Python Final Project

Team 3
Eshitha, Paramesh, Nidhi and Sharon

Table of Contents

01 Introduction

02 Project Pipeline

03 Data Points and Feature Selection

04 Database

05 ML Algorithms used

06 Video & Code Demo

07 Challenges faced

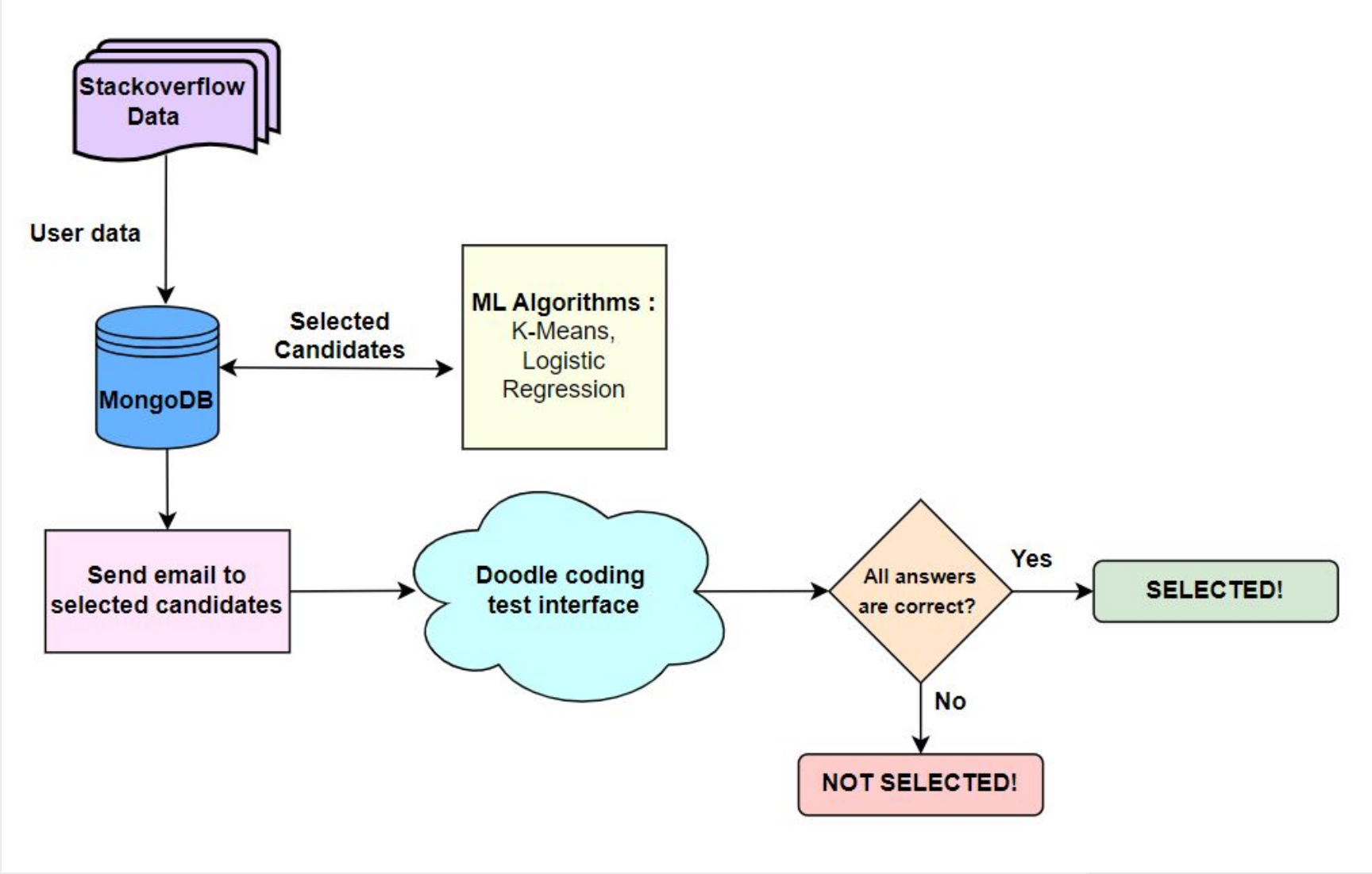
08 Improvements

Introduction

Doodle, a global software company, is dedicated to innovation and excellence. Now they are trying to transform Talent Acquisition.

The Objective: To to automate our candidate selection process, making it both efficient and discreet, while still ensuring the highest quality of talent acquisition.

Project Pipeline



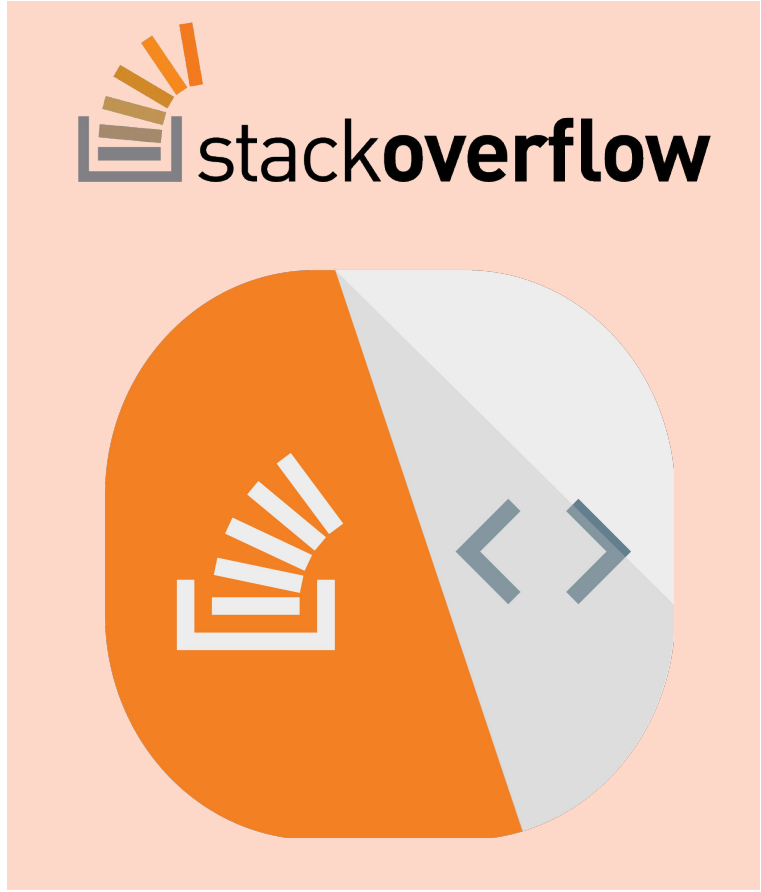
Data Points and Feature Selection

Why Stack overflow?

- Stack Overflow provides publicly available data is ideal for candidate scraping due to rich technical profiles, demonstrated skills, and a global talent pool.

For Selection criteria

- Here we used the features of “**badge_count**”, **“repetition”, “view count”, “answer_count” and “question_count”** to determine the best candidate for the position.



MongoDB

- We choose to use MongoDB as our database as it is designed to scale out horizontally. It can handle large amounts of data and high read/write workloads

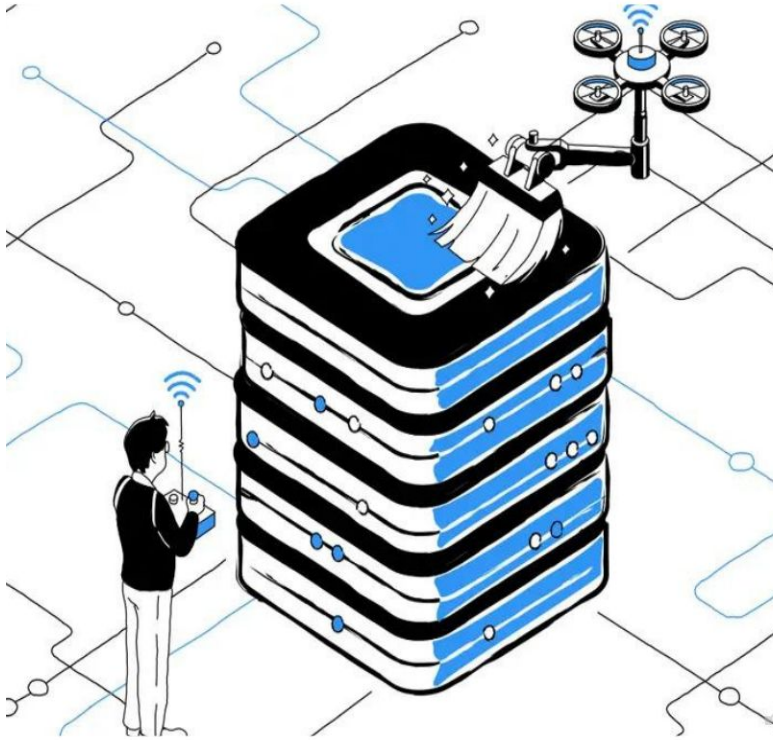
- The other purpose we used it was because MongoDB is a NoSQL based database, which means it's schema-less. We can add records in MongoDB without defining a strict schema beforehand. This flexibility is useful when dealing with semi-structured or unstructured data.

Data Preprocessing techniques

Correlation: Assists in feature selection, avoiding strong relationships between variables & data validation, enhancing data reliability & usability.

- In our case we dropped all the values with higher correlation to avoid data redundancy and overfitting.
- Low correlated values are less linearly dependent on each other and are more independent allowing to capture different patterns within data.

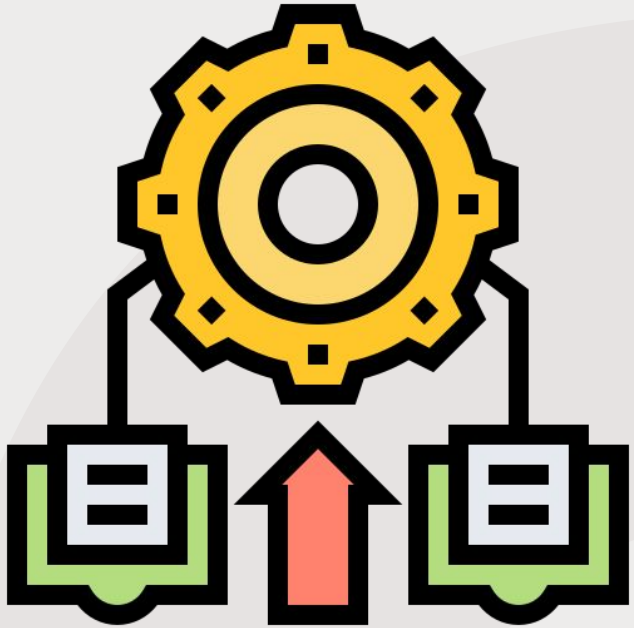
Standardization: Equalizes the scale enabling equitable comparisons.



ML Algorithms

- **K-means**: We used this unsupervised technique for the purpose of candidate selection which will group candidates into clusters based on similarities in their qualifications, aiding in the identification of distinct talent pools and improving the efficiency of recruitment processes.

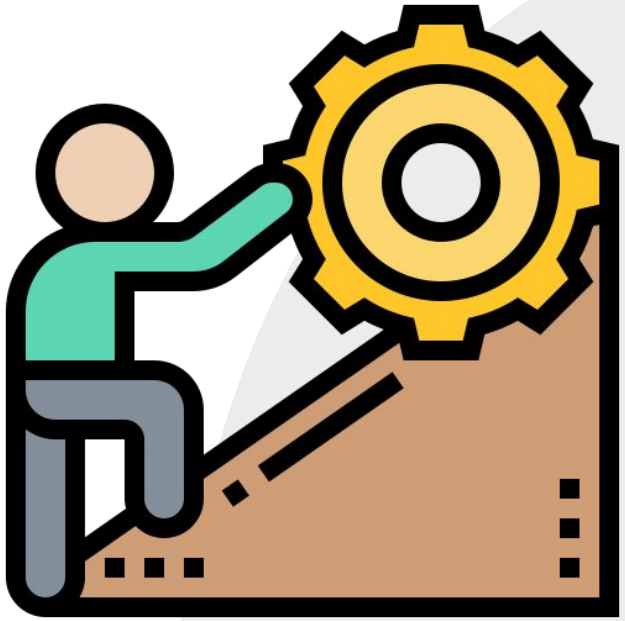
- **Logistic Regression**: We used this supervised technique for the purpose of candidate selection to predict whether candidates meet specific criteria or not, enabling automated screening and efficient identification of suitable candidates for further evaluation in recruitment processes.



Video & Code Demo

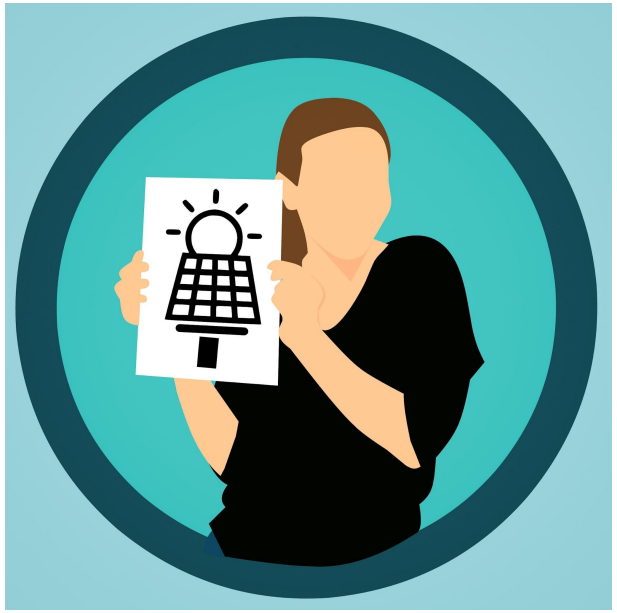
Challenges Faced

- Scraping certain datasource due to API limitations.
- Lack of skills.
- Time management.



Improvements & Future Implementations

- Scrape other sites that contain more user information like Linkedin to make selection based on skills and experience.
- Have a more interactive and better UX design implemented.
- Include a dashboard for recruiting managers to view answers and make more accurate decisions.
- Every coding test should be different and unique for every selected candidate.



Vielen Dank für deine Aufmerksamkeit!