Date: 10/06/2025

PREDICTING JOB APPLICANT SUITABILITY

Using Machine Learning to Support Smarter Hiring

Presented by: Vanessa Wambui

OVERVIEW

- Hiring teams often review hundreds of resumes — a time-consuming, biased process
- This project uses data science to help identify the most suitable candidates quickly
- By predicting applicant suitability, recruitment can become faster, fairer, and more effective

BUSINESS UNDERSTANDING

Problem: Resume screening is slow, inconsistent, and costly

Goal: Build a reliable system to automatically flag suitable candidates

03

Value:

- **Reduce time-to-hire**
- **Improve quality of shortlisted candidates**
- **Empower HR platforms with smart tools**

DATA UNDERSTANDING

- **Total Records: ~38,000 applicants**
- **Key Information Used:**
  - **Education (level, major, university tier)**
  - **Experience (years of experience, current role, previous jobs)**
  - **Skills & Industry**
  - **Gender & Company Size**
- **Outcome: Whether a candidate is suitable (1) or not (0)**

05

WHAT MAKES A CANDIDATE SUITABLE?

Most important predictors from the model:

Years of experience

- More experience = more likely to be suitable

Relevant major and education level

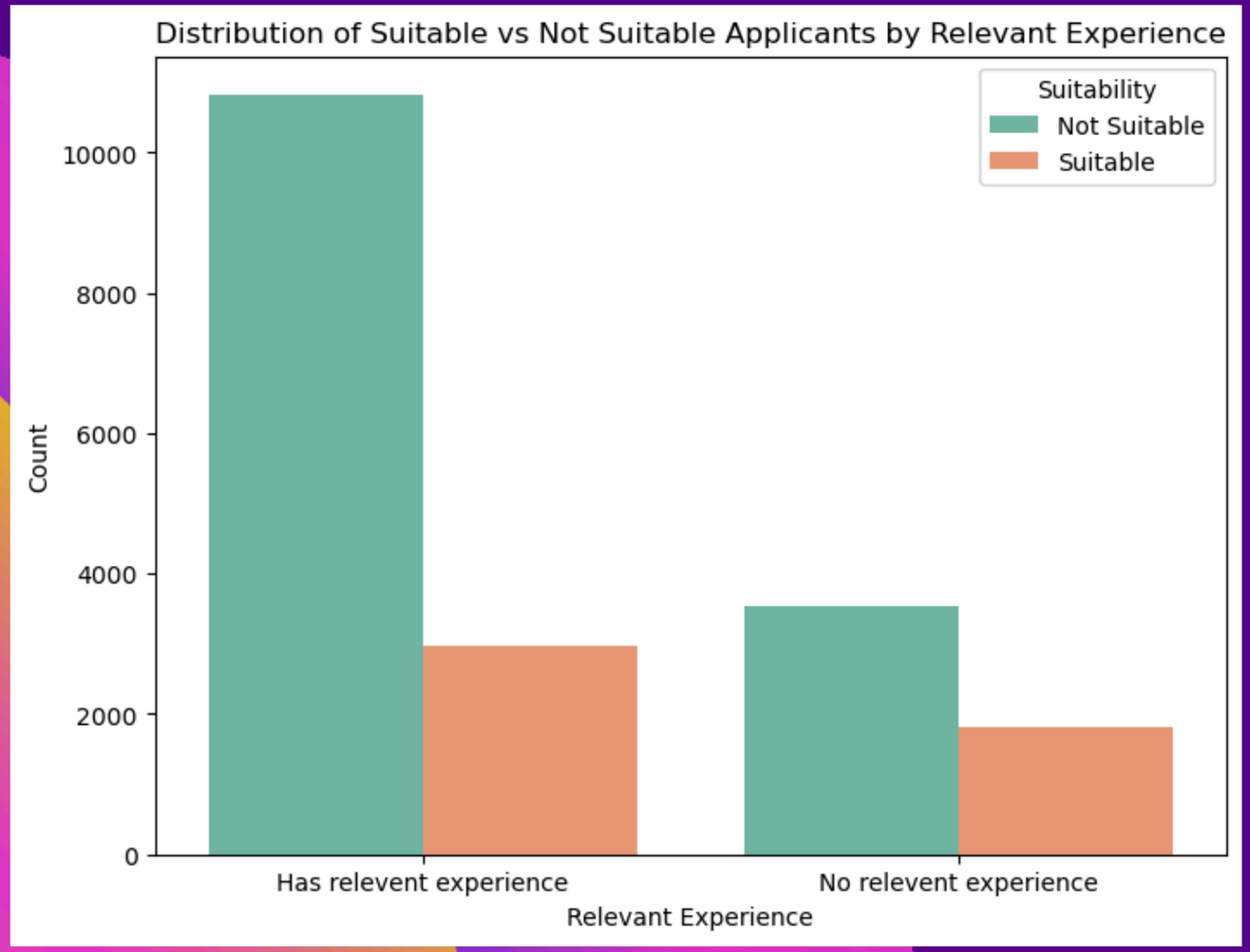
Strong academic background matters

Company type and industry

Some backgrounds align better with role requirements

Key Data Insights

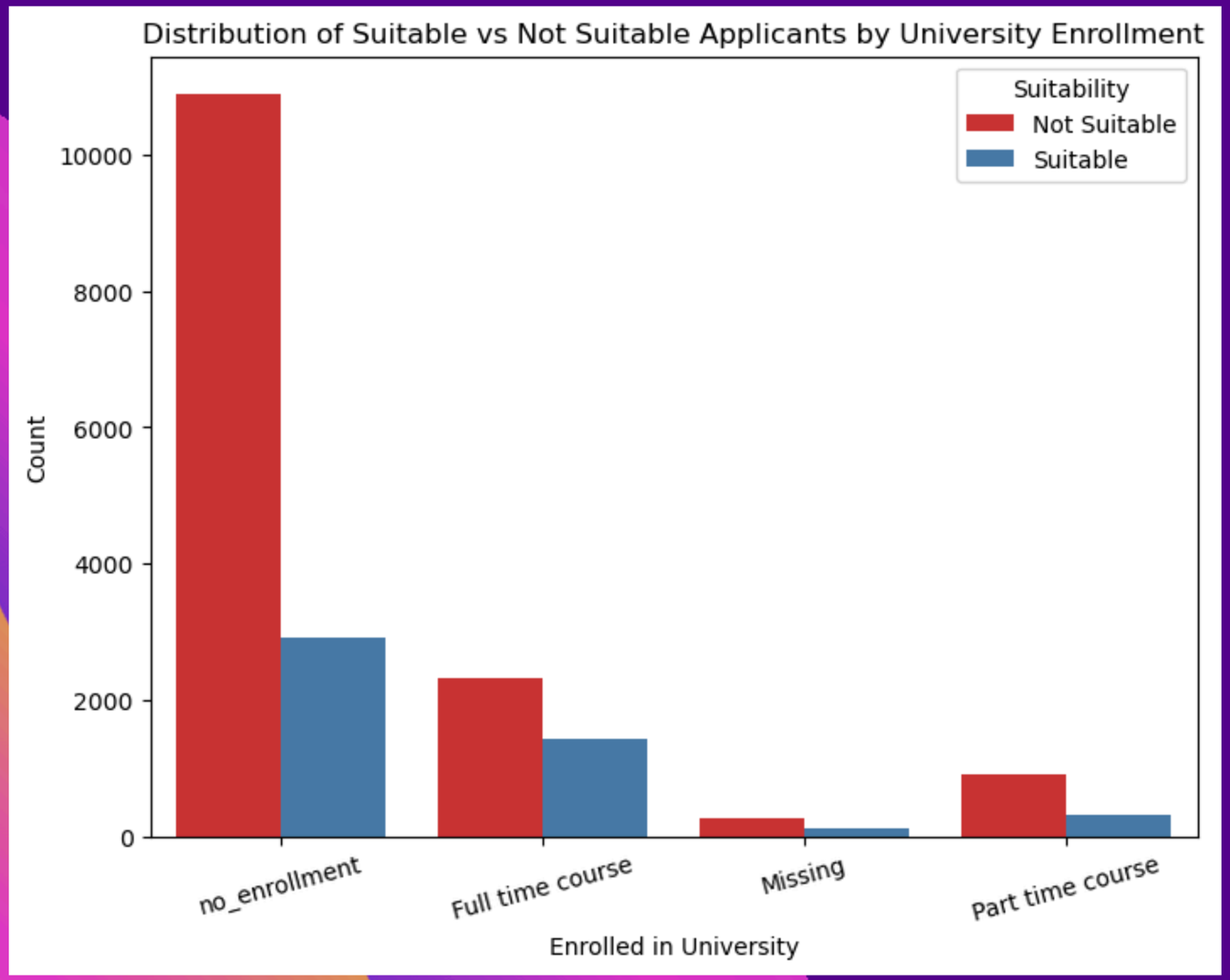
Relevant Experience vs Suitability



Candidates with relevant experience are significantly more likely to be marked as suitable.

Key Data Insights

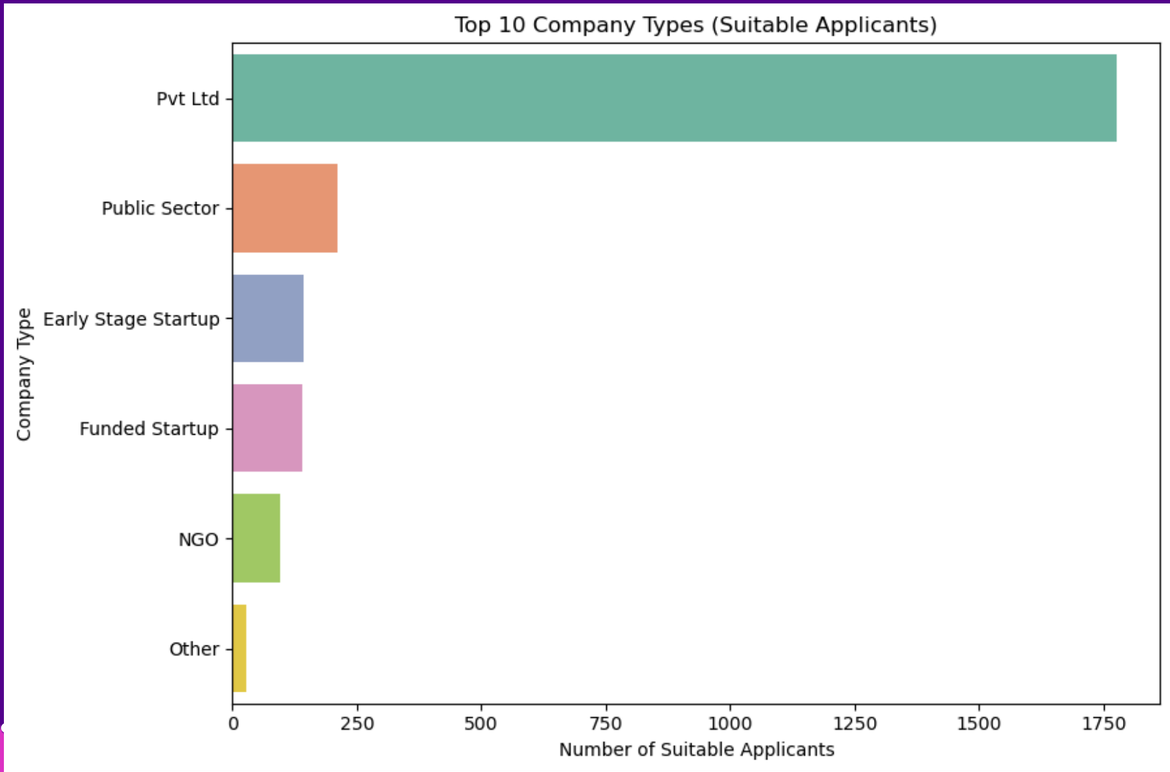
University Enrollment vs Suitability

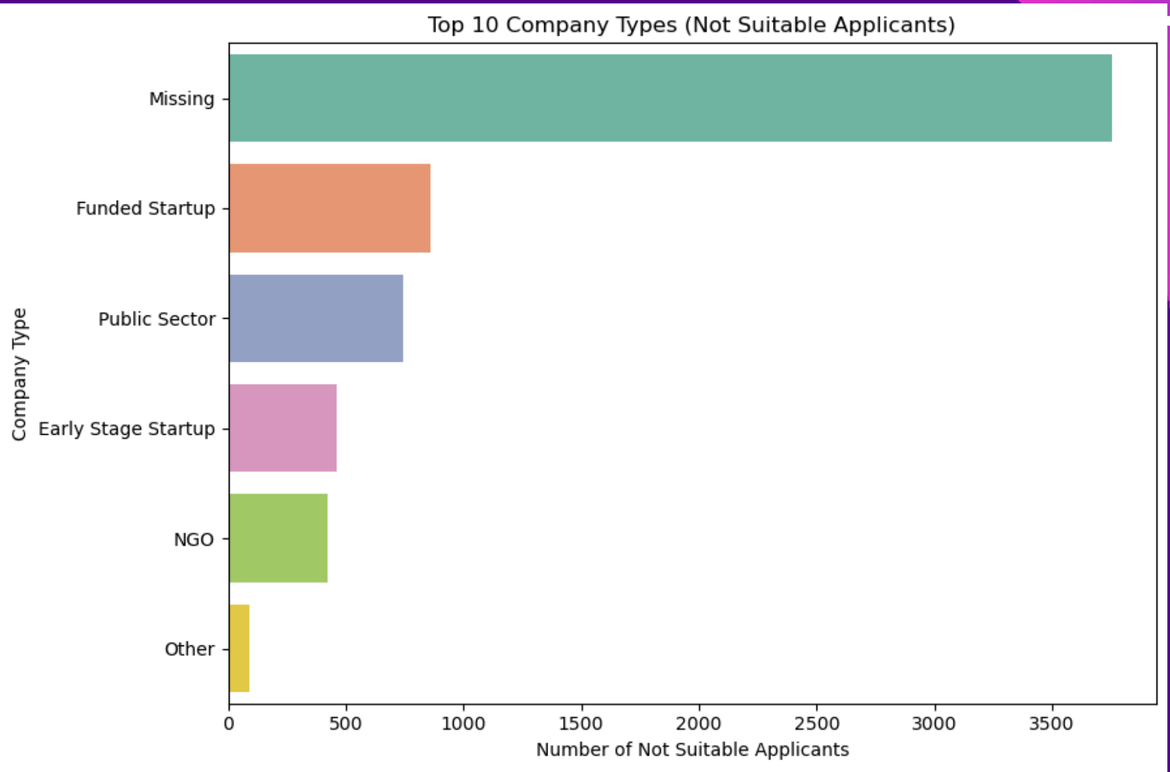


Suitability is highest among candidates enrolled in full-time courses, but most candidates are not currently enrolled meaning majority are either done with or out of school.

Key Data Insights

Company Type vs Suitability





Most suitable applicants come from private companies, while unsuitable ones often lack company information.

MODEL COMPARISON SUMMARY

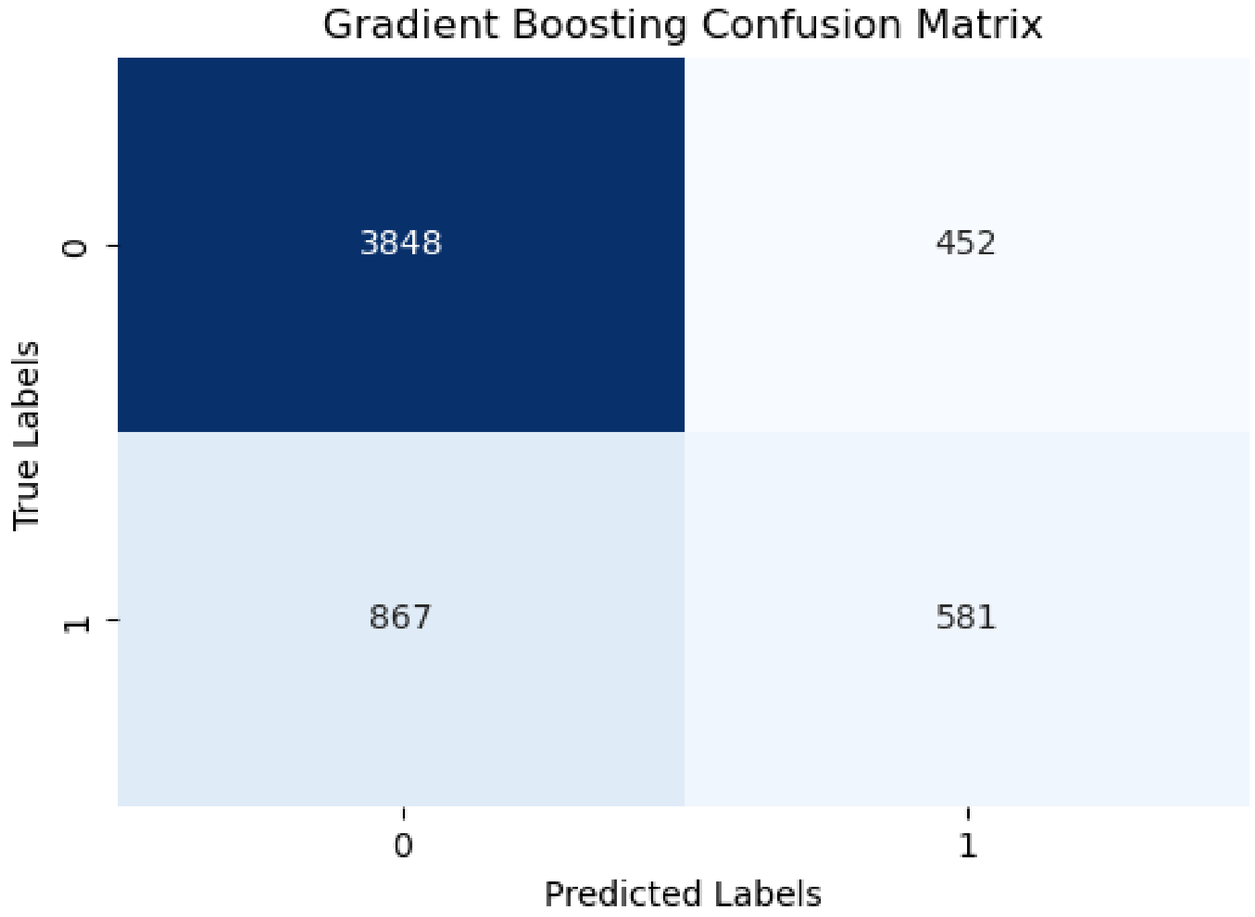
Gradient Boosting performed best across all key metrics, offering the most balanced trade-off between identifying suitable and unsuitable candidates with a recall of 89% for class 0.

Model	Accuracy	Precision (Class 1)	Recall (Class 1)	ROC-AUC
Logistic Regression	75.20%	53%	36%	0.7
Decision Tree	71.80%	49%	38%	0.66
Random Forest	76.30%	54%	39%	0.71
Gradient Boosting	**77.10%**	**56%**	**40%**	**0.72**

EVALUATION

09

BEST PERFORMING MODEL



Gradient Boosting Model Performance

- Accuracy: 77.1%
- ROC-AUC: 0.72
- Recall (Not Suitable - Class 0): 89%
- Precision (Suitable - Class 1): 56%
- Recall (Suitable - Class 1): 40%

Recommendations

What we recommend

- Adopt the Gradient Boosting model to assist in the early stages of recruitment
- Combine with human review for edge cases and final decisions
- Highlight key attributes (e.g. years of experience, education) on applicant dashboards
- Educate recruiters on the model’s limitations and strengths

VANESSA WAMBUI

+254 702 386 548

LINKEDIN

Vanessa Wambui

THANK YOU! LET'S CONNECT