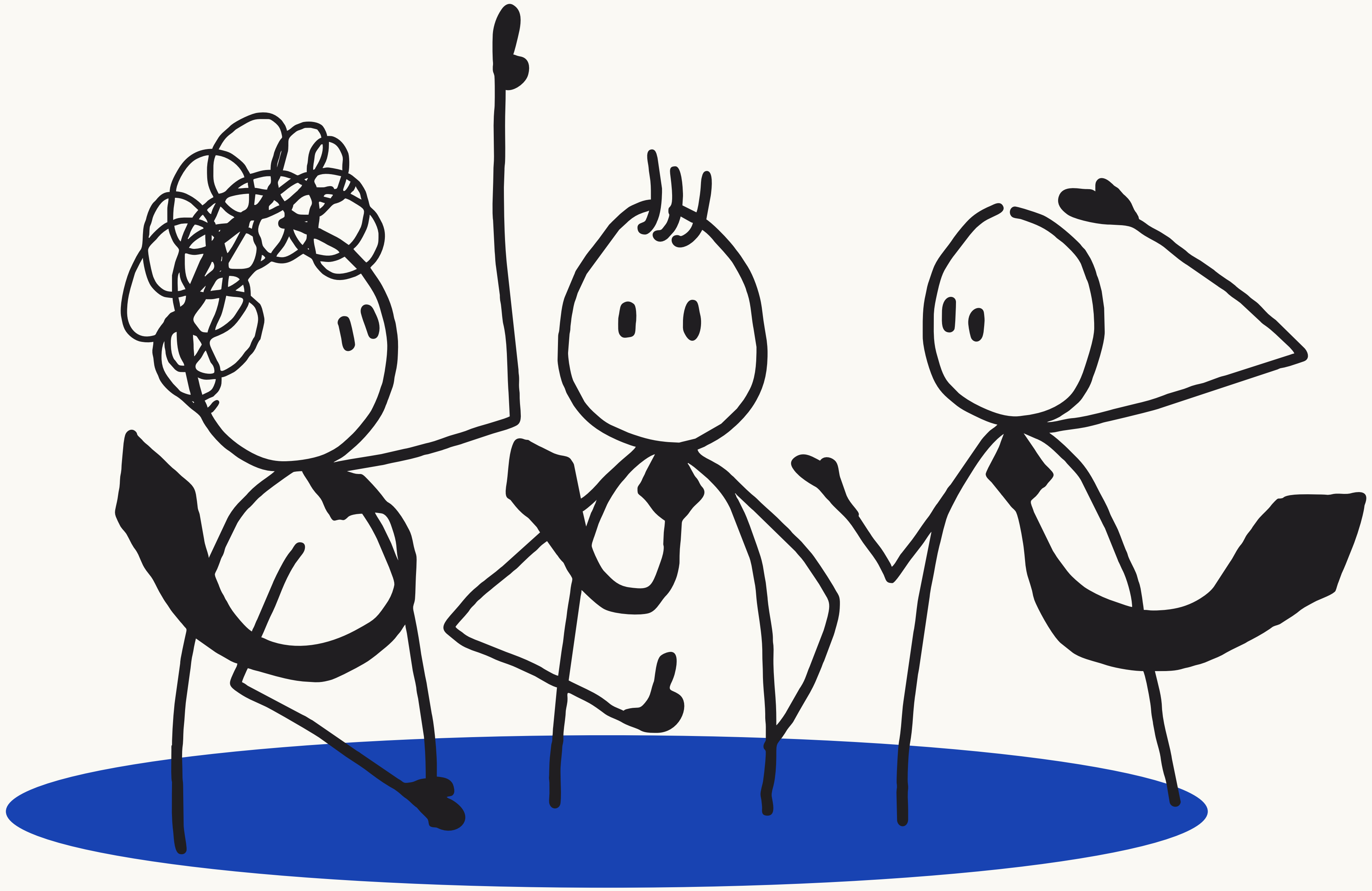
Cracking the "**Efficiency Paradox**" in Talent Acquisition

Rakamin

Optimizing Cost, Speed, and Quality through Predictive



TriMedian

Presented by **TriMedian**
Future Data Scientist

Meet Our Team

Project Manager &
Business Analyst
Naufal Althaf R

Business Analyst
Hutomo Dwiki L

Data Engineer
Rifki Muhammad F

Data Scientist
Kevin Medika P

The Background

When More Investment Delivers Less Results

Financial Perfomance

Quality hires increase productivity, **poor decisions cost 30%** of annual salary.

Operational Speed

Fast, effective hiring maintains momentum, vacancies slow output and morale.

Strategic Advantage

Right talent = competitive edge. Efficient process = stronger employer brand.

$4.700
Average Cost per Hire
(SHRM)

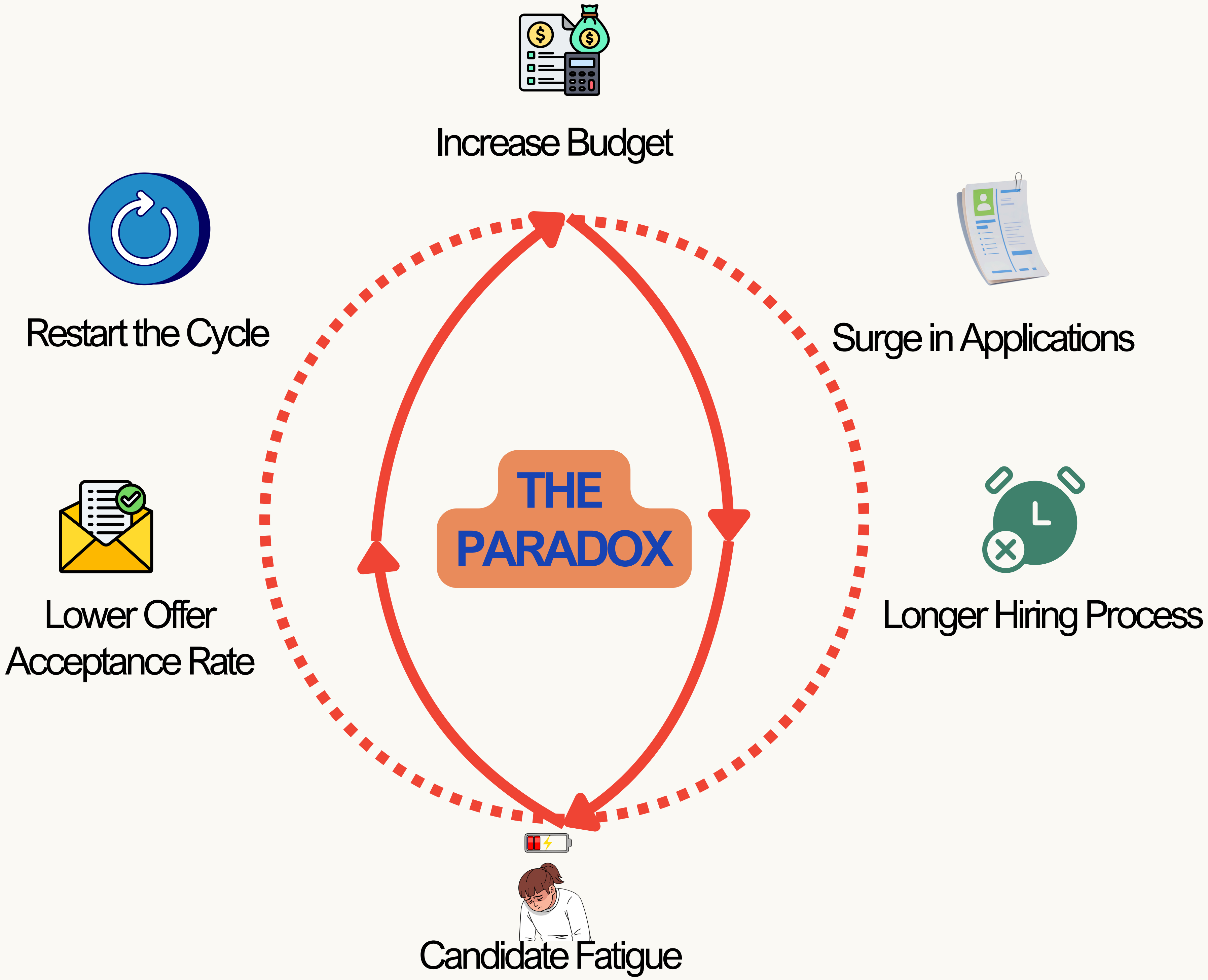
44 Days
Global Time to Hire
(HeroHunt)

58%
Decline Due to Poor Experience
(HeroHunt)

91%
Need Positive Experience to Accept
(CareerPlug)

When recruitment works well, companies grow. When it fails, costs spiral.

The Efficiency Paradox:
When More Effort Yields Less Results

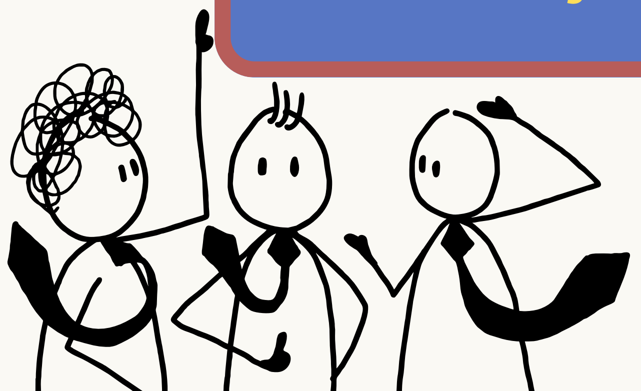


The Paradox:

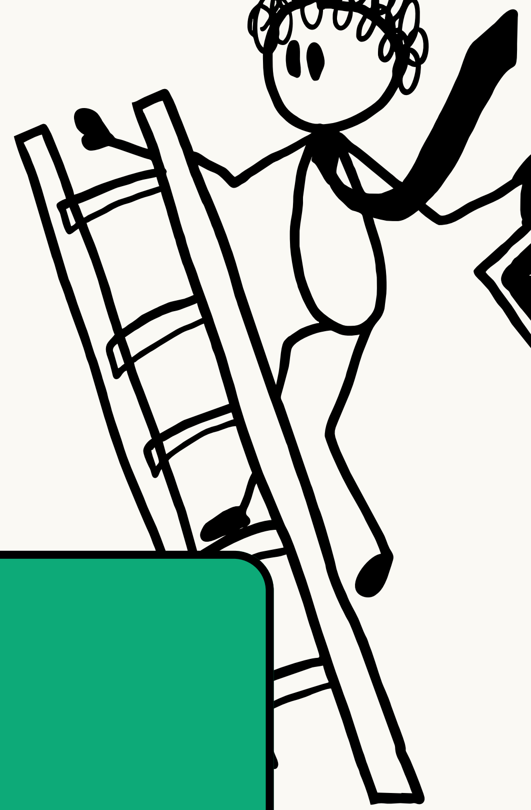
Companies invest **MORE** in recruitment, but see **FEWER quality** hires accepting offers.

Our Data Shows:

156 applicants per role. 47 days to hire.
Despite scale and effort, conversion efficiency remains capped at ~65% — translating into **~USD 9.1M (estimated) of structurally inefficient recruitment spend annually**.



How We Measure Victory



BUSINESS IMPACT

<15%

Reduction in agency spending through channel optimization

MODEL PERFORMANCE

>80%

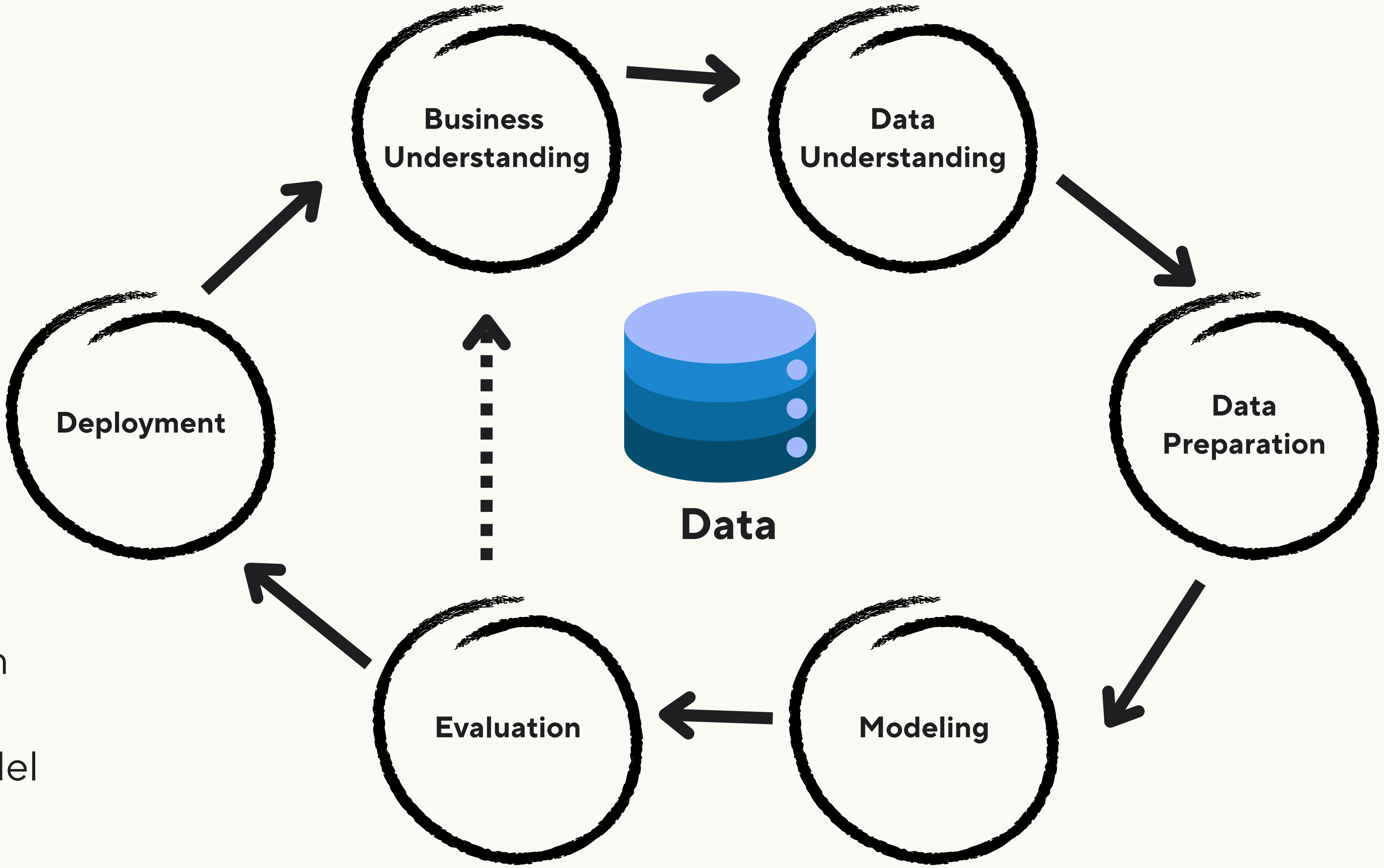
Precision in predicting "Likely to Reject" candidates

USER ADOPTION

100%

Usability for non-technical HR managers via Dashboard

METHODOLOGY



WHY CRISP-DM?

- **Business-First**: Every step starts with business objectives, not just code.
- **Iterative**: Allows us to refine the model as we understand the data better.

Feasibility, Risk, & Mitigation

Techincal Feasibility

using open-source tools & existing data; clean dataset supports modeling.

+

Resource Feasibility

Skilled team (PM, BA, DS, DE); no external hires needed.

+

Schedule Feasibility

7-week plan with Gantt; MVP focus ensures on-time delivery.

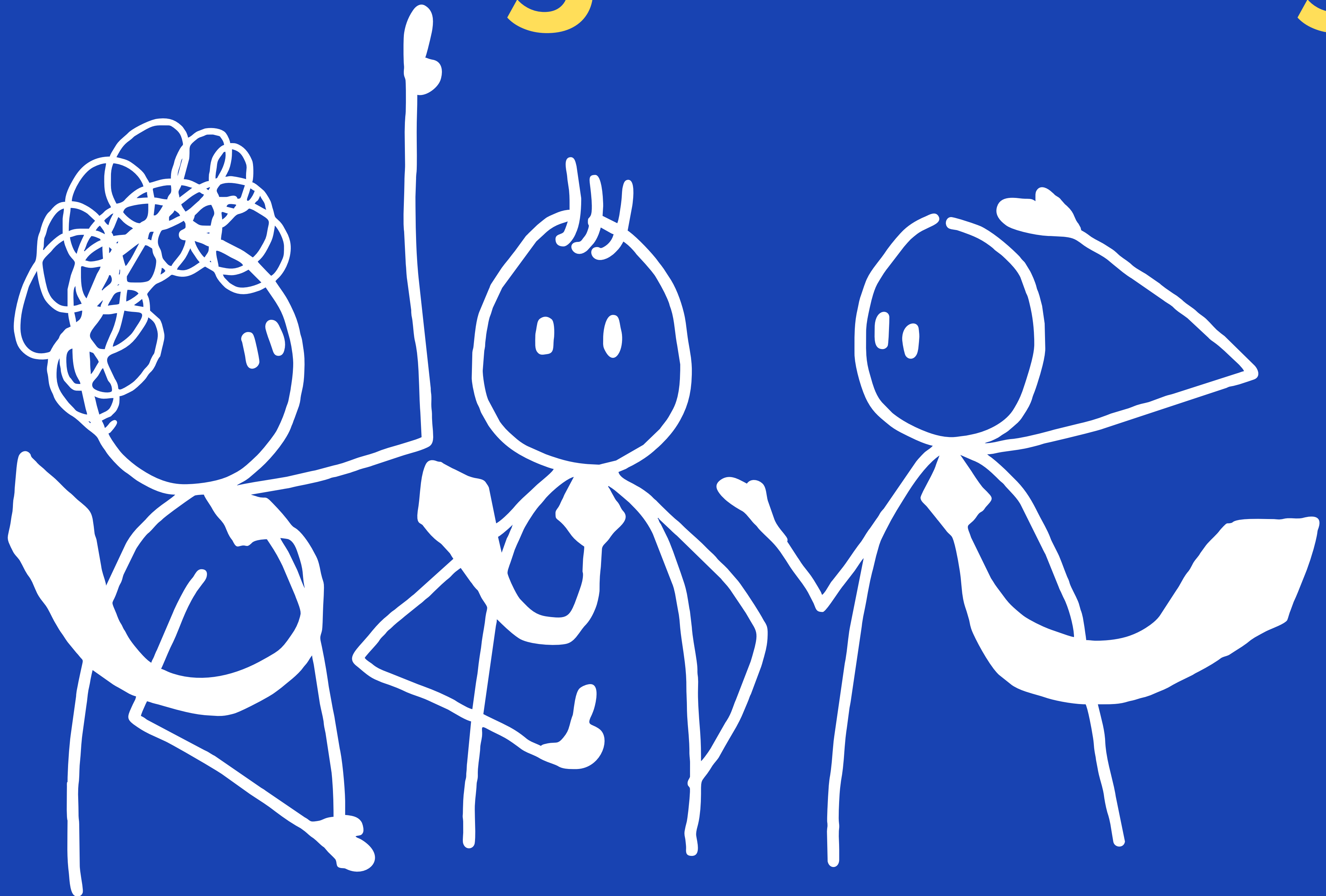
Risk

- Small Dataset Size
- Data Leakage
- Model Overfitting
- Model Interpretability
- Insight Bias
- Tight Deadline

Mitigation

- Low-Complexity Models
- Strict Feature Selection
- Cross-Validation
- SHAP/LIME Explanation
- Data Normalization
- MVP Focus

Exploratory Data Analysis



Data Handling

Stadarization Categorical

Convert categorial to numerical

Label Encoding Categorical

Drop Feature

Split Data

SMOTE

Scalling

Data Set Before and After Feature Engineering

Before

Coloumn Name
reqruitment_id
department
job_title
num_applicants
time_to_hire_days
cost_per_hire
source
offer_acceptance_rate

After

Coloumn Name	
department	applicant_pressure_index
job_title	cost_efficiency_daily
num_applicants	cost_per_applicant
source	hire_days_per_applicant
time_to_hire_category	difficult_index
cost_bucket	acceptance _time_pressure
acceptance_cost_pressure	offer_acceptance_category

Our Data Shows:

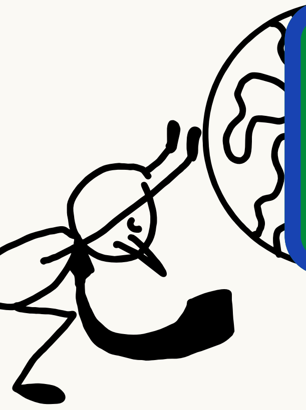
- There are **5000 rows** and **8 column**
- There are **5 numeric** and **3 categorial** data
- There is **no duplicated data**
- There is **no missing values**

Dropped Feature

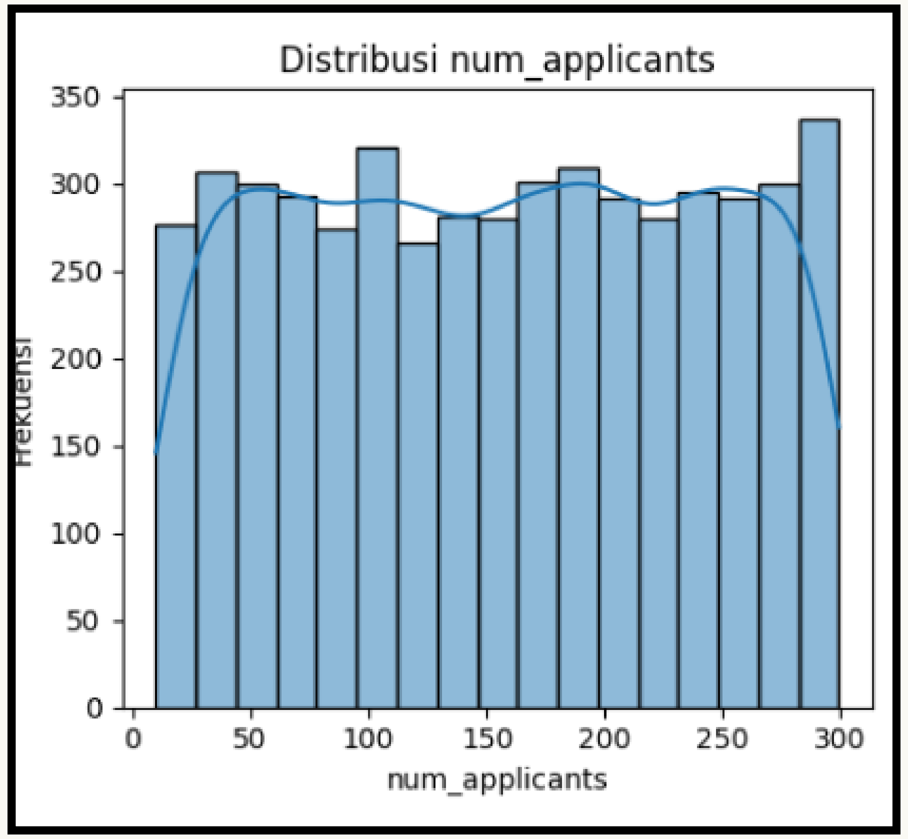
- reqruitment_id
- time_to_hire_days
- offer_acceptance_rate
- cost_per_hire

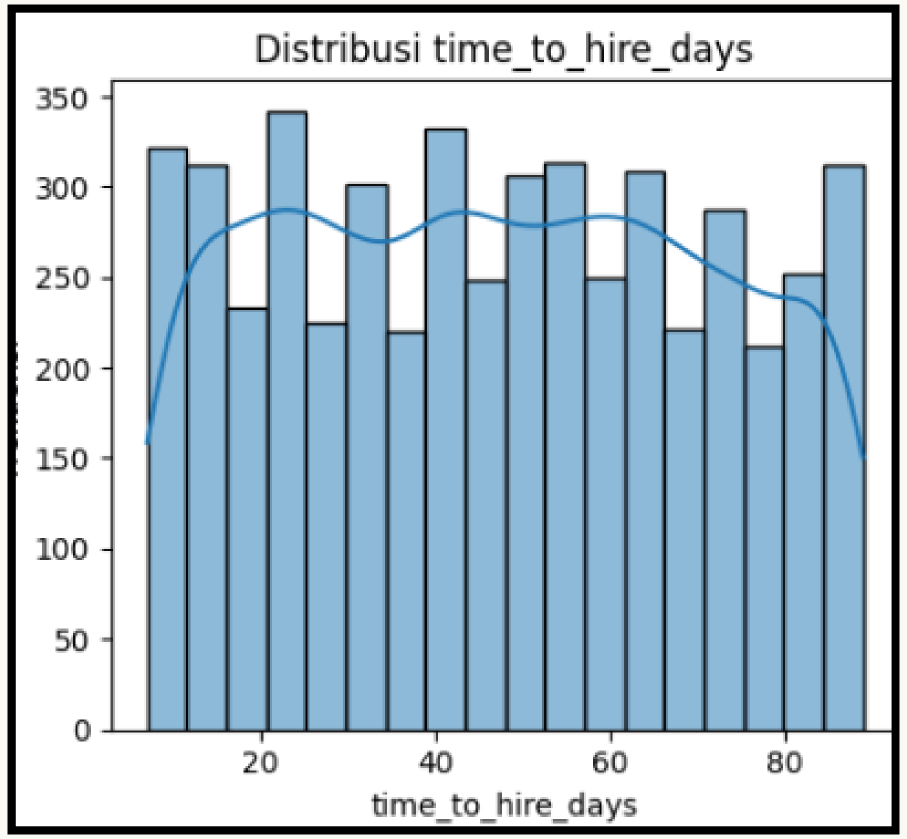
Recruitment Outcomes Are Driven by Variability Not Single Bottlenecks

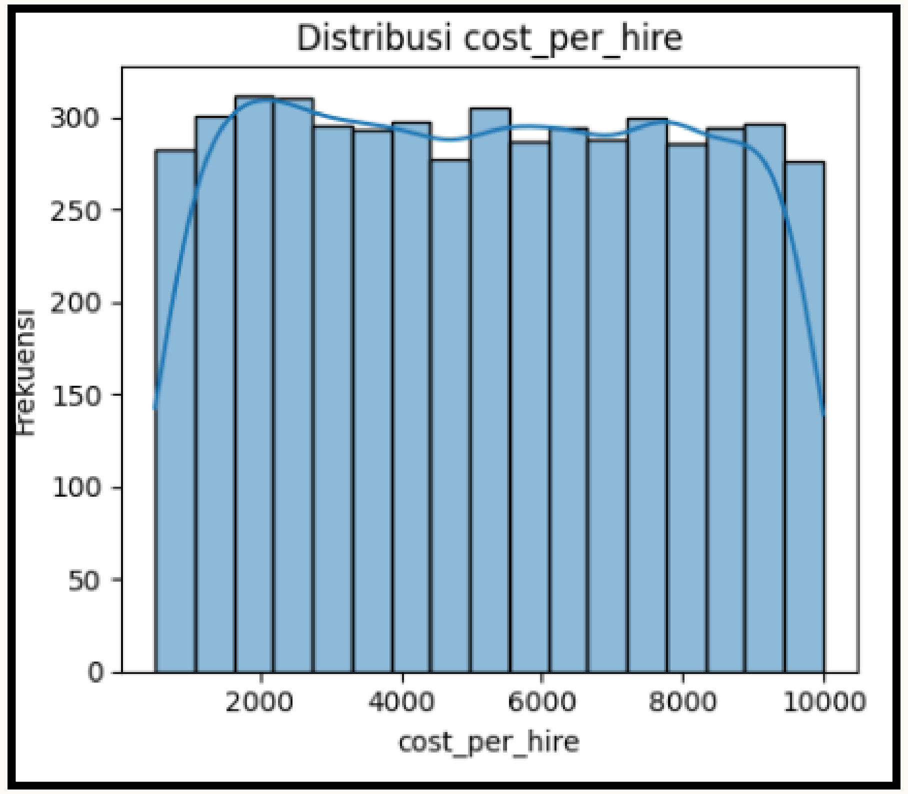
- Applicant volume, time, cost, and acceptance vary widely
- No single extreme bottleneck dominates the process
- Inefficiency is systemic, not isolated

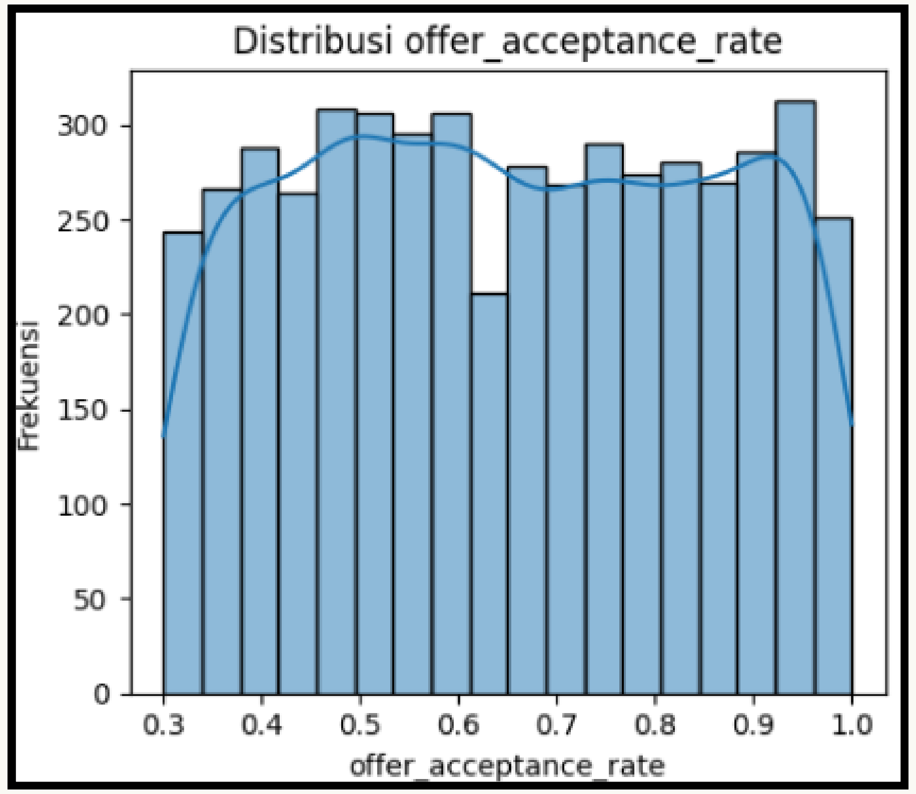


Wide distributions → multiple interacting factors



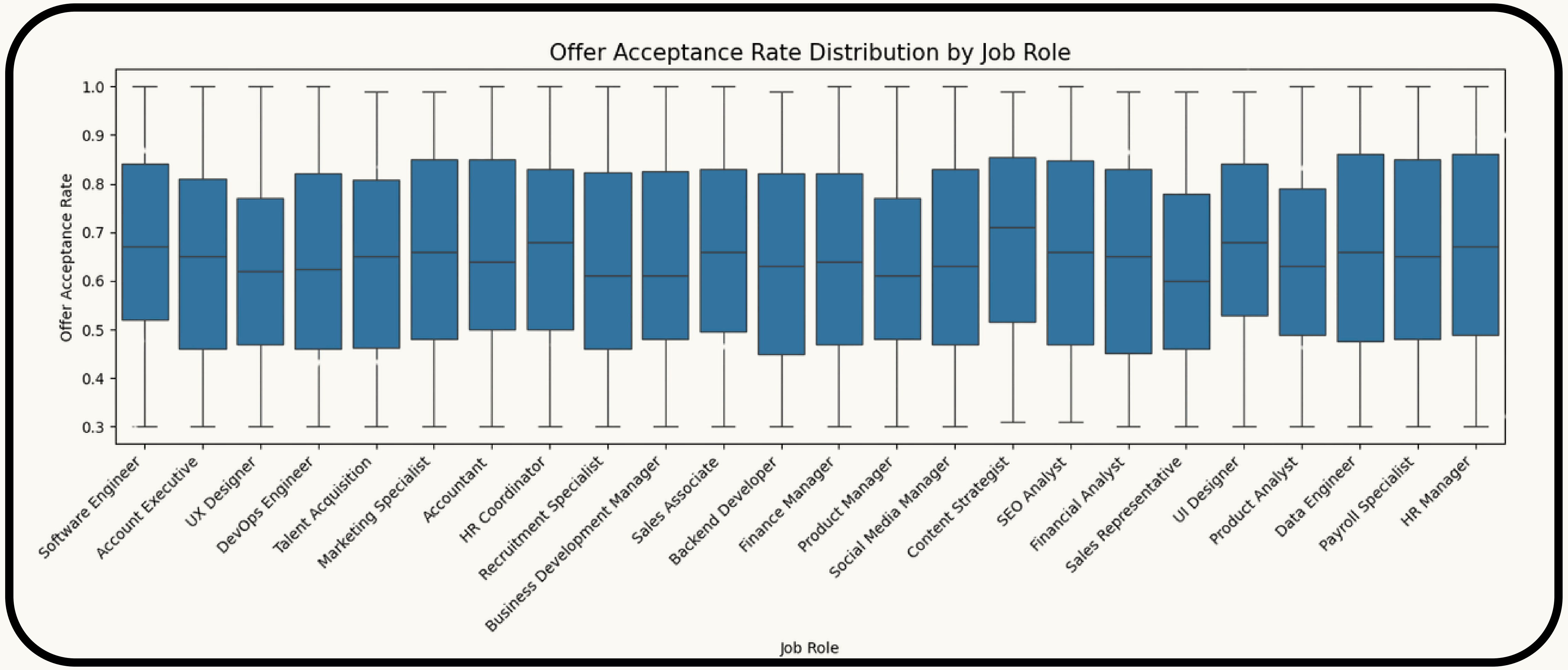


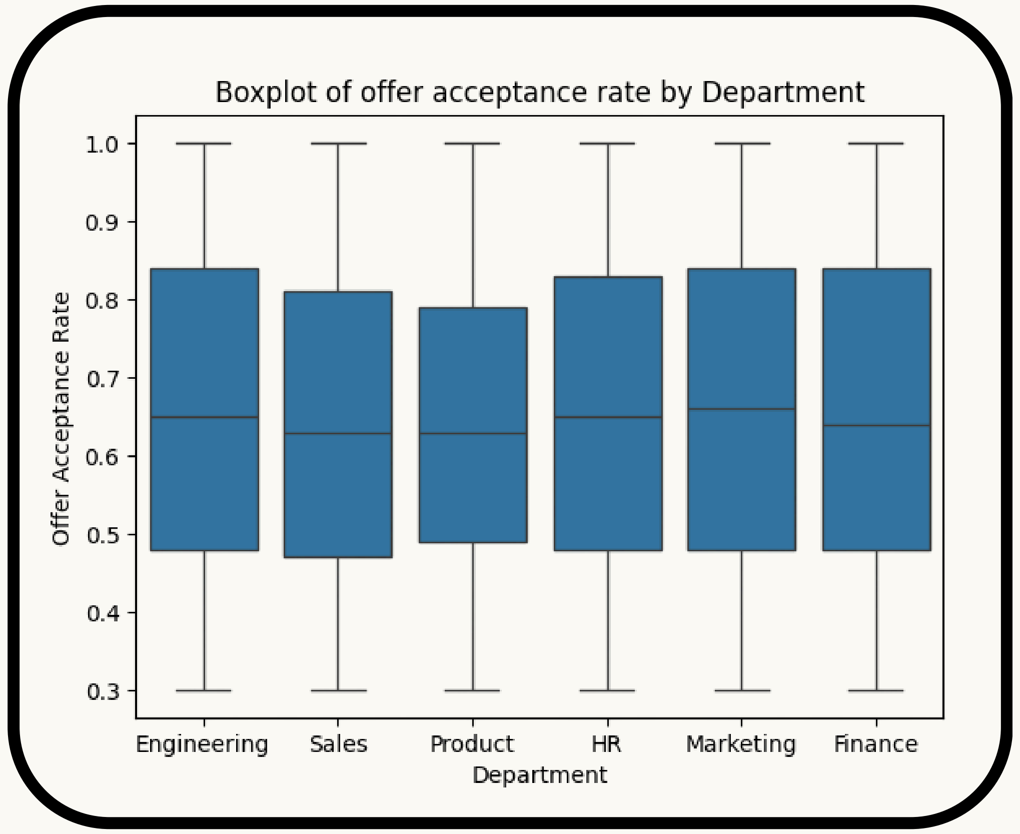


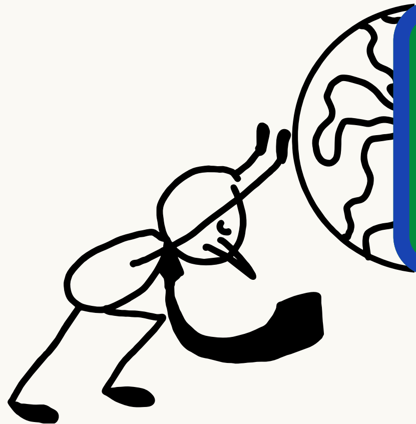


KEY EDA INSIGHT

Recruitment Outcomes Vary More by Job Role Than by Department





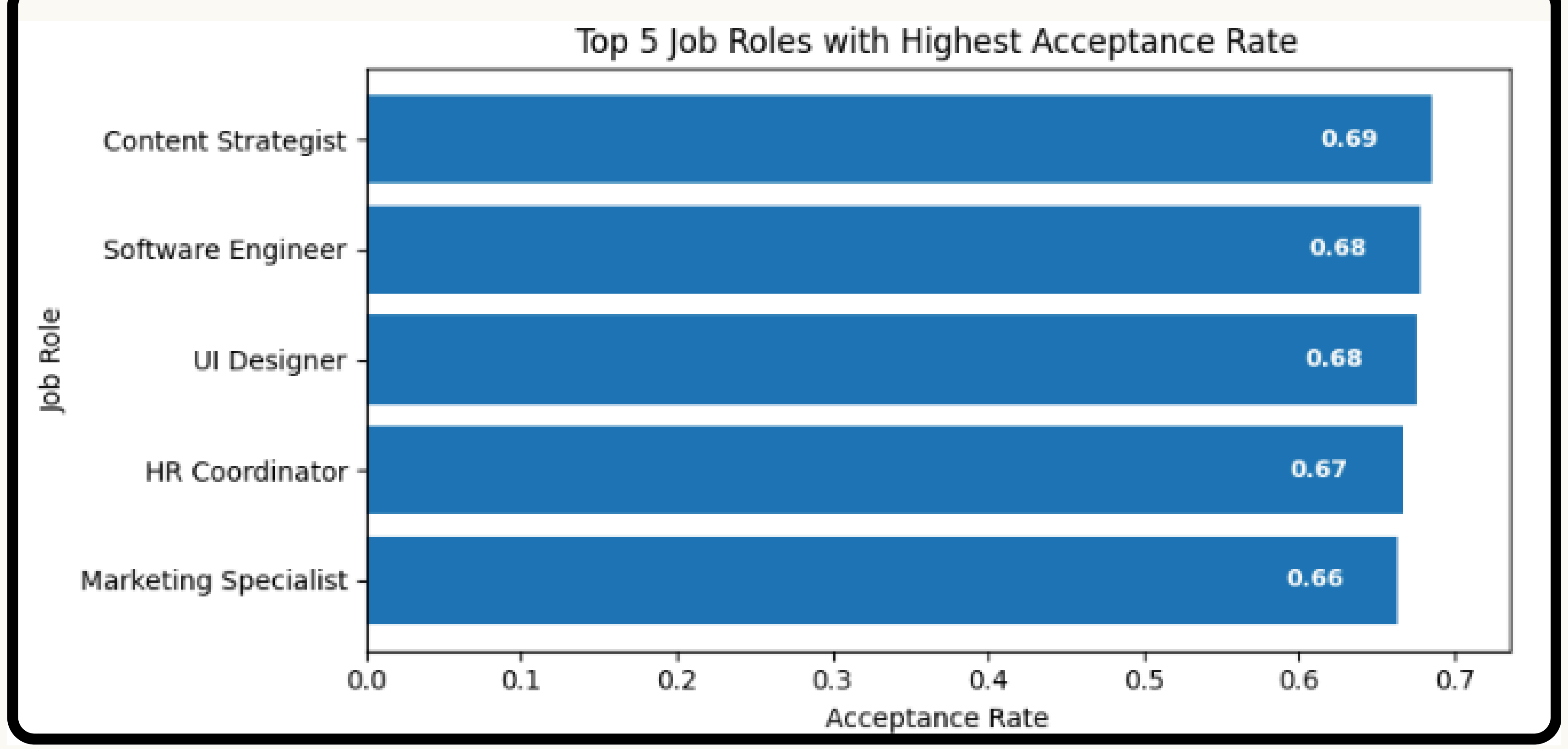


- Acceptance rates cluster in a narrow range
- No role consistently over- or under-performs

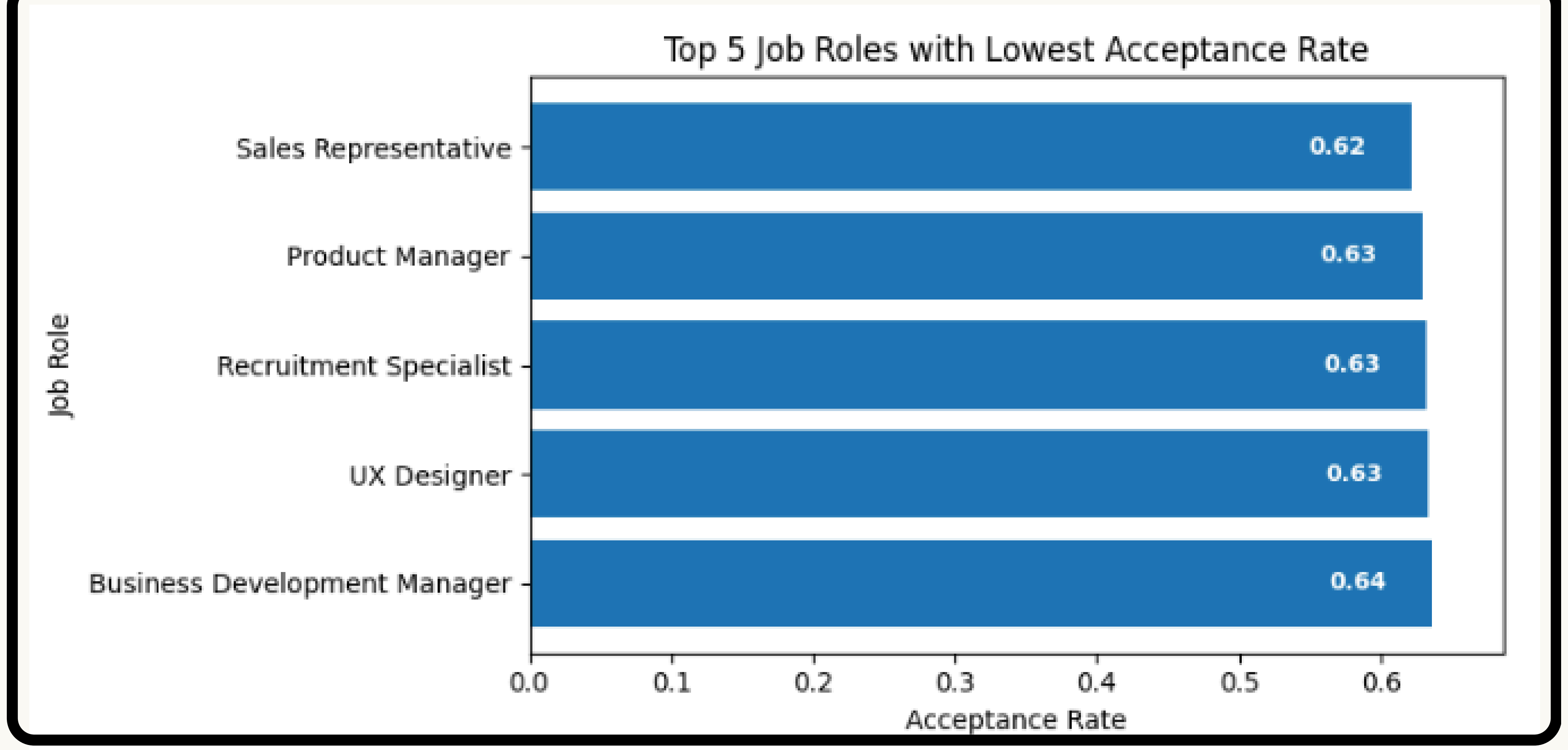
KEY EDA INSIGHT

Offer Acceptance Rates Are Largely Consistent Across Job Roles

Top 5 Job Roles With Highest Acceptance Rate

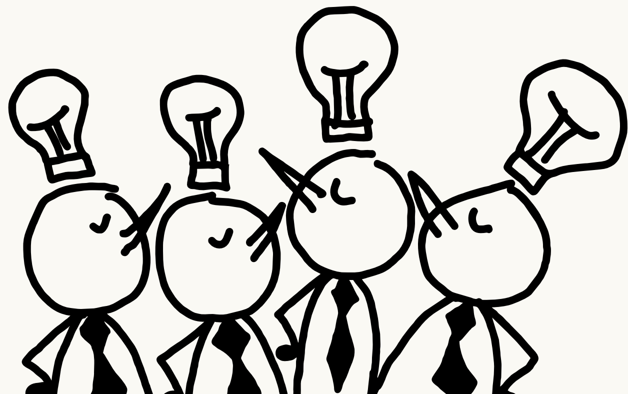


Top 5 Job Roles With Lowest Acceptance Rate



- Acceptance rates cluster in a narrow range
- No role consistently over- or under-performs

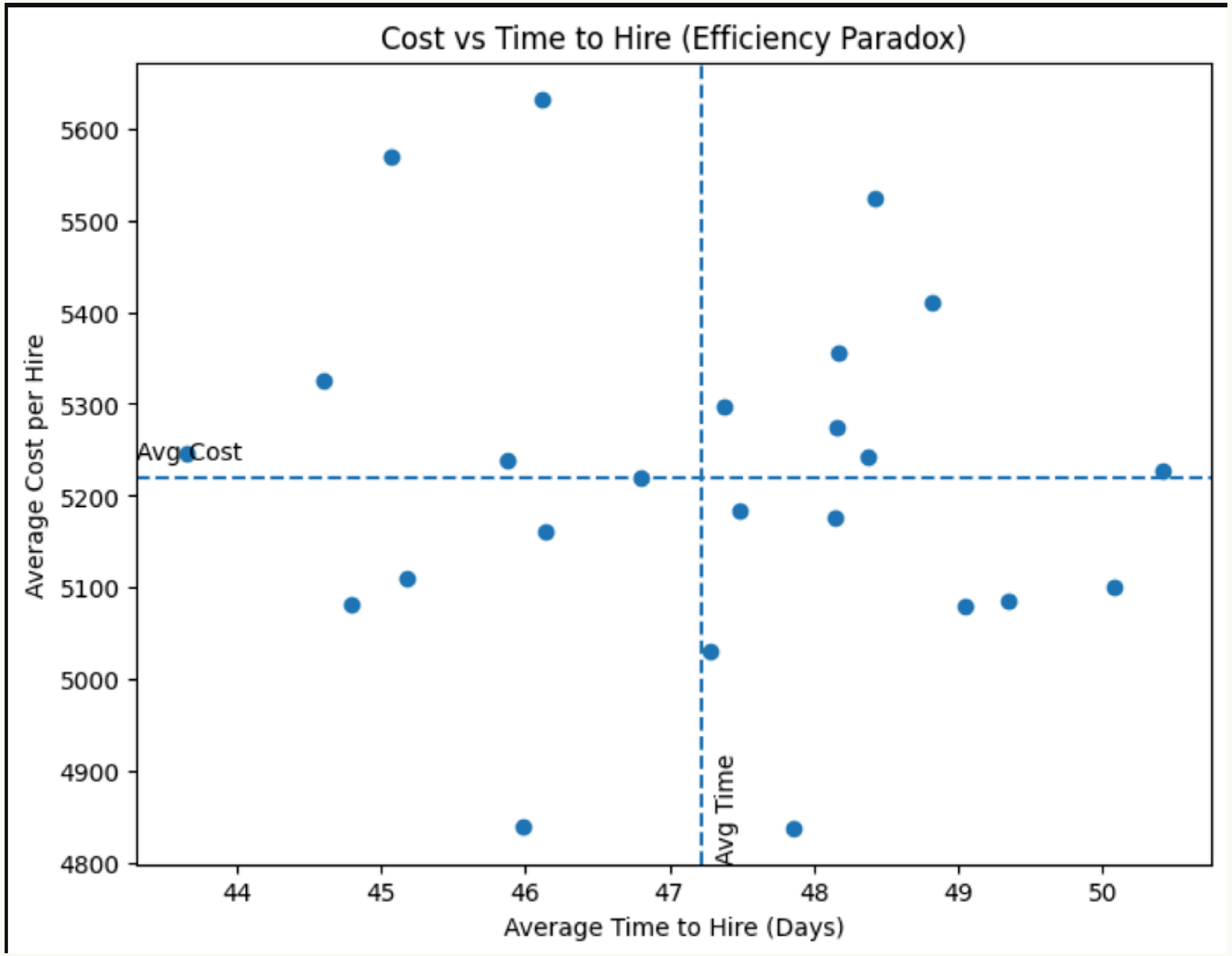
Difference < ~10%



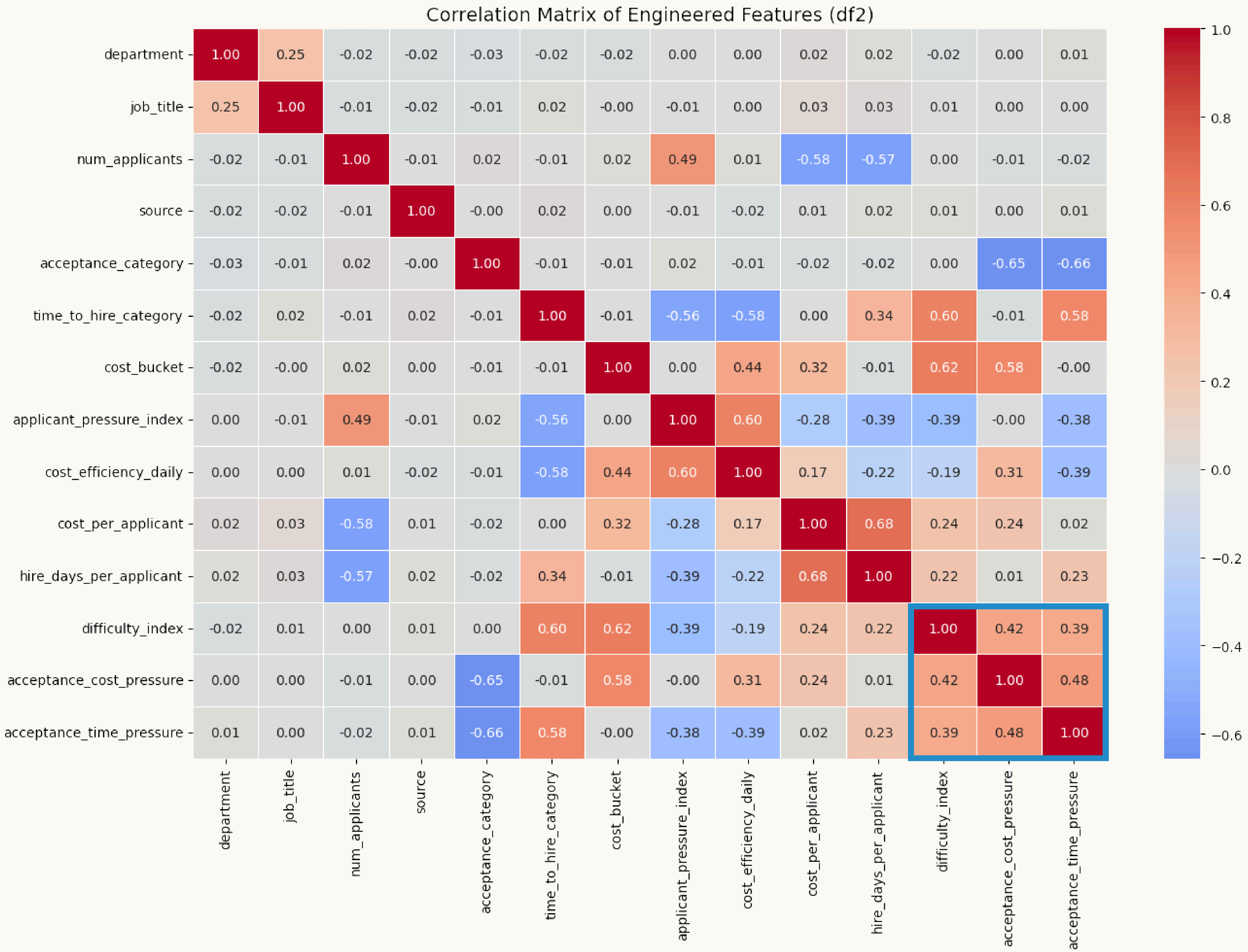
Spending More Does Not Mean Hiring Faster

- No clear correlation between cost and speed
- **Higher spend ≠ better outcome**

"Efficiency is about decisions, not budget."



Pressure & Role Difficulty Drive Recruitment Outcomes



- Recruitment outcomes are more strongly influenced by **applicant pressure and role difficulty than by department or sourcing channel.**
- Offer acceptance is driven by a **combination** of cost pressure, time pressure, and role difficulty, **rather than by a single factor**.

Introducing:
Recruitment Efficiency Predictor

Candidate Data

XGBoost Model

Prediction

What It Does

Predicts offer acceptance probability BEFORE making an offer

How It Helps

Identifies high-risk rejections early & optimizes budget allocation

Five Models, Ten Experiments, One Winner

Model	Test F1 Macro		Improvement
	Baseline	Tuned	
XGBoost	0.953	0.958	+0.005
Random Forest	0.921	0.941	+0.020
Logistic Regression	0.927	0.928	+0.001
Decision Tree	0.909	0.918	+0.009
KNN	0.704	0.746	+0.042

All models were evaluated using the same **5-fold** stratified cross-validation

Same class balancing strategy (**SMOTE**)

Same **train–test split**

Same primary evaluation metric (**F1 Macro**)

Winner: XGBoost with Randomized Search CV + SMOTE | 94% Precision for "Likely Reject" | Train-test gap < 5%

Proven Accuracy:
94%+ Precision Where It Matters

94% +

Precision for "**Likely Reject**"
Prevents Wasted Recruitment Effort

Model Type
XGBoost

Data Handling
SMOTE

Validation
Cross-val

Tuning
Randomized Search CV

Prevents Wasted Recruitment Effort

A Day in HR's Life:

From $9.1M Problem to Measurable Savings

Before

Time to Hire	Cost Per Hire
47 Days	$5,215

Annual Inefficiency

USD 9.1M

Process

Manual screening → Guesswork → Wasted effort

After

Time to Hire	Cost Per Hire
38 Days	$4,850

Annual Savings

USD 1.82M

Process

AI prediction → Smart prioritization → Focused effort

Powered by AI,
Built for HR

Real-Time Predictions
Get instant offer acceptance probability for any candidate profile

Performance Insights
Department & channel analytics to optimize recruitment strategy

Cost-Time Tradeoff Viz
Visualize efficiency metrics to balance speed and budget

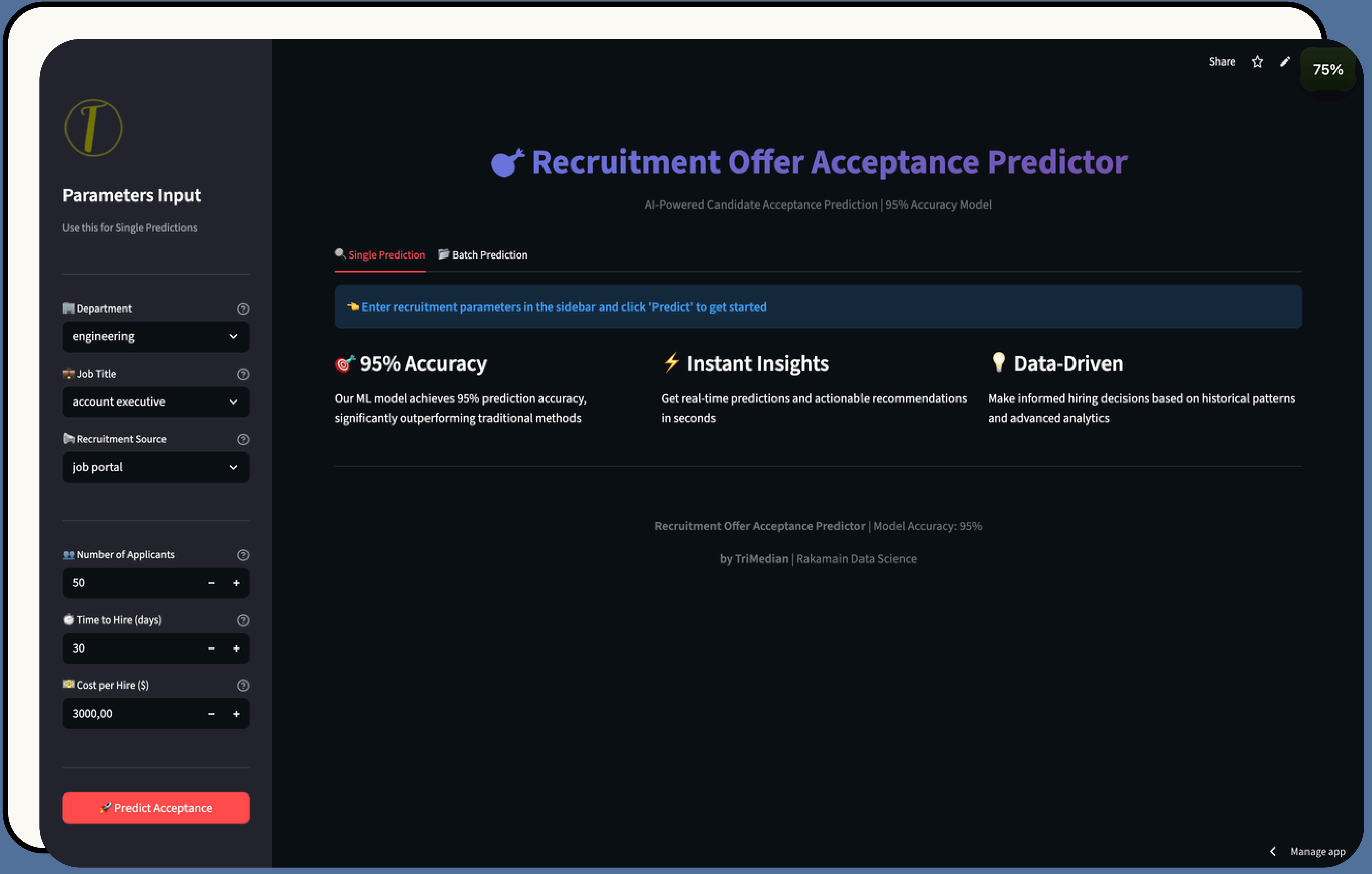
Candidate Prioritization
AI-powered recommendations on which candidates to focus efforts

User-Friendly Interface
100% usability for non-technical HR managers—no coding required

Explainable AI
SHAP values show WHY the model makes each prediction

Meet Your New Recruitment Assistant

Live Streamlit Application



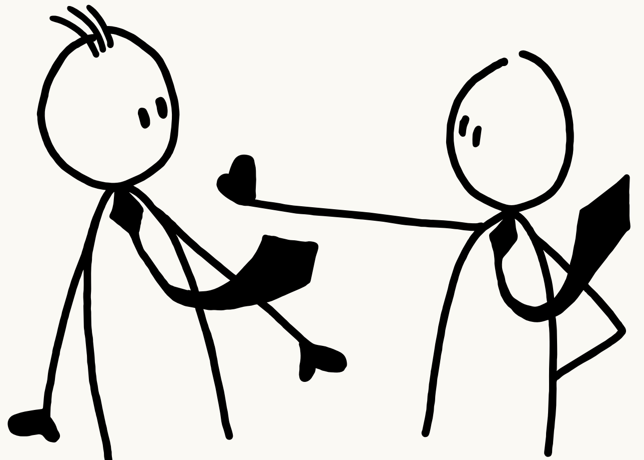
https://trimedian-app.streamlit.app

Thank you

Stop Hiring by Guesswork. Start Hiring by Data

Rakamin

TriMedian



Appendix

Built Right: Production-Ready
Tech Stack

Python

XGBoost

Streamlit

Pandas

Numpy

Scikit-learn

Joblib

SHAP

ML Pipeline Architecture

End-to-End Prediction Flow

Raw Data Input → Feature Engineering → Label Encoding → SMOTE → Standard Scaling → XGBoost Prediction

Saved Artifacts for Deployment

- Model (.joblib)
- Encoders (.joblib)
- Scaler (.joblib)
- Metadata (.joblib)

DataFrame Structure & Data Describe

Data Type

Coloumn	Data Type
reqruitment_id	int64
department	object
job_title	object
num_applicants	int64
time_to_hire_days	int64
cost_per_hire	float64
source	object
offer_acceptance_rate	float64

Data Description

	recruitment_id	num_applicants	time_to_hire_days	cost_per_hire	offer_acceptance_rate
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
mean	2500.500000	155.621000	47.191400	5214.826236	0.650832
std	1443.520003	84.164264	23.864934	2730.999185	0.202052
min	1.000000	10.000000	7.000000	507.160000	0.300000
25%	1250.750000	83.000000	26.000000	2820.597500	0.480000
50%	2500.500000	157.000000	47.000000	5218.290000	0.650000
75%	3750.250000	229.000000	67.000000	7611.412500	0.830000
max	5000.000000	299.000000	89.000000	9998.910000	1.000000

Our Data Shows:

- There are **5000 rows** and **8 column**
- There are **5 numeric** and **3 categorial** data
- There is **no duplicated data**
- There is **no missing values**

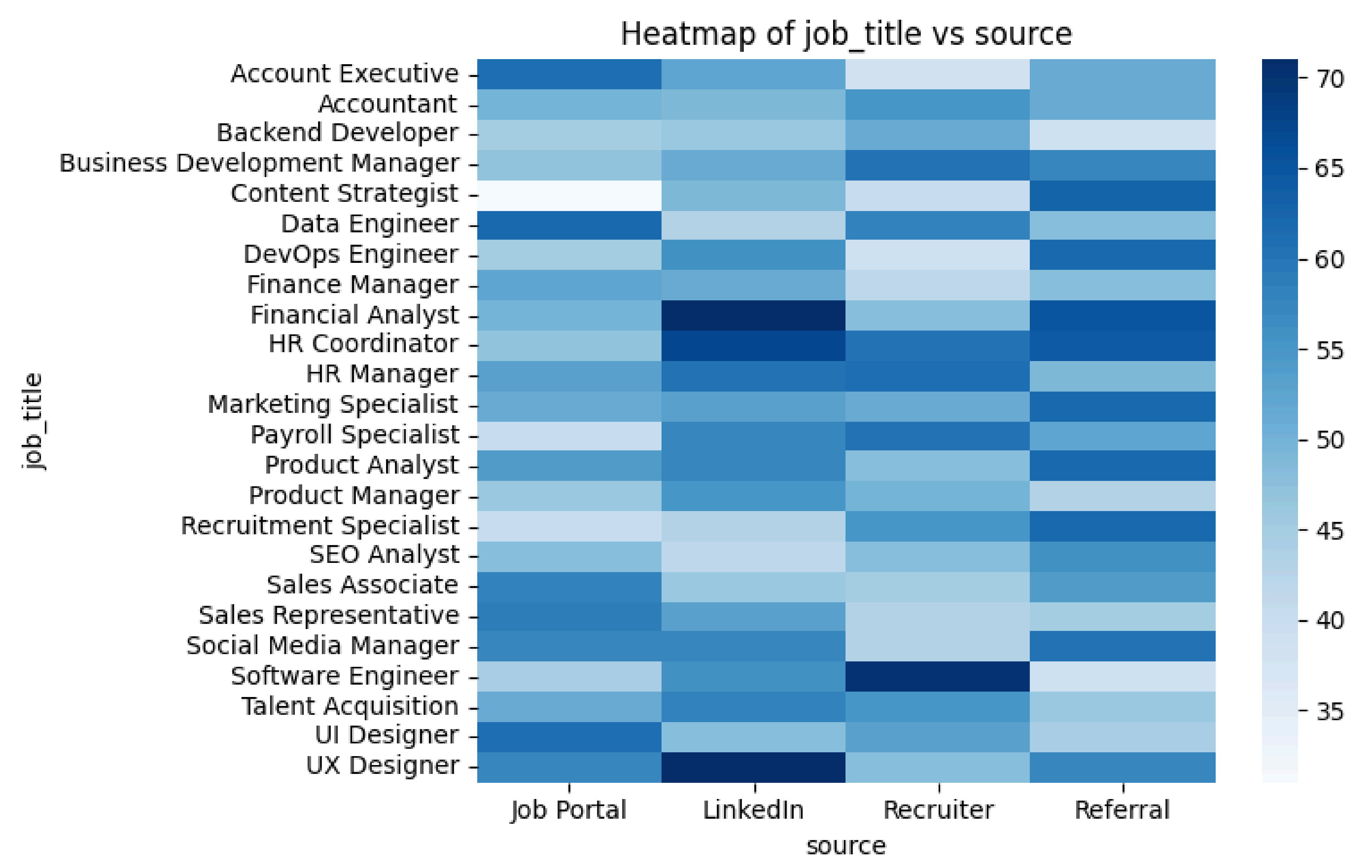
5000
Recruitment Records Analyzed
Complete hiring funnel from sourcing to acceptance

6 Departments
Engineering, Sales, Product, HR, Marketing, Finance.

24 Job Titles
From entry-level to strategic positions.

4 Chanels
LinkedIn, Job Portal, Recruiter, Referral

Recruitment Channel Pattern by Job Title



INSIGHT :

- **Job Portals** are dominant for operational and entry–mid level positions such as Sales Associate, Account Executive, and Content Strategist.
- **LinkedIn** is highly effective for professional and specialized roles, including Software Engineer, Data Engineer, HR Manager, and UX Designer.
- **Recruiters** are widely used for strategic or highly specialized positions, such as Finance Manager, Recruitment Specialist, and Product Manager.
- **Referrals** contribute significantly to certain roles, particularly in Marketing, Product, and selected technical positions.

CONCLUSION :

Each job title exhibits a distinct recruitment source pattern, reflecting the effectiveness of specific channels for different types of roles.

Feature Selection

Feature Used

- Department
- Job Title
- Num Applicants
- Source
- Time To Hire Category
- Cost Bucket
- Acceptance Cost Pressure
- Applicant Pressure Index
- Cost Efficiency Daily
- Cost Per Applicant
- Hire Days per Applicant
- Difficult Index
- Acceptance Time Pressure

Feature Dropped

- Recruitment ID
- Offer Acceptance Rate
- Time To Hire Days
- Cost Per Hire

Target

- Acceptance Category

- **Selected features** capture role context, recruitment effort, and process efficiency to predict offer acceptance.
- **Recruitment ID** is removed as it provides no predictive value. We deliberately drop **Offer Acceptance Rate** after feature engineering to avoid target leakage.
- We Drop also **Time To Hire Days and Cost Per Hire** to avoid redundant or leakage data.
- The target feature, **Acceptance Category** directly aligns with the business objective of improving recruitment outcomes.

WHY FEATURE ENGINEERING?

Why Raw Data Is Not Enough

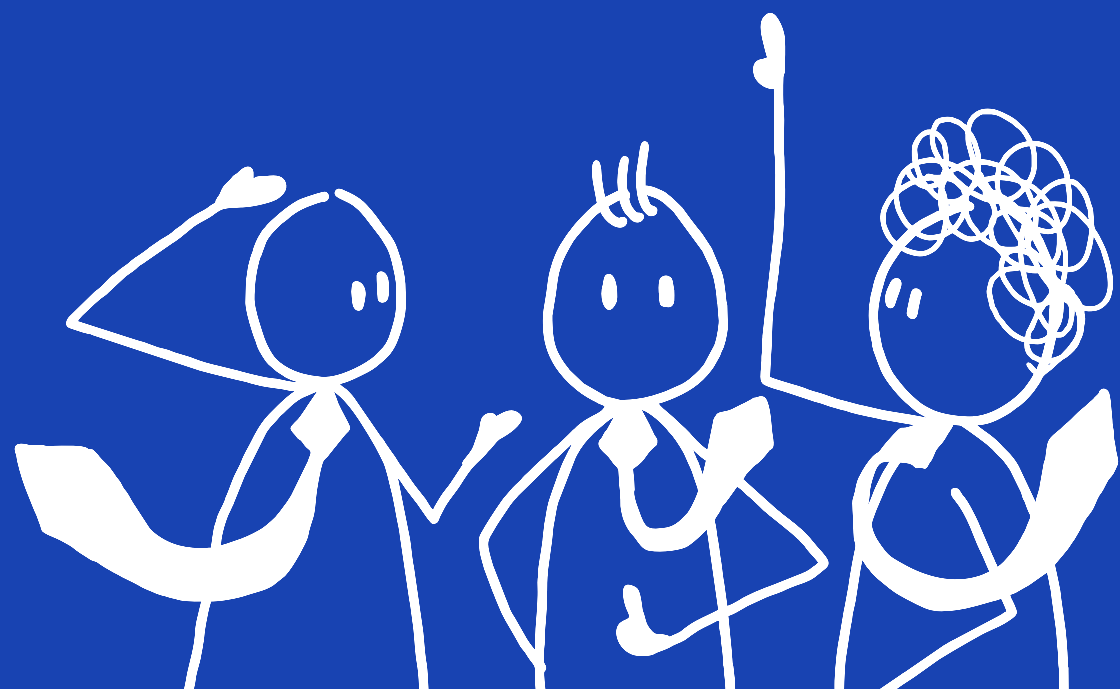
Raw recruitment data

- Tells **what happened**
- Does NOT explain **why offers are accepted or rejected**

Feature engineering helps us

- Capture **pressure**
- Capture **difficulty**
- Capture **efficiency**

Raw Data → **Engineered Signals** → **Better Decisions**



Feature Engineering

New Features

Time-to-Hire Category

Classifies hiring duration (days) into **Fast / Medium / Slow** using time cutoffs.

Acceptance Category

Groups acceptance rate (%) into **Likely Acc/ Likely Rej. / Uncertain** Rebased on predefined thresholds.

Cost-per-Hire Bucket

Segments cost per hire into **Low / Medium / High** based on data distribution (e.g. quantiles).

Applicant Pressure Index

Measures selection pressure = **Applicant Volume ÷ Time-to-Hire**.

Feature Engineering

New Features

Cost Eficiency Ratio

Cost Eficiency Ratio
Measures the cost per day of hiring, indicating whether the process is **expensive but fast** or **cheap but slow**.

Hire Days Per Aplicant

Hire Days Per Aplicant
Measures the number of days required per incoming application

Sourcing Effort Intensity

Sourcing Effort Intensity
Sourcing Efficiency indicates how effective candidate acquisition is, where a **lower value means more efficient sourcing**

Difficult Index

Difficult Index

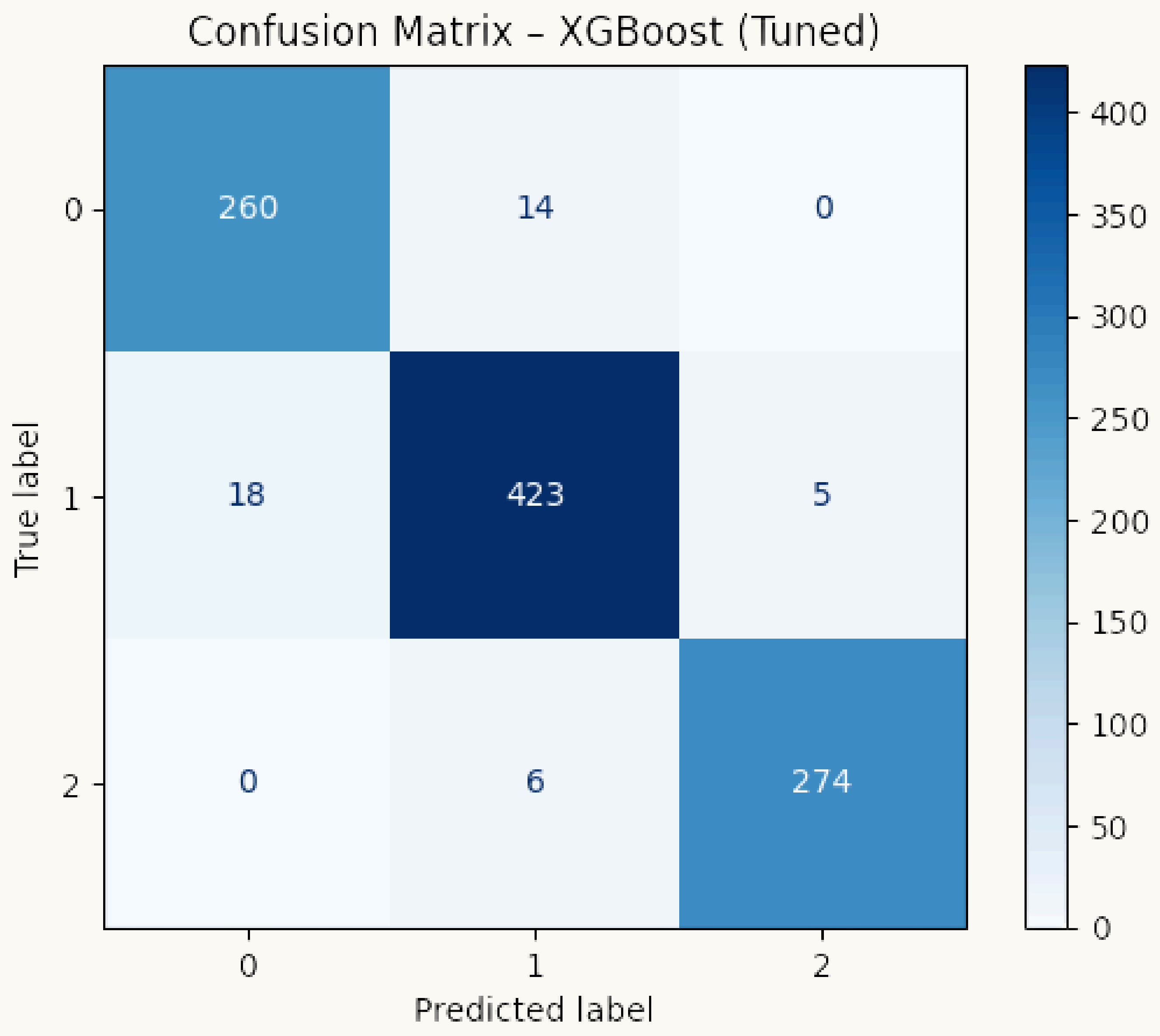
How difficult a position is to fill. **Higher means harder to fill.**

Acceptance Cost/Time Pressure

Acceptance Cost/Time Pressure

Hiring pressure caused by high cost or long hiring time combined with low offer acceptance.

Confusion Matrix



1. Class 0 — "Likely Reject"

- 255 out of 274 candidates correctly classified
- Recall: 93.06%
- Insight: The model is highly effective at identifying candidates who are likely to reject the offer.

→ This prevents wasted time and resources on high-risk candidates.

2. Class 1 — "Uncertain"

- 424 out of 446 candidates correctly classified
- Recall: 95.06%
- Insight: The model demonstrates strong sensitivity in detecting candidates with ambiguous acceptance behavior.

→ This group requires strategic communication and closer engagement from HR.

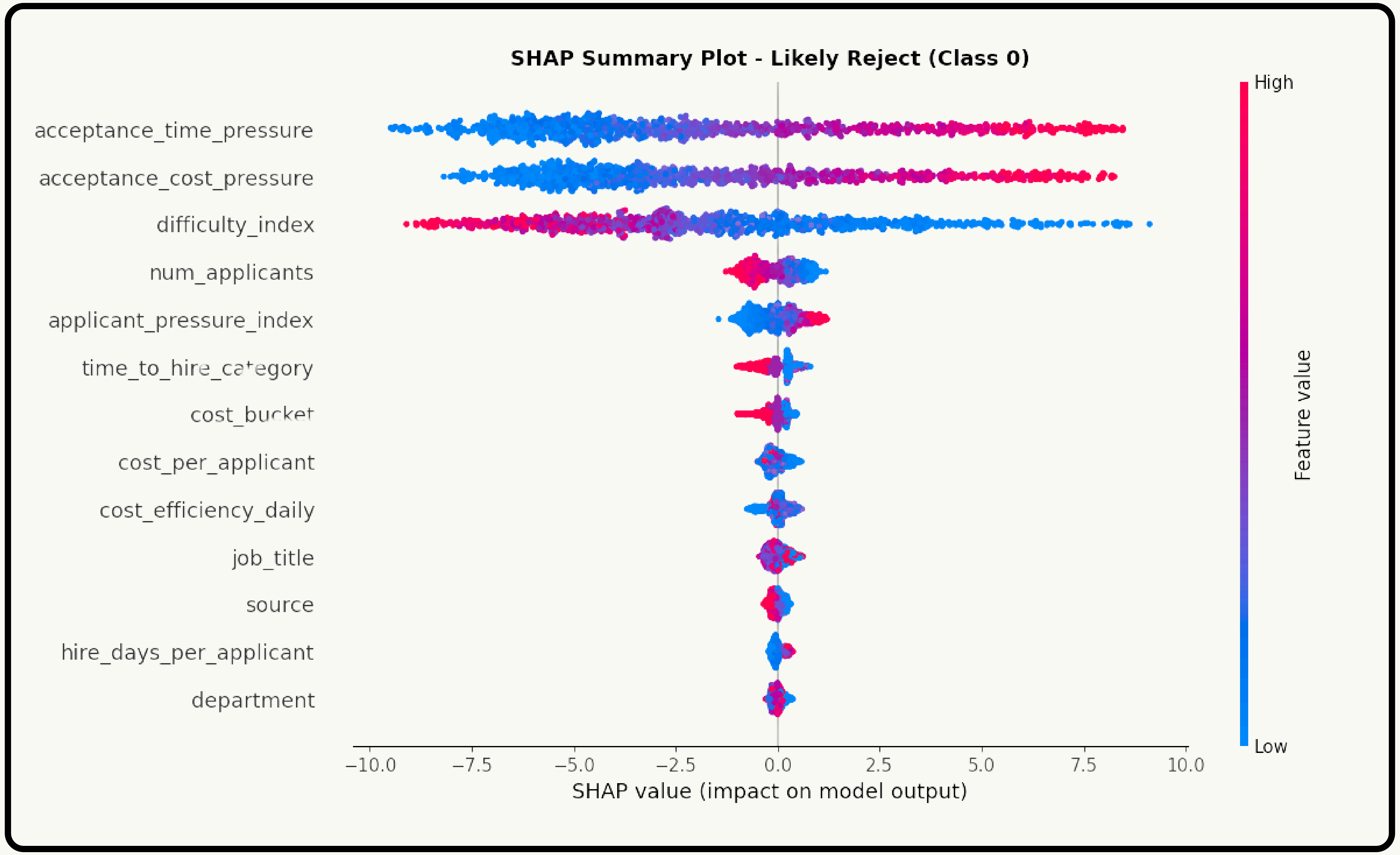
3. Class 2 — "Likely Accept"

- 272 out of 280 candidates correctly classified
- Recall: 97.14%
- Insight: The model is highly accurate in predicting candidates who are very likely to accept the offer.

→ HR can fast-track these candidates to accelerate hiring and increase conversion rates.

EXPLAINABILITY ANALYSIS

SHAP - *Why the Model Makes Its Decision*



Likely Reject (Class 0)

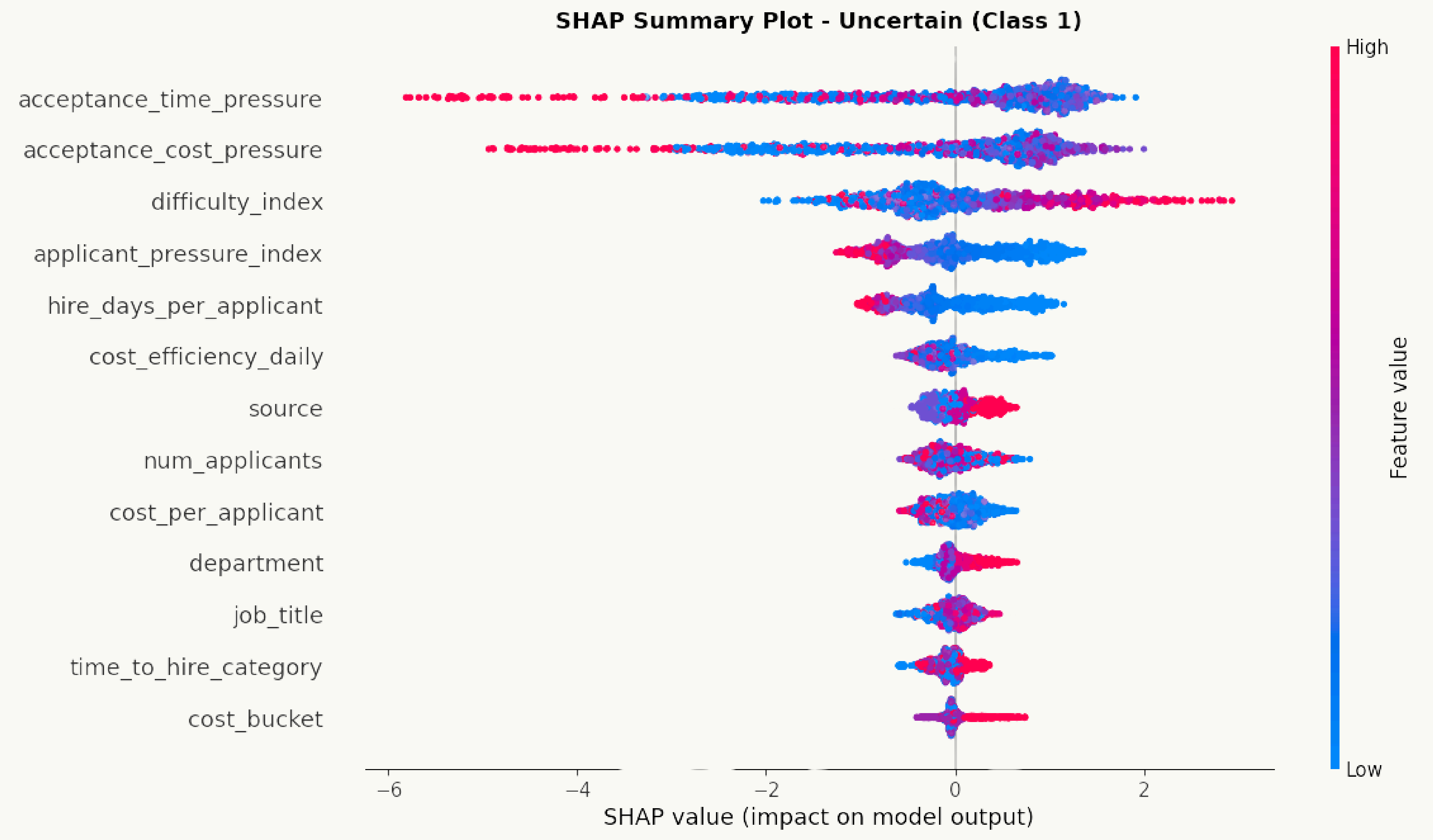
Why candidates tend to REJECT the offer

- A long hiring process reduces candidate interest
- High cost pressure indicates an inefficient recruitment process
- High role difficulty makes candidates more selective

The model identifies time pressure, cost pressure, and role difficulty as the main drivers of rejection

EXPLAINABILITY ANALYSIS

SHAP - *Why the Model Makes Its Decision*



Uncertain (Class 1)

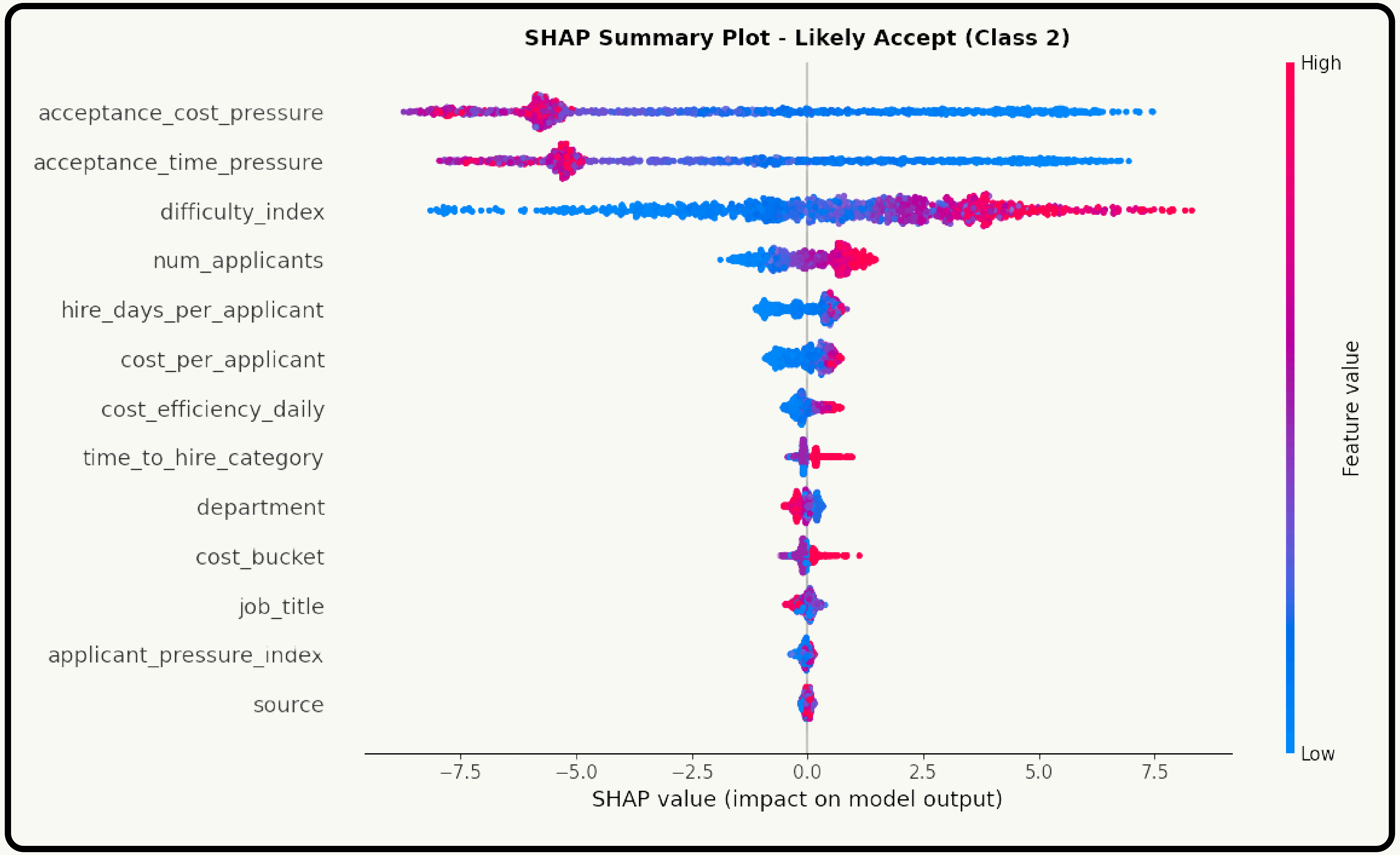
Candidates in the grey area

- Moderate time and cost pressure
- No extreme role difficulty
- Outcomes strongly depend on HR follow-up and engagement

This class represents the highest opportunity for HR intervention, where actions can change the final decision

EXPLAINABILITY ANALYSIS

SHAP - *Why the Model Makes Its Decision*



Likely Accept (Class 2)

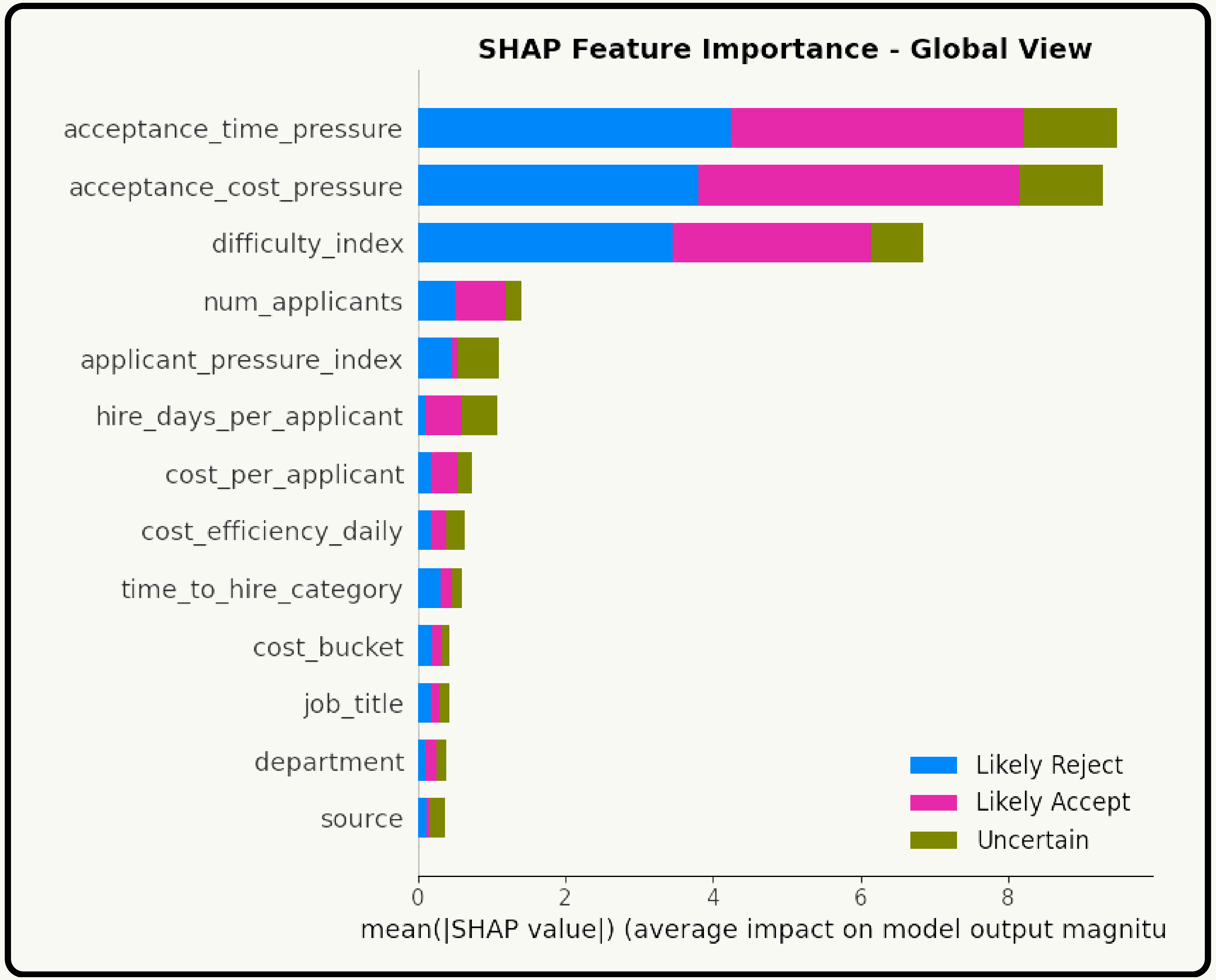
Why candidates tend to ACCEPT the offer

- A fast and efficient hiring process
- Low cost pressure, indicating good process management
- Roles that are realistic and aligned with candidate expectations

Process efficiency is the key driver of offer acceptance

EXPLAINABILITY ANALYSIS

Feature Importance



The "Big Drivers"

1. **time_to_hire_category** (Score: ~0.45)
- The length of the hiring process (fast, medium, or slow) is the strongest predictor of the target variable. How quickly a candidate is hired plays a decisive role in the model's outcome.
2. **cost_bucket** (Score: ~0.20)
- Recruitment cost (low vs high) shows a strong correlation with prediction results, indicating that financial efficiency is a key decision factor.
3. **acceptance_time_pressure & acceptance_cost_pressure** (Combined score: ~0.23)
- Situations involving time pressure or budget constraints during offer acceptance significantly influence the final outcome.

Significant Feature Importance Drop-off

- After the top four features, there is a sharp decline in importance scores.
- Demographic or administrative features such as source (candidate source), department, or job_title contribute minimally compared to efficiency-related metrics.
- The XGBoost model prioritizes process efficiency (time and cost) over positional or departmental attributes.

Business Comparassion

Before

Metric	Value	Calculation Basis	Data Source
Total recruitment records / year	5000	Total rows in dataset	Dataset
Avg. time to hire	47 days	Mean(time_to_hire_days)	Dataset
Avg. cost per hire	USD 5,215	Mean(cost_per_hire)	Dataset
Offer acceptance rate	65%	Observed acceptance ratio	EDA
Offer rejection rate	35%	1 − acceptance rate	Derived
Total annual recruitment spend	USD 26.1M	5,000 × 5,215	Derived
Annual inefficient spend	USD 9.1M	35% × 26.1M	Derived
Process characteristic	Manual, intuition-based	No predictive filtering	Observed

After

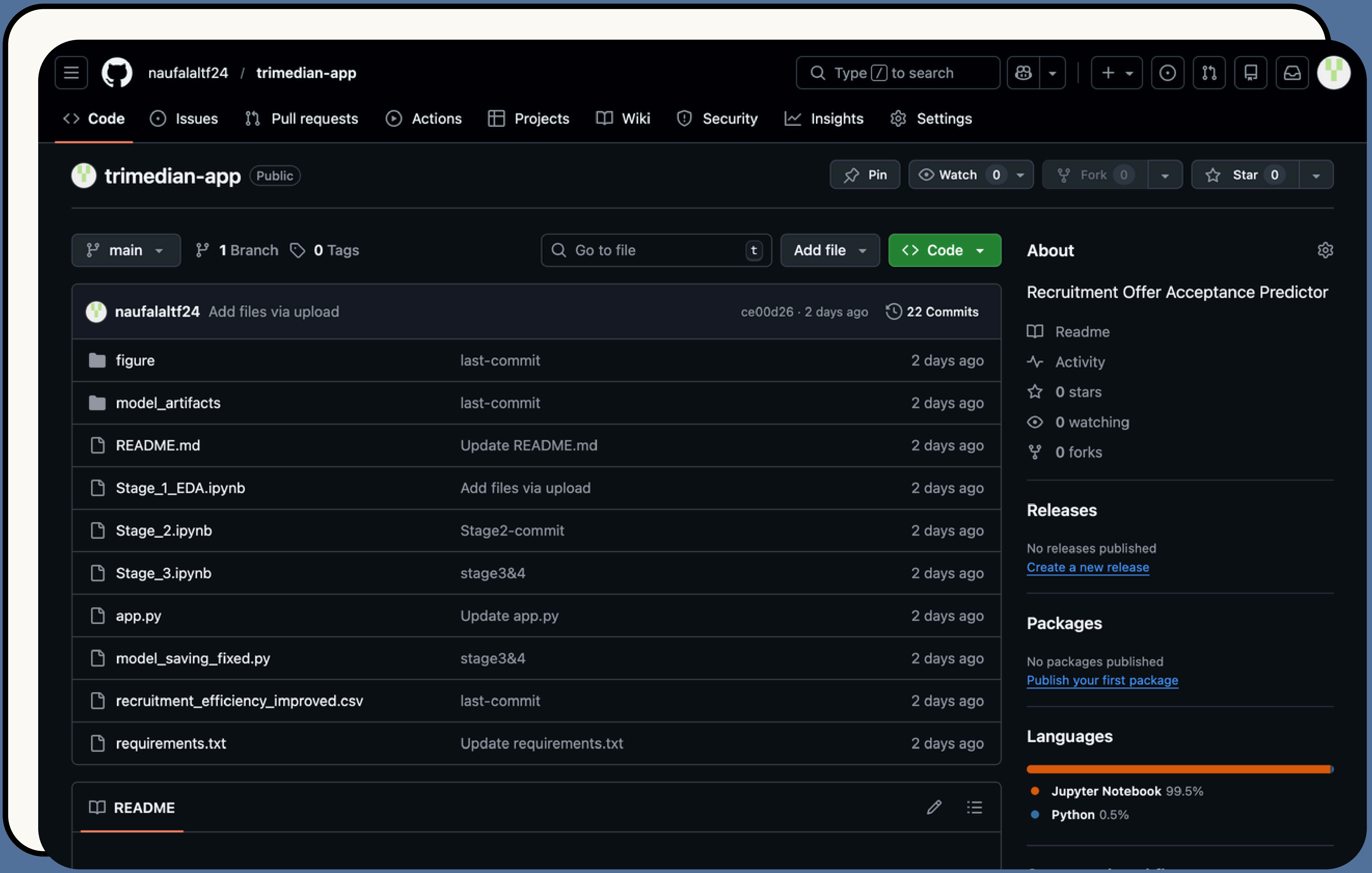
Metric	Value	Calculation Basis	Evidence
Model used	XGBoost	Best F1-macro & precision model	Notebook
Precision (Likely Reject)	94%	TP / (TP + FP)	Model evaluation
HR time saved per candidate	50–95 mins	Benchmark avg task reduction (link)	HR benchmark
Avg. HR time saved per role	≈ 55 hours	47 candidates × 70 mins	Dataset + benchmark
Time saved per hire	7–12 days	HR hours → process delay reduction	Process mapping
Avg. time to hire	35–40 days	47 − (7–12)	Derived
Annual recruitment cost saved	USD 1.82M	20% × 9.1M rejected spend	Conservative scenario
Avg. cost saved per hire	≈ USD 360	1.82M ÷ 5,000	Derived
Process characteristic	AI-assisted prioritization	Early rejection risk filtering	Deployment

FINAL EFFICIENCY SUMMARY

Dimension	Before AI	After AI	Efficiency Formula
Avg. time to hire	47 days	35–40 days	47 − (7–12)
HR effort per candidate	90–150 mins	35–55 mins	Baseline − AI-assisted effort
Avg. cost per hire	USD 5,215	≈ USD 4,850	5,215 − 360
Annual inefficient spend	USD 9.1M	USD 7.3M	9.1M − 1.82M
Decision quality	Intuition-based	Data-driven	Qualitative structural shift

Github

Github Repository



https://github.com/naufalaltf24/trimedian-app

PROJECT ROADMAP

8-Week Sprint to Value Delivery

Start Date: 12/1/2025
Display Week: 1

Stage	PIC	ROLE	Sprint Task	Description	Plan to Start	Plan to Finish	Status	Link
0 - Project Initiation	Naufal	PM	Kickoff Review	Memahami brief, dataset, rubric, dan tujuan proyek	12/6/25	12/9/25	DONE	
	Hutomo & Naufal	BI	Industry Research	Riset talent intelligence & recruitment efficiency	12/6/25	12/9/25	DONE	
	Naufal	PM	Quick Analyst	Merumuskan problem statement, menenutkan sucess	12/6/25	12/9/25	DONE	
	Naufal	PM	Risk Assesment	Risk data, modelling, technical, timeline	12/6/25	12/9/25	DONE	
	Hutomo & Naufal	BI	Workflow	Membuat CRISP-DM diagram	12/6/25	12/9/25	DONE	
	Naufal	PM	Timeline	Membuat gantt chart & pembagian tugasa	12/6/25	12/10/25	DONE	
	Naufal	PM	Compile Proposal Deck	Membuat PPT stage 0	12/6/25	12/10/25	IN PROGRESS	
	All Team	All Team	Mentoring Session	Present stage 0 ke mentor	12/11/25	12/11/25	TO DO	Link Meet
1- Data Aquisition & Preparation	Rifki	DE	Load Raw Data	import seluruh dataset, cek struktur	12/6/25	12/18/25	TO DO	
	Rifki	DE	Data Quality Check	cek missing values, outliers, duplicates	12/6/25	12/18/25	TO DO	
	Rifki	DE	Cleaning Missing Values	Fill/remove missing values	12/6/25	12/18/25	TO DO	
	Rifki	DE	Remove Duplicates	Hapus duplikat	12/6/25	12/18/25	TO DO	
	Rifki	DE	Outlier Handling	IQR/z-score, etc	12/6/25	12/18/25	TO DO	
	Rifki	DE	Data Standarization	Normalisasi data	12/6/25	12/18/25	TO DO	
	Rifki	DE	Encoding Categorial	One-hot, label encoding	12/6/25	12/18/25	TO DO	
	Rifki	DE	Feature Engineering	Buat fitur baru	12/6/25	12/18/25	TO DO	
	Rifki	DE	Univariate EDA	Distribusi fitur	12/6/25	12/18/25	TO DO	
	Rifki	DE	Bivariate EDA	korelasi antar fitur, heatmap, dll	12/6/25	12/18/25	TO DO	
	Hutomo & Naufal	BI	Insght Sumary	BI menerjemahkan insght ke bisnis	12/6/25	12/18/25	TO DO	
	Rifki	DE	Feature Selection Strateg	Tentukan fitur untuk modeling	12/6/25	12/18/25	TO DO	
	Rifki	DE	Final Clean Dataset	Dataset siap untuk modeling	12/6/25	12/18/25	TO DO	
	Rifki	DE	EDA Report	Membuat PPT stage 1	12/6/25	12/18/25	TO DO	
	All Team	All Team	Mentoring Session	Present stage 1 ke mentor	12/18/25	12/18/25	TO DO	Link Meet
2 - Model Development	Kevin	DS	Define Modeling Strateg	Tentukan Clasification/regression, baseline metrics	12/18/25	1/10/26	TO DO	
	Kevin	DS	Train Baseline Model	Logistic baseline	12/18/25	1/10/26	TO DO	
	Kevin	DS	Train ML Model 1	melatih model a	12/18/25	1/10/26	TO DO	
	Kevin	DS	Train ML Model 2	melatih model b	12/18/25	1/10/26	TO DO	
	Kevin	DS	Train ML Model 3	melatih model c	12/18/25	1/10/26	TO DO	
	Kevin	DS	Train ML Model 4	melatih model d	12/18/25	1/10/26	TO DO	
	Kevin	DS	Train ML Model 5	melatih model e	12/18/25	1/10/26	TO DO	
	Kevin	DS	Hyperparameter TUnning		12/18/25	1/10/26	TO DO	
	Kevin	DS	Cross Validation	K-Fold Experiment, etc	12/18/25	1/10/26	TO DO	
	Kevin	DS	Model Ranking	Bandingkan semua model	12/18/25	1/10/26	TO DO	
	Kevin	DS	Model Report	Membuat PPT stage 2	12/18/25	1/10/26	TO DO	
	All Team	All Team	Mentoring Session	Present stage 2 ke mentor			TO DO	Link Meet
3 - Model Evaluation	Kevin	DS	Compute Metrics		1/10/26	1/17/26	TO DO	
	Kevin	DS	Confusion Matrix		1/10/26	1/17/26	TO DO	
	Kevin	DS	ROC	Model perfomance curves	1/10/26	1/17/26	TO DO	
	Kevin	DS	SHAP Analysis	Feature importance	1/10/26	1/17/26	TO DO	
	Hutomo & Naufal	BI	BIAS & Fairness Check	Check bias ke gender/role/etc	1/10/26	1/17/26	TO DO	
	Hutomo & Naufal	BI	BUsiness Interpretation	Translansi insght model	1/10/26	1/17/26	TO DO	
	Kevin	DS	Model Evaluation Repor	Membuat PPT stage 3	1/10/26	1/17/26	TO DO	
	All Team	All Team	Mentoring Session	Present stage 3 ke mentor	1/10/26	1/17/26	TO DO	Link Meet
4 - Deployment & Business Integration			Choose Deployment Me	Dashboard / API / Streamlit			TO DO	
			Buat API / Model Pipeline				TO DO	
			Buat dashboard	power BI?			TO DO	
			Final Presentation Deck	Membuat PPT FInal Presentation			TO DO	
			Pitching Simluation	Simulai pitching	1/24/26	1/24/26	TO DO	
			Final Day Presentation D	Present ke judge	1/31/26	1/31/26	TO DO