Interview Status Prediction

Author:- Rohith Narayanan S
Batch:- CIT-PGA-001

The Business Problem

Problem Abstract

A data scientist of a big MNC who usually hires more than 10k candidates every year. To complete the task they conduct more than 1 lakhs of interviews every year.

Problem statement

Need to predict the status of the interviews so that recruiters can check the sanity of the interview and find if the interview was biased.

Target Approach

To predict the “Status” column using Classifier Machine Learning Algorithms.

Challenges Faced (Deep-Dive)

Challenge 1

Challenge 2

Challenge 3

Columns

The dataset consists of many features consisting of both numerical (Major) and categorical(Minor) values, hence feature understanding need to be done efficiently.

Feature Selection

The Correlation the columns were unique because it consists of very low positive values and high negative values except S.L.R.C & S.L.R.I Columns.

Model Predictions

By Trailing and Testing the data with ML models to get a close to perfect prediction score.

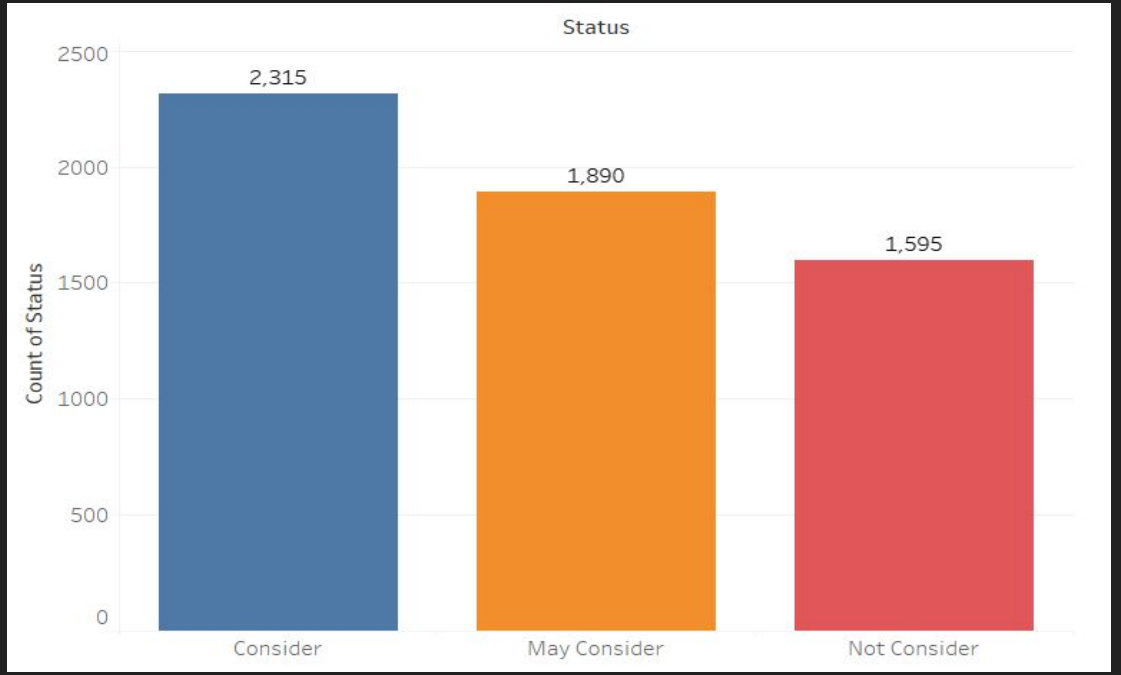
Solution

SVM Classifier

Support Vector Machines Classifier ML Algorithm was the best model when comparing with other models and considering this dataset.

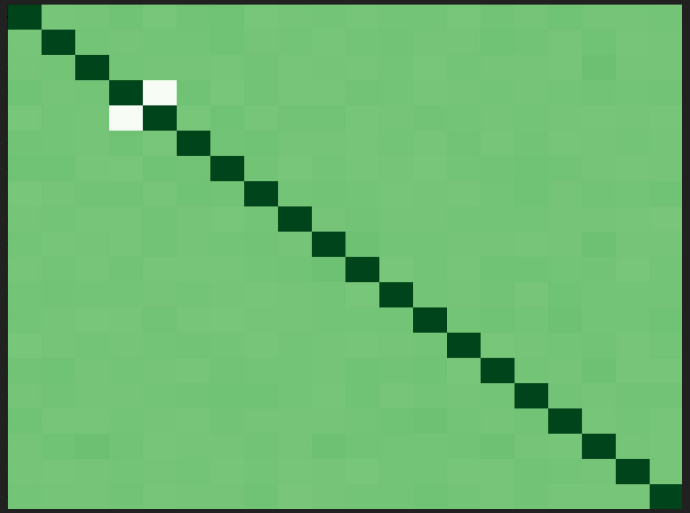
Target Feature - Status

- ➔ The Target feature “Status” has 3 unique values in it, namely ‘Consider’, ‘May Consider’ & ‘Not Consider’.
- ➔ The “Status” feature is has poor correlation with each and every feature in the dataset.
- ➔ The Majority of the candidates were considers in this company.

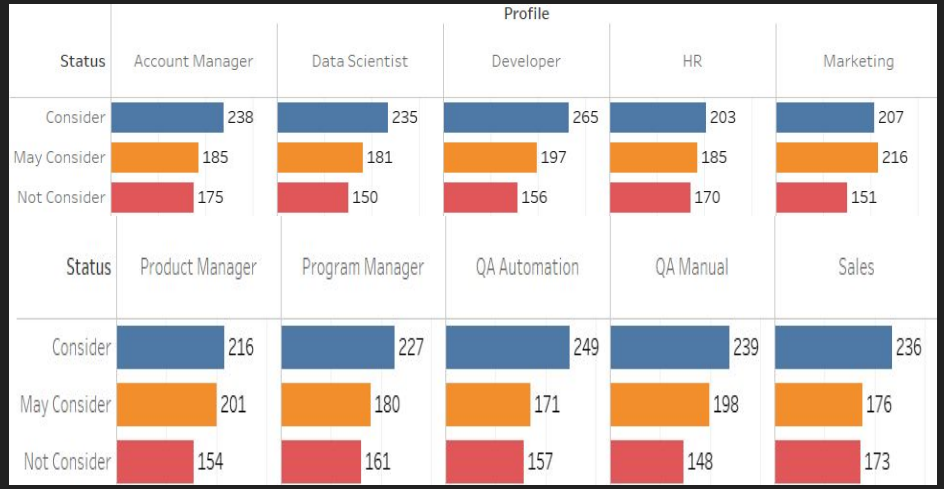


Interesting Findings

Low Data Correlation



Diversity in Data



Implementation of the Solution

Feature Understanding of the dataset

Preparing the dataset for Modeling by Encoding the columns

Splitting the dataset into train and test datasets

Cleaning the dataset of noises and outliers

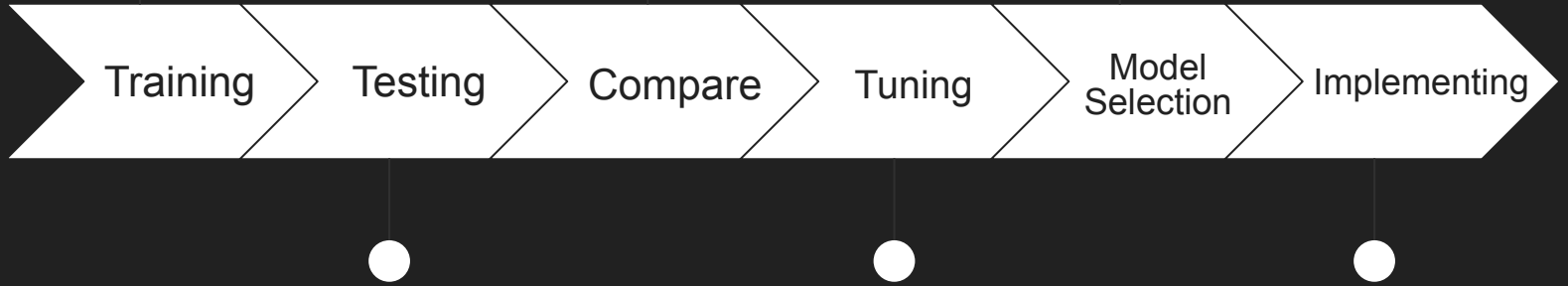
Selecting the Features based on relevance and Correlation

Building and initiating several models

Training the Models with the training dataset

Comparing each Model's training & testing score

Selecting the best Model



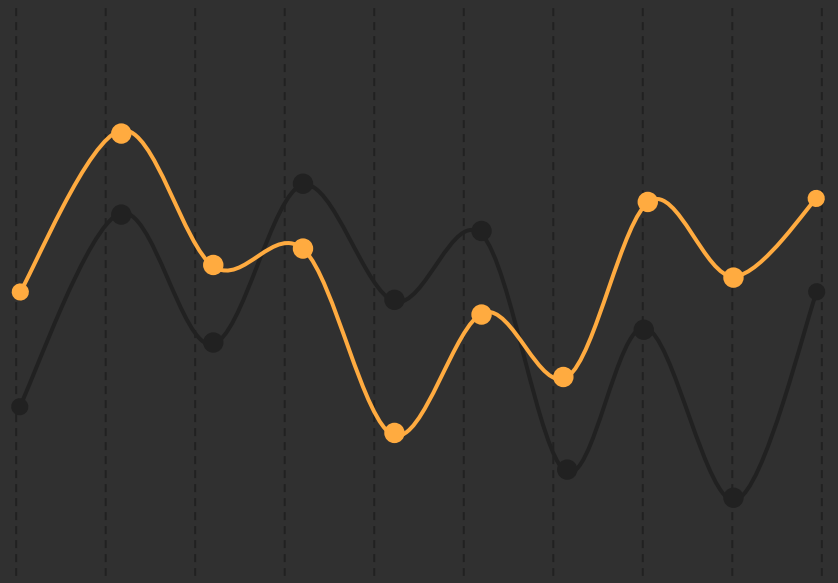
Testing the Models for accuracy

Change the Hyper Parameters of all the Models for better performance

Implementing the selected Model

Final Accuracy Score

Train Score: 0.9982758620689656
Test Score: 0.9942528735632183
Submission Score: 0.708

Source Code:
RohithNarayananS/Imarticus-Hackathon-Round-3 (github.com)

LinkedIn: linkedin.com/in/rohith-narayanan-61630318b