Final Report

Prepared by: **Bilal Hassan Nizami**

Results and Evaluation

Model Interpretation:

We trained and evaluated multiple models to understand the key drivers of employee attrition and to accurately predict whether an employee is likely to leave. The analysis helped uncover relationships between key HR metrics (such as satisfaction level, number of projects, evaluation scores, and working hours) and attrition behavior.

- **Logistic Regression** provided a straightforward, interpretable baseline to identify linear relationships.
- **Tree-based models** (Decision Tree and Random Forest) were employed to capture non-linear relationships and feature interactions for higher accuracy and deeper insight.

Model Performance:

Logistic Regression:

- **Precision**: 90%
- **Recall**: 82%
- **F1-score**: 84%
- **Accuracy**: 85%

These metrics were calculated on the **test set** using **weighted averages**, which account for class imbalance. While logistic regression performed well, the model struggled with some of the more complex patterns in the data.

Tree-Based Machine Learning:

After additional **feature engineering**:

- **Decision Tree**
  - **AUC**: 98.1%
  - **Precision**: 97.6%
  - **Recall**: 92.5%
  - **F1-score**: 95.0%
  - **Accuracy**: 97.7%
- **Random Forest**
  - Slightly outperformed the decision tree with improved generalization and robustness.

These results indicate excellent model performance, particularly in identifying employees likely to leave (high recall) and minimizing false positives (high precision).

Visualizations and Key Findings:

The following visual analyses supported the modeling:

- **Boxplots** revealed that employees who left generally worked longer hours and were involved in more projects.
- **Bar plots of feature importances** (Random Forest and Decision Tree) showed:
  - **Satisfaction level**, **number of projects**, **average monthly hours**, and **evaluation score** were top predictors.
  - Employees with high evaluation scores but low satisfaction were more likely to leave.
- **Histograms** of number of projects and satisfaction levels further supported these patterns.

These visualizations helped translate statistical findings into actionable insights for stakeholders.

Conclusion, Recommendations, and Further Steps:

Conclusion:

Both modeling and feature importance analyses confirmed that **overwork and misalignment in performance expectations** are key drivers of employee turnover. The Random Forest model provided the best predictive power and reliability for stakeholder use.

Recommendations for better employee retention and avoiding new hiring to save cost:

To address employee attrition, we recommend the following actions:

1. **Limit project load** Cap the number of simultaneous projects an employee can be assigned to in order to avoid burnout.
2. **Review promotion timelines** Promote employees who have stayed for at least 4 years or explore dissatisfaction trends among them.
3. **Compensate long hours appropriately** If long work hours are necessary, implement clear rewards or overtime benefits.
4. **Clarify overtime policies and workload expectations** Ensure transparency in HR policies related to compensation, workload, and time off.
5. **Foster a healthy work culture** Conduct regular team- and company-wide discussions to improve employee satisfaction and morale.
6. **Rethink evaluation metrics** Avoid equating high evaluation scores with working over 200+ hours per month. Design a proportional recognition system that values both quality and effort.

Further steps that can be taken for better results:

1. **Check for potential data leakage** Investigate how model performance changes when removing features like `last_evaluation`. This may identify over-reliance on potentially retrospective or infrequent metrics.

2. **Explore alternate modeling targets** If features like `last_evaluation` or `satisfaction_level` are strong indicators of attrition, consider developing models to **predict evaluation scores or satisfaction levels** as a way to intervene earlier.

3. **Deploy and monitor the model** Integrate the best-performing model into an HR analytics pipeline. Begin tracking live predictions and real attrition outcomes for iterative improvement.