Project Report

(Project Id: 2017CSEPID18)

on

AI Recruiter

Submitted in Partial Fulfillment of the Requirement

For the Degree of

Bachelor of Technology

In

Computer Science and Engineering

By

TARUN AGARWAL (1729010175)
SAHYOG SAINI (1729010137)
SANDEEP KUMAR SHUKLA (1729010140)

Under the Supervision of

Prof. Avinash Kumar Sharma
(Assistant Professor)

Department of Computer Science & Engineering
ABES INSTITUTE OF TECHNOLOGY, GHAZIABAD

AFFILIATED TO
Dr A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY, UTTAR PRADESH, LUCKNOW (July-2021)

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature:
Name : TARUN AGARWAL
Roll No. : 1729010175
Date :

Signature:
Name : SANDEEP KUMAR SHUKLA
Roll No. : 1729010140
Date :

Signature:
Name: SAHYOG SAINI
Roll No.: 1729010137
Date:

CERTIFICATE

This is to certify that Project Report entitled “AI Recruiter” which is submitted by Tarun Agarwal (1729010175), Sandeep Kumar Shukla (1729010140) and Sahyog Saini(1729010137) in partial fulfilment of the requirement for the award of degree B. Tech. in Department of Computer Science & Engineering of A.K.T.U is a record of the candidate own work carried out by him under my supervision. The matter embodied in this project is original and has not been submitted for the award of any other degree.

Date:

Supervisor

Prof. Avinash Kumar Sharma

Department of Computer Science

& Engineering

ACKNOWLEDGEMENTS

It gives us a great sense of pleasure to present the report of the B. Tech Project undertaken during B. Tech. Final Year. We owe special debt of gratitude to Prof. Avinash Kumar Sharma, Department of Computer Science & Engineering, ABES Institute of Technology, Ghaziabad for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavours have seen light of the day.

We also take the opportunity to acknowledge the contribution of his for his full support and assistance during the development of the project.

We also would not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our parents and friends for their constant support throughout the project.

Signature:
Name: TARUN AGARWAL
Roll No.: 1729010175
Date:

Signature:
Name : SANDEEP KUMAR SHUKLA
Roll No. : 1729010140
Date :

Signature:
Name : SAHYOG SAINI
Roll No. : 1729010137
Date :

ABSTRACT

The purpose of this study is to develop an artificial intelligence based system, which we call "AI Recruiter" that can effectively screen candidates. In this study, we have enhanced the classification and natural language processingalgorithms to accurately classify resumes and evaluate personality of candidateswith the aim to get most suitable candidates out of all the applications. Thisis designed to bring efficiency, transparency and reliability to the process ofrecruitment in an organization and to reduce the efforts of human resourcedepartment in manually screening through all the applications. In this paper, wehave presented the architecture and design of our project, highlighting the resultsof using ensemble learning to improve resume classification human personalityevaluation.We have presented a detailed study of AI based tools presently in usein recruitment industry.We also get into details of our project through the use ofvarious diagrams tables to show the scope of our project, its uniqueness and howit is innovative in its own and better than existing solutions in unleashing the truepotential of AI in the whole process of recruitment.

Table of contents

Chapter 1

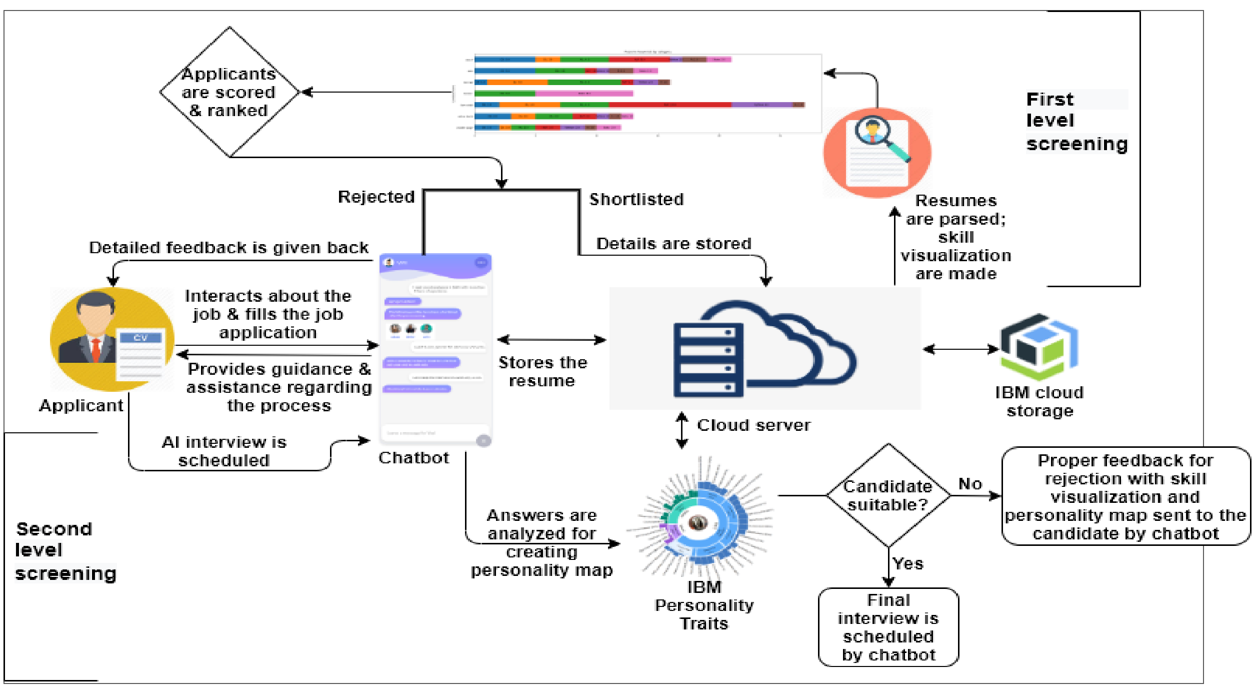
Introduction

With the increasing number of applications for each vacancy in industry, the job of recruiters are becoming tiresome day by day. In today's scenario, there is a need for development of an AI - based platform that can interact with every candidate, and can assist them in the whole recruitment process adding transparency to the process. Moreover, on the recruiter side, this platform should be able to reduce their workload by identifying potential candidates for the job role based on their skills, experience and personality evaluation.

This type of platform will add

- Accuracy to the whole process by analysing every resume and selecting most suitable candidate.
- Efficiency by reducing time and efforts of the Recruiters.
- Transparency since candidates can see every ongoing process and receive thorough feedback.

1. Shortlisting of candidates in hirings



Level I Screening -

Level I screening in our project is basically the first screening provided by us. It is basically a resume visualization based candidate screening taking into consideration relevant experience and skills as per the job role. After this level,

- Shortlisted candidates will get to the database for further screening.
- Non-shortlisted candidates will get immediate feedback via chatbot.

Level II Screening -

Level II screening is the virtual AI based interview screening that will involve analysis of personality of the candidates and selection done based on that.
In this level also,

- Shortlisted candidates data will be shared with the organization for further process.

Non-shortlisted candidates will get immediate feedback which will help them improve.

Chapter 2

Literature Survey

The use of AI in recruitment began in 2015 with the emergence of the era of *Digital Recruiting 3.0.* Around that time, due to the digital revolution in recruitments, there was a boost in the number of applications for each vacancy increasing both

- Number of Unqualified Candidates applying
- The workload on Recruiters analyzing and segregating 1000s of applications.

To be a helping hand in such situations, Digital Recruitment 3.0 came into the picture that witnessed many AI-based platforms providing a path-breaking reduction in time in various recruitment phases. These platforms are basically the base of our project. Some main players of this domain are -

- HireVue
- Mya Chatbot
- Pando Logic
- PyMetrics

Such software tools are being used today by most big small companies for a smooth recruitment process. But a big limitation that can be seen in all these products is that neither of them is a complete solution to seamless and effective shortlisting of candidates out of the whole crowd of thousands of candidates (suitable and unsuitable; qualified and unqualified) which is our problem statement for this project.

For being a complete solution, multiple existing software tools have to be used together integrated with an Application Tracking System (ATS), that

- involves a high cost to the company per hire;
- is a hassle for the HR team to manage different platforms for different levels of hiring;

A detailed study of these major existing solutions is given further in this chapter.

1. HireVue

➔ Problem statement solved -

"Video Interviews and personality analysis based on it."

➔ Introduction and working -

HireVue is an enterprise video interviewing technology vendor that allows recruiters and hiring managers to screen candidates and conduct live interviews over the Internet. HireVue's products include game-based cognitive assessments and artificial intelligence (AI) driven assessments based on job-specific competencies.

Screening of candidates using HireVue requires that the candidate login to the HireVue talent interaction interface (which can be customized to match the hiring company's brand). Candidates then look into a webcam as the system records their answers to pre-set questions. Applicants can be interviewed using a desktop computer, laptop, smartphone or tablet.

Along with voice recognition and facial recognition software, HireVue uses a software ranking algorithm to determine which candidates are ideal for a specific job. An ideal candidate is defined by a combination of traits including body language, tone, and keywords. The interview data is then compared to approximately four million video interviews of top candidates in that specific role. HireVue's intelligent assessment service analyzes an average of 25,000 data points for a single fifteen-minute video interview.

➔ Limitations -

Being a really good platform with an average rating of 6.4 out of 10 (source []), HireVue is widely used by clients like L'oreal for video interviews based assessment of candidates. But this system is useful at the interview level of the process. The beforehand process of resume-based shortlisting is still a burden for the recruiter. Secondly, for organizations, the cultural fit of candidates matters a lot which is given a backseat in this platform as evaluation is more around the stability of the candidate rather than assessing their personality traits.

2. Mya Chatbot

➔ Problem Statement solved -

"Conversational Interface that analyzes the candidate based on questions."

➔ Introduction & Working -

Mya Systems, formerly known as FirstJob, was co-founded in 2012 by Eyal Grayevsky, CEO, and James Maddox, CTO.

The San Francisco-based company launched Mya, its conversational AI-recruiter, in July of 2016 to address major inefficiencies in recruiting and improve the process for both job candidates and recruiters.

The Mya chatbot is able to ask candidates factual questions to determine whether their profiles match the position requirements:

- Are you available at the internship start date and throughout the entire internship period?
- What year are you currently in and how long is your degree course?
- Do you need us to make any reasonable adjustments in order for you to be able to complete your application?
- Do you have any questions for me regarding the company culture, application process, or position details?

User-friendly, Mya is accessible via mobile phone anytime and anywhere. From the start, Mya presents itself as a chatbot. At the end of this first phase, qualified applicants are put in contact with recruiters. Especially valuable in a world where many job seekers do not have the necessary level of information regarding their application follow-up, this technology provides ongoing reminders and updates for a more considerate, more people-centric experience.

➔ Limitations -

Mya chatbot is a pioneer chatbot in the hiring domain, provides an interactive interface & follow-ups for candidates to make their candidature smooth. But a major limitation in this solution is that it does not take into account a candidate's resume/CV which is an important aspect for any HR.

3. Pando Logic

➔ Problem Statement solved -

"AI-Based Networking platform."

➔ Introduction & working -

PandoLogic's programmatic recruitment advertising platform helps employers source talent faster and more efficiently than ever thanks to predictive algorithms, machine learning, and AI. PandoLogic enables employers to source quality applicants faster and more efficiently through the use of big data, artificial intelligence, and proprietary campaign algorithms that fully automate and optimize the job advertising process from job classification and targeted distribution to budget allocation and dynamic bidding across diverse job categories.

Their talent acquisition solutions connect publishers, employers, job board operators, and job seekers across a shared talent network using the most advanced programmatic platform available.

PandoIQ **intelligently automates and optimizes job advertising spend.** Companies faced with complex hiring needs now can source quality applicants faster, smarter, and more efficiently with our unique proprietary technology. Target relevant candidates across all the major search recruitment sites including Indeed, CraigsList, LinkedIn, Monster, CareerBuilder, ZipRecruiter, and hundreds of others and spend accordingly. pandoIQ provides an **end-to-end job advertising solution that delivers a significant increase in job ad performance without any wasteful spending**.

➔ Limitations -

Pando Logic is a solution to smartly advertise the jobs & to reach desired candidates. This solution neither gives any feature to smartly shortlist or select the most suitable candidate out of the crowd applied, nor it stops any unsuitable candidate from applying.

4. PyMetrics

- Problem Statement solved -

  "Neuroscience games based candidate personality evaluation."

- Introduction & working -

  Pymetrics is a series of behavior-based neuroscience games that objectively measure cognitive, social, and behavioral attributes. Pymetrics leverages advanced technology to help look beyond your resume and understand more about your natural strengths to identify which roles you are most likely to succeed in.

  Candidates play a series of 12 games that measure the attributes that make you, you! There are no right or wrong answers - we're looking at how you approach problems, not specific solutions.

  After playing the pymetrics games, he/she receives a detailed, personalized report for the talents you already possess. Pymetrics gives everyone the opportunity to shine, beyond what's on paper.

- Limitations -

  Pymetrics is really an interesting tool used by top recruiters including Unilever, Nielsen, LinkedIn, Accenture, KraftHeinz, MasterCard, and Boston Consulting Group. It is amazing gamification in the hiring process, which helps select the most analytical candidates based on scores in its analytical and psychometric games. But this platform is again not a complete solution but just a subset of solutions that has to be used with other leading solutions and an ATS for supporting a complete effective hiring experience. This includes the hassle of managing different platforms and involves high costs.

Chapter - 3

Problem Formulation

Current Process of Shortlisting candidates -

Current hiring practices make use of software tools to shortlist the most suitable candidates. With tools such as Pymetrics, selecting the best candidates based on their psychometric skills is really a revolutionizing step.

An overview of a general process of hiring is -

- A job post for a given number of vacancies for a particular role is posted on the company's career page and major professional networking platforms by the HR department.

- Anyone who sees that post & wants to apply for the same can do so by submitting basic details and their resumes / CVs (which is considered as the crux of all the information needed to evaluate whether a candidate has the right set of skills for the job).

- Based on resumes, recruiters shortlist a subset of all the applied candidates which then has to go through a set of assessments and interviews.

- Finally, the candidates shortlisted are sent their offer letters

This process of recruitment(with minor changes) is being used by all the major firms of the world for filling their vacancies.

But this process is actually the “at the wit’s end” for the recruiters. This system has some basic yet major drawbacks which are being discussed in the next subsection.

Limitations & Challenges faced by the current system -

With the rapid digitization of the recruitment process, any job vacancy can now reach a larger audience through platforms like Linkedin and since it costs nothing to apply to a job, everyone (qualified or unqualified) applies to it by submitting their resumes. Some stats in support of this argument are -

- In 2013, Walmart, the largest private employer on the planet, received on average 23,000 applications for 600 positions when it opened a new store.
- In 2017, Johnson Johnson generated over 1 million applications for 28,000 positions.
- In 2017, Google generated an estimated 2 million applications for just 14,500 jobs [4], meaning that it was nearly 10 times more difficult to get a job at Google than to get into Harvard University.

The main challenges faced by recruiters with this digitized current system of recruitment are -

1. A large number of applications for fewer vacancies, which directly implies that the organization either has to take a longer time to review applications or hire an army of recruiters to review them;
2. Studies show that nearly 75% - 80% of all applications received for a job are unqualified for it, which means that the time spent on reviewing these resumes is spent in vain;
3. Recruiters do not consider all the resumes, rather take some initial 100 resumes or only those resumes with someone's referral attached to it, for others, no feedback is ever given or an automated mail of rejection is sent;
4. To speed up the process of hiring, many software products are available in the market but none of them are complete solutions, i.e. they only solve the part of the problem and hence multiple tools have to be used together which requires high costs to the organization and recruiters also need to understand the working of all these platforms;
5. Even after all these attempts, this process is not transparent at all; Candidates who are not shortlisted do not get any rejection mail, or a report as to why they are not selected, which can be crucial for their self-development;

This background and context turned AI-enabled recruiting from **nice-to-have to necessary-to-employ**. AI-enabled recruiting tools have primarily been employed across four general sets of activities: **outreach, screening, assessment, and coordination**. Understanding this problem, we have decided to develop an AI-based platform that can help in all 4 defined recruitment activities and provide a seamless experience for recruiters.

Problem statement & Objective of this project -

➔ Problem Statement -

With the increasing number of applications for each vacancy in the industry, the job of recruiters is becoming tiresome day by day. In today's scenario, there is a need for the development of an AI-based platform that can interact with every candidate and can assist them in the whole recruitment process adding transparency to the process. Moreover, on the recruiter side, this platform should be able to reduce their workload by identifying potential candidates for the job role based on their skills, experience, and personality evaluation.

This type of platform will add

- **Accuracy** to the whole process by analyzing every resume and selecting the most suitable candidate.
- **Efficiency** by reducing time and efforts of the Recruiters.
- **Transparency** since candidates can see every ongoing process and receive thorough feedback.

➔ The objective of this project -

"AI Recruiter" is a system designed and developed to **assist organizations in their recruiting process**. It connects both the recruiting team and the candidates in an efficient manner so that the whole **process gets more robust and seamless**.

A recruiter is the **combination of AI-powered chatbot and backend algorithms** that performs the following objectives-

- **Create first interaction with each candidate through some basic questions to capture and analyze his/her impression about the organization and job role.**
- **Screening and shortlisting of candidates based on their first interaction & resumes is another important task done by it.**
- **AI recruiters will also interview the shortlisted candidates and try to bring out their key personality traits.**
- **Another task of importance is providing feedback to non-shortlisted candidates thus satisfying them too.**

<u>Chapter - 4</u>

AI-based decision making

<u>Introduction to ML-based classification</u> -

Basic Definition

Classification or decision-making is an activity we perform most often. Humans tend to make decisions, what to purchase, what to see, what to eat, what to wear etc. Making efficient decisions needs a proper evaluation of the problem and choices and thus data plays a major role in making a decision.

ML (abbreviation for Machine Learning) is a technique in which we use mathematical formulae to train a machine-based model on input and output, such that it will be able to identify data patterns and is used to predict output for unforeseen inputs. And classification is a major application of ML.

ML-based classification is the process of predicting the class of given data points. Classes are sometimes called **targets/ labels or categories**. Classification predictive modeling is the task of <u>approximating a mapping function (f) from input variables (X) to discrete output variables (y)</u>". For example, spam detection in email service providers can be identified as a classification problem. This is binary classification since there are only 2 classes as spam and not spam. A classifier utilizes some training data to understand how given input variables relate to the class. In this case, known spam and non-spam emails have to be used as the training data. When the classifier is trained accurately, it can be used to detect an unknown email. Classification belongs to the category of **supervised learning** where the targets are also provided with the input data. There are many applications in classification in many domains such as in credit approval, medical diagnosis, target marketing, etc.

In recruitment, recruiters have candidate's data in the form of resumes so that hiring decisions can be made but such a large amount of data poses a problem. Thousands of resumes are very tiresome to be processed thoroughly so that the most suitable candidates can be chosen out of the crowd. Hence additional help in the form of ML-based classification algorithms came to the rescue in such situations.

Key Terminology

- Classifier: An algorithm that maps the input data to a specific category.
- Classification model: A classification model tries to draw some conclusions from the input values given for training. It will predict the class labels/categories for the new data.
- Feature: A feature is an individual measurable property of a phenomenon being observed.
- Binary Classification: Classification task with two possible outcomes. Eg: Gender classification (Male / Female).
- Multiclass Classification: Classification with more than two classes. In multi-class classification, each sample is assigned to one and only one target label. Eg: An animal can be a cat or dog but not both at the same time.
- Multilabel Classification: Classification task where each sample is mapped to a set of target labels (more than one class). Eg: A news article can be about sports, a person, and a location at the same time.

Steps involved in building a Classification model

➔ **Initialize** the classifier to be used.

➔ **Train the classifier:** All classifiers fit the model(training) for the given train data X and train label y.

➔ **Predict the target:** Given an unlabeled observation X, the predict(X) returns the predicted label y.

➔ **Evaluate** the classifier model.

ML algorithms for classification -

From time to time, many algorithms have been developed based on different mathematical techniques such as probability theory.
All the classification algorithms can be mainly of 2 types: **Lazy Learners & Eager Learners.**

Lazy Learners - Lazy learners simply store the training data and wait until testing data appears. When it does, classification is conducted based on the most related data in the stored training data. Compared to eager learners, lazy learners have less training time but more time in predicting.
E.g.: *K-nearest Neighbours*

Eager Learners - Eager learners construct a classification model based on the given training data before receiving data for classification. It must be able to commit to a single hypothesis that covers the entire instance space. Due to the model construction, eager learners take a long time to train and less time to predict.
E.g.: *Decision Trees, ANN, Naive-Bayes Classifier*.

There are 7 main classification algorithms -

1. Logistic Regression
2. Naive Bayes Classifier
3. Stochastic Gradient Descent
4. K-nearest neighbor (KNN)
5. Decision Tree
6. Random Forest
7. Artificial Neural Network (ANN)
8. Support Vector Machine (SVM)

These algorithms are discussed in brief further;

1. Logistic Regression

It is a classification algorithm in machine learning that uses one or more independent variables to determine an outcome. The outcome is measured with a dichotomous variable meaning it will have only two possible outcomes.

The goal of logistic regression is to find a best-fitting relationship between the dependent variable and a set of independent variables. It is better than other binary classification algorithms like nearest neighbor since it quantitatively explains the factors leading to classification.

Logistic Regression equation: $\boldsymbol{p = 1 / 1 + e^{-(\beta 0 + \beta 1X1 + \beta 2X2 \ldots. + \beta nXn)}}$

Logistic regression is specifically meant for classification, it is useful in understanding how a set of independent variables affect the outcome of the dependent variable.

The main disadvantage of the logistic regression algorithm is that it only works when the predicted variable is binary, it assumes that the data is free of missing values and assumes that the predictors are independent of each other.

The main use cases are -

- Identifying risk factors for diseases
- Word classification
- Weather Prediction
- Voting Applications

2. Naive Bayes Classifier

It is a classification algorithm based on Bayes's theorem which gives an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

Even if the features depend on each other, all of these properties contribute to the probability independently. The naive Bayes model is easy to make and is particularly useful for comparatively large data sets. Even with a simplistic approach, Naive Bayes is known to outperform most of the classification methods in machine learning.
Following is the Bayes theorem to implement the Naive Bayes Theorem.

$$\boxed{P(C_i|\ x_1, x_2 \dots, x_n) = \frac{P(x_1, x_2 \dots, x_n|C_i).P(C_i)}{P(x_1, x_2 \dots, x_n)}\ for\ 1 < i < k}$$

The Naive Bayes classifier **requires a small amount of training dat**a to estimate the necessary parameters to get the results. They are **extremely fast in nature** compared to other classifiers.

The only disadvantage is that **they are known to be bad estimators**.

The main use cases are -

- Disease Predictions
- Document Classification
- Spam Filters
- Sentiment Analysis

3. Stochastic Gradient Descent

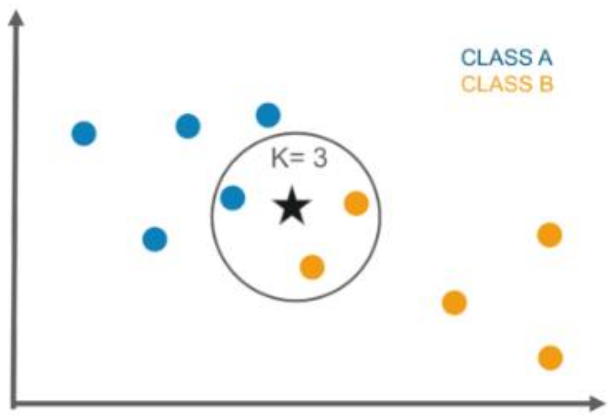
It is a very effective and simple approach to fit linear models. Stochastic Gradient Descent is particularly useful when the sample data is in a large number. It supports different loss functions and penalties for classification.

Stochastic gradient descent refers to calculating the derivative from each training data instance and calculating the update immediately.
The only advantage is the ease of implementation and efficiency whereas a major setback with stochastic gradient descent is that it requires a number of hyper-parameters and is sensitive to feature scaling.

The main use cases are -
- Internet Of Things
- Updating the parameters such as weights in neural networks or coefficients in linear regression

4. K - Nearest Neighbour (KNN)

It is a lazy learning algorithm that stores all instances corresponding to training data in n-dimensional space. It is a lazy learning algorithm as it does not focus on constructing a general internal model, instead, it works on storing instances of training data.



Out of 3 nearest points, 2 belongs to class B, So, through KNN, class of an unknown star is class B

Classification is computed from a simple majority vote of the k nearest neighbors of each point. It is supervised and takes a bunch of labeled points and uses them to label other points. To label a new point, it looks at the labeled points closest to that new point also known as its nearest neighbors. It has those neighbors vote, so whichever label most of the neighbors have is the label for the new point. The “k” is the number of neighbors it checks.

This algorithm is quite simple in its implementation and is robust to noisy training data. Even if the training data is large, it is quite efficient. The only disadvantage with the KNN algorithm is that there is no need to determine the value of K and computation cost is pretty high compared to other algorithms.

The main use cases are -

- Industrial applications to look for similar tasks in comparison to others
- Handwriting detection applications
- Image recognition
- Video recognition
- Stock analysis

5. Decision Tree

The decision tree algorithm builds the classification model in the form of a tree structure. It utilizes the if-then rules which are equally exhaustive and mutually exclusive in classification. The process goes on with breaking down the data into smaller structures and eventually associating it with an incremental decision tree. The final structure looks like a tree with nodes and leaves. The rules are learned sequentially using the training data one at a time. Each time a rule is learned, the tuples covering the rules are removed. The process continues on the training set until the termination point is met.

The tree is constructed in a top-down recursive divide and conquer approach. A decision node will have two or more branches and a leaf represents a classification or decision. The topmost node in the decision tree that corresponds to the best predictor is called the root node, and the best thing about a decision tree is that it can handle both categorical and numerical data.
A decision tree gives an advantage of simplicity to understand and visualize, it requires very little data preparation as well. The disadvantage that follows with the decision tree is that it can create complex trees that may not categorize efficiently. They can be quite unstable because even a simplistic change in the data can hinder the whole structure of the decision tree.
The main use cases are -

- Data exploration
- Pattern Recognition
- Option pricing in finances
- Identifying disease and risk threats

6. Random Forest

Random decision trees or random forests are an ensemble learning method for classification, regression, etc. It operates by constructing a multitude of decision trees at training time and outputs the class that is the mode of the classes or classification or mean prediction(regression) of the individual trees.

A random forest is a meta-estimator that fits a number of trees on various subsamples of data sets and then uses an average to improve the accuracy of the model's predictive nature. The sub-sample size is always the same as that of the original input size but the samples are often drawn with replacements.
The advantage of the random forest is that it is more accurate than the decision trees due to the reduction in over-fitting. The only disadvantage with the random forest classifiers is that it is quite complex in implementation and gets pretty slow in real-time prediction.

The main use cases are -

- Industrial applications such as finding if a loan applicant is high-risk or low-risk
- For Predicting the failure of mechanical parts in automobile engines
- Predicting social media share scores
- Performance scores

7. Artificial Neural Network(ANN)

A neural network consists of neurons that are arranged in layers, they take some input vector and convert it into an output. The process involves each neuron taking input and applying a function which is often a nonlinear function to it and then passes the output to the next layer.

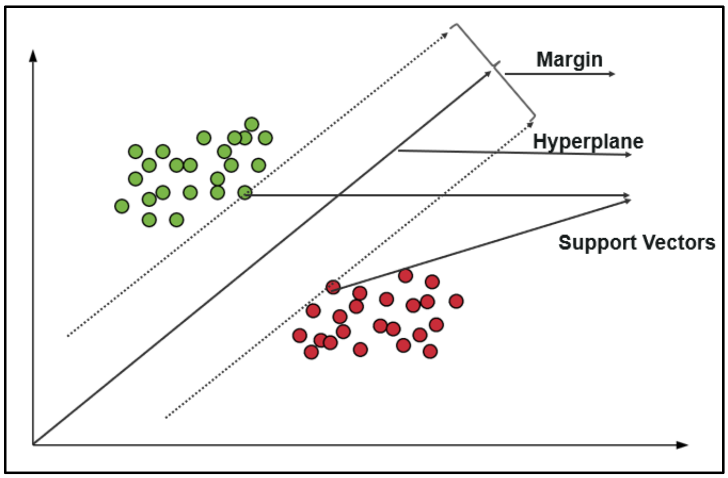
In general, the network is supposed to be feed-forward meaning that the unit or neuron feeds the output to the next layer but there is no involvement of any feedback to the previous layer.
Weighings are applied to the signals passing from one layer to the other, and these are the weights that are tuned in the training phase to adapt a neural network for any problem statement.
It has a high tolerance to noisy data and is able to classify untrained patterns, it performs better with continuous-valued inputs and outputs. The disadvantage with the artificial neural networks is that it has poor interpretation compared to other models.

The main use cases are -

- Handwriting analysis
- Colorization of black and white images
- Computer vision processes
- Captioning photos based on facial features

8. Support Vector Machine (SVM)

The support vector machine is a classifier that represents the training data as points in space separated into categories by a gap as wide as possible. New points are then added to space by predicting which category they fall into and which space they will belong to.



It uses a subset of training points in the decision function which makes it memory efficient and is highly effective in high dimensional spaces. The only disadvantage with the support vector machine is that the algorithm does not directly provide probability estimates.

The main use cases are -

- Business applications for comparing the performance of a stock over a period of time
- Investment suggestions
- Classification of applications requiring accuracy and efficiency

Classifier Evaluation Methods -

Evaluation of the created classifier model is necessary since the accuracy of prediction is the ultimate goal. Thus, it is important to understand how well the model is performing with known and unknown data.

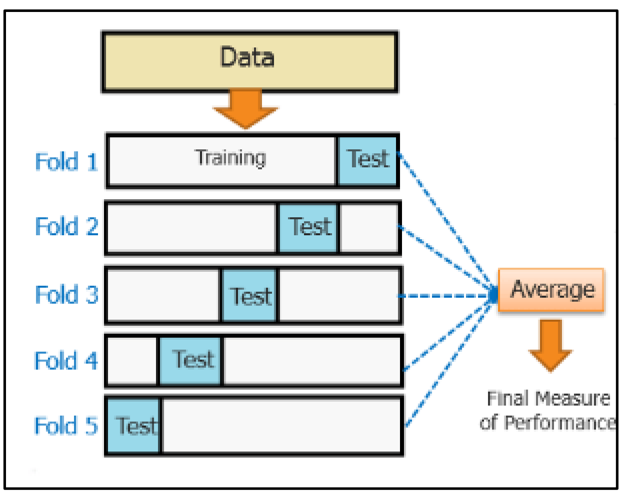
The main evaluation methods are -

Holdout Method

This is the most common method to evaluate a classifier. In this method, the given data set is divided into two parts as a test and train set 20% and 80% respectively. The train set is used to train the data and the unseen test set is used to test its predictive power.

Cross-Validation

Overfitting is the most common problem prevalent in most machine learning models. K-fold cross-validation can be conducted to verify if the model is overfitted at all. In this method, the data set is randomly partitioned into k mutually exclusive subsets, each of which is of the same size. Out of these, one is kept for testing and others are used to train the model. The same process takes place for all k folds.



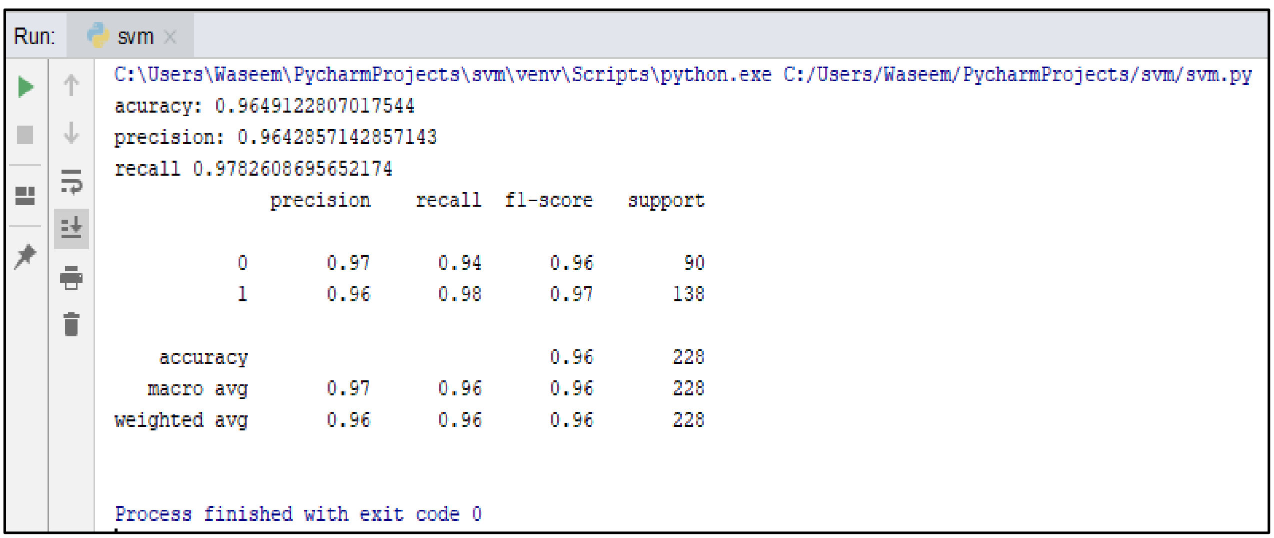
Classification Report

A classification report gives the following 3 measures of model performance -

1. Accuracy
    a. Accuracy is a ratio of correctly predicted observation to the total observations
    b. True Positive: The number of correct predictions that the occurrence is positive.
    c. True Negative: Number of correct predictions that the occurrence is negative.

2. F1 - Score
    a. It is the weighted average of precision and recall

3. Precision & Recall
    a. Precision is the fraction of relevant instances among the retrieved instances, while recall is the fraction of relevant instances that have been retrieved over the total number of instances. They are basically used as the measure of relevance.



```
Run:   svm
C:\Users\Waseem\PycharmProjects\svm\venv\Scripts\python.exe C:/Users/Waseem/PycharmProjects/svm/svm.py
acuracy: 0.9649122807017544
precision: 0.9642857142857143
recall 0.9782608695652174
              precision    recall  f1-score   support

           0       0.97      0.94      0.96        90
           1       0.96      0.98      0.97       138

    accuracy                           0.96       228
   macro avg       0.97      0.96      0.96       228
weighted avg       0.96      0.96      0.96       228


Process finished with exit code 0
```

A sample classification report of an SVM classifier using a cancer_data dataset.

ROC Curve

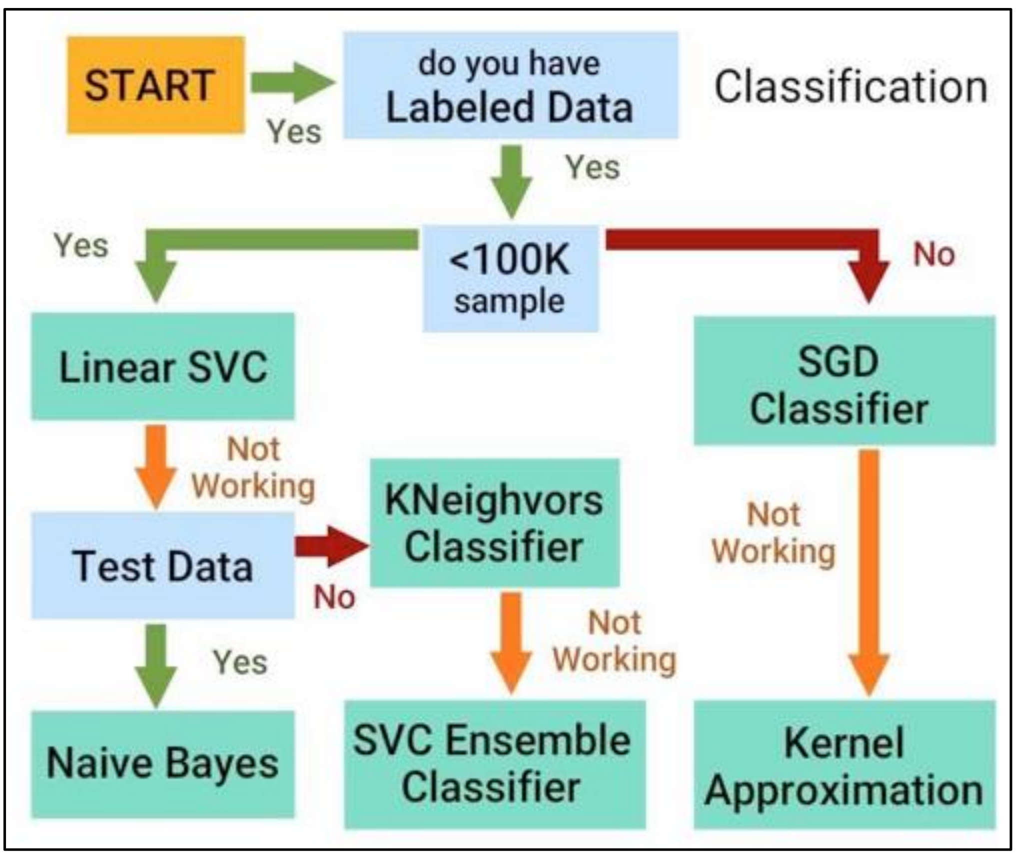
Receiver operating characteristics or ROC curve is used for visual comparison of classification models, which shows the relationship between the true positive rate and the false positive rate. The area under the ROC curve is the measure of the accuracy of the model.

Which classifier to use?

Apart from the above approach, We can follow the following steps to use the best algorithm for the model

- Read the data
- Create dependent and independent data sets based on our dependent and independent features
- Split the data into training and testing sets
- Train the model using different algorithms such as KNN, Decision tree, SVM, etc
- Evaluate the classifier
- Choose the classifier with the most accuracy.

Although it may take more time than needed to choose the best algorithm suited for your model, accuracy is the best way to go forward to make your model efficient.



Flowchart representing flow that should be followed while selecting classifier algorithm based on data

Our classifier model - Ensemble Learning

Our resume scoring and classification model is based on ensemble learning. By definition, it means, "**Ensemble learning** is the process by which multiple models, such as classifiers or experts, are strategically generated and combined to solve a particular computational intelligence problem. **Ensemble learning** is primarily used to improve the (classification, prediction, function approximation, etc.)"
The classification system we are using currently is based on these 2 major steps -

1. Calculating suitability score for each resume for a job.
2. Selecting top resumes based on that score.

In order to accomplish step 1 i.e. to calculate the suitability score of a resume for a job profile, we use many different similarity matching algorithms starting from basic ones like cosine similarity and experimenting with more statistical ones like the Jaccard index.
Following are the main scoring algorithms we have considered -

1. Jaccard Index
2. Sorensen-Dice coefficient
3. Tversky Index
4. Cosine similarity

Out of these, our results showed that the Jaccard index gives the most relevant scoring of a resume based on its similarity and aptness to the applied job profile.

Furthermore, we used an ensemble learning-based classification algorithm combining these models -

- Naive-Bayes
- ANN
- SVM

That ensemble model gives our system the highest accuracy on training data (95%) and on testing data (90%).
Hence, our model is able to predict the most suitable candidates with an accuracy of nearly 90%.

Chapter - 5

Personality Evaluation Model

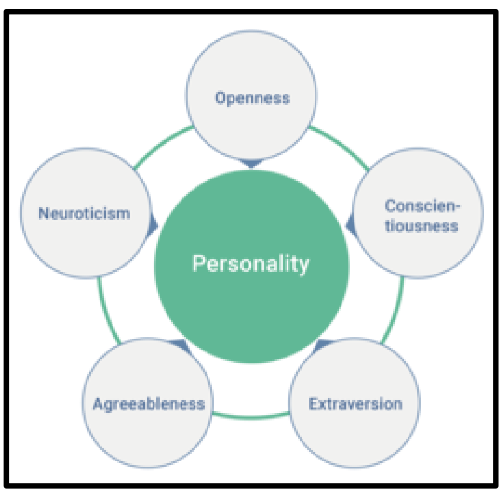
Introduction, concept, and intuition -

The personality of a candidate is the most important trait that makes him/her suitable for an organization. Every organization has its own culture, values, and principles. And it is really important that a candidate's personality matches with the personality of the organization they are going to work in. Hence, many organizations nowadays use psychometric tests to evaluate a person's cultural fit.

In our system, we included this very key evaluation in the second level of screening. After getting selected based on their resumes, candidates have to answer some questions based on which we will evaluate their personality (in a big 5 personality model) and assign a score to every candidate. Finally, top scorers will get ahead for further rounds and others simply get feedback that will help for their self-development.

The model we are using for personality evaluation is the Big 5 model. The Big Five personality traits were the model to comprehend the relationship between personality and academic behaviors] This model was defined by several independent sets of researchers who used factor analysis of verbal descriptors of human behavior.

The **Big Five** evaluates **personality** by measuring—as the name suggests—**five personality** traits on a continuous scale.



The Big 5 personality evaluation model -

The **Big Five personality traits** is a suggested taxonomy, or grouping, for personality traits,[1] developed from the 1980s onwards. The theory identifies five factors:

Openness to experience :

This trait defines curiosity, imagination, and openness to new ideas. People who are open to experience are intellectually curious, open to emotion, sensitive to beauty, and willing to try new things. They tend to be when compared to closed people, more creative and more aware of their feelings.

Conscientiousness :

This trait highlights the tendency to display self-discipline, act dutifully, and strive for achievement against measures or outside expectations. High conscientiousness is often perceived as being stubborn and focused. Low conscientiousness is associated with flexibility and spontaneity, but can also appear as sloppiness and lack of reliability.

Extraversion :

The trait is marked by pronounced engagement with the external world. Extraverts enjoy interacting with people and are often perceived as full of energy. They tend to be enthusiastic, action-oriented individuals. Introverts have lower social engagement and energy levels than extroverts. They tend to seem quiet, low-key, deliberate, and less involved in the social world. This does not mean that they are unfriendly or antisocial; rather, they are reserved in social situations.

Agreeableness :

This trait highlights how well a person can agree with other’s ideas and can work collaboratively in a team. Agreeable individuals value getting along with others. They are generally considerate, kind, generous, trusting and trustworthy, helpful, and willing to compromise their interests with others. Disagreeable individuals place self-interest above getting along with others. They are generally unconcerned with others' well-being and are less likely to extend themselves for other people.

Neuroticism :

Neuroticism is the tendency to experience negative emotions, such as anger, anxiety, or depression.[49] It is sometimes called emotional instability or is reversed and referred to as emotional stability. Those who score high in neuroticism are emotionally reactive and vulnerable to stress. At the other end of the scale, individuals who score low in neuroticism are less easily upset and are less emotionally reactive. They tend to be calm, emotionally stable, and free from persistent negative feelings.

Personality evaluation in our system -

In our proposed system, in the second level of candidate screening, we evaluate the personality of the candidates to see if they can fit culturally into the organization.

The challenge

Accurate personality estimation is difficult and can be falsified to some extent by a candidate. Given that employment is often associated with significant financial and social benefits, job seekers are incentivized to portray themselves in a way that best reflects the characteristics of an ‘ideal candidate. This can lead candidates to mask their true personalities and results in inaccurate test assessments, particularly when carried out in formal environments. In contrast, the assessment of personality in relaxed or informal environments may lead to more representative estimations of an individual’s personality.

Our solution

To this end, we created a personality trait estimator that takes informal written text as input and gives personality estimates based on the Big 5 personality model. The original test consists of a self-reported questionnaire that seeks to categorize respondents into 5 personality type profiles and is commonly used to gauge a person’s preference towards self-expression, information processing, planning, and decision making. We have treated personality trait detection as a supervised learning problem. Our process was to train four classifiers to predict a binary outcome for each of the 5 traits. So essentially for an excerpt of text, each classifier would predict one of the 5 dimensions of the model.
A two-part questionnaire was developed to identify a respondent’s personality using the LR classifier. Users were first asked to complete the Myers-Briggs questionnaire and then provide written responses to the following questions:

1. Tell me something about yourself?
2. What is something you are good at?
3. What do you love about your job?

These responses were then used to estimate their personality with the classifier.

Technical description of the developed model

We followed a method consistent with that found in the literature and represented the text as a weighted term-document matrix using word-unigrams. Preliminary experimentation found that (word) bi-grams and trigrams representations degraded predictive performance and introduced sparsity in the term-document matrix. Similarly, character n-grams produced unfavorable results, and so were not used in the term weighting process. Finally, a TF-IDF (term frequency-inverse document frequency) weighting process was used to convert the term-document matrix into a normalized feature representation of the text used by the classifiers. We used a combination of a regular expression, SpaCy and scikit-learn methods to achieve these transformations. Preliminary experimentation identified that only the top 2,000 most frequently occurring unigrams were needed for adequate classification performance. In addition to the TF-IDF matrix, several other features were added to try and capture the nuance of emotional expression and personality in text, which is described below.

Sentences of extroverted types also tend to be simpler with fewer words, lower lexical diversity, and fewer negations. Conversely, introverts generally use more negating words and use a wider vocabulary. To identify these differences, a formality score (Heylighen and Dewaele, 1999) was calculated for each user post:
Sentences of extroverted types also tend to be simpler with fewer words, lower lexical diversity, and fewer negations. Conversely, introverts generally use more negating words and use a wider vocabulary. To identify these differences, a *formality score (Heylighen and Dewaele, 1999)* was calculated for each user post:

$$\begin{aligned}\text{Formality} &= \frac{1}{2}\left(\text{noun}_{\text{freq}} + \text{adjective}_{\text{freq}} + \text{preposition}_{\text{freq}} + \text{article}_{\text{freq}}\right) \\ &\quad - \frac{1}{2}\left(\text{pronoun}_{\text{freq}} + \text{verb}_{\text{freq}} + \text{adverb}_{\text{freq}} + \text{interjection}_{\text{freq}}\right) + 50\end{aligned}$$

where the frequency is a part of speech measured as a fraction of the total number of words written in a sentence. Higher formality values indicate a less expressive tone in the text. Formality scores were calculated for each user post and recorded as an average formality score. The SpaCy POS (Parts of Speech)tagger was used to identify the parts of speech required in the equation.
Additionally, several other features were also computed for each post to capture lexical diversity and word usage:

1. Type to token ratio: the number of unique words divided by the total number of words.
2. Average words per sentence: a proxy for sentence complexity
3. Average word length: a proxy for vocabulary use

So basically given a text, we would try to predict 4 different outcomes from it, i.e.

1. Introverted or extroverted, **I-E**
2. Intuitive or sensing, **N-S**
3. Thinking or feeling, **T-F**
4. Judging or perceiving, **J-P**

Five classifiers were constructed and used to determine each of the four personality dimensions:

1. Random Forest (RF)
2. Support Vector Machine (SVM)
3. Logistic Regression (LR)
4. Bernoulli Naive Bayes (NB)
5. Majority Class Classifier (MCC)

Classifier	Hyperparameter settings
LR	• LBFGS solver
SVM	• RBF kernel • C = 1 • Gamma = scaled
RF	• 100 trees • Gini information criterion • 50 max features/tree • Max-depth=None
NB	• Default settings, no priors supplied

A random forest classifier had been chosen because the tree structure permits an understanding of the decision logic used to arrive at predictions. SVM and LR classifiers are frequently reported in the literature as popular choices for personality prediction from the text. A (Bernoulli) Naïve Bayes classifier is used as a control to test the independence assumption between features (or tokens). Since emotion is often derived from not just words but also their relation to other words in a sentence, one would expect a Naïve Bayes classifier to perform poorly at text classification. Finally, a majority class classifier is used as a baseline classifier to compare performance against the other four classifiers.

<u>Chapter 6</u>

Design & Implementation of AI Recruiter

<u>Unique features</u> -

Key features highlighting the uniqueness of the idea are:-

- **<u>First interaction before the application</u>** -
  The chatbot will first interact with the applicant before filling up the application so as to provide recruitment process guidance and create first impression analysis of the applicant towards the organization and the job role.
  Questions -
  "Why do you want to apply for this role?",
  "What do you expect from the organization?" etc.

- **<u>AI-based visualization for shortlisting</u>** -
  A visualization of the applicant's skills will be created using AI resume parsers and using along with the first impression analysis, the system will shortlist the candidate.

- **<u>Personality traits evaluation</u>** -
  Shortlisted candidates will be interviewed by the chatbot with more questions and some psychological games to build a personality traits map for each candidate and find out the most suitable candidates for further HR rounds.

- **<u>Feedback for non-shortlisted candidates</u>** -
  The chatbot will give proper feedback to all the non-shortlisted candidates by providing them their skill visualizations and personality traits map highlighting the reasons for their rejection.

These features will add a novelty factor to the whole integrated system and will also make the whole process fair, transparent, and trustworthy.

Technologies used -

The primary language we are going to use for this project will be **Python** since it is robust, has a
A large number of supporting modules and is highly flexible in terms of cross-platform

application development.

Various technologies employed are selected taking into consideration the full flexibility and scalability of the project.

Chatbot Interface

- Flutter for mobile application
- Firebase for hosting
- MongoDB as dynamic database
- Google cloud DialogFlow
- IBM Watson assistant

Resume Visualization

- Python data science & visualization modules such as
    - SkLearn, Keras
    - Seaborn
    - Pandas
    - Bokeh
    - Plotly
- NodeJS

Personality Evaluation

- GPT-3
- IBM Personality Traits

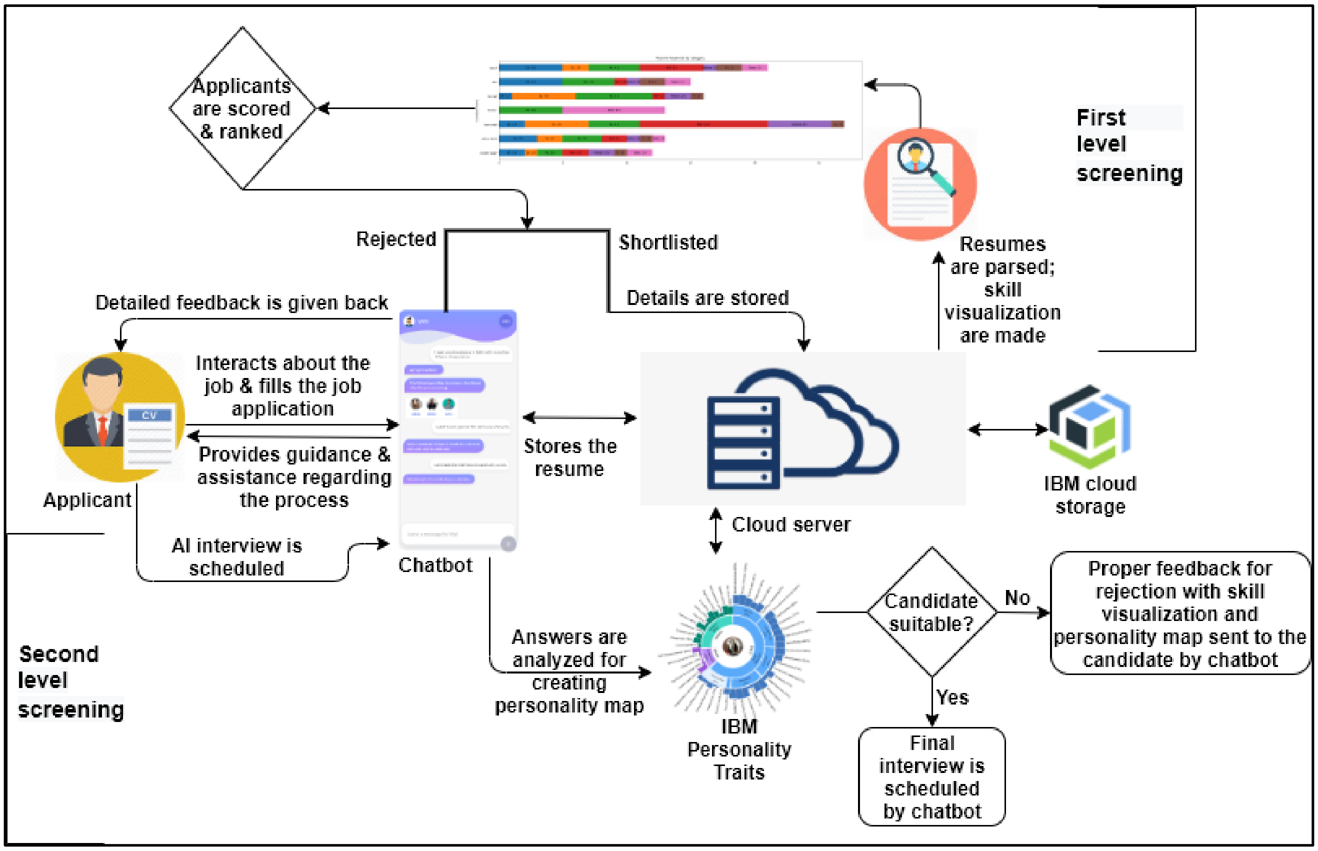
Recruiter Dashboard

- Angular
- Typescript

Backend APIs & Hosting

- Django
- Heroku

Workflow of the system -

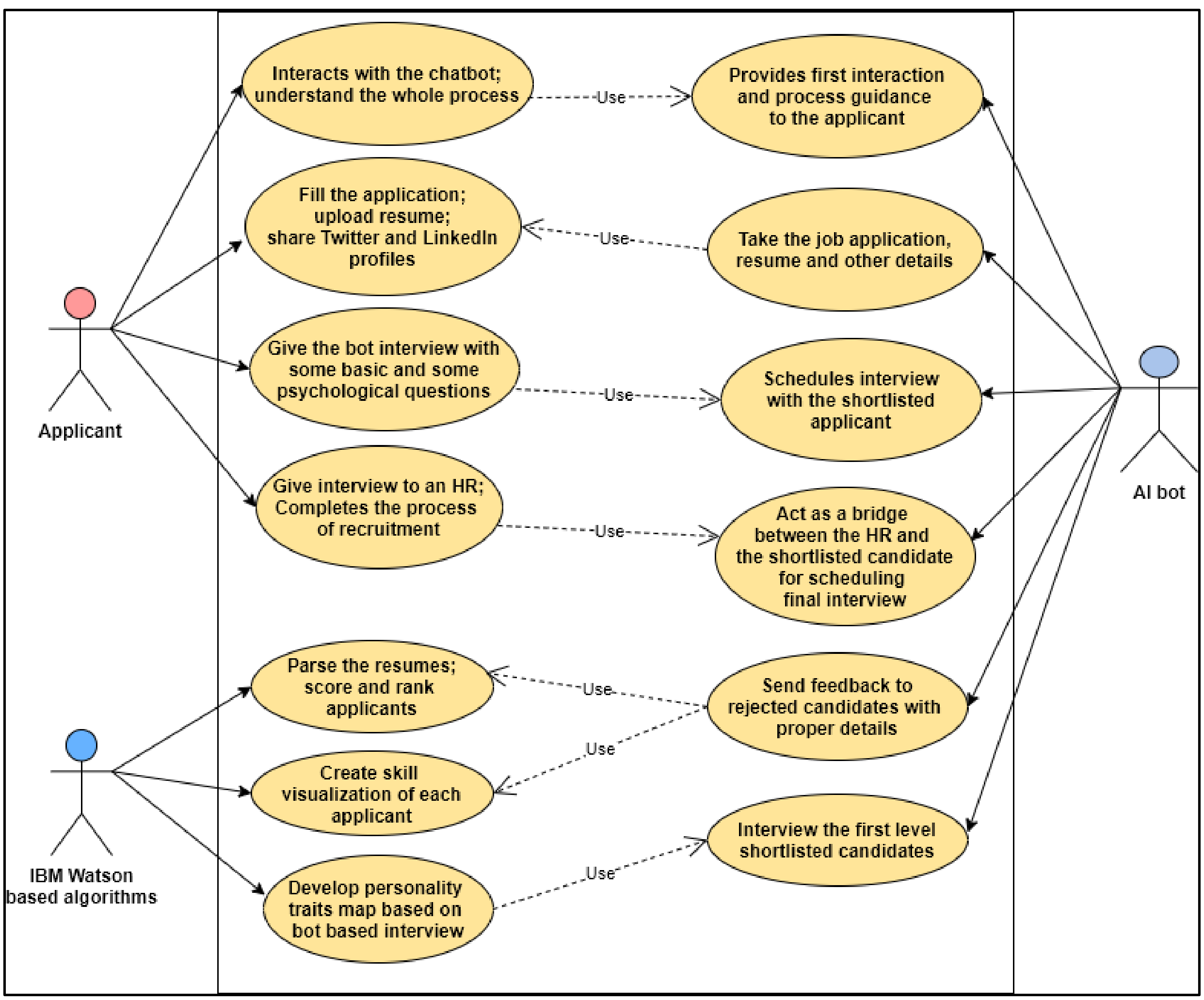


- ➢ The Chatbot will be the conversational interface between the company and the applicant. It will assist the applicant by giving JD, an overview of the process, follow-up on application status, and proper timed feedback.
- ➢ Applicants can directly fill their applications through the chatbot. All the collected details and resumes will go to the cloud storage.
- ➢ In Level I Screening, all the resumes will go through a *resume_visualizer.py* Will parse all the resumes and create a visualization map of skills and experience.
- ➢ Based on the visualization map, another file *screening1.py* will shortlist the most suitable candidates synchronizing the map with the job role's qualifications.
- ➢ Shortlisted candidates will go for the next level of screening and will be notified about that. Non-shortlisted candidates will get proper feedback via chatbot regarding their shortcomings and suggestions for improvement.
- ➢ For Level II Screening, shortlisted candidates can schedule an interview in the desired slot, and then can have a video interview based on some neuroscience questions.
- ➢ All the answers of the candidate will go through the Speech-to-Text module and then that text is fed to IBM Personality Traits. The output will be a personality map of that person.

➢ Maps of all the candidates will be compared to already existing samples of similar job roles and finally, recommendation algorithms will suggest the final group of candidates to the company.

➢ Since AI is not 100% reliable, we are also adding the provision of adding a little human bias in Level II Screening using Google's What-If tool.

Use case diagram -



In our project, there are 3 main entities interacting with each other in the system to make it a reliable, transparent, and robust system. These entities and their roles are -

Applicant :

- ➔ Can interact with a chatbot, and seeks its assistance;
- ➔ Fill the application form, submit a resume, take follow-up on application status;
- ➔ Get proper feedback from the system with suggestions for improvement;
- ➔ Select slot for an interview, & give interview through chatbot;

AI Algorithm :

- ➔ *resume_visualizer.py* will parse all the resumes and create a visualization map out of it selecting relevant skills and experiences;
- ➔ *screening_1.py* will rank resumes & perform shortlisting;
- ➔ Develop personality map based on interview answers;

AI Chatbot :

- ➔ Provides assistance to the applicants for the whole process of recruitment;
- ➔ Take applications and pass them on to cloud-based server;
- ➔ Acts as a bridge between the applicants and the recruiters and provide a medium of communication between the two;
- ➔ Help schedule interviews by the shortlisted candidates in desired slots;
- ➔ Take video interviews of candidates, convert speech-to-text and pass it on to cloud server;
- ➔ Provide proper feedback to the candidates rejected with suggestions of improvement;

The architecture of the system -

Backend System (Django based) :

The backend of the AI Recruiter is developed using the Python-based framework Django. Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It’s free and open source.

These are the primary advantages of using Django for backend development :

- Fast
- Secure
- Scalable

The backend has a set of APIs that supports both candidate’s mobile application and recruiter dashboard. The code has the following architecture -

➢ There is a main folder named “API” which is the base to which all requests are redirected.

➢ Inside the API folder, each entity has its own folder with its own set of routes and controllers. These entities are :
  - Company
  - HR
  - Job
  - Candidate

➢ The folder structure is shown in Fig. 10

➢ ER diagram is shown in Fig. 11

The backend has 2 key algorithms implemented namely, Resume based Classification & Personality Evaluation.

Code snippet for resume based classification is shown in Fig.(s) 12, 13
Code notebook for personality evaluation is shown in Fig. 14

Mobile Application for Candidates (Flutter)

TBD by Sahyog Saini

Recruiter Web Dashboard (Angular)

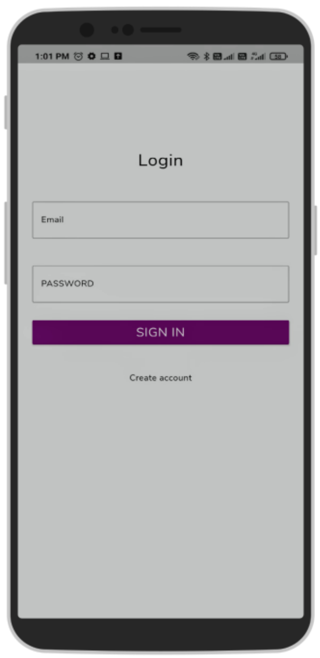
TBD by Sandeep Kumar Shukla

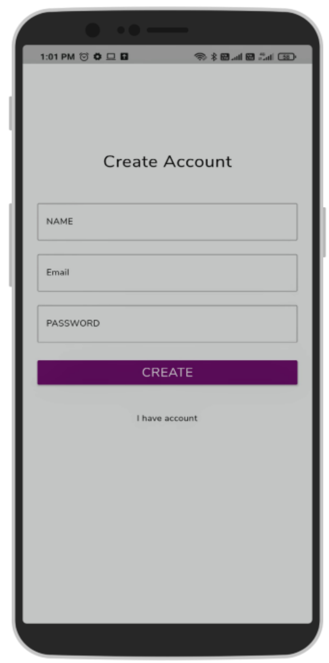
Chapter - 7

Results and Software Developed

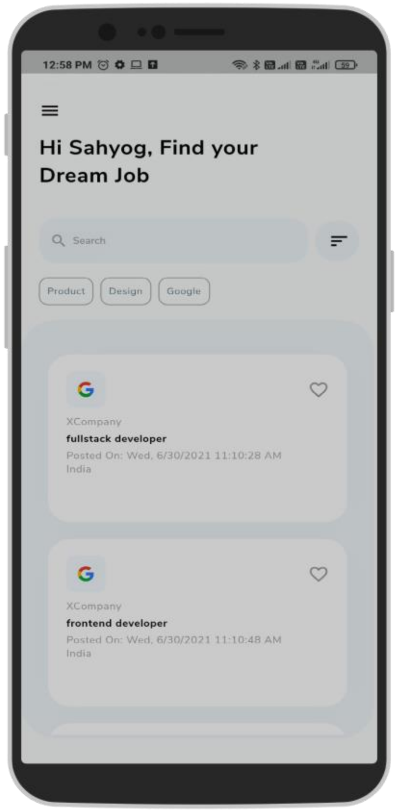
Mobile Application for candidates -

1. Login and register Screen

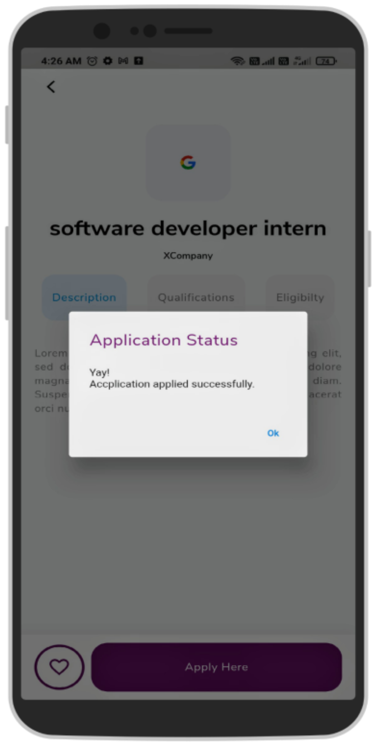




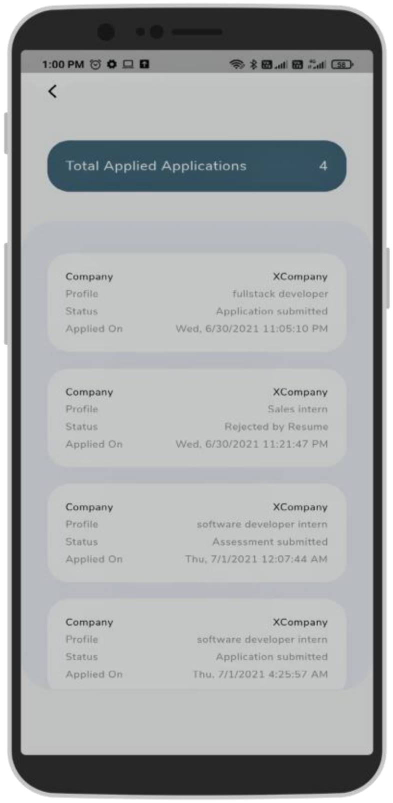
2. View all jobs and search jobs



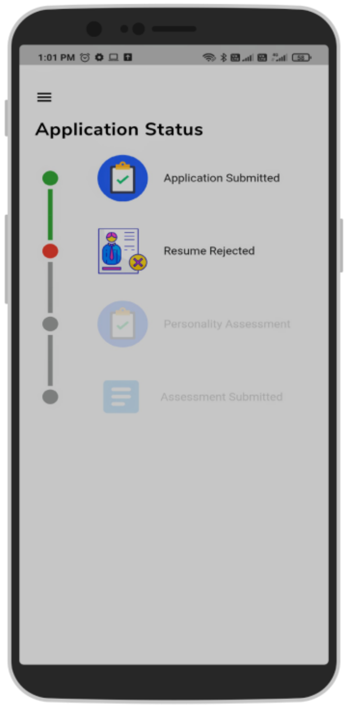
3. Apply to a job

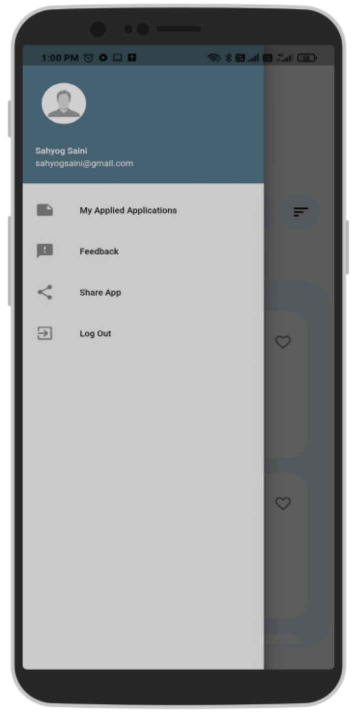


4. My applied applications

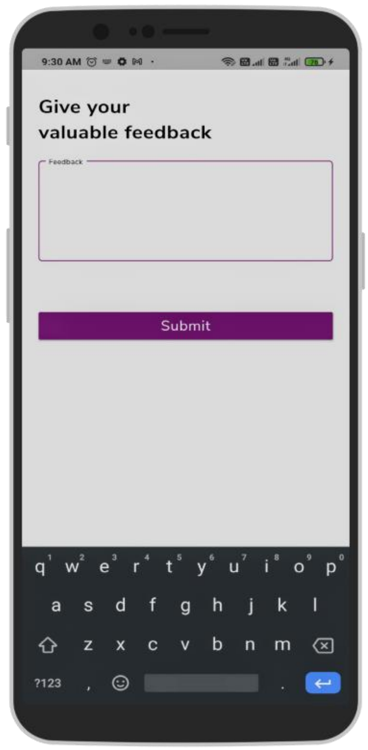


5. Application details with status



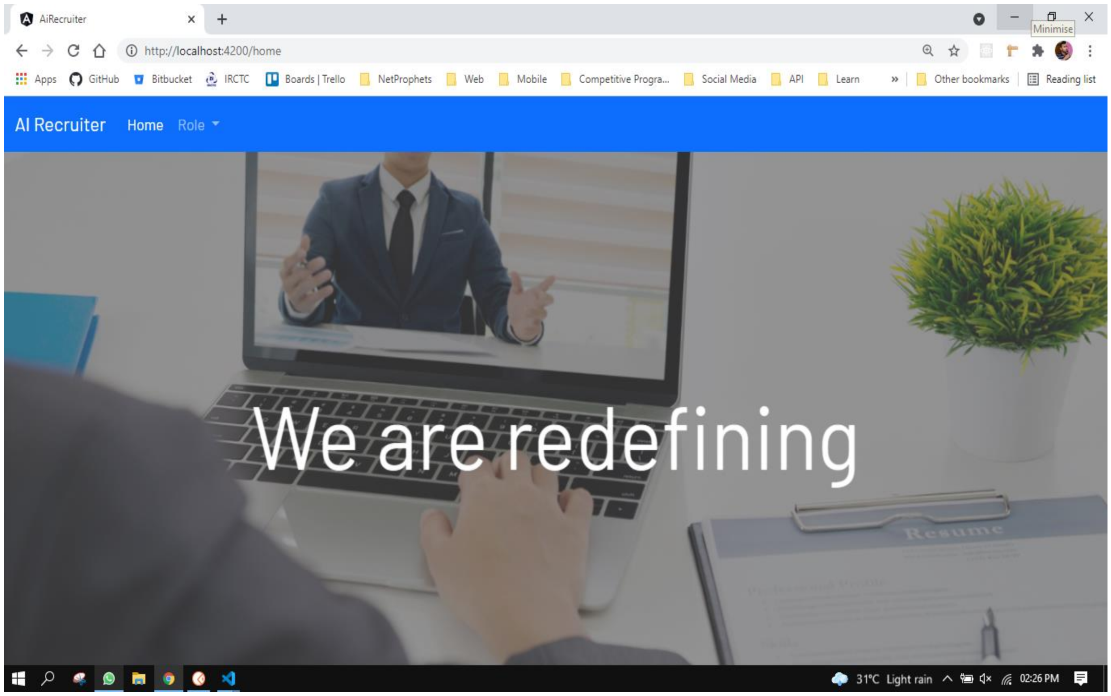


6. Feedback page

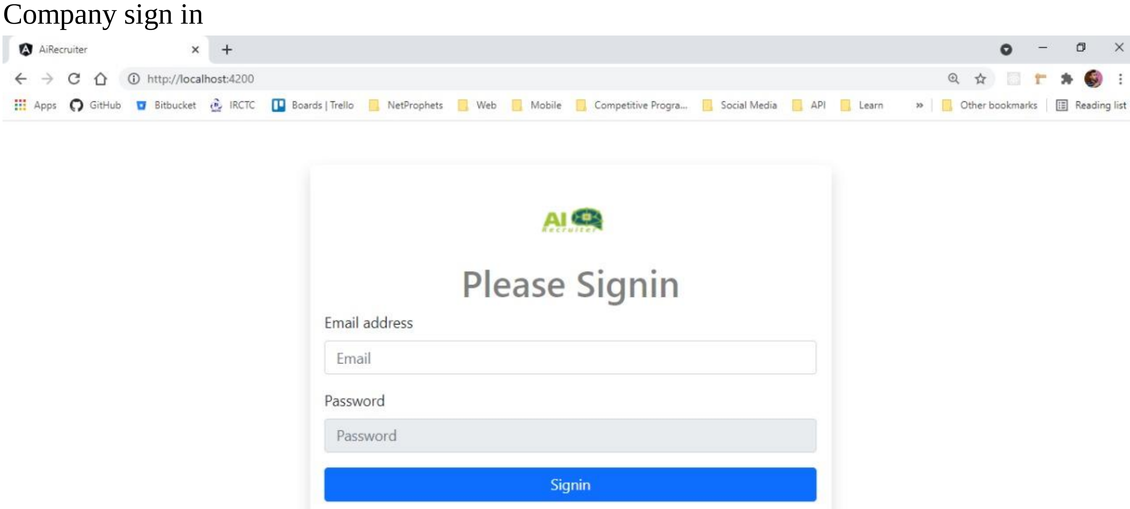


Recruiter Web Dashboard

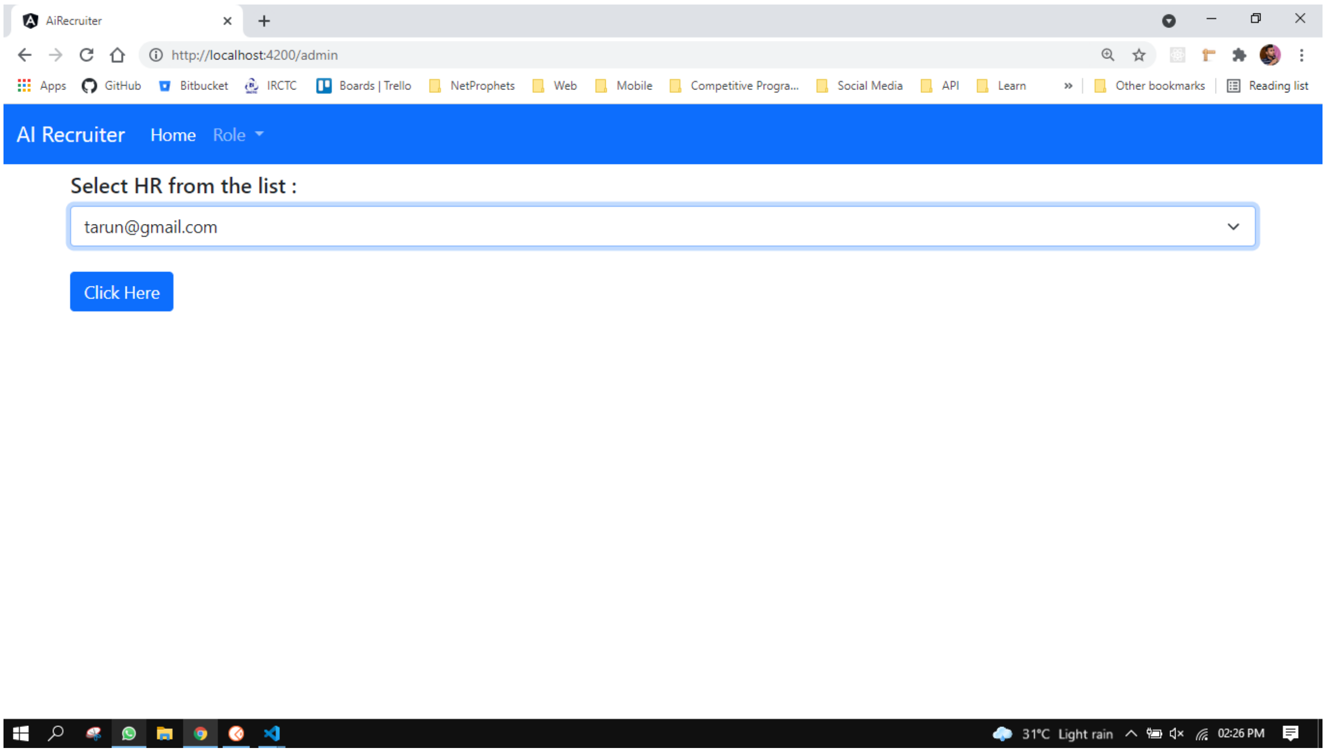
1. Landing page



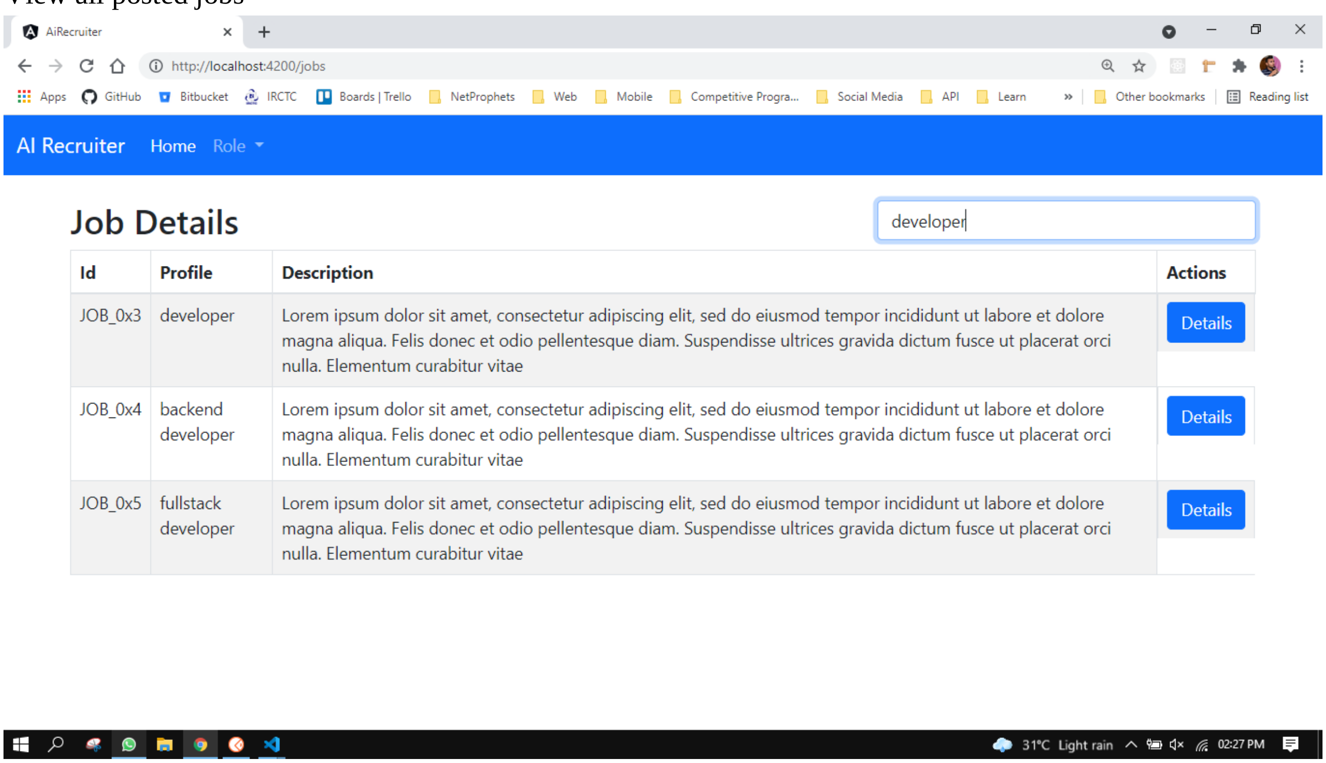
2. Company sign in



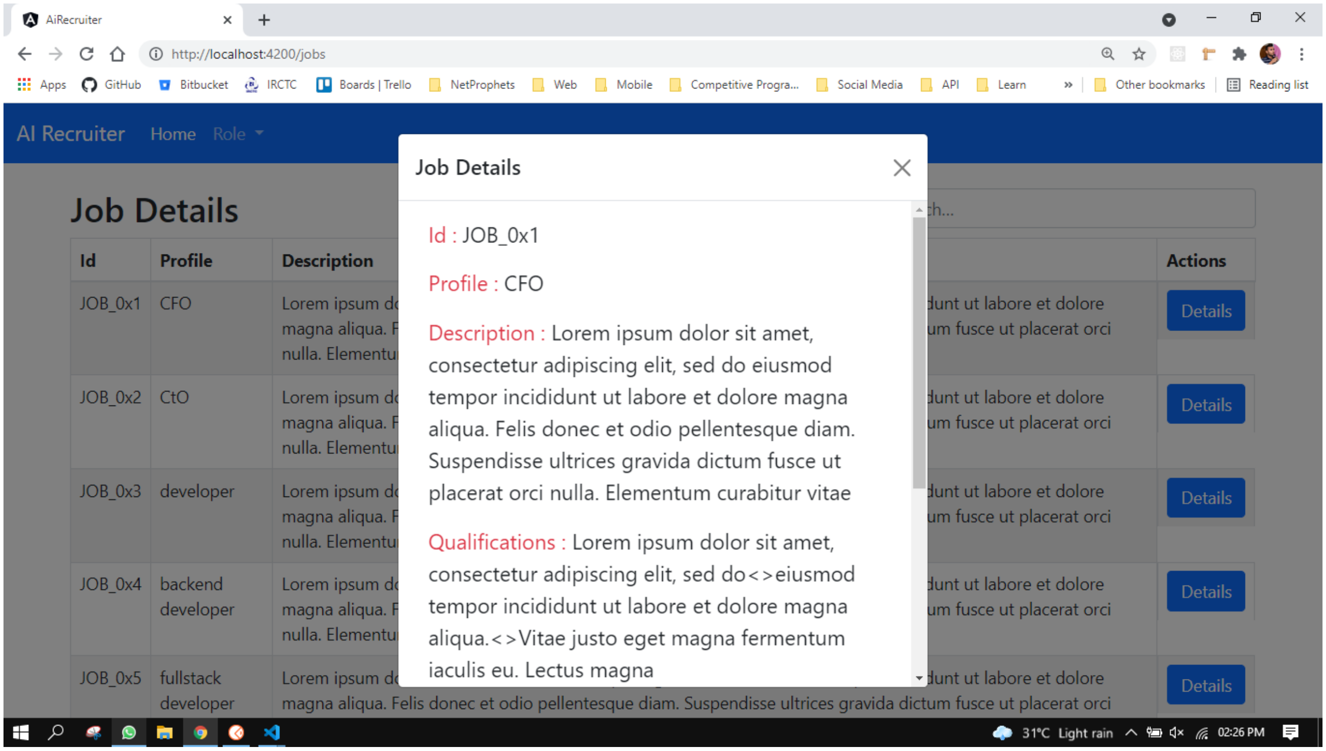
3. HR sign in



4. View all posted jobs



5. Job details



Chapter - 8

Results & Conclusions

Results -

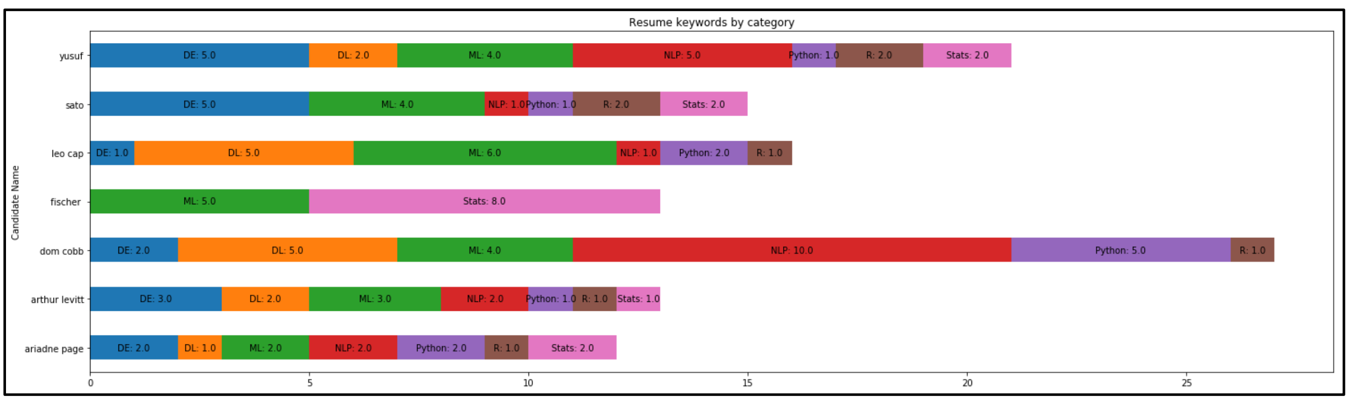
Any developed system has to be tested to collect results that prove its usefulness and aptness of the solution to the problem stated. For testing our proposed system, the following datasets are used -

Results obtained from resume based classifier :

For training and testing of resume based classifier -

1. Kaggle dataset with a total of 8,653 entries of applicant experiences and 80,000 job listings. This dataset is first cleaned, feature selection is done, text processing is done and then the ensemble model is trained over 80% of data, and the rest 20% data was used for testing.

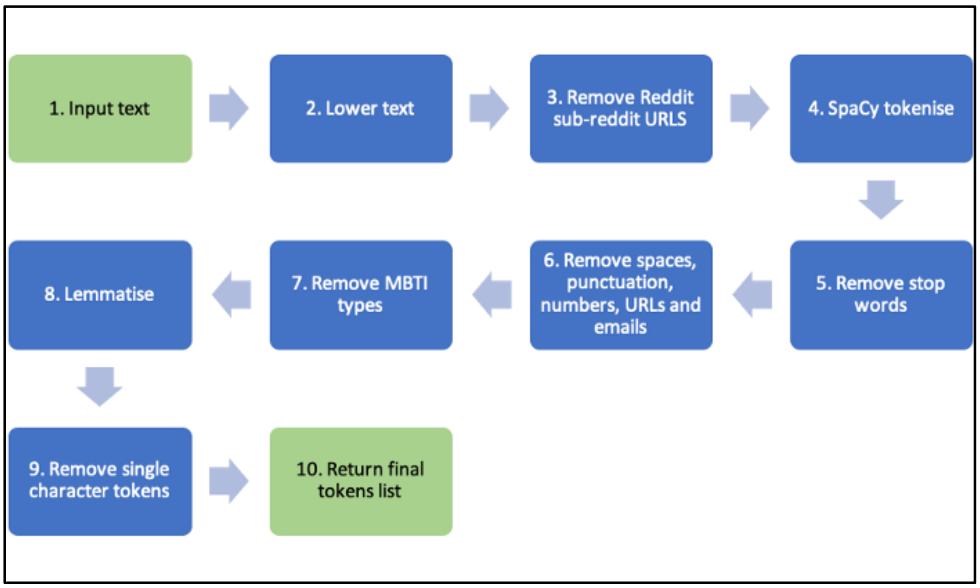
Example: For a data science job profile, we screen nearly 100 resumes through our resume visualizer and this is the visualization of top candidates

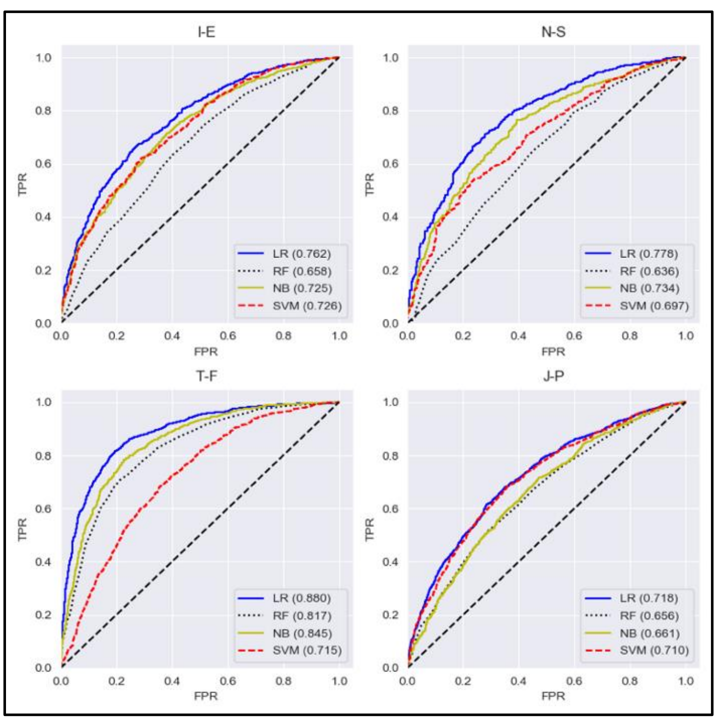


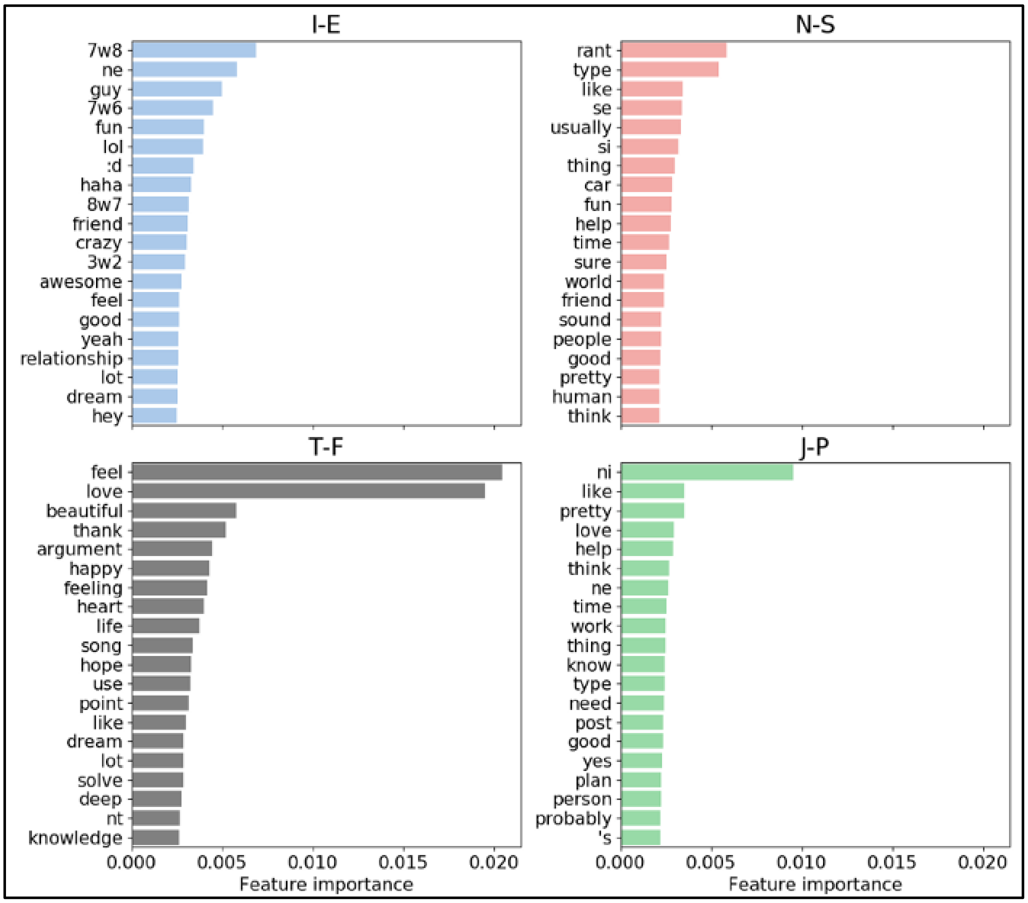
Results obtained from personality evaluation :

For training and testing personality evaluation classifier -

1. *Training Set 1:* A Kaggle dataset consisting of 8,600 labeled entries and posts, written in a semi-casual tone.
2. *Training Set 2:* A Reddit dataset consisting of 9,300 posts, written in an informal tone.
3. *Test Set:* A LinkedIn dataset scraped from 12,000 employees across 10 companies, written in a casual to a professional tone.

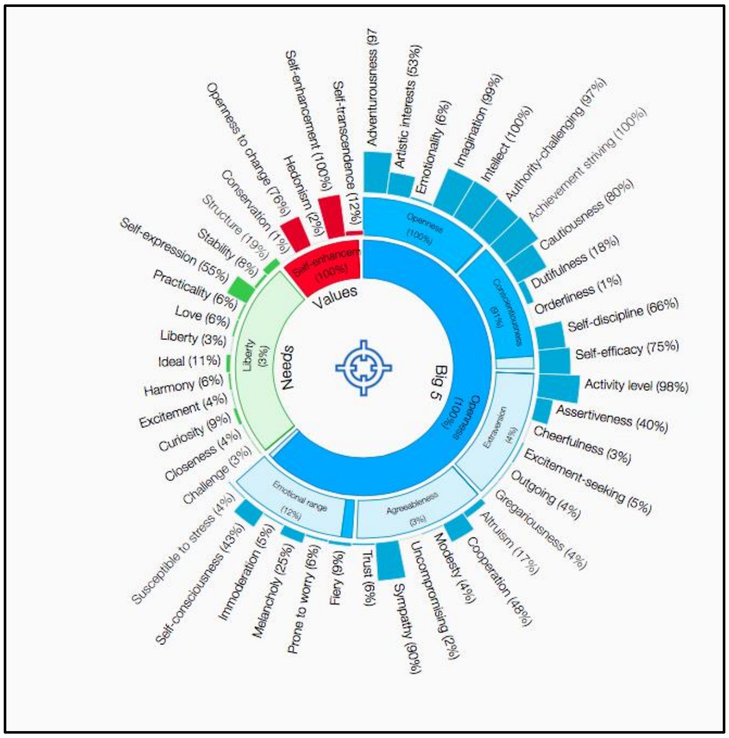






Consider a sample response: "*A very passionate and enthusiastic individual who believes in living life to its fullest. Absolutely love the performing arts and fashion. I'm good at being able to empathize with those around me and also very persistent in achieving my goals. Love being able to influence such a variety of individuals about decisions that directly impact their lives. Being able to be the balance between a company's vision and its biggest asset, the people and their needs.*"

For this example, the following will be the big 5 model evaluation results produced by GPT-3-



<u>Conclusions</u> -

After development and testing of the project with different datasets, we conclude that -

1. Automated resume selection can reduce the time to hire by 85%, thus saving organizations both in time and resources;

2. The use of AI in recruitment can save up to 14 hours in a week for HRs that is spent on manually screening a large number of applications;

3. By giving automated and detailed feedback, our system gives rejected candidates an opportunity to understand the gap between their skills and industry requirements, at the same time helps in their self-development;

4. AI recruiter also evaluates personality of candidates based on a psychometric questionnaire, which thus helps select candidates that are more culturally fit to the organization;

5. With a chatbot-powered mobile application, it facilitates candidates in a better way and provides 2-way communication unlike current mediums of communication such as emails which are generally automated and hence one-way only.

Hence, AI Recruiter is a complete system not just a part of the process, making candidates screening efficient, transparent, fast, and effective. This thus gives both organizations and the candidates trust in the process.

Chapter - 9

Future Scope in the Project

In our current system, we have tried to provide an AI-based solution to the whole process and not just a part of it. The 2 main algorithms that we have improved are -

1. Automated resume classification and skill visualization.
2. Personality evaluation

For both the algorithms, we have tested many different ML and Neural Network-based models, combined them in ensemble learning to get better results and indeed we succeeded to some extent. Still, there is more that can be done to improve this system and make it to be used in the industry. In the future, we would like to work on these some points -

- Testing the system with actual candidates resumes in actual hiring;
- Working with different HR recruiters to apply their feedback and create some sort of feedback loop in a recommendation engine;
- Implement unsupervised learning techniques to resume classifier so that it can work with different formats and newer sections;
- Add more traits to be predicted in the personality evaluation algorithm;
- Improving scalability and robustness of the overall architecture of the system so that it can work with multiple requests at the same time:

These are some of the improvements we think can be done to the system to make it more useful in the practical world, though more suggestions and improvements are always welcome;

References

[1] J. Stewart Black, Patrick van Esch, AI-enabled recruiting: What is it and how should a manager use it?, Business Horizons, Volume 63, Issue 2, 2020, Pages 215-226, ISSN 0007-6813, http://www.sciencedirect.com/science/article/pii/S0007681319301612

[2] Lutz, A. (2013, November 19). Applicants for jobs at the new D.C. Walmart face worse odds than people trying to get into Harvard. Business Insider. Available at https://www.businessinsider.com/wal-mart-receives-23000-applications-2013-11

[3] McIlvaine, A. (2018). Data in the driver's seat. Human Resource Executive. Available at https://hrexecutive.com/talentacquisitions-leaders-use-ai-to-improve-hiring/

[4] Torres, M. (2017). 2 million candidates are desperate to work at Google. Why? Ladders. Available at https://www.theladders.com/career-advice/2-million-job

[5] Ideal. (2018). AI for recruiting: A definitive guide for HR professionals. Available at https://ideal.com/ai-recruiting/

[6] Paschen, J., Wilson, M., and Ferreira, J. (in press). Collaborative intelligence: How human and artificial intelligence create value along the B2B funnel. Business Horizons.

[7] Herman, A., O'Boyle, E. (2012). The best and the rest: Revisiting the norm of normality in individual performance. Personnel Psychology, 65(1), 79e119.

[8] Keller, S., and Meaney, M. (2017). Attracting and retaining the right talent. McKinsey and Company. Available at https://www.mckinsey.com/businessfunctions/organization/ourinsights/attracting-and-retaining-the-right-talent

[9] Jeanne Meister (2019). Forbes featured "Ten HR Trends In The Age Of Artificial Intelligence". Available at https://www.forbes.com/sites/jeannemeister/2019/01/08/ten-hr-trends-in-the-age-ofartificial-intelligence

[10] Josh Bersin (2017). Forbes featured "Robotics, AI And Cognitive Computing Are Changing Organizations Even Faster Than We Thought".

[11] D. Çelik et al., "Towards an Information Extraction System Based on Ontology to Match Resumes and Jobs," 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops, 2013, pp. 333-338, doi: 10.1109/COMPSACW.2013.60. https://ieeexplore.ieee.org/document/6605812

[12] Y. Luo, H. Zhang, Y. Wang, Y. Wen and X. Zhang, "ResumeNet: A LearningBased Framework for Automatic Resume Quality Assessment," 2018 IEEE International Conference on Data Mining (ICDM), 2018, pp. 307-316, doi:10.1109/ICDM.2018.00046. https://ieeexplore.ieee.org/document/8594855

[13] C. H. Ayishathahira, C. Sreejith and C. Raseek, "Combination of Neural Networks and Conditional Random Fields for Efficient Resume Parsing," 2018 International CET Conference on Control, Communication, and Computing (IC4), 2018, pp. 388-393, doi: 10.1109/CETIC4.2018.8530883. https://ieeexplore.ieee.org/document/8530883

[14] M. Fallahnezhad, M. Vali and M. Khalili, "Automatic Personality Recognition from reading text speech," 2017 Iranian Conference on Electrical Engineering (ICEE), 2017, pp. 18-23, doi: 10.1109/IranianCEE.2017.7985447.
https://ieeexplore.ieee.org/document/7985447

[15] S. Jothilakshmi and R. Brindha, "Speaker trait prediction for automatic personality perception using frequency domain linear prediction features," 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2016, pp. 2129-2132, doi: 10.1109/WiSPNET.2016.7566518.
https://ieeexplore.ieee.org/document/756651