Capstone Three - Project Proposal - Hector Sanchez

Optimizing Candidate Selection Using Recruitment Data

Problem Statement:

Hiring and Operations managers can often find it difficult to find high quality candidates to join their teams. Recruiting top talent requires a lot of energy, time, and attention to detail. In many cases, hiring decisions are made out of desperation due to staffing issues which can often have adverse effects on the performance of businesses on the front line, regardless of the industry. These hiring and operations managers would benefit greatly from having a predictive model that can determine which candidates are most likely to be hired based on their profiles. This would allow leaders to focus their efforts on finding high-potential candidates during the recruitment process.

Proposed Solution:

Develop a predictive model that assesses candidates based on key features, which will help leaders prioritize those candidates who align well with the company's hiring patterns.

Context:

Job boards and other hiring platforms have made it much easier for candidates to complete job applications. While the application process has been streamlined, it also means that hiring and operations managers now tend to be faced with significantly larger applicant pools. As a result, it can be incredibly difficult to find which candidates are best fit for the role in an efficient manner. In most cases, hiring decisions need to be made quickly in order to fill critical front line roles, and also to ensure that highly

applicants aren’t hired elsewhere. Due to this, it is incredibly valuable to improve the speed of recruitment and overall hiring success rates. Streamlining this process would offer increases in overall productivity and team performance/morale.

Criteria for Success:

The success of this project will be measured by the following:

- The model’s ability to accurately predict which candidates are most likely to be hired with a high degree of precision and recall.
- Actionable insights derived from the model that can help hiring and operations managers optimize their hiring strategies/efficiency.
- A clear and interpretable model that can be integrated into real-world companies that are facing staffing issues and need to hire quickly and effectively.

Scope of Solution Space:

The solution will focus on developing an end-to-end machine learning pipeline that includes data preprocessing, feature engineering, model selection, and evaluation. We will focus on using machine learning algorithms such as logistic regression, decision trees, or ensemble methods to classify hiring likelihood. This model will provide hiring teams with candidate recommendations and insights on which candidates' features influence hiring decisions the most.

Constraints:

- Data quality could harm the performance of our model. We could face issues such as having missing values due to incomplete records. We would need to address this by applying data cleaning techniques.
- Data imbalance could also be an issue regarding the values found in the “hiring decision” column. If we have an imbalance between candidates that were hired and not hired, we’ll have to consider using techniques like SMOTE to balance the classes. This would improve the model’s reliability.

- Feature consistency could also be an issue if the values found in “Interview Score” and “Skill Score” were measured inconsistently or with a lot of subjectivity.
- We would need to ensure that the model is simple enough for Hiring and Operations managers to be able to navigate and apply the model to their recruitment strategies.

Stakeholders:

The primary stakeholders include:

- Hiring Managers, Operations Managers, and Talent Acquisition Teams.
- Field level leaders and general employees since they would highly benefit from having high quality, high performing team members.
- Data science and analytics teams responsible for building and maintaining predictive models.

Datasets of Interest:

- Predicting Hiring Decisions in Recruitment Data
- **Author: Rabie El Kharoua**
- **This dataset is synthetic, is original, and is owned by the author listed above.**
    - This dataset contains several features/columns that relate to the hiring process. Here is a list of the features found in the data:
        - Age
        - Gender
        - EducationLevel
        - ExperienceYears
        - PreviousCompaniesWorked
        - DistanceFromCompany
        - InterviewScore
        - SkillScore
        - PersonalityScore
        - RecruitmentStrategy
        - HiringDecision
    - The dataset’s features ultimately make it ideal for exploring machine learning models that can help companies optimize recruitment processes.

Methods/Approach:

- **Data Collection/Acquisition & Cleaning**:
    - The initial step will involve gathering and cleaning the relevant datasets
    - Ensure that missing data is handled
    - Features are structured in a meaningful way
    - Normalize/standardize features, and one hot encode categorical variables if needed.
- **Exploratory Data Analysis (EDA)**:
    - Perform EDA to uncover trends, correlations, and patterns in candidate features in relation to hiring decisions.
    - Visualize the distributions of the key features that could strongly influence hiring decisions.
    - Leverage correlation heatmaps and scatterplots to distinguish any relationships that can help us with feature selection and engineering.
- **Data Preprocessing**:
    - Divide the data into training and test sets.
    - Reserve a portion of the data to use for validation.
    - Improve the model's performance by using cross-validation.
- **Model Development**:
    - Begin by using models such as Logistic Regression or decision trees.
    - Implement other more complex models such as random forest or gradient boosting if evaluation metrics aren't ideal with the first two models.
    - Perform hyperparameter tuning to optimize each model's performance. Cross validation can be used to avoid overfitting.
    - Ensure that the model is easy for recruitment teams to understand.
- **Model Evaluation**:
    - Evaluate each model based on accuracy, precision, and recall scores to select the best performing model.
    - Precision and recall will help with balancing false positives and false negatives in candidate selection.
    - Compare models to identify the top performer.
- **Insights & Recommendations**:

    - Generate actionable insights and provide recommendations to recruitment teams on how to leverage the model's predictions for optimizing their hiring decisions.

Goals/Deliverables:

- A model that predicts hiring decisions baked on candidate profiles with relevant performance metrics.
- Provide completed Jupyter notebooks that display code, data processing pipelines, and analysis (do this all via a GitHub repository).
- Provide a detailed final project report that highlights the technical approach, methodology, findings, and business implications related to our hiring predictive model.
- Create a slide deck that is easy to follow highlights the project's problem, approach, key results, and key insights for recruitment teams.
- Put together a detailed project report and compelling, easy to follow slide deck that presents the findings, methodologies, and business implications.