Employee Retention Prediction Report

Objective

The purpose of this project is to create a predictive model using Logistic Regression to determine the likelihood of an employee staying with the company. The intent is not just predictive accuracy, but also to generate interpretable insights that HR can use to understand and influence employee behavior.

Business Context

Employee retention is a critical issue for many organizations due to the high cost of turnover and its impact on productivity and morale. The company in focus is a mid-sized technology firm that wants to reduce voluntary attrition by identifying the attributes of employees who are likely to stay. Understanding these characteristics can help in developing targeted retention strategies.

Dataset Description

The dataset includes a variety of features categorized into:

- **Demographics**: Age, gender, marital status, and educational background.

- **Job Details**: Department, role, salary band, job level, and years at the company.

- **Performance Metrics**: Annual review ratings, number of promotions, and training scores.

- **Engagement Scores**: Job satisfaction, work-life balance, and relationship with supervisors.

- **Retention Outcome**: A binary variable indicating whether the employee stayed or left.

Data Preprocessing

Before modeling, several data preparation steps were carried out including:

- Handling missing values using imputation techniques.

- Encoding categorical variables such as gender, department, and education level.

- Normalizing numerical features for consistent model performance.

- Splitting the dataset into training and testing subsets to evaluate model generalizability.

Modeling Approach

Logistic Regression was chosen due to its interpretability and efficiency in binary classification problems. It calculates the probability that a given input belongs to a certain class—in this case, whether an employee will stay or leave.

The model was trained using the training set and evaluated on the test set using metrics such as:

- Accuracy: Proportion of correctly predicted instances.

- Precision: Correctness of positive predictions.

- Recall: Coverage of actual positives.

- F1-Score**: Harmonic mean of precision and recall.

Key Findings

Some of the most important features influencing employee retention were:

- High job satisfaction and positive relationships with supervisors were strong indicators of retention.

- Employees who had received recent promotions or training were more likely to stay.

- Tenure and performance scores were positively correlated with loyalty.

- Departments with high workload but low recognition had higher turnover.

Recommendations

Based on the model insights, the following actions are recommended:

- Invest in employee development through promotions and training programs.

- Monitor and improve job satisfaction scores regularly.

- Pay attention to departments or teams with higher attrition rates.

- Use predictive scores to identify at-risk employees and intervene early.

Conclusion

This project demonstrates the effectiveness of logistic regression in predicting employee retention. By understanding the drivers of employee loyalty, the company can make proactive decisions to improve workplace culture and reduce costly turnover.