Optimizing Candidate Selection Using Recruitment Data

Data Science Capstone Three Project

Hector Sanchez

Problem Statement and Context

Problem:

- Hiring high quality candidates can be difficult and extremely time consuming for Hiring/Operations Managers
- Recruitment Teams would benefit from having a predictive model that can determine which candidates are most likely to be hired based on their profiles.
- The goal of this project is to build a predictive model that can accurately classify a candidate's likelihood of being hired.

Stakeholders:

- Hiring Managers, Operations Managers, Talent Acquisition Teams, Human Resources.

Success Metrics:

- Evaluation metrics such as Accuracy, Precision, Recall, and F1-Score will be used to ensure that the model accurately predicts hireable candidates while minimizing false positive (hiring unsuitable candidates).

Dataset Description

Objective:

- This Dataset contains essential recruiting information which can be used to predict hiring outcomes

Dataset Overview

- Source: Predicting Hiring Decisions in Recruitment Data
- Key Features Included:
  - Demographic: Age, Gender, DistanceFromCompany
  - Qualifications: EducationLevel, SkillScore, PersaonalityScore, InterviewScore
  - Experience: ExperienceYears, PreviousCompaniesWorked
  - Target: HiringDecision

Data Wrangling and Preprocessing

Data Cleaning

- Verified the absence of missing values or duplicates

Preprocessing:

- One-hot encoded categorical RecruitmentStrategy
- Applied code mapping to RecruitmentStrategy after one-hot encoding
- Applied code mapping to Gender and EducationLevel to improve interpretability

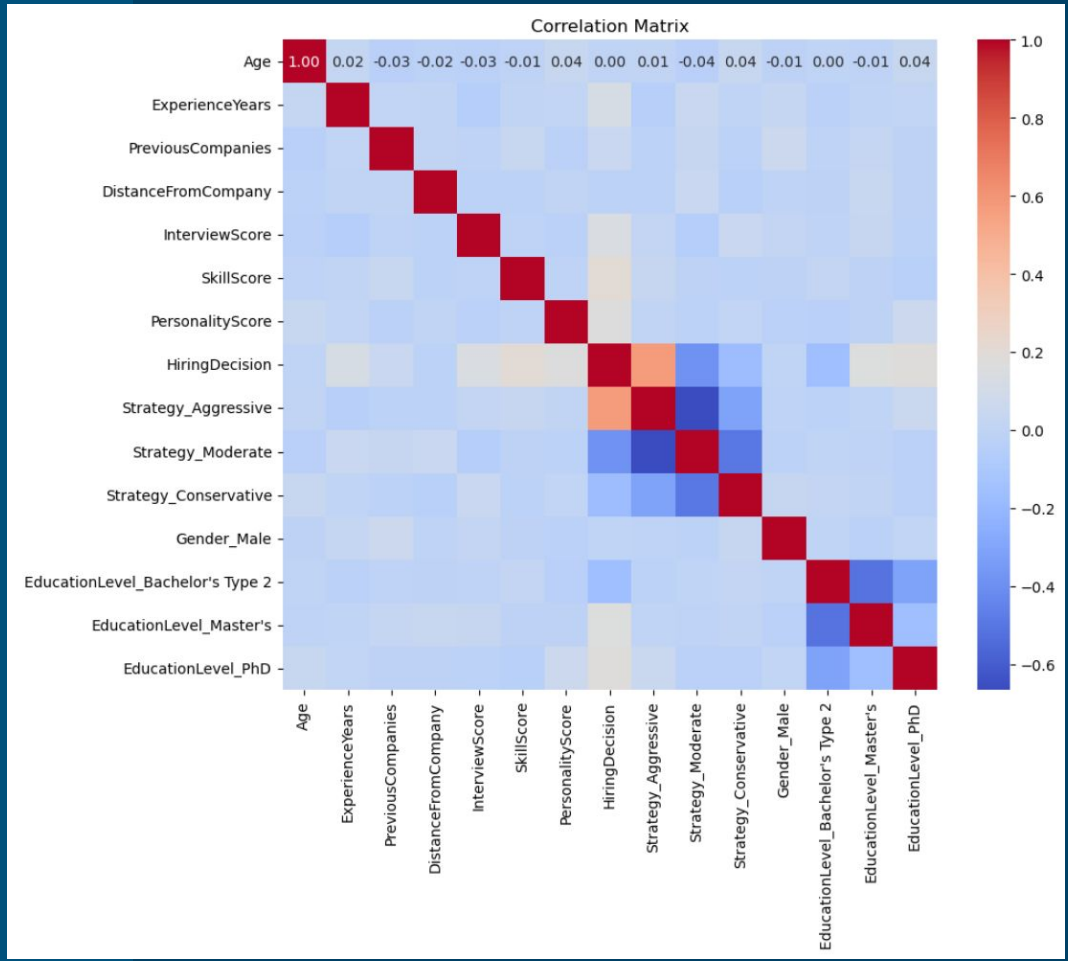
Exploratory Data Analysis

Key Visualizations and Insights

- Features such as Strategy_Aggressive, InterviewScore, PersonalityScore, and SkillScore have a positive correlation with HiringDecision
- Recruiters should place an emphasis on evaluating candidates performance in interview/personality/skill metrics.
- Aggressive recruitment strategies demonstrates the highest success
    - Conservative strategies could be useful for senior/niche roles.

Correlation Matrix

- **Positively Correlated:** Strategy_Aggressive
- **Slightly Postively Correlated:** EducationLevel_Master's, EducationLevel_PhD, InterviewScore, SkillScore, PersonalityScore, ExperienceYears
- **Negatively Correlated:** Strategy_Moderate, Strategy_Conservative, EducationLevel_Bachelor's Type 2



Modeling Approach

Selected Models

- Logistic Regression:
  - Baseline Model
- Random Forest Classifier
  - Versatile Model that handles both categorical/numerical features
- Support Vector Classifier
  - Flexible Model that reduces overfitting

Hyperparameter Tuning

- GridSearchCV to optimize Random Forest
- RandomSearchCV to optimize Support Vector Classifier

Random Forest showed the most balanced performance metrics

Model Evaluation Metric

Comparison Table

- Random Forest provides the highest Accuracy (0.923) and F1-Score (0.861) which makes it the most balanced choice

Comparison of Model Evaluation Metrics				
Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Logistic Regression	0.913	0.839	0.859	0.849
Best Random Forest	**0.923**	**0.887**	**0.835**	**0.861**
Best Support Vector Classifier	0.907	0.835	0.835	0.835

Results and Recommendations

- Apply model to Applicant Tracking Systems to help recruiters sceen candidates
- Focus recruitment efforts on candidate's scoring with interview/personality/skill metrics
- Adjust Canddidate selection criteria based on the role and a candidate's level of education
- Begin efforts for employee retention by considering candidate's commutes

Future Work

Next Steps:

- Collect more data to expand the dataset and account for shifts in hiring trends
- Wrok on incorporating features such as TrainingCapabilites or RemoteWork to provide a more comprehensive snapshot of current hiring trends
- Consider performing a cost analysis to measure the financial impact of hiring unsuitable candidates (false positives)
- Explore more Advanced models by implementing ensemble methods or neural networks
- Gather feedback from stakeholders to determine utility and usability of the model

Conclusion

In this analysis we

- Identified key features that influence hiring decisions
- Built 3 predictive models to help in the recruitment process
- Random Forest was the most balanced and suitable model for predicting hiring decisions

Closing Note

- Utilizing data to help streamline the recruitment process proves to be useful for the long term success of organizations that seek to build strong/effective teams

Thank You!