition observed among younger employees, those with less experience, and employees with longer commuting distances.
* Over Time and Job Satisfaction showed significant relationships with attrition.

**4.3 Feature Selection**

* Features such as *Employee Count*, *Over18*, and *Standard Hours* were dropped due to lack of variability.
* Department had minimal impact and was considered for removal in some models.

**4.4 Model Building**

* Applied machine learning algorithms including:
  + **Logistic Regression**
  + **Decision Tree Classifier**
  + **Random Forest**
  + **Gradient Boosting Classifie**

**5. Model Evaluation**

* **Logistic Regression**: Provided baseline performance.
* **Decision Tree Classifier**: Moderate performance with overfitting tendencies.
* **Random Forest**: Improved generalization with better accuracy than Decision Trees.
* **Gradient Boosting Classifier**: Achieved the **highest accuracy of 0.8878 (~88.8%)**, making it the best-performing model.

**6. Findings and Insights**

* Younger employees and those with shorter tenure are more likely to leave.
* High distance from home correlates with higher attrition.
* Over Time workers showed significantly higher attrition rates.
* Job satisfaction and work-life balance play crucial roles in employee retention.

**7. Conclusion**

The project successfully demonstrated how machine learning can be used to predict employee attrition with high accuracy.

* **Best Model**: Gradient Boosting Classifier, with an accuracy of **88.8%**.
* **Business Implications**: HR teams can use these insights to develop targeted retention strategies such as:
  + Improving work-life balance policies.
  + Offering flexible work arrangements to employees with long commutes.
  + Creating career development opportunities for younger employees.

**8. Recommendations**

1. Implement retention initiatives targeting at-risk employee groups.
2. Regularly monitor employee satisfaction and workload distribution.
3. Use predictive models in HR analytics dashboards to identify employees at high attrition risk in real time.