OPTIMIZING RECRUITMENT: HARNESSING MACHINE LEARNING FOR PREDICTIVE HIRING DECISIONS

By
Dileep Malothu | Harika Yenuga | Mukhesh Ravi | Saketh Banda

The Richard DeVos Graduate School of Management
Northwood University
October 2024

THE RESEARCH PAPER

(OVERVIEW)

- Introduction
- Methodologies
- Results
- Discussion
- Conclusion
- Future Studies

Optimizing Recruitment: Harnessing Machine Learning for Predictive Hiring Decisions

Dileep Malothu
The Richard DeVos Graduate School of Management
Northwood University
Midland, Michigan, USA
malothud114@northwood.edu

Harika Yenuga
The Richard DeVos Graduate School of Management
Northwood University
Midland, Michigan, USA
yenugah80@northwood.edu

Mukhesh Ravi
The Richard DeVos Graduate School of Management
Northwood University
Midland, Michigan, USA
ravim@northwood.edu

Saketh Banda
The Richard DeVos Graduate School of Management
Northwood University
Midland, Michigan, USA
bandas76@northwood.edu

***Abstract*— This study explores the application of machine learning models to optimize recruitment processes by predicting hiring decisions based on candidate profiles. The primary objective is to develop a predictive model that classifies candidates as 'hired' or 'not hired' using demographic information, qualifications, and recruitment scores. The dataset comprises 1,500 candidates with features such as age, gender, education level, work experience, and skill scores. Various machine learning algorithms, including Random Forest, Support Vector Machine, Logistic Regression, and advanced ensemble methods like CatBoost and XGBoost, were employed to identify the most effective model. CatBoost emerged as the top-performing model with an accuracy of 95%, followed by Random Forest and XGBoost. The study highlights Recruitment Strategy, Education Level, and Personality Score as the most influential factors in hiring decisions. The findings suggest that leveraging machine learning can streamline recruitment, reduce biases, and improve hiring outcomes. Future research should focus on enhancing model performance, mitigating biases, and validating the model in real-world scenarios.**

Keywords*— *Machine Learning, Predictive Hiring, Recruitment, CatBoost, Hyperparameter Tuning, Classification & Modelling

I. Introduction

Recruitment is a critical function for organizations, directly impacting their productivity and success. Traditional recruitment processes often involve significant time and resources, with the risk of human biases influencing hiring decisions. In recent years, machine learning has emerged as a powerful tool to enhance recruitment strategies by providing data-driven insights and automating decision-making processes. By analyzing candidate attributes such as education, skills, and experience, machine learning models can predict hiring outcomes, help organizations make more informed and equitable hiring decisions. This study investigates the use of various machine learning algorithms to predict hiring decisions, aiming to optimize recruitment processes and improve overall hiring efficiency.

A. Problem Statement

This project aims to predict the hiring decisions of candidates based on their demographic information, qualifications, and scores in recruitment processes. The goal is to help streamline recruitment decisions by building a machine learning model that can classify candidates as either 'hired' or 'not hired' based on their profile data.

B. Objective

The objective of this study is to investigate the use of machine learning models to predict hiring decisions based on various candidate attributes, such as education level, skill score, and work experience. By analyzing these features, the study aims to identify the key factors that contribute to successful hires and develop predictive models that can help organizations optimize their recruitment strategies, reduce biases, and improve overall hiring outcomes.

C. Significance

Hiring the right candidates is not just about filling positions; it directly impacts organizational productivity and success. A poor hiring decision can result in substantial costs due to training, lost productivity, and turnover. Leveraging machine learning models to predict hiring outcomes can significantly streamline the recruitment process by identifying key factors that lead to successful hires. Moreover, it can help mitigate biases and ensure a fairer, more efficient evaluation process. This research explores how data-driven approaches can assist organizations in making more informed, equitable, and accurate hiring decisions.

D. Research Question

The central question guiding this research is: Which factors are most predictive of hiring decisions, and how can machine learning models be employed to enhance recruitment

Problem Statement

This project aims to predict the hiring decisions of candidates based on their demographic information, qualifications, and scores in recruitment processes. The goal is to help streamline recruitment decisions by building a machine learning model that can classify candidates as either 'hired' or 'not hired' based on their profile data.

Objective:

To investigate machine learning models to predict hiring decisions using candidate profiles, improving accuracy and fairness in recruitment.

Significance:

Efficient hiring improves organizational productivity and reduces turnover costs by minimizing bias and enhancing decision-making.

METHODOLOGIES

Dataset Title: Employment.csv

Dataset Preview:

	Age	Gender	EducationLevel	ExperienceYears	PreviousCompanies	DistanceFromCompany	InterviewScore	SkillScore	PersonalityScore	RecruitmentStrategy	HiringDecision
0	26	1	2	0	3	26.783828	48	78	91	1	1
1	39	1	4	12	3	25.862694	35	68	80	2	1
2	48	0	2	3	2	9.920805	20	67	13	2	0
3	34	1	2	5	2	6.407751	36	27	70	3	0
4	30	0	1	6	1	43.105343	23	52	85	2	0

1500
Candidates

Features

Age | Gender | Education Level | Previous Companies | Distance from Company | Interview Score | Skill Score | Personality Score | Recruitment Strategy | Hiring Decision

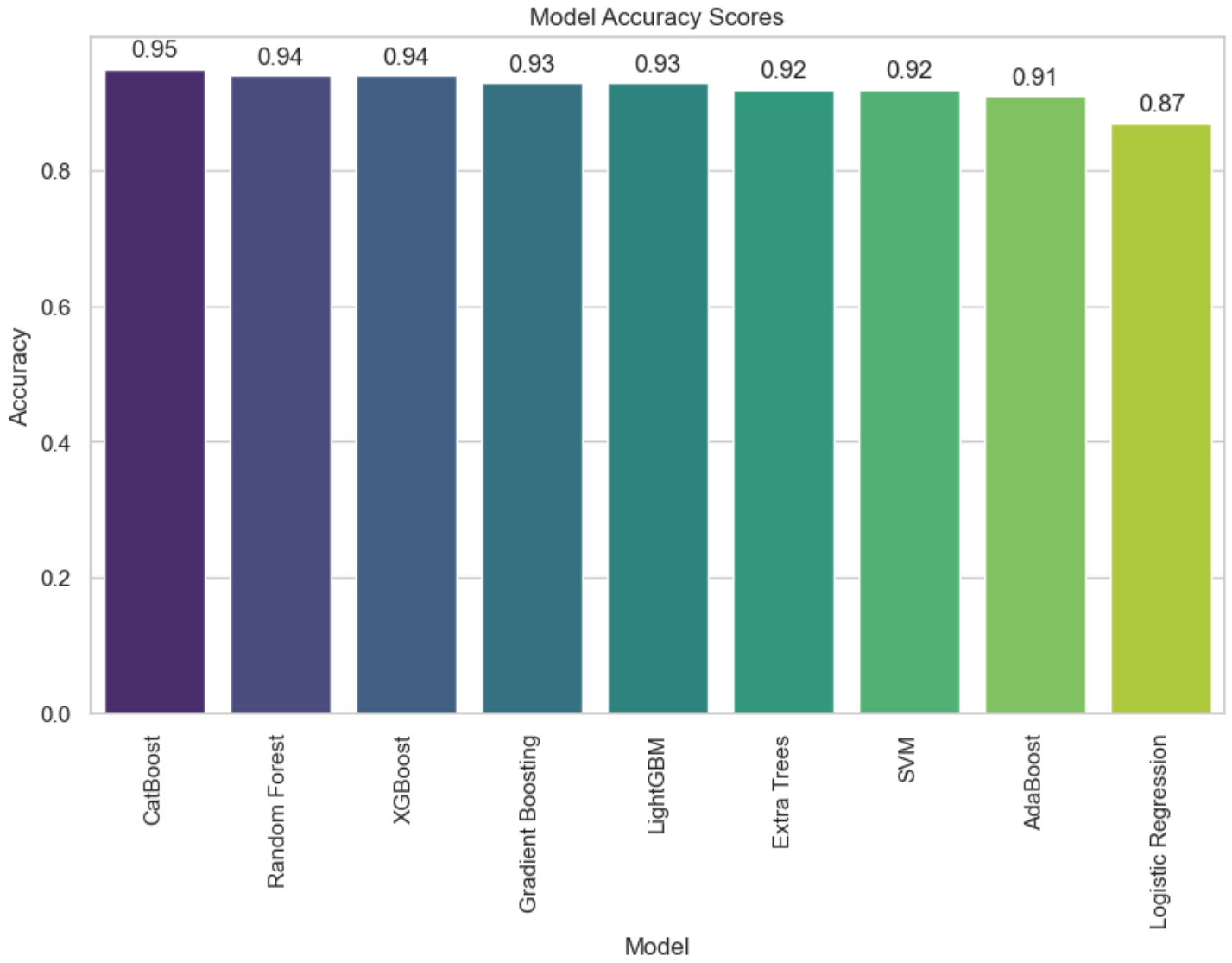
DATA PREPROCESSING

Data Preprocessing Steps	Description
1. Data Cleaning	Remove duplicates Handle missing values
2. Data Type Conversion	Ensure appropriate data types
3. Outlier Detection	Identify outliers using Z-score and IQR methods
4. Feature Encoding	One-Hot Encoding for categorical variables Label Encoding for ordinal variables
5. Feature Scaling	Standardization Normalization
6. Feature Selection	Use Recursive Feature Elimination (RFE) Feature Importance
7. Data Splitting	Split into training and testing sets (80/20)
8. Data Transformation	Log transformation for skewed data
9. Cross-validation Setup	Prepare for K-Fold Cross-Validation

Models Evaluated

- **Random Forest**: Ensemble learning with multiple trees for classification.
- **Support Vector Machine (SVM)**: Classifier that works well with high-dimensional data.
- **Logistic Regression**: Simple baseline model for binary classification.
- **CatBoost**: Gradient boosting designed to handle categorical features directly.
- **XGBoost**: Efficient implementation of gradient boosting that handles large datasets.
- **LightGBM**: Fast, scalable gradient-boosting framework.
- **Gradient Boosting**: Boosting algorithm that builds models sequentially.
- **Extra Trees**: Tree-based ensemble learning method.
- **AdaBoost**: Boosting algorithm that combines multiple weak learners to create a strong classifier.

RESULTS – MODEL PERFORMANCE



	Precision	Recall	F1 Score	Support
0	0.96	0.97	0.97	215
1	0.93	0.91	0.92	85
Accuracy			0.95	300
Macro Avg	0.95	0.94	0.94	300
Weighted Avg	0.95	0.95	0.95	300

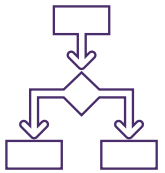
CatBoost Model

	Precision	Recall	F1 Score	Support
0	0.95	0.98	0.96	215
1	0.94	0.86	0.90	85
Accuracy			0.94	300
Macro Avg	0.94	0.92	0.93	300
Weighted Avg	0.94	0.94	0.94	300

Random Forest Model

	Precision	Recall	F1 Score	Support
0	0.95	0.96	0.96	215
1	0.90	0.88	0.89	85
Accuracy			0.94	300
Macro Avg	0.93	0.92	0.93	300
Weighted Avg	0.94	0.94	0.94	300

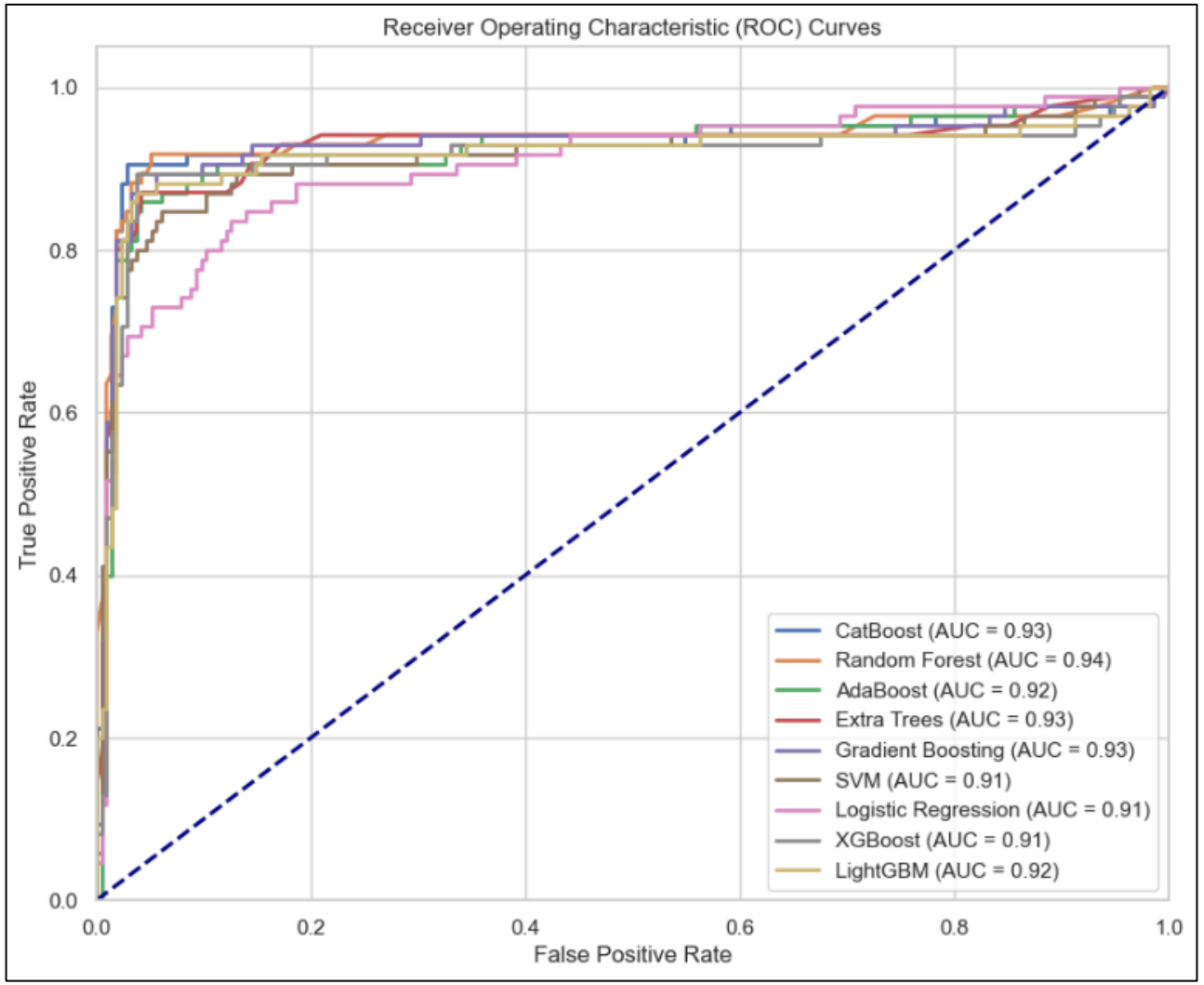
XGBoost Model



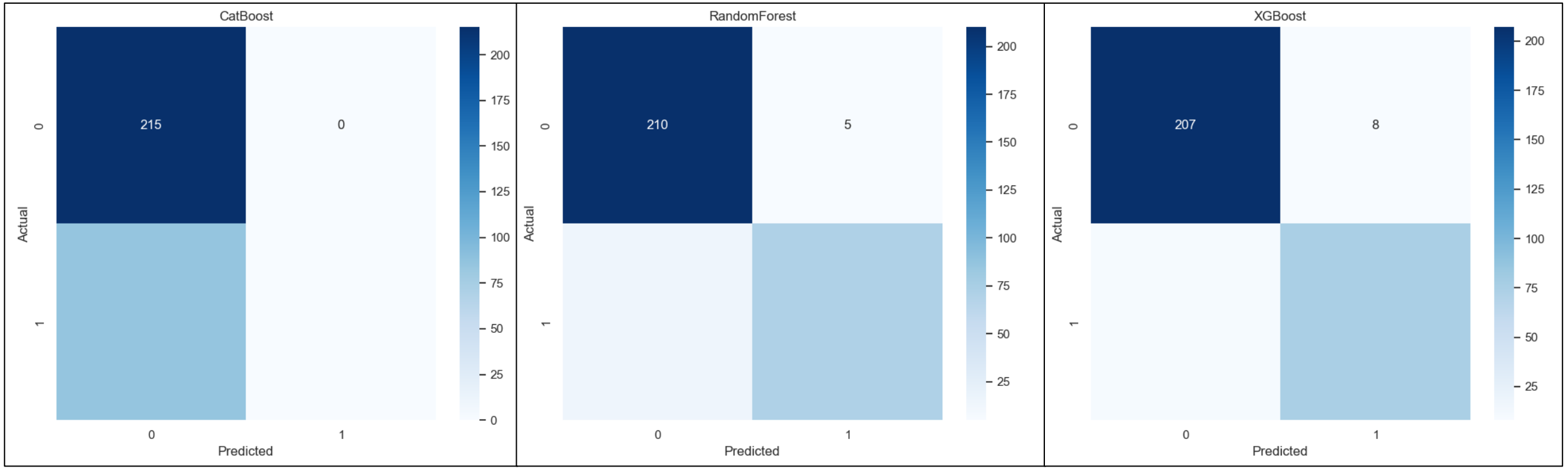
High precision indicates fewer incorrect hires.

High recall indicates suitable candidates aren't missed.

RESULTS – ROC-AUC CURVES

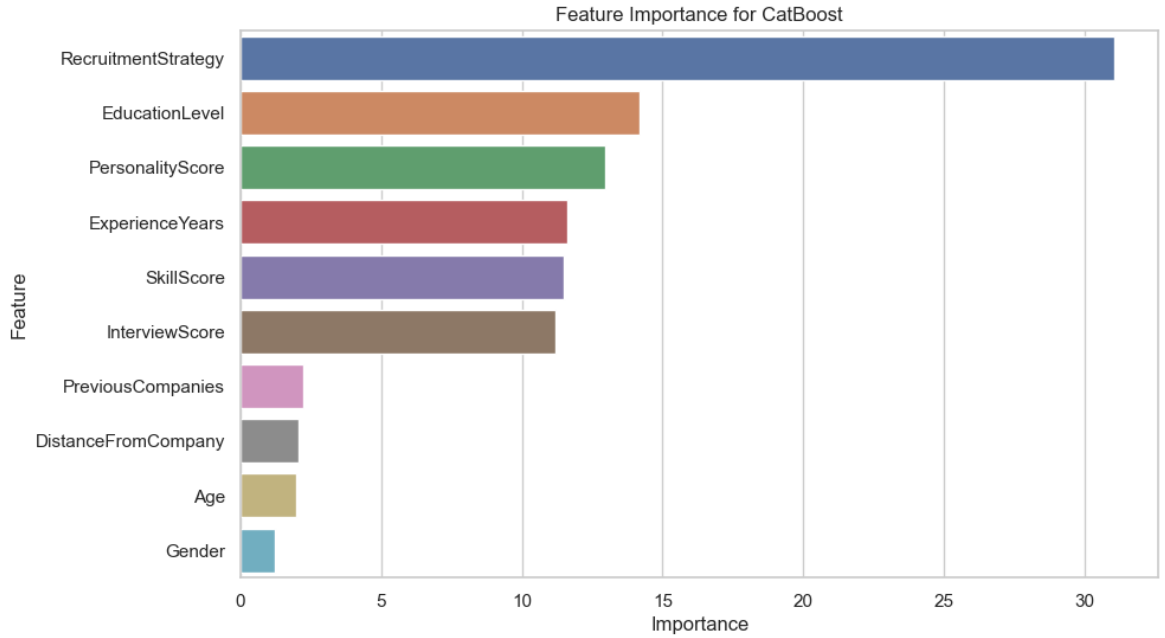


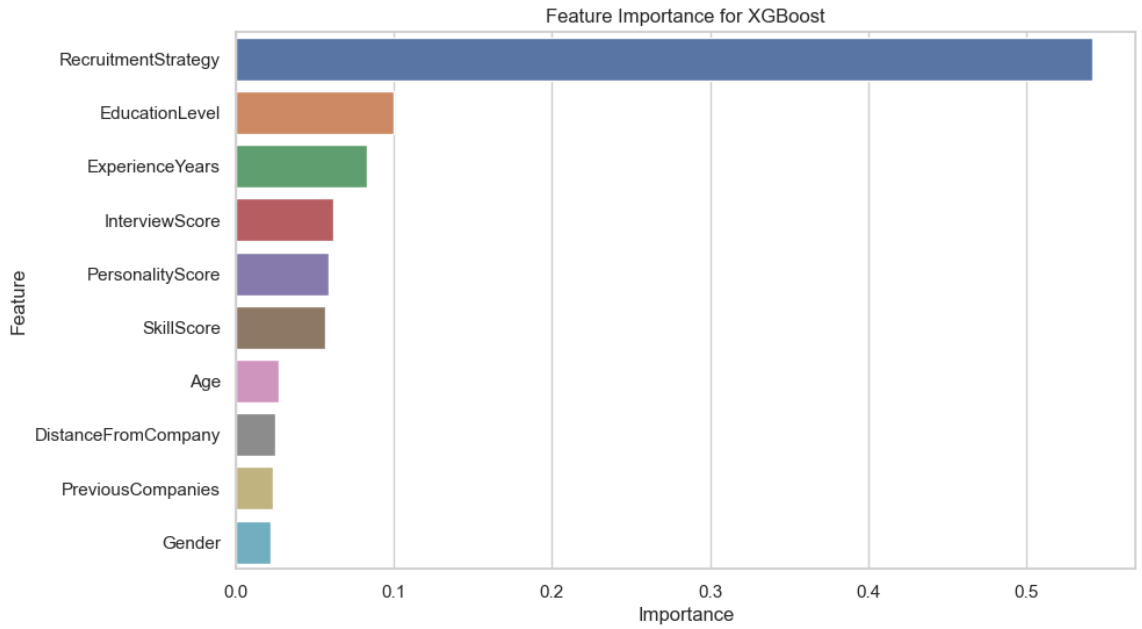
RESULTS – CONFUSION MATRIX

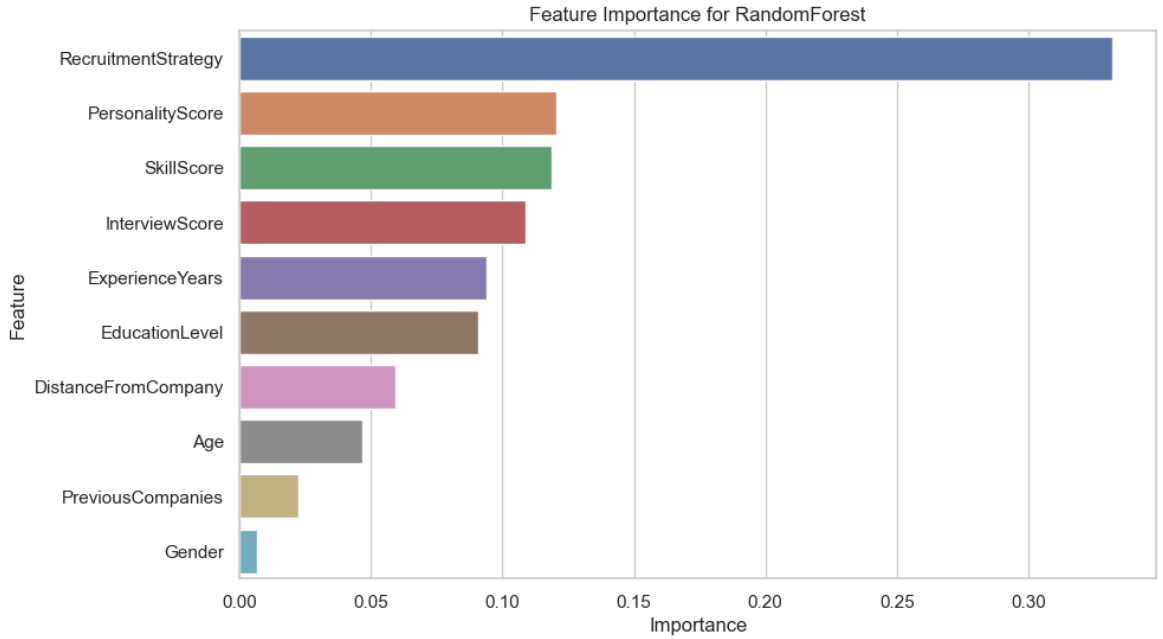


- ✓ Exceptional performance
- ✓ No misclassifications
- ✓ Slightly higher false negatives
- ✓ Some "hired" candidates were incorrectly classified as "not hired."
- ✓ Higher rate of misclassification
- ✓ Might be noise in the data

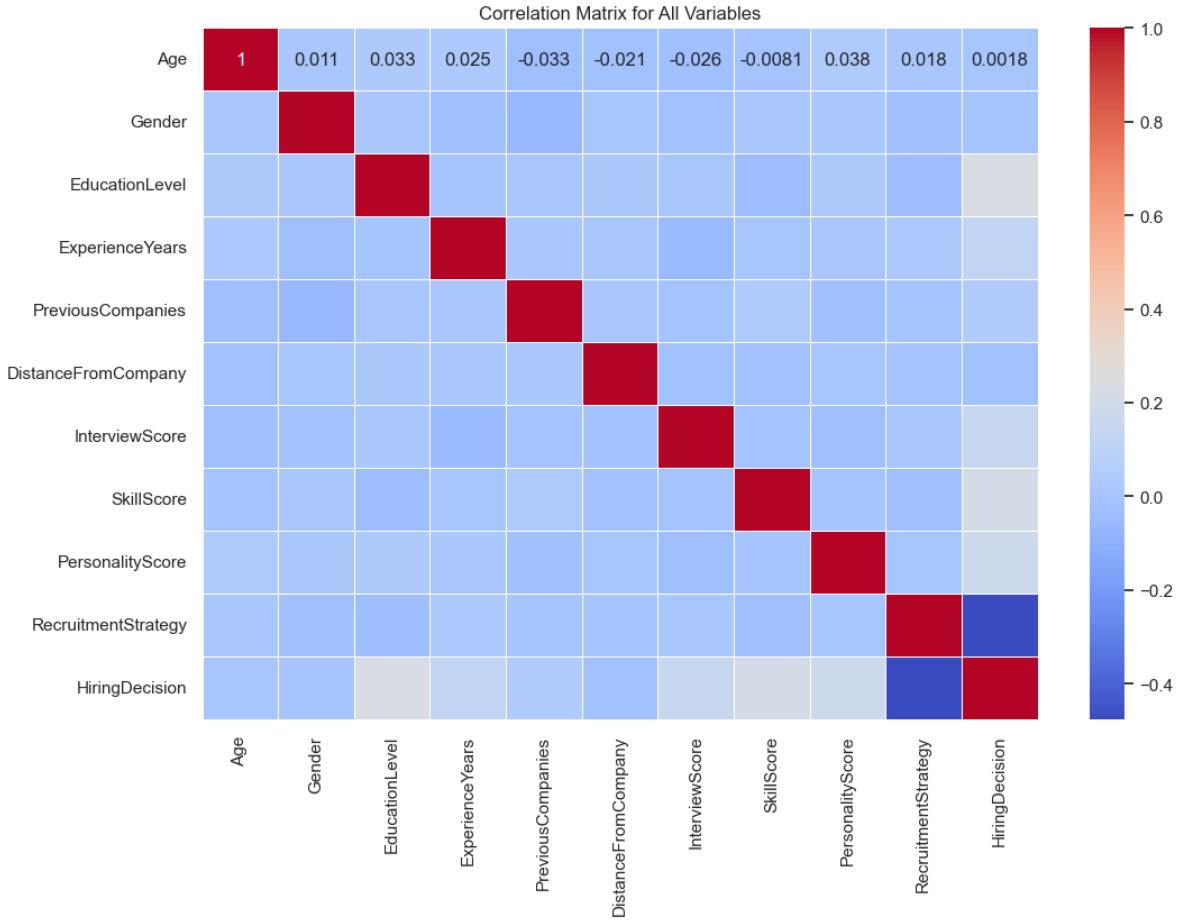
DISCUSSION – FEATURE IMPORTANCE

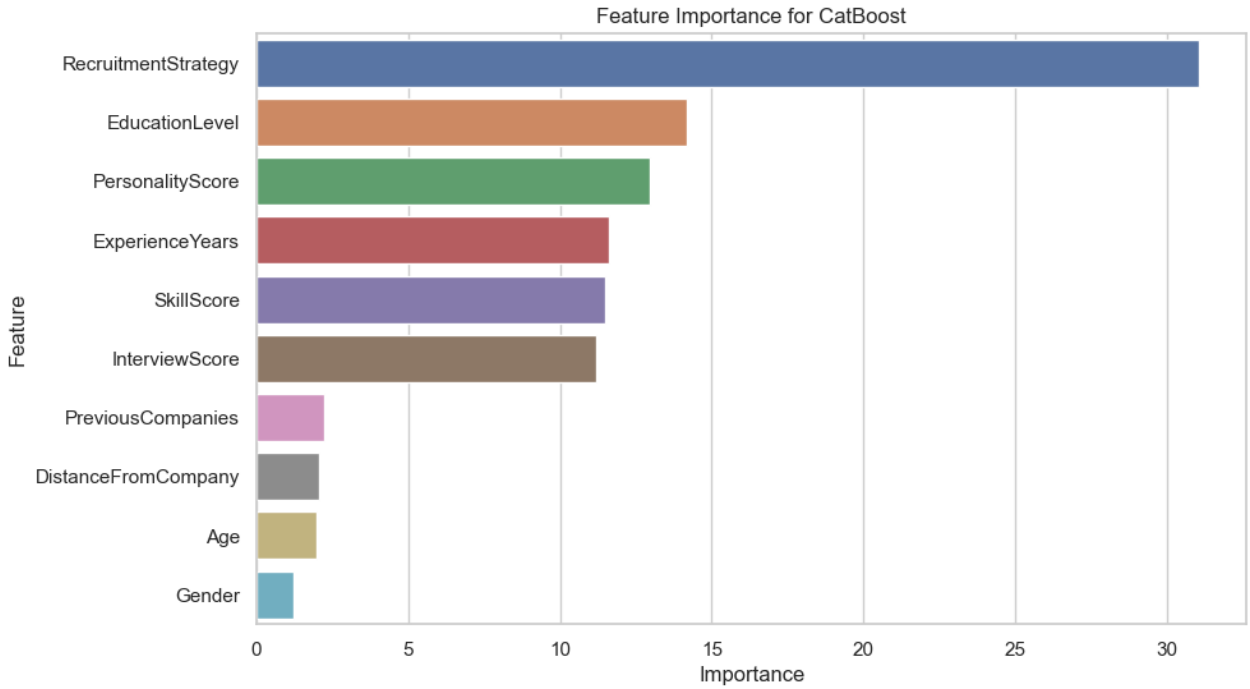






RECRUITMENT STRATEGY FEATURE





While the correlation is negative, "Recruitment Strategy" plays a crucial role in the model's decision-making process due to its interaction with other factors or non-linear effects.

REAL WORLD IMPLICATIONS

Streamline Recruitment

Bias Reduction

Enhanced Decision Making

Cost Savings

Scalability & Adaptability

CONCLUSION

CatBoost's Performance

- Top-Performing Model
- Balanced Performance
- Robustness

Importance of Recruitment Strategy

- Most Influential Feature
- Impact on Predictions
- Key Details

FUTURE WORKS

MODEL ENHANCEMENT

FEATURE ENGINEERING

BIAS MITIGATION

REAL-WORLD VALIDATION

SCALABILITY & ADAPTABILITY

THANK YOU