Employee Performance Prediction using ML

By Umar Farooq

UpGrad Capstone Project

Project Objective & Dataset

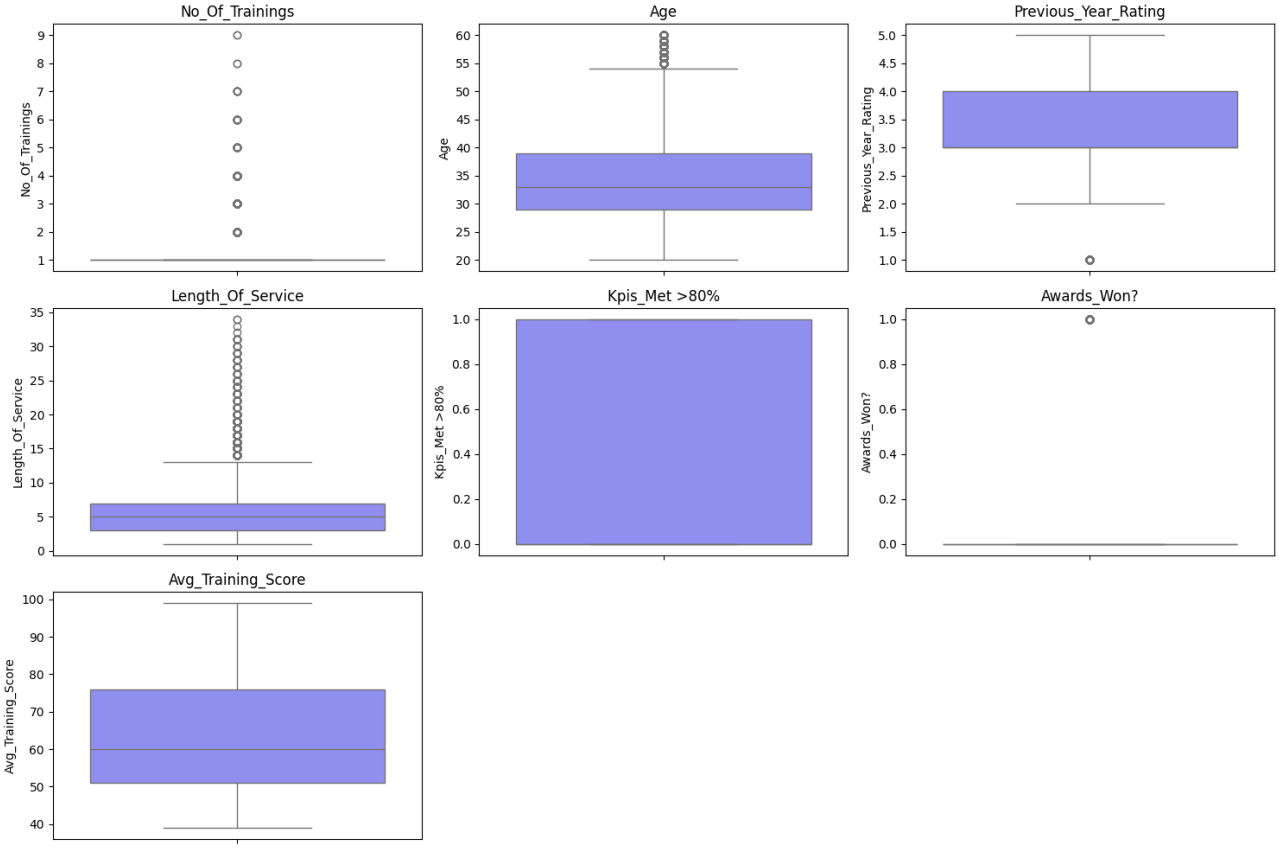
The goal of this project is to predict employee performance using various ML models.

The dataset includes demographic details, performance ratings, awards, and training records of employees.

We aim to use ML to help HR focus training efforts and improve productivity.

Data Preprocessing & EDA

- Checked and handled missing values.
- Converted categorical variables using encoding.
- Visualized correlations using a heatmap.

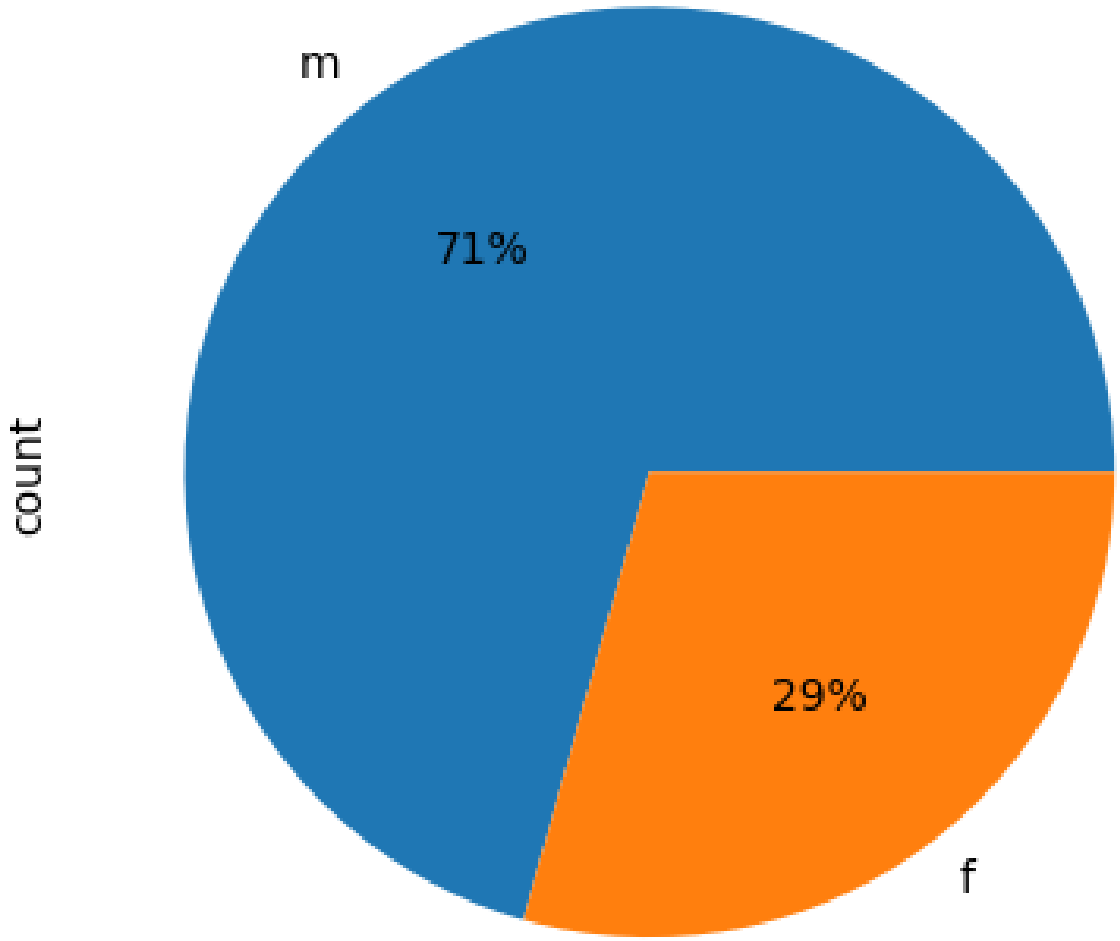


Scaler Comparison

Tested four scalers: StandardScaler, MinMaxScaler, RobustScaler, MaxAbsScaler.

Different models were trained with each scaler and evaluated for performance.

The best scaler was selected based on accuracy and F1 score.

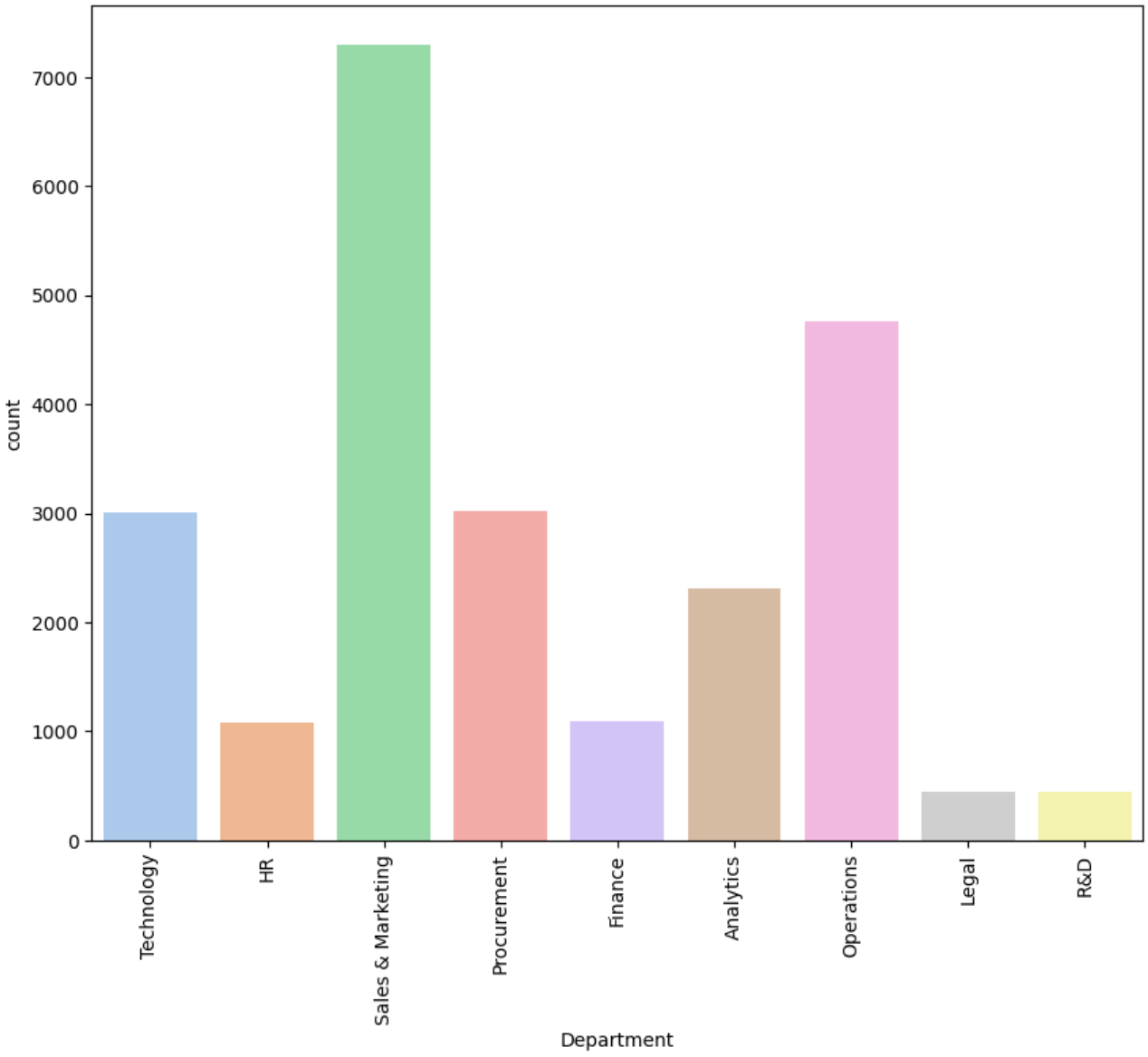


Model Comparison

Compared XGBoost, LightGBM, and CatBoost classifiers.

Metrics: Accuracy, Precision, Recall, F1 Score, ROC AUC.

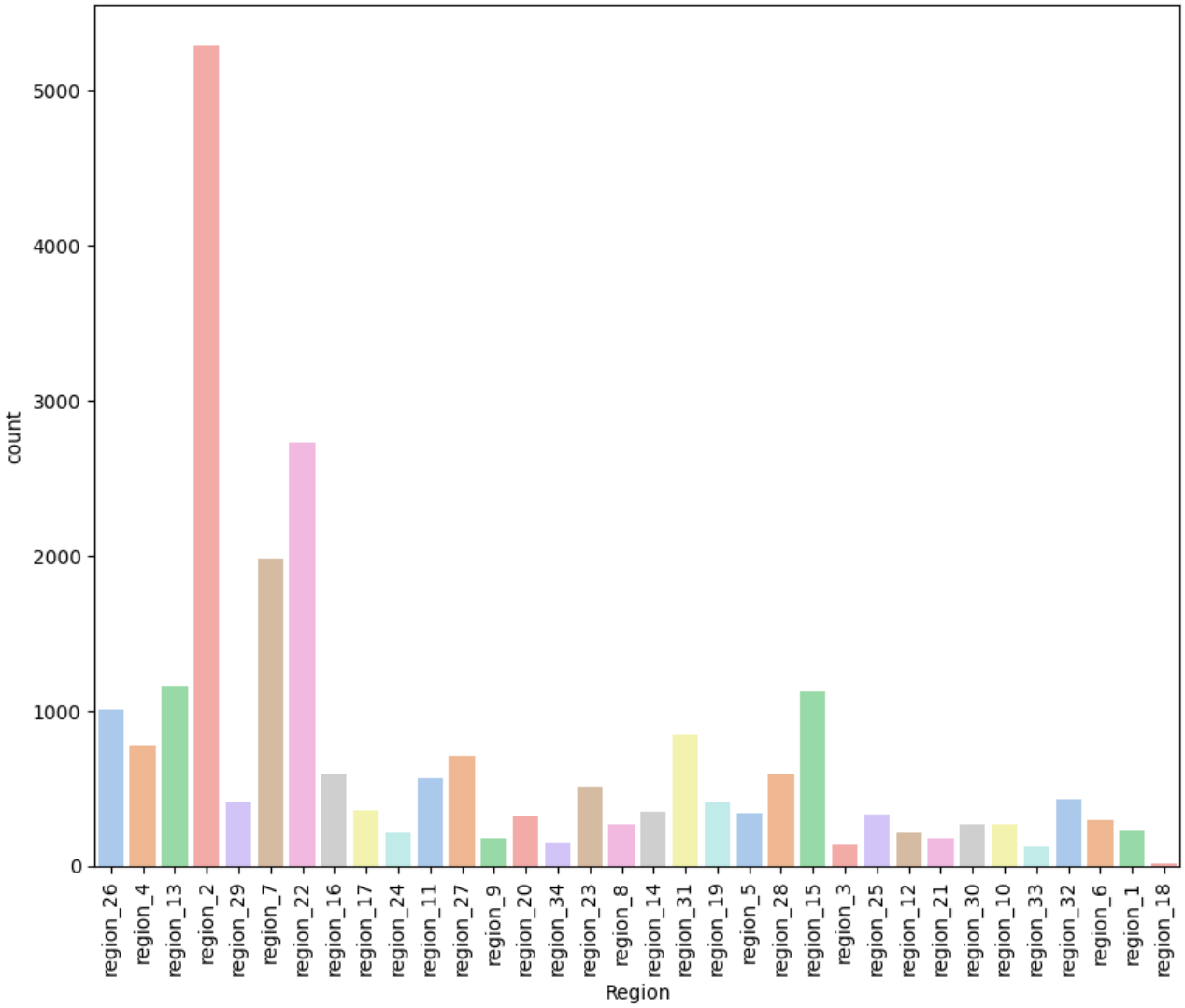
CatBoost showed the best performance consistently across all metrics.



ROC Curve Comparison

ROC curve comparison to evaluate classifier performance.

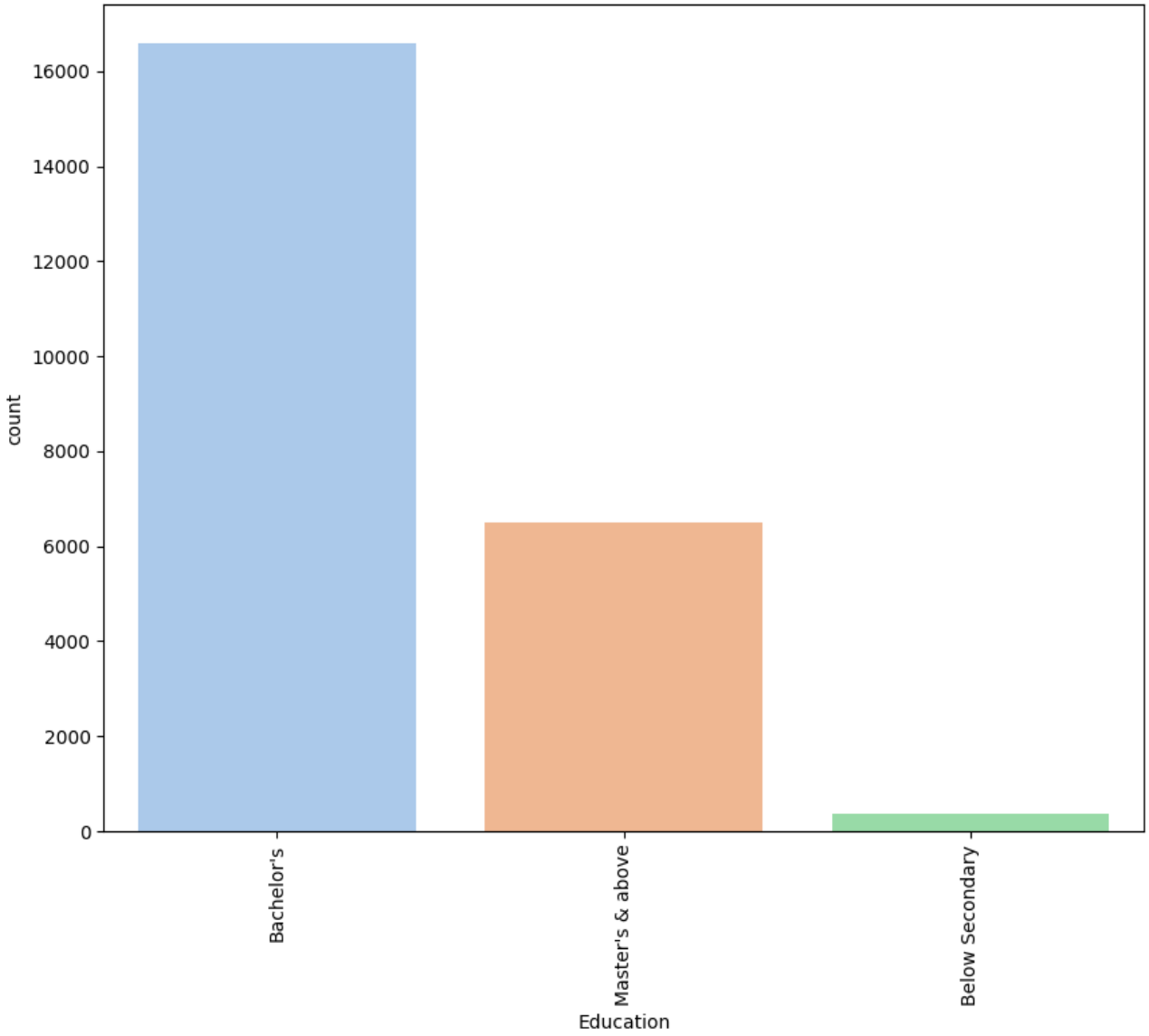
AUC Score was highest for CatBoost, confirming model robustness.



Confusion Matrices

Visual comparison of confusion matrices across models.

CatBoost produced a balanced classification with minimum false positives.



Best Model: CatBoost Classifier

CatBoost outperformed others in all key metrics.

- Handles categorical data natively.
- Robust to overfitting.
- High F1 and ROC AUC score.

Selected as the final model for business recommendations.

Business Recommendations

- Focus training on employees predicted to underperform.
- Consider award history and past performance in evaluation.
- Use model insights to guide promotions and interventions.
- Continuously retrain the model with new data for accuracy.

Thank You

For more information, contact Umar Farooq.

Project submitted as part of UpGrad AI/ML Capstone.