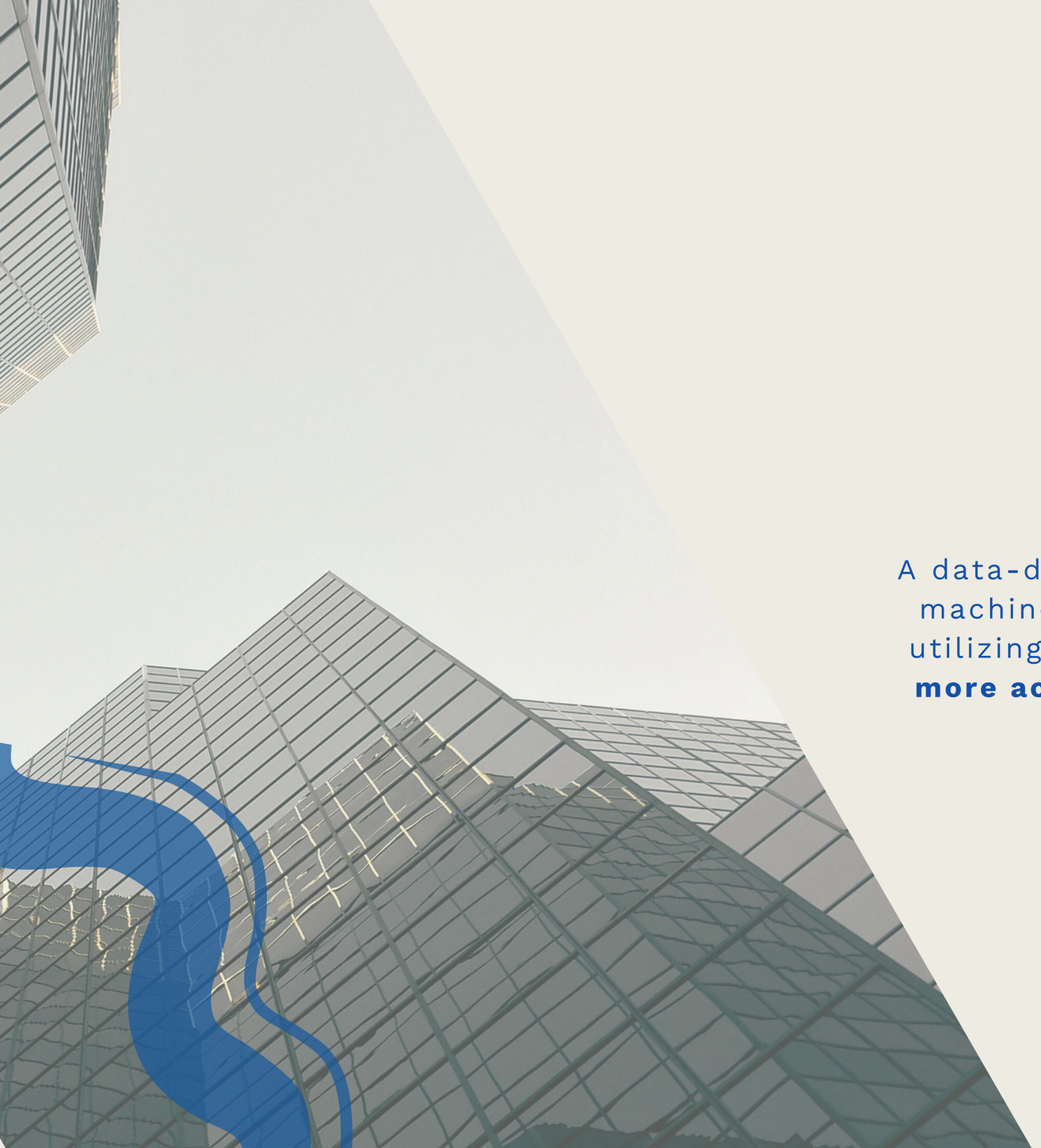
RECRUITMENT PREDICTION

MENTOR : MUHAMMAD HANIF FAJARI

FINAL PROJECT BY HIRESIGHT
DATA SCIENCE BATCH 55
RAKAMIN ACADEMY

AUGUST 9, 2025



HireSight

A data-driven HR consulting firm that leverages machine learning to optimize recruitment. By utilizing historical candidate data, We support **more accurate**, **objective**, and **efficient hiring decisions.**

OUR TEAM

AZIZ PRABOWO

Project Manager

ESA HAERUNISA

Data Scientist

RAYI LENTERA SURYA

Data Analyst

MARISA ADIRATNA K.

Data Engineer

Rakamin Academy HireSight

OUTLINE

- PROBLEM STATEMENT
- EXPLORATORY DATA ANALYSIS & INSIGHT
- MACHINE LEARNING MODEL
- BUSINESS IMPACT
- BUSINESS RECOMMENDATION

Rakamin Academy HireSight

PROBLEM STATEMENT

CURRENT PROBLEM

Most companies are currently facing inefficiencies in its recruitment strategy, which relies heavily on structured assessments and interviews to screen candidates.

1. The recruitment process incurs **high costs and time investment**, with each candidate undergoing structured interviews (~IDR 320K), technical skill tests (~IDR 75–225K), and personality assessments (~IDR 50–150K), leading to **total costs of approximately IDR 450K–700K per candidate**.
2. Despite this investment, **69% of applicants** are ultimately identified as **less suitable hires**, resulting in significant resource waste and highlighting inefficiencies in both early-stage screening and final selection processes.

GOALS & OBJECTIVES

Goals

To enhance the recruitment process using **machine learning models that detect key patterns in candidate success.** These models aim to support more consistent and **objective hiring decisions**, minimize human bias, and streamline candidate evaluation — particularly in the pre- and post-interview stages. Ultimately, we aim to reduce recruitment costs and time-to-hire while ensuring alignment between hires and long-term company goals.

Objectives

Leverage machine learning to optimize candidate screening and support data-informed decisions throughout recruitment.

01 PRE-INTERVIEW MODEL TARGETS

↑ **113%** increase in the success rate of shortlisting suitable candidates from model recommendations.

↓ **70%** reduction in low-fit candidates in shortlist (from 60% to 18%).

2.12× lift in shortlisting rate for model-recommended candidates.

02 POST-INTERVIEW MODEL TARGETS

↑ **184%** increase in the success rate of hires suitable candidates from model recommendations.

↓ **88%** reduction in low-fit hires (from 69% to 8%).

2.80× lift in hire conversion rate for model-selected candidates.

BUSINESS METRICS

REDUCE RECRUITMENT COST

RECRUITMENT CONVERSION RATE

Rakamin Academy HireSight

EXPLORATORY DATA ANAYSIS & INSIGHT

DATA OVERVIEW

Rakamin Academy HireSight

Dataset contains **1.500 rows & 11 columns**

Including **7 numerical features, 3 categorical features, and 1 target variable.**

List of Variabels:

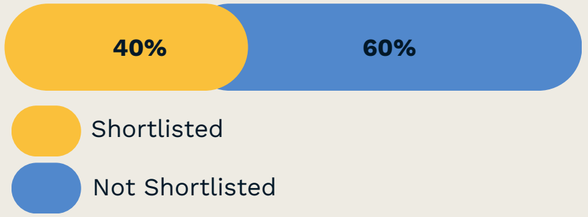
AGE	INTERVIEWSCORE
GENDER	SKILLSCORE
EDUCATIONLEVEL	PERSONALITYSCORE
EXPERIENCEYEARS	RECRUITMENTSTRATEGY
PREVIOUSCOMPANY	HIRINGDECISION
DISTANCEFROMCOMPANY	

Data Quality Insight

- Missing Values: **0%**
- Duplicates: **0%**
- Outliers: **0%** detected by **IQR**

Pre-Interview

Baseline



This label is based on model-driven (Random Forest) feature engineering, using a **60th percentile** cutoff to shortlist the **top 40% of candidates** (602 out of 1500) objectively.

Target Feature

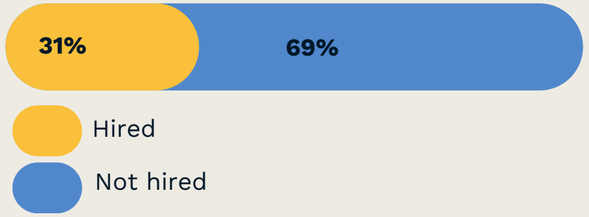
SHORTLISTINGDECISION

SHORTLISTED
The candidate is considered **suitable to proceed to the interview stage** based on initial screening data

NOT SHORTLISTED
Candidate is **filtered out before interview** due to low potential fit

Post-Interview

Baseline



Target Feature

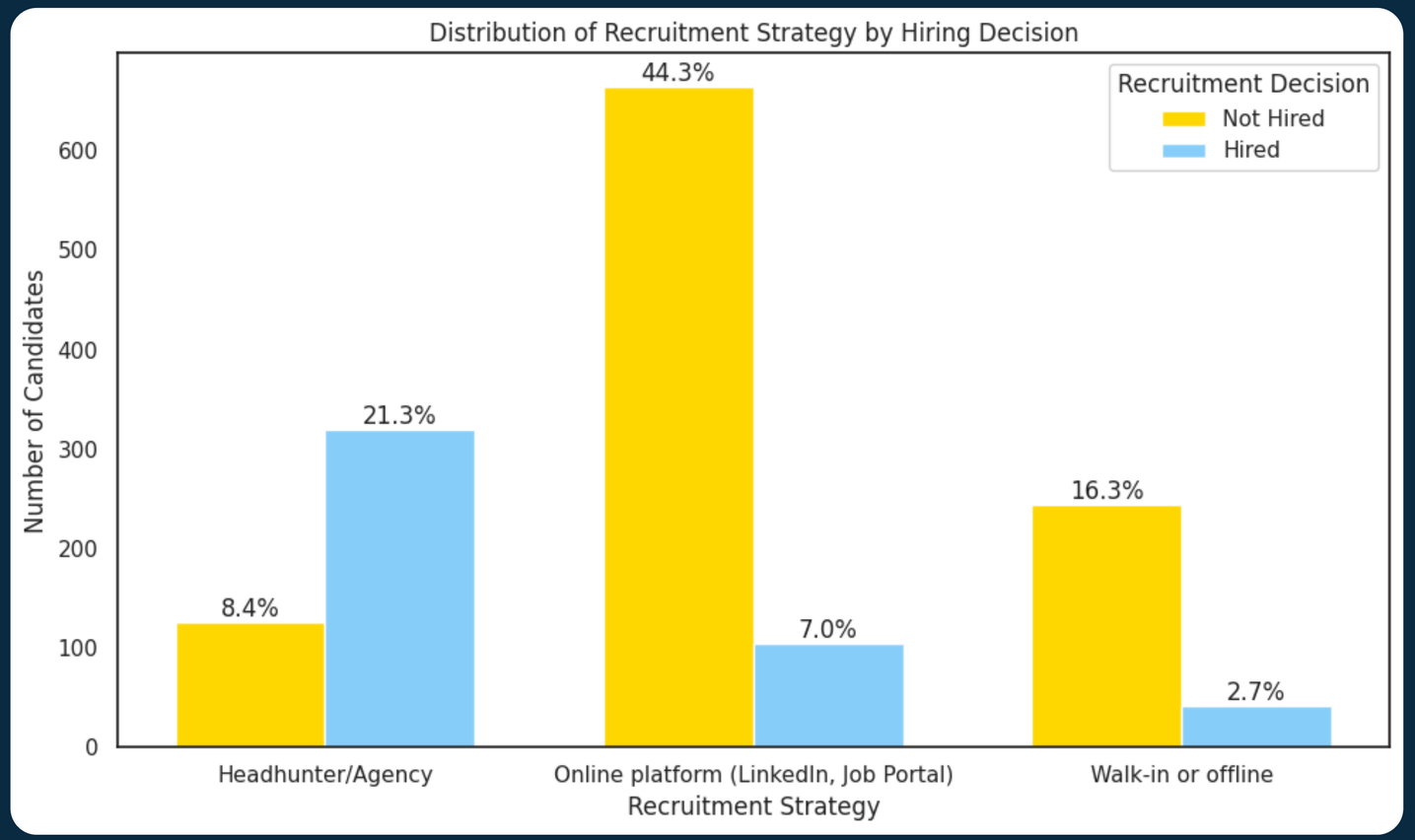
HIRINGDECISION

HIRED
The candidate is officially **accepted** after completing the full selection process

NOT HIRED
Candidate **not approved** for hiring

DATA INSIGHT

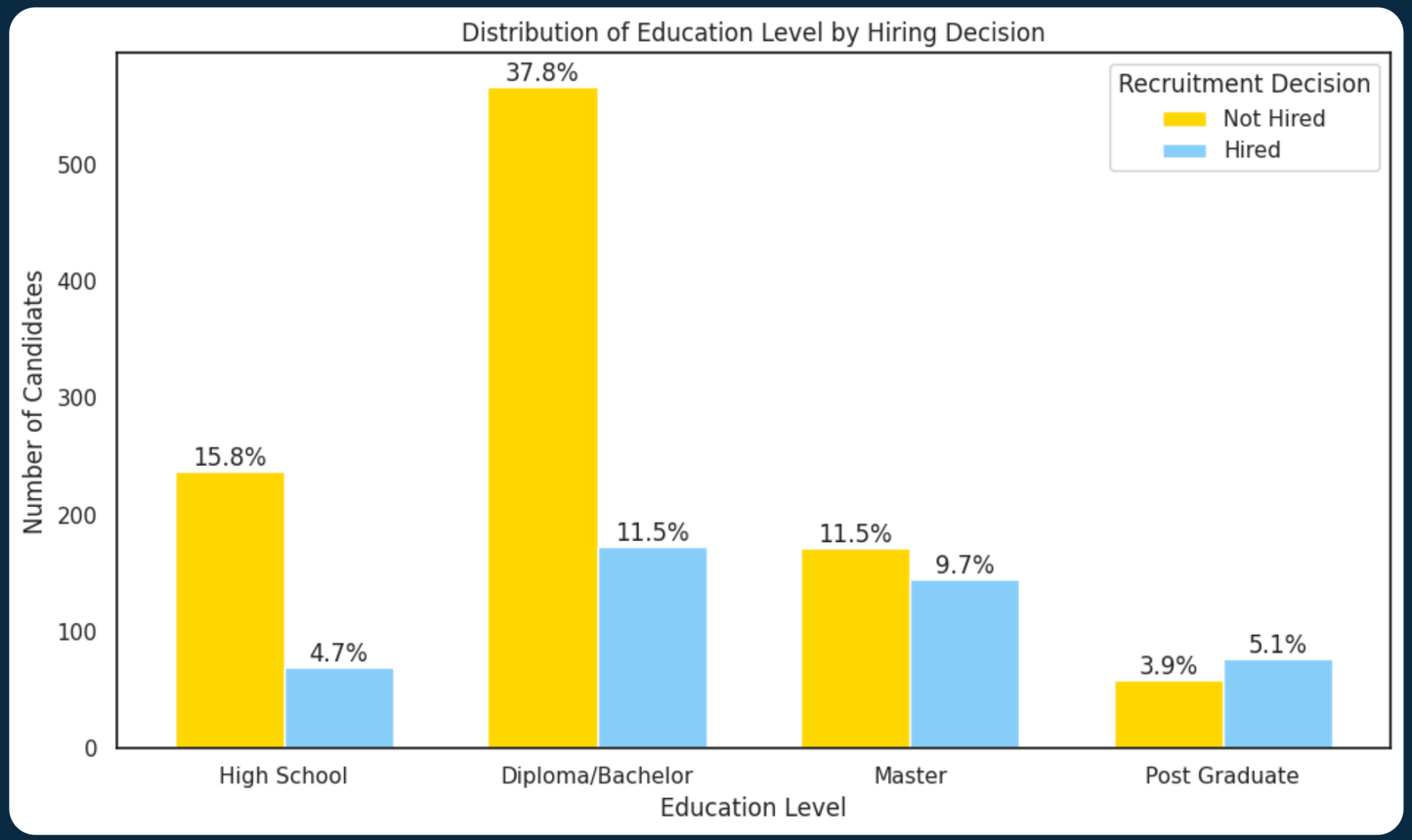
HIGH REJECTION RATES IN CERTAIN RECRUIMENT STRATEGY



Online recruitment platforms account for **51% of total applications**, yet over **44% of those applicants are rejected** at the initial screening stage. This suggests a need to refine both **the sourcing strategy and candidate filtering criteria** to reduce **CV "waste"** and improve recruitment efficiency.

DATA INSIGHT

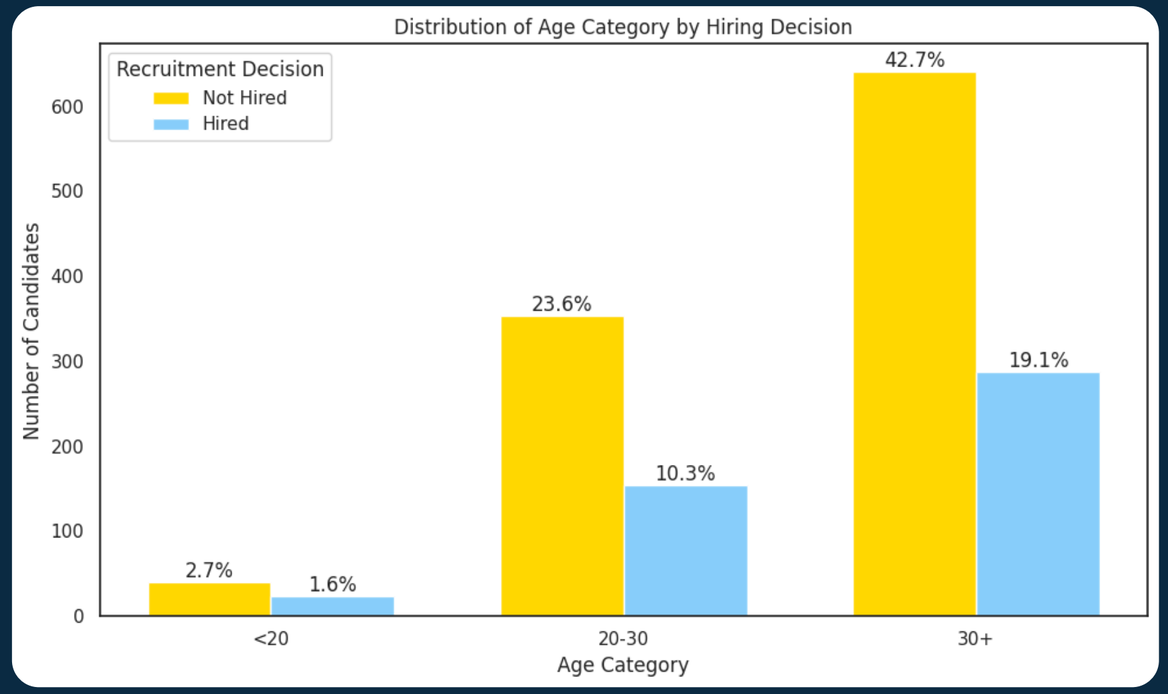
HIGH REJECTION RATES AMONG DIPLOMA/BACHELOR

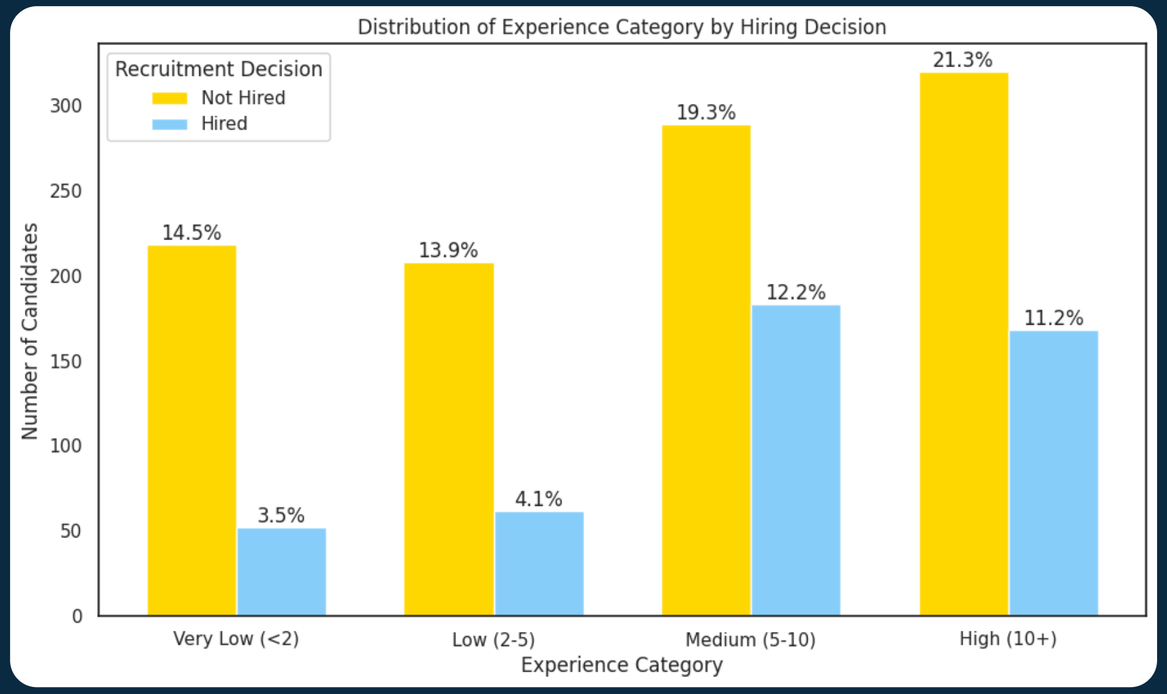


The majority of applicants (**49%**) come from **Diploma and Bachelor degree** backgrounds, but **only 11% of them are hired**, highlighting a mismatch between applicant profiles and job requirements.

DATA INSIGHT

AGED 30+ AND MEDIUM EXPERIENCE DRIVE HIGHER HIRING SUCCESS

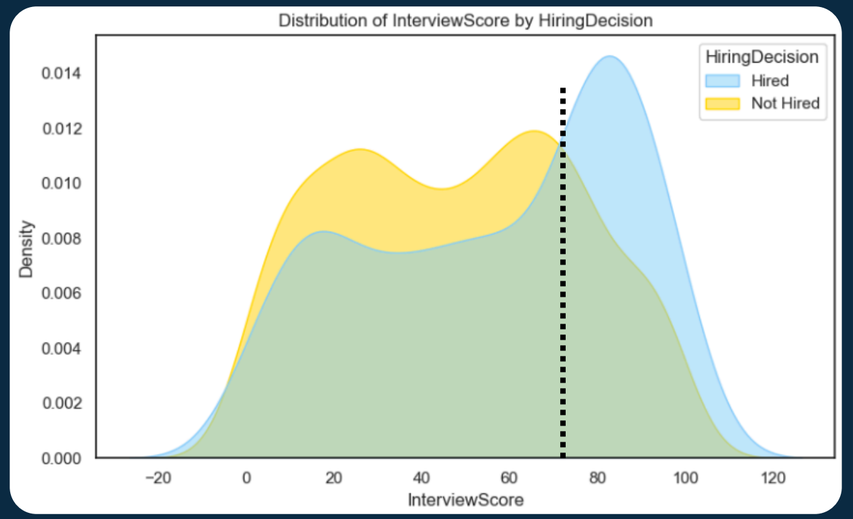


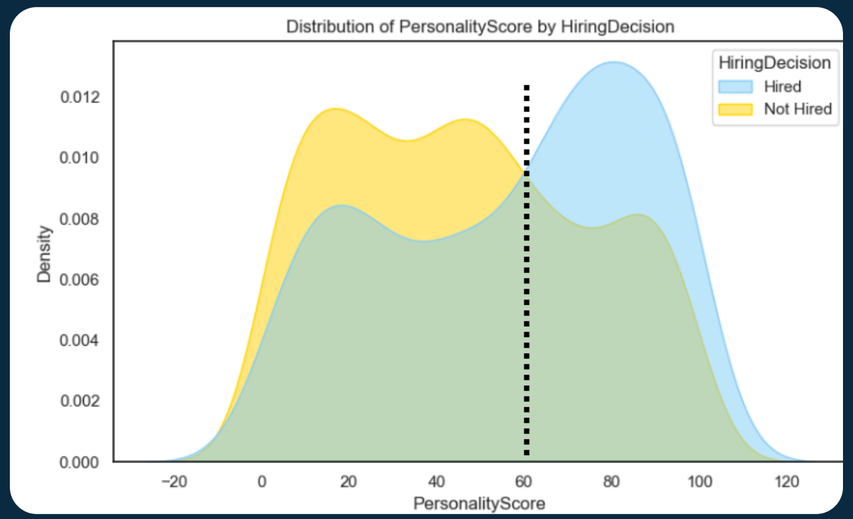


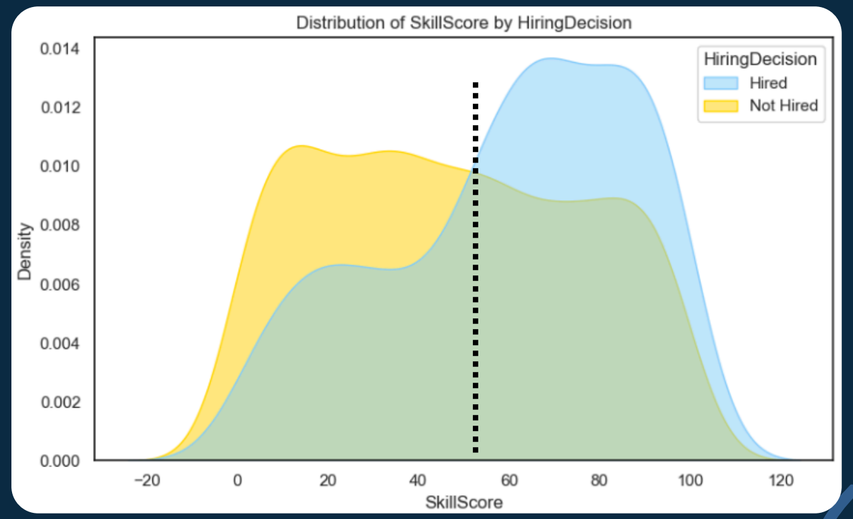
- Candidates **aged 30 years and above** have the **highest acceptance** counts within the age group category, indicating a stronger likelihood of being hired compared to younger applicants.
- Candidates with **Medium (5–10 years) of experience** also **lead in acceptance** numbers within the experience category, outperforming both junior and senior-level candidates.

DATA INSIGHT

HIGH INTERVIEW, PERSONALITY, AND SKILL SCORES STRONGLY PREDICT ACCEPTANCE



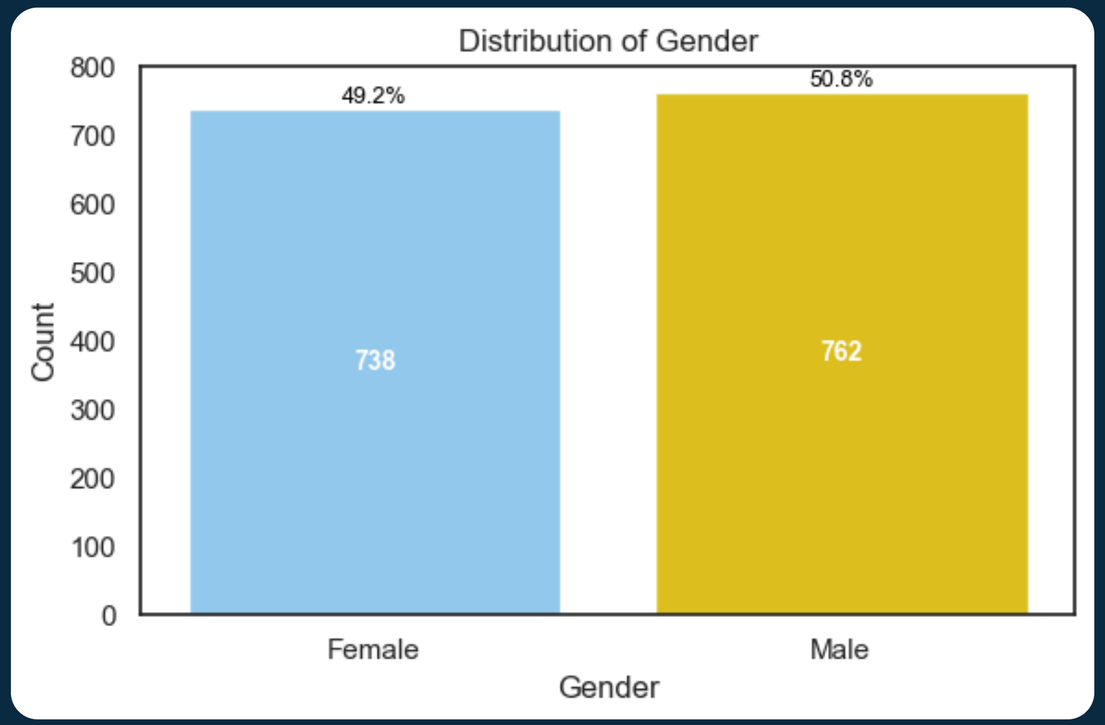


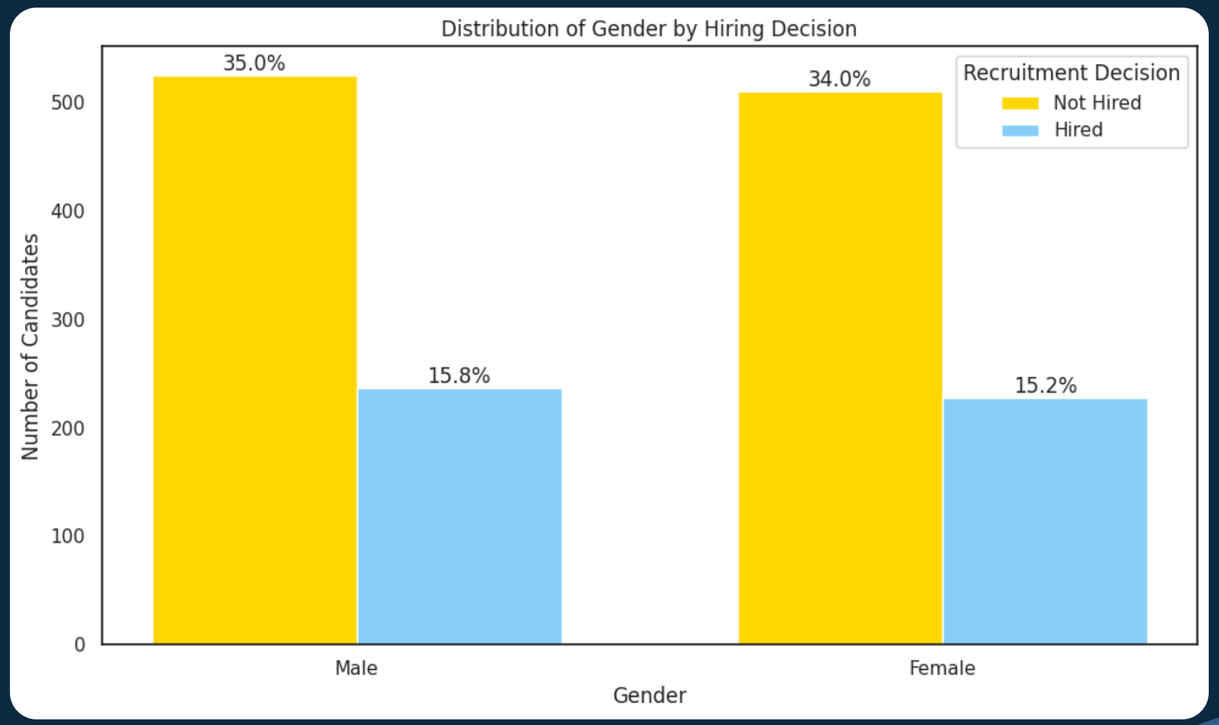


- The majority of accepted applicants have **Interview Scores above 70**, indicating strong performance in structured evaluations.
- **Skills Scores above 50** are common among accepted candidates, reinforcing their practical competency as a key hiring factor.
- **Personality Scores tend to exceed 60** for accepted candidates, suggesting a positive correlation between personality fit and hiring success.

DATA INSIGHT

GENDER-NEUTRAL TRENDS OBSERVED IN APPLICATIONS AND HIRES





Hiring rates are consistent across genders — Male: 15.8%, Female: 15.2%. This indicates **no significant gender disparity** in the recruitment outcome, suggesting that **gender does not influence hiring decisions**.

DATA INSIGHT

SKILLS, PERSONALITY, AND INTERVIEW PERFORMANCE ARE KEY DRIVERS OF HIRING DECISIONS

- **SkillScore** (+0.20): The higher the skill score, the greater the chance of being hired.
- **PersonalityScore** (+0.17): There is a positive impact, but not as strong as Skill. Still important to consider.
- **InterviewScore** (+0.15): Moderate impact. This could be because interview scores are subjective
- **Other features**: Very low correlation (around 0.00–0.12), meaning these features do not have a direct linear effect on hiring decisions. However, they could be useful in non-linear interactions/features.

PRE-PROCESSING

Rakamin Academy HireSight

Feature Encoding

- **One Hot Encoding**
- **Ordinal Encoding**

Split Data Train & Test (80:20)

- Data Train: 1200
- Data Test: 300

Feature Scaling

- **Standard Scaling**

Handling Class Imbalance

- Random Oversampling

MACHINE LEARNING MODEL

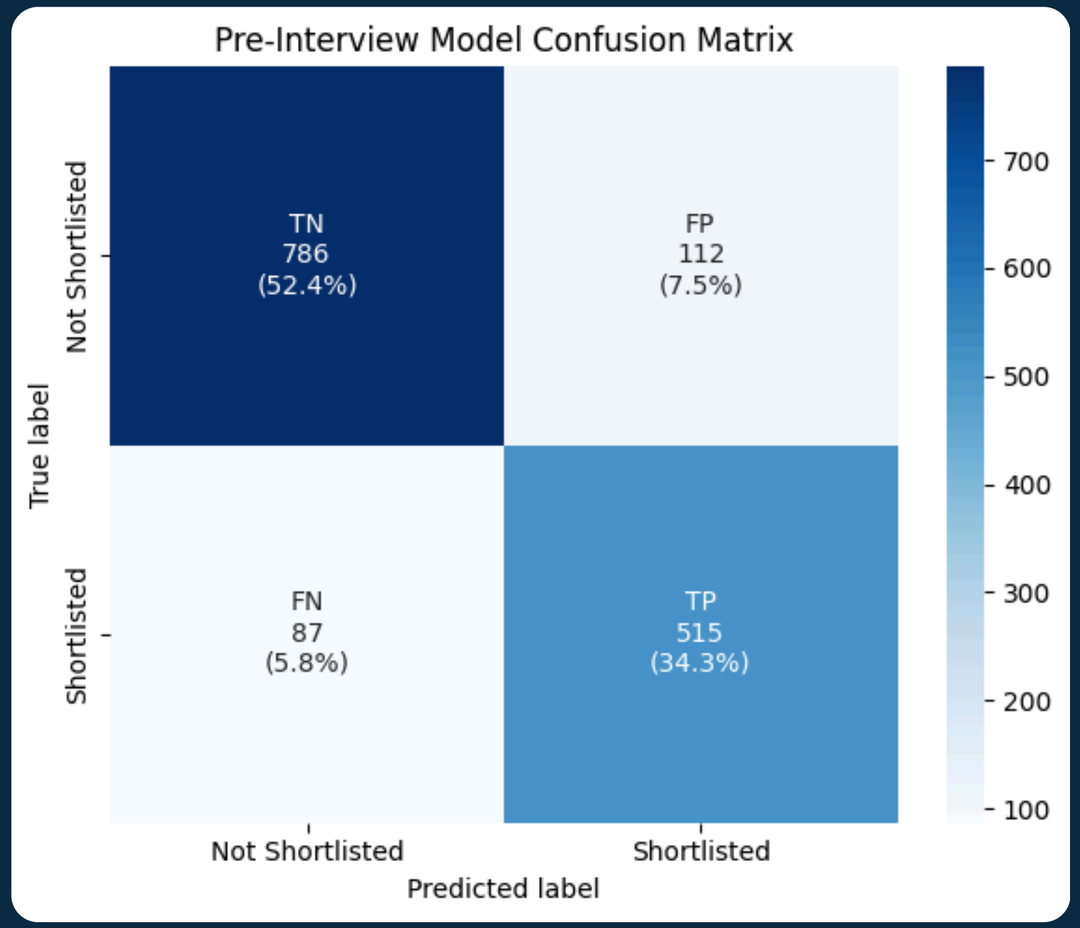
Rakamin Academy HireSight

MACHINE LEARNING MODEL

MODELING RESULT

PRE-INTERVIEW

Model	Training		Testing	
	F1 Score	ROC-AUC	F1 Score	ROC-AUC
DecisionTree	0.7800	0.8159	**0.8136**	**0.8444**
RandomForest	0.8434	0.9377	0.8053	0.8375
XGBoost	0.8504	0.9306	0.7946	0.8292
GradientBoost	0.8368	0.9248	0.7932	0.8278
SVM	0.8409	0.9046	0.7706	0.8097
KNN	0.7964	0.8961	0.7577	0.8000
LogReg	0.8239	0.9118	0.754	0.7972

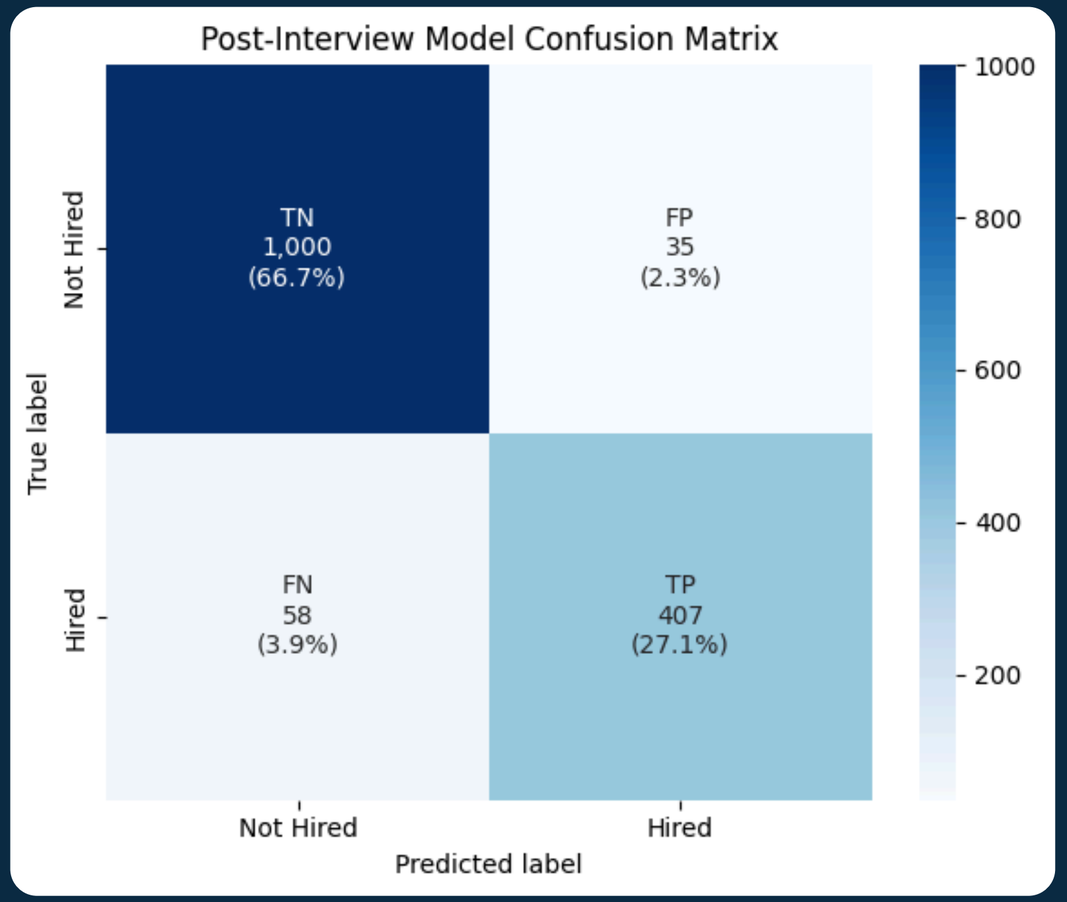


The Decision Tree model achieved the best performance with a **F1 Score of 0.8136** and **ROC-AUC of 0.8444** on the testing set after hyperparameter tuning.

MODELING RESULT

POST-INTERVIEW

Model	Training		Testing	
	F1 Score	ROC-AUC	F1 Score	ROC-AUC
DecisionTree	0.7862	0.8505	0.8511	0.8939
RandomForest	0.8495	0.9234	0.8778	0.9054
XGBoost	0.8945	0.9331	**0.8989**	**0.9180**
GradientBoost	0.8520	0.9293	0.8602	0.8987
SVM	0.7908	0.9072	0.7978	0.8514
KNN	0.6965	0.8470	0.6762	0.7706
LogReg	0.7734	0.9095	0.7897	0.8536



The XGBoost model achieved the best performance with a **F1 Score of 0.8989** and **ROC-AUC of 0.9180** using **default parameters** on the testing set.

MODELING EVALUATION

Rakamin Academy HireSight

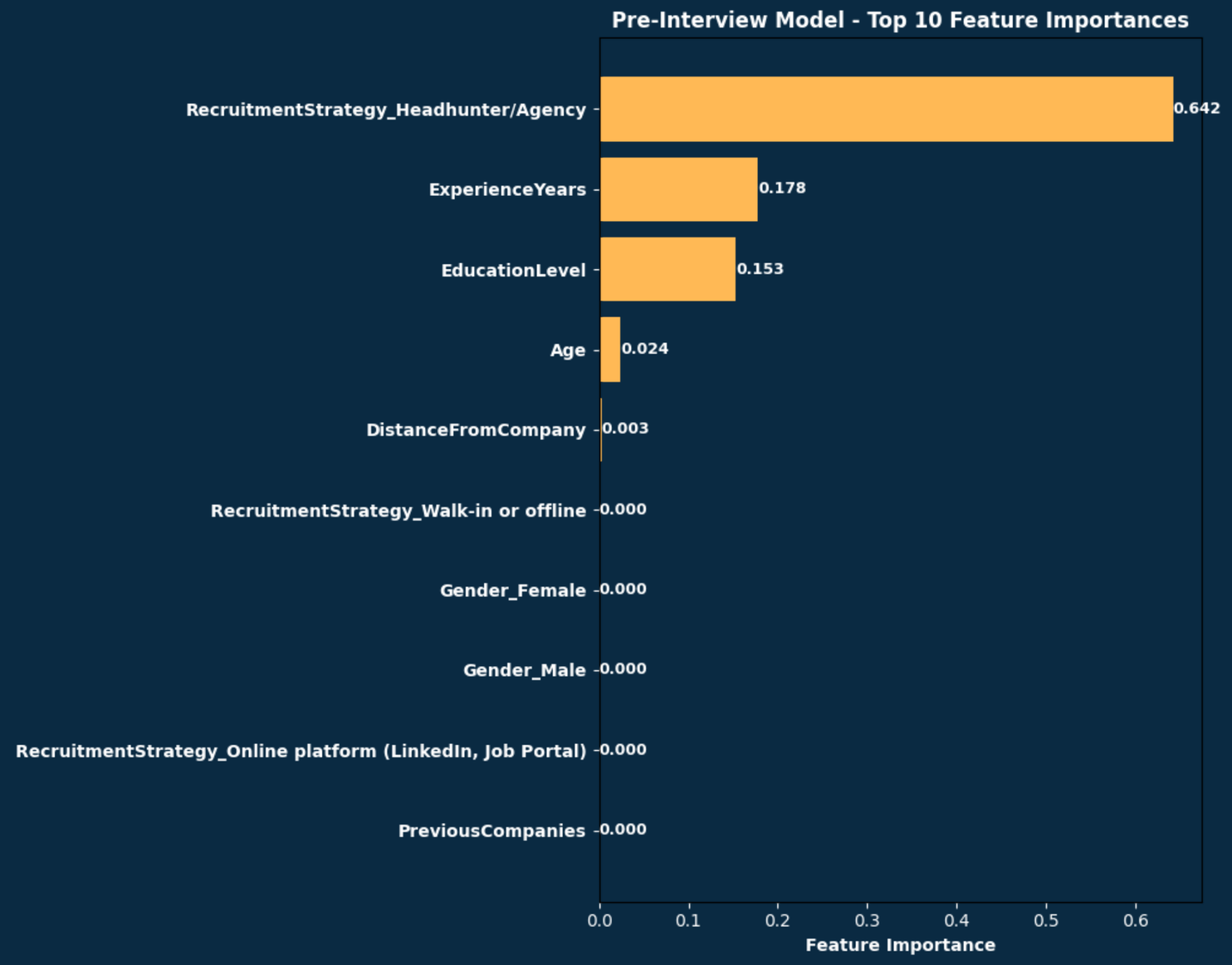
Pre-Interview Model — Best: Decision Tree

- F1 Score: 81% | Precision: 82% | Recall: 85% | ROC-AUC: 84%
- Confusion Matrix: FP: 112 (7.5%) | FN: 87 (5.8%)
- Good for initial filtering: Low FN helps retain more potential talent.
- Acceptable trade-off: FP still tolerable if cost impact remains low.

Post-Interview Model — Best: XGBoost

- F1 Score: 90% | Precision: 92% | Recall: 88% | ROC-AUC: 92%
- Confusion Matrix: FP: 35 (2.3%) | FN: 58 (3.9%)
- Good for final decision: Low FP reduces risk of hiring mistakes.
- Trade-off: Missing a few good candidates (FN) is acceptable when post-hiring quality is maintained.

FEATURE IMPORTANCE

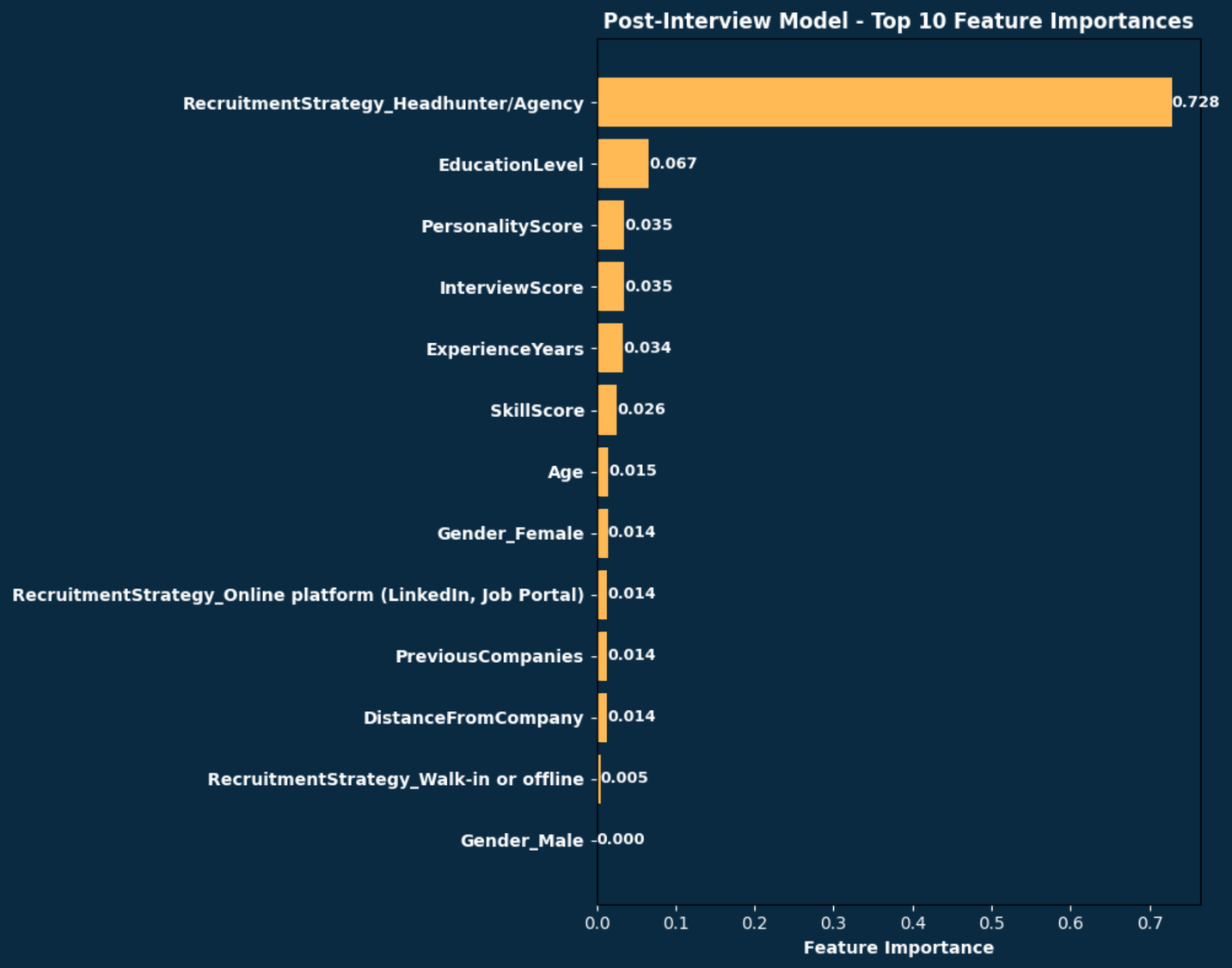


PRE-INTERVIEW MODEL

Headhunter/Agency Recruitment Strategy is by far the most influential predictor accounting for **nearly 64% of the model's predictive power**, followed by **ExperienceYears** and **EducationLevel**, while several features contribute little to no predictive power.

Business Recommendation

FEATURE IMPORTANCE



POST-INTERVIEW MODEL

Headhunter/Agency Recruitment Strategy dominating as the most influential factor, while other features like **EducationLevel**, **PersonalityScore**, and **InterviewScore** contribute moderately, and the remaining features have minimal impact.

↓

Business Recommendation

MODEL USAGE SCENARIO

Rakamin Academy HireSight

Data Collection

HR collect data from each recruitment strategy (online portal/agency/walk-in, etc)

Prediction & Output

Implementation of Pre-interview & Post-interview model prediction

Risk Review by HR

HR filters suitable candidates and reviews key drivers behind their predicted risk

Intervention & Action

HR makes policy adjustments if necessary

Monitoring & Feedback

Actual recommendation outcomes are tracked, feedback used for evaluation

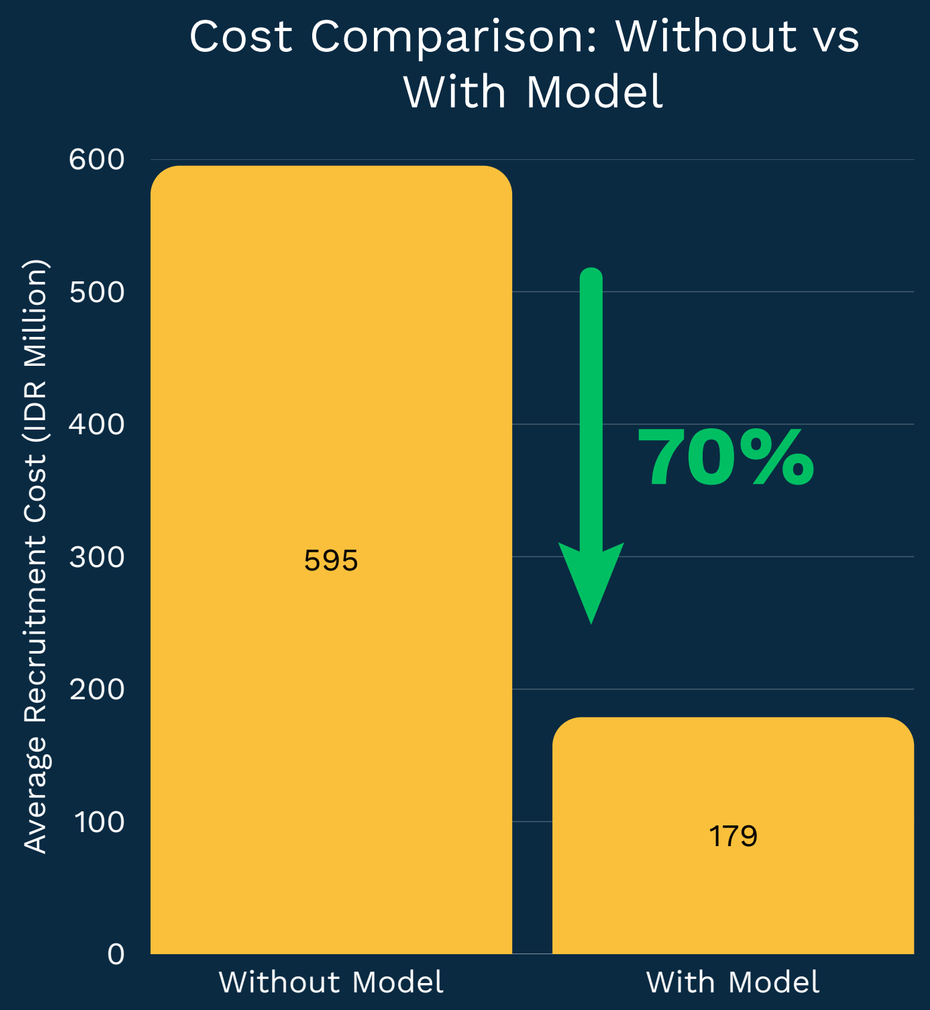
Model Maintenance

Retraining is done periodically or based on model performance degradation

Rakamin Academy HireSight

BUSINESS IMPACT

FINANCIAL IMPACT ANALYSIS



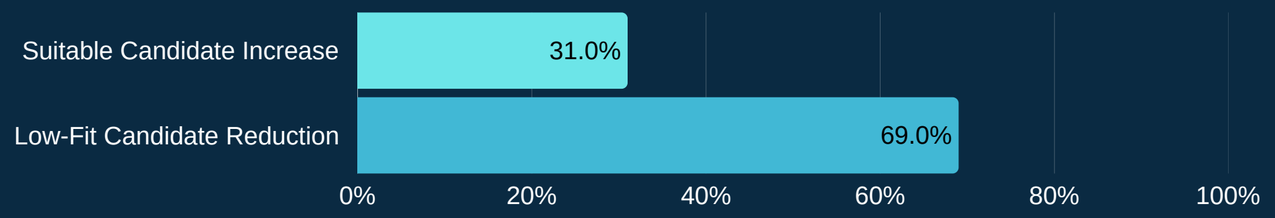
IDR 416.3 M savings

(1.500 candidates)

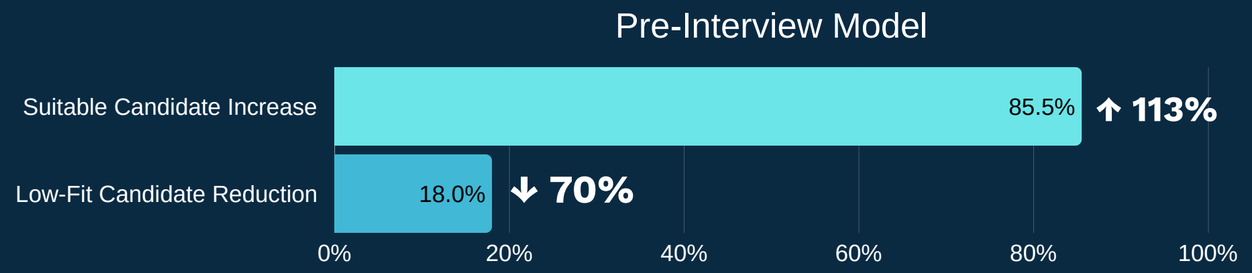
- 70% Cost Reduction
- Significant improvement in hire quality
- Data-driven process reduce bias & inefficiency

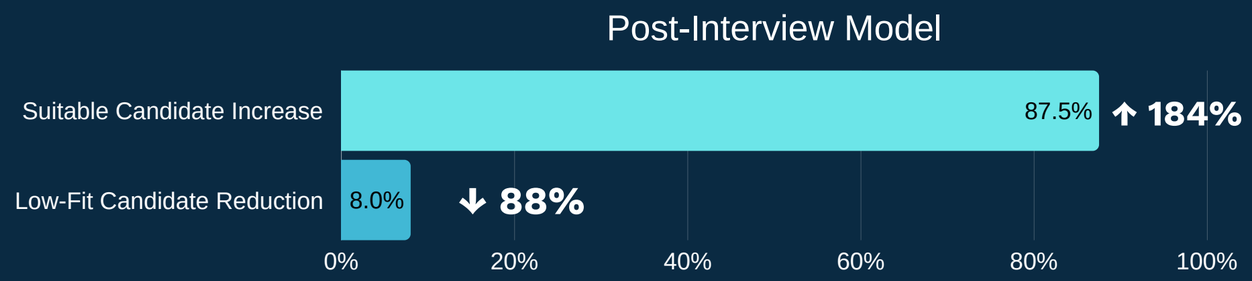
CONVERSION RATE ANALYSIS

Without Model

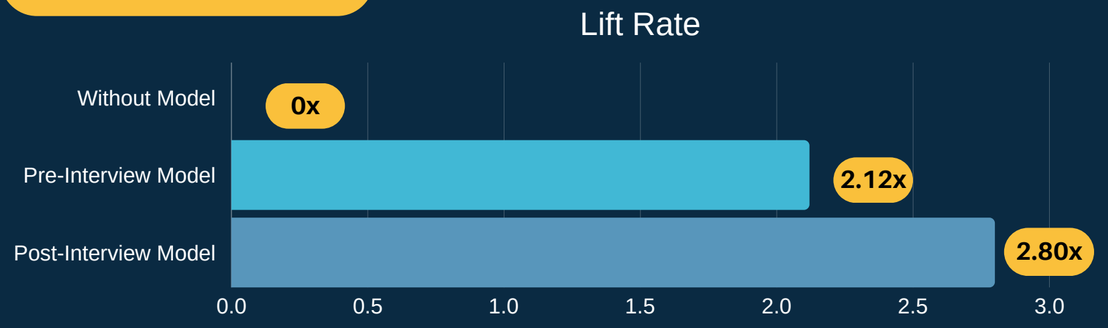


With Model





Lift Rate



Summary

Conversion Rate	Without Model	With Model	
		Pre-Interview	Post-Interview
Suitable Candidate	31%	85.5%	87.5%
Low-Fit Candidate	69%	18%	8%
Lift Rate	0x	2.12x	2.80x

Rakamin Academy HireSight

BUSINESS RECOMMENDATION

BUSINESS RECOMMENDATION

Rakamin Academy HireSight

Impact Estimation

Audit Headhunter/ Agency Partnerships

- **Review historical data** for each agency partner
- Terminate partnerships with agencies **consistently producing rejected candidates**
- Negotiate performance-based contracts with **top-performing agencies**

→

- Helps **identify underperforming partners** and **improve candidate quality**.
- Removing agencies with high rejection rates **reduces the number of low-fit candidates**.
- Shifting to performance-based contracts incentivizes agencies to submit **higher-quality candidates**.

Address Experience Bias Systematically

- Create **a Junior Fast-Track** by moving strong junior candidates with high test scores into a separate hiring path
- Use **skills over experience** by giving more weight to assessment results in jobs where experience matters less
- **Adjust scoring for junior roles** by lower the penalty for low experience in jobs meant for early-career talent

→

- Enables faster hiring of high-potential junior candidates, **increasing fill rate for entry-level roles**.
- Ensures strong candidates are not overlooked due to lack of tenure, **improving overall candidate suitability**.
- **Increases fairness** and expands the qualified talent pool for early-career positions.

THANK YOU

Github Repository: [**click here**]

Streamlit: [**click here**]

Dataset: [**click here**]

Laporan Final Project: [**click here**]