RECRUIT BEST CANDIDATES WITH MACHINE LEARNING

Selvantharajah Thushanth

168272D

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

June 2018

RECRUIT BEST CANDIDATES WITH MACHINE LEARNING

Selvantharajah Thushanth

168272D

Thesis submitted in partial fulfillment of the requirements for the degree Master of Science in Computer Science specializing in Data Science, Engineering and Analytics

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

June 2018

DECLARATION

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: ……………………..… Date:……………….

Selvantharajah Thushanth

The above candidate has carried out research for the Masters of Science thesis under my supervision.

Name of the supervisor: Dr. Charith Chitraranjan

Signature of the supervisor: …………………………. Date: ……………….

Abstract

In this research, I propose a robust approach for predicting personality traits of job candidates using machine learning. Relationship between personality traits and job performance has been studied extensively during the past few decades and thus this relationship can be utilized to overcome limitations in choosing the right candidates.

The proposed approach uses scenario-based analysis using machine learning techniques. Candidates will be asked to take part in scenario-based written conversations and their personality traits will be extracted from these conversations using machine learning techniques. Exacted personality traits of the candidates will be compared with the required job related characteristics in order to evaluate the fitness for the position for which candidates are applying. In order to categorize personality traits of candidates, the Five Factor model is used. Existing methods of evaluating personality traits such as standard set of questionnaires are susceptible to candidates providing false information and also time consuming.

Besides candidates’ qualifications, knowledge and experience, candidates’ personality traits also used to rank the candidates and shortlist them for face-to-face interviews. Thus, this technique not only allows recruiting right candidates to right position but also reduces significant amount of time and cost spent on evaluating candidates’ suitability for given a job position by reducing the number of interviews to conduct. Further, this proposed system can be incorporated into existing e-recruitment system thus leveraging its effectiveness. Therefore, it is beneficial for companies since the proposed system helps to reduce cost and time consumption in the recruitment process while assisting them to choose more suitable candidates for a particular job position.

ACKNOWLEDGEMENTS

I would like to acknowledge the support and motivation given by Dr. Charith Chitraranjan, my supervisor, on successfully completing my research work. I also would like to thank him for his continuous suggestions and guidance throughout my research work. I wish to express my sincere gratitude to Dr. Surangika Ranathunga, for her valuable feedbacks during the initial phase of my research work.

Many thanks and appreciations go to all the staff members of Department of Computer Science and Engineering, University of Moratuwa for their generous help and motivation. Finally yet importantly, I would like to thank to my family for continuous support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	**Description**
MAE	Mean Absolute Error
MBTI®	Myers-Briggs Type Indicator®
MSE	Mean Squared Error
POS	Part-Of-Speech
RAE	Relative Absolute Error
RMSE	Root Mean Squared Error
RSE	Relative Squared Error
SMO	Sequential Minimal Optimization

1. INTRODUCTION

1.1 Problem and motivation

Recruiting the most suitable candidates for a given job position is a key to the success of any organization around the world. This research addresses the problem of developing an efficient and effective tool to predict the personality traits of job applicants through short written conversations on a casual topic such as restaurants. Currently available standard set of questionnaires [1], [2], [3], [4] for evaluating personality traits are susceptible to candidates providing fake information. Thus, they are less reliable. Further these standard set of questionnaires contain significantly large number of items thus it is time consuming and users might lose interest/patience in answering all questions. Therefore, these standard set of questionnaires cannot be used in the candidate evaluation process. Other proposed techniques require data such as applicants' posts or comments on social media to be made available [5].

Beside educational qualifications and experience, personality traits of a job applicant have great impact on his/her job performance. There are many studies [6], [7], [8], [9] have identified the relationship between individual's personality traits and job performance based on Big Five personality dimensions and concluded that personality trait is a good predictor in evaluating candidates for different job positions.

Companies spend more time and resources on evaluating suitability of each candidate. For example, Written/oral examinations and series of face-to-face interviews to ensure candidates' educational skills, technical knowledge and candidates' characteristics match the job demands. In the modern business environment, different jobs require different characteristics to excel in job performance.

So that, it opens new set of opportunities to improve existing recruitment process by incorporating effective candidates' personality evaluation.

1.2 Objectives

Main objective is to extract candidates' personality traits and include them in the evaluation process to hire suitable candidates for a particular job position. So that the proposed system should able to extract personality traits of candidates' with high accuracy using the popular "Big-Five Factors" and machine learning techniques. Further, the system should use those extracted candidate personality traits to evaluate the suitability of the candidate for a particular job position. This evaluation should be done along with evaluation of general criteria such as educational qualifications required and experience needed for a particular job. Therefore, the proposed system should assist recruiters to hire suitable candidates for available job positions.

Furthermore, the evaluation process should not be too time consuming and costly yet perform better than other existing e-recruitment systems in terms of choosing the right candidates. This system should be able to integrate with existing e-recruitment systems used by companies. Therefore, the proposed system should not replace the entire e-recruitment system but rather assist the recruitment process to hire suitable candidates.

1.3 Organization of the thesis

Rest of this thesis is structured as follows. Chapter 2 explains the evolution of recruitment process from traditional recruitment process to modern e-recruitment systems and their advantages as well as drawbacks. This chapter also talks about famous personality models and linguistic markers for each of the personality traits in Big Five from previous studies. Finally, various methods and techniques used in literature to extract personality traits from different sources, correlation between personality traits in Big Five and job performance along with some of the well-recognized personality measures are discussed in detail.

In Chapter 3 talks about the processes involved in extracting predictive feature sets as well as construction of the prediction model.

Chapter 4 evaluates the model discussed in this thesis and compares the results with baseline method and existing model. This chapter includes the details of experimental methods carried out and follows with a discussion.

Finally, chapter 5 concludes this thesis and discusses on future work directions.

2. LITERATURE REVIEW

2.1 Background

2.1.1 Recruitment process

Recruitment is one of the vital processes for any organization in the world. Any organization’s success heavily depends on human resource factors such as skills and competence of employees. Today, many organizations have outsourced their recruitment process to recruitment agencies. Companies conduct many personal development programs such as trainings to improve their employees’ productivity and make sure they cope up with current business practices. Companies laid out well standard procedures and policies towards recruitment process to ensure that candidates fulfill certain requirements needed to carry out job roles effectively.

For a very long period, newspapers are used to advertise job vacancies since newspapers are one of the few primarily mediums for communicating with the vast population in cost effective way. Once a job vacancy is published in newspapers, interested applicants will apply for that vacancy by mailing their curriculum vitae through postal services. The company will run an initial scan by going through each of these collected curriculum vitas and filter out curriculum vitas, which don’t fulfill basic requirements. Then rank the applicants based on predefined mechanism such as applicants’ qualifications, skills and experiences. Finally, shortlisted candidates will be asked to sit for a series of interviews such as technical interviews and formal interviews. Conducting interviews are very time and cost consuming task. However, this is important for a company to have interviews because this provides a good opportunity for them to talk with the candidates face to face.

Since interviews are very time and cost consuming tasks, companies cannot afford to conduct interviews for all the suitable candidates who fulfill the basic requirements for a given job position. Companies limit the number of interviews by picking up limited number of candidates whom they considered best. Technical knowledge of the candidate will be tested by oral or written technical question paper with limited time duration. Thus, interviewer can able to identify candidates' knowledge on technical subjects, background and the related skills.

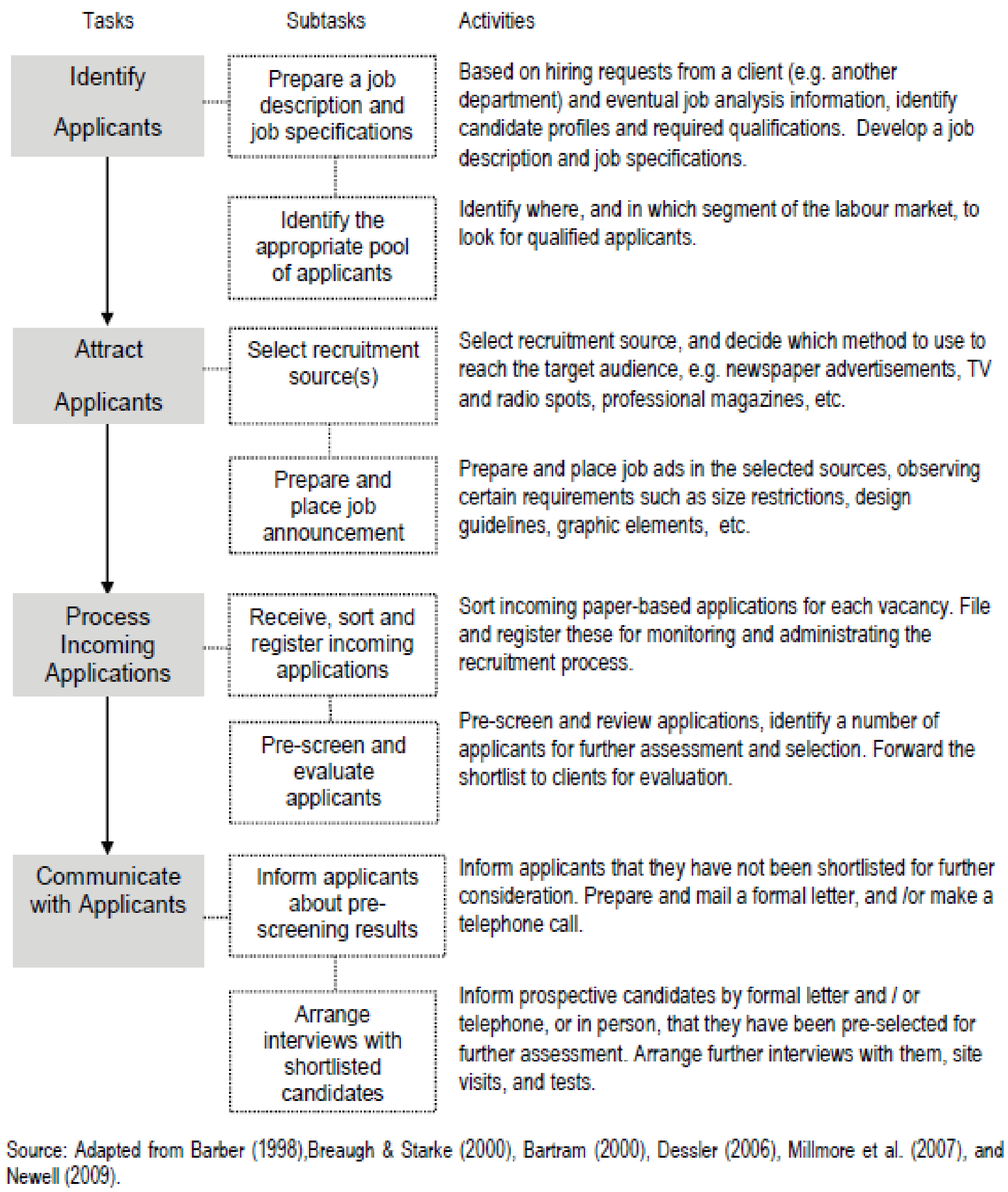

Figure 1.1: Traditional paper-based recruitment process using job advertising

Source: [10]

2.1.2 E-recruitment systems

Later due to rapid growth of internet accessibility, e-recruitment systems are now being widely used by companies throughout the world to target massive population [11], [12]. There are many e-recruitment systems available which automates the traditional recruitment process with the help of web. Most of the online recruitment web sites [13], [14], [15], [16] allow recruiters to advertise available job vacancies and help them to search for matching candidates based on general criteria such as educational qualifications, job experience. On the other hand, job seekers are allowed to search for matching job positions. Some of these online recruitment web sites [13], [16] provide users to have user accounts for better user management. However, applicants are free to search for matching job vacancies without having user account. Search for both jobs as well as candidates are done using simple key word matching. Therefore, the basic functionality of these e-recruitment systems was that it processes and retrieves matching job proposals as well as candidates based on general criteria selected by the user. Thus, facilitate both applicants and employers to engage with the system. Further these online recruitment web sites provide other facilities such as enable job applicants to apply for job positions via their site.

These e-recruitment systems were become popular in past two decades and replaced traditional recruitment process. E-recruitment systems eliminated hazards in traditional recruitment process and made it very effective [17]. Two major benefits are there in using e-recruitment systems. Job vacancies are well communicated with very large pool of candidates with very cost effective way and since it is automated, it collects applications and then filters based on basic requirements specified by recruiter.

Beside the educational qualifications and technical knowledge, there are important job related characteristics required to perform a particular job well. Recruiters are more emphasized on job characteristics because the nature of job roles have changed due to dynamic business environment. Therefore, whenever companies publish for job vacancies, they not only expect required educational qualifications but also job related characteristics from the candidates. Candidates' characteristics cannot be evaluated easily and existing basic e-recruitment systems are not design to extract candidates' personality traits. Thus, many companies fail to recruit best suitable candidates for job vacancies even with e-recruitment system in place.

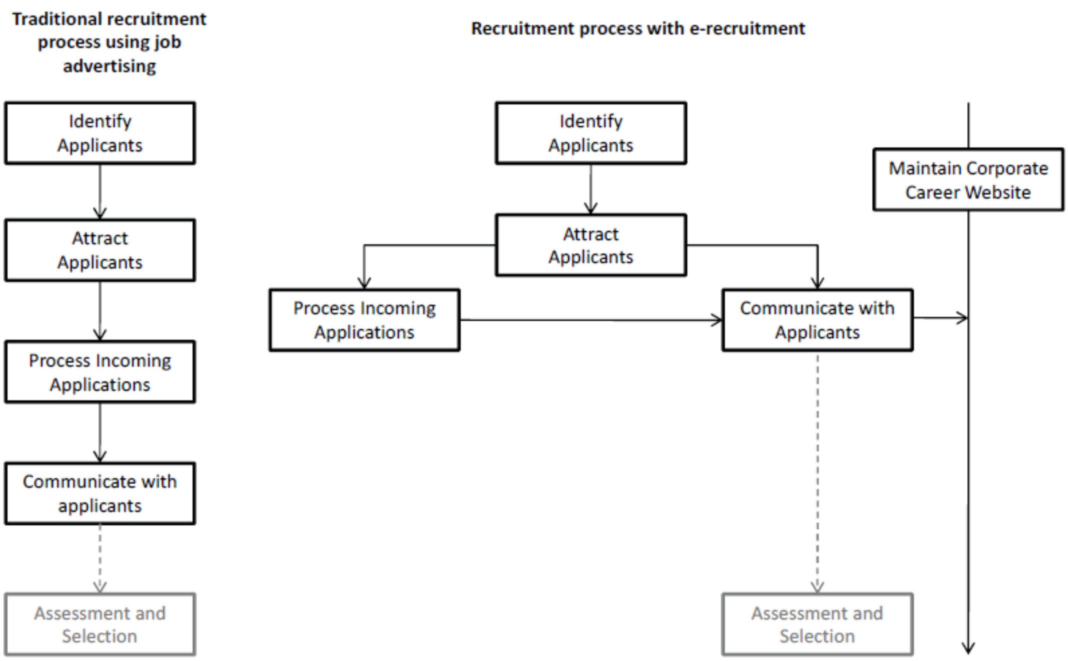

Figure 1.2: The design and sequence of tasks in traditional paper-based recruitment process vs. the (new) recruitment process using e-recruitment

Source: [10]

2.1.3 Personality models

In past six decades, several scientific studies were conducted to explore why an individual is unique and different from every other individual. Some of the famous personality models are Myers-Briggs Type Indicator® (MBTI®) [18], [19], Holland Occupational Themes [20] and Big Five Factors [21].

Myers-briggs type indicator® (MBTI®)

This theory is based on the "Psychological Types" described by Swiss psychiatrist Carl G. Jung. This theory identifies differences between individuals based on their personal preferences. This theory categorizes the preferences based on four different dichotomies. They are Introversion or Extraversion, Sensing or Intuition, Feeling or Thinking and Perceiving or Judging.

Each of the four different dichotomies captures different aspects of individual preferences. Introversion/Extraversion captures how people prefer to interact with the world and prefer to gain energy. This means that individual prefers to gain energy by spending time with people or by having time to themselves. Sensing/Intuition focuses on how people prefer to collect information. This means that whether people prefer to rely on facts or imaginations. Feeling/Thinking focuses on how people prefer to make decision. This means whether people prefer to make decision on what is reality or make decision by considering others' feelings. Finally, Perceiving/Judging captures how people prefer to live their life. This means whether people prefer to do things in much organized way or do things on an ad-hoc basis. Following describes each of these dichotomies.

Extraversion – Referring to people who know lot of people, interact with others a lot, very talkative and these people get motivated by socializing with people and avoid loneliness. For example, this group of people who like to go parties and participate in outdoor activities, which engage in lot of interaction with many people. Further, these people dominate conversations.

Introversion – Referring to people who is opposite of extrovert. These people are quiet and more like to be themselves. They do not start or actively take part in conversations. Rather these people are great listeners and do not get easily distracted by external environment.

Sensing – Referring to people who actually rely on facts and details based on what they sense and from their experiences. These people are more concern about the present rather figuring out future. For example, these people belief what they sense and not from unknown facts. These people are more practical than theoretical.

Intuition – Referring to people who focus on future based on their ideas and subconscious. These people are anxious about deriving possibilities about how the future may look like from insights and imaginations. For example, these people try to see the overall big picture and general concepts than looking thing separately in detail.

Thinking – Referring to people who make decisions based on scientific and logical aspects. These people give more importance to reality since these people deal with reality and make fair decisions. Thus, their decisions do not concern about others' feelings.

Feeler – Referring to people who give importance to others. For example, these people try to understand others' feelings and try to incorporate them when making decisions.

Judger – Referring to people who are well organized and try to schedule their works beforehand. For example, these people list out their work plan and finish each task to completion in an ordered manner.

Perceiver – Referring to people who do their work spontaneously and as and when they emerge. These people do not plan their work beforehand and can easily carry away with other external matters.

Figure 1.3 shows sixteen personality types based on these four dichotomies.

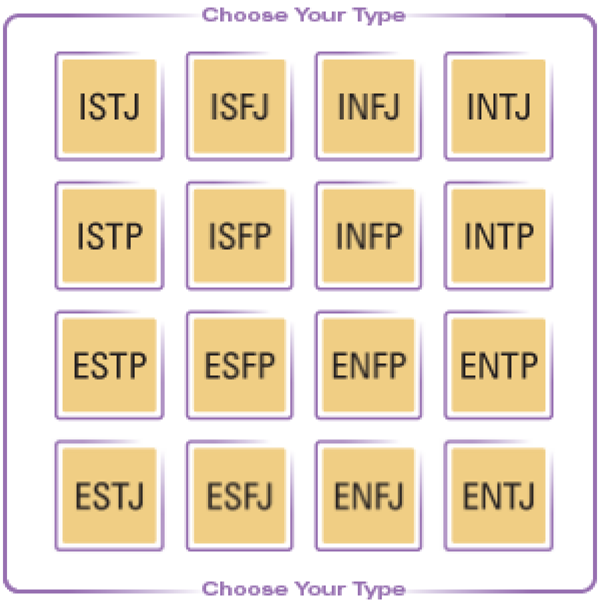

Figure 1.3: 16 personality types of the Myers-Briggs Type Indicator®

Source: [22]

Holland's codes

Based on John L. Holland's theory, most of the people can be categorized under six different personality types. They are Realistic, Investigative, Artistic, Social, Enterprising and Conventional.

Realistic – Those who like to work with real objects such as machineries, plants and animals. Thus, tend to be more practical. Realistic type of people is commonly known as "Doers".

Investigative – Those who are having high analytical skills such as observing, learning and solving problems. Investigative type of people is commonly known as "Thinkers".

Artistic – Those who are having high creative skills. They work in unstructured environment. Artistic type of people is commonly known as "Creators".

Social – Those who like to engage with other people. Social type of people is commonly known as “Helpers”.

Enterprising – Those who good at leading, influencing and managing people. Enterprising type of people is commonly known as “Persuaders”.

Conventional – Those who like to carry out the tasks in systematic and ordered way. Conventional type of people is commonly known as “Organizers”.

Big five factors

This famous scientific study explains the differences in individuals’ behavior. This theory is based on lexical hypothesis where personality dimensions were described based on natural language terms derived from dictionary. Allport and Odbert conduced lexical study where they have categorized 18,000 terms as relevant terms to describes personalities from second edition of Webster’s Unabridged Dictionary. Later Raymond B. Cattell [21] reduced this huge lexicon to 4,500 terms.

These Big Five factors are very broad in nature and these five dimensions are as follows.

Neuroticism – This indicates emotional stable of an individual. This focuses on how an individual would behave in a stressful environment. High emotional stable people are calm and relaxed whereas low emotional stable people are nervous, depressed and showing low performance with stress.

Extraversion – This indicates how individual interact with world. This means that extroverts are very talkative and participate in various activities, which involve lot of people. Introverts are reserved, quiet and comfortable with themselves.

Openness to experience – This dimension focuses on how individuals are open to try out new things. This means whether he/she welcome changes or reluctant to changes and try to follow similar way of doing things.

Agreeableness – This dimension describes what extend an individual respect others feelings. For example, agreeableness determines whether a person is co-operative or not.

Conscientiousness – Focuses on how an individual is organized and structured. High conscientious people are well organized whereas low conscientious people are less self-controlled and they do not plan their tasks well.

2.1.4 Data mining

Data mining is referred as discovering the unknown knowledge from large amounts of data. Over the years, very large volumes of data have been collected and this exponential growth of data became an important source for exacting useful patterns, which are not previously known. Processes involved prior to data mining are data selection, cleaning, integration and transformation and these are time-consuming processes. Once these processes are completed, then data mining is the application of various algorithms and techniques to derive unknown patterns.

The processes involved in knowledge discovery are as follows:

1. Data cleaning and integration
2. Selection and transformation
3. Data mining
4. Pattern exaction
5. Evaluation and presentation

2.1.5 Supervised vs unsupervised learning

Machine learning can be broadly divided into two sub categories. They are, supervised learning and unsupervised learning. If the corresponding output variable of the dataset is already known then it is categorized under supervised learning and if not then it is categorized as unsupervised learning. Classification problem and regression problem can be further categorized under supervised learning whereas clustering problem can be further grouped under unsupervised learning.

2.2 Traditional e-recruitment systems

E-recruitment systems mainly have two set of users. One is job applicants or job seekers and other is employers or recruiters. This means that job seekers can search through available jobs in e-recruitment system and on the other hand, employers can search for available candidates in the same e-recruitment system. In both of these scenarios, job seeker will be looking for suitable job vacancies and on the other hand, employers will be looking for best suitable candidates for available job vacancies in their companies. In order to find the relevant items, both job applicants and employers should depend on keywords search.

Several improved systems were proposed [12], [23], [24], [25], [5] in last fifteen years to overcome challenges in the e-recruitment systems. One of the major challenges is depending solely on keyword search matching which ends up with too many "hits".

2.3 Agent based e-recruitment systems

Some of the proposed systems incorporate agent technology in e-recruitment system. The concept of agent technology is all about assisting users by automating tasks through learning. In this context, agent technology can be used to enhance web browsing experience based on the user preferences [26] which are derived automatically from user's search behavior such as user's search history, stored reference to documents and user's decision to follow links further down.

Pasquale De Meo et al, presented approach called XML-based Multi-Agent system to support online recruitment system [12]. In their study, they combined both XML and multi-agent technology to build a robust system, which can support already existing e-recruitment systems. They have used XML since it is light weighted, versatile and easy to exchange and store data in different platforms. Further, they used multi-agent technology to personalized search based on the job seeker's preferences in addition to traditional key word matching. This approach is centered on job seekers. In their approach, they used "Multi-agent" technology where two or more agents will be interacting and collaborating with each other. This proposed system contains four agents. They are "User Agent", "Recruitment Agent", "Company Agent' and "Wrapper Agent".

One of the problems with is approach is that unlike web browsing, users of this system do not access it often. This means that when a job seeker searches for a relevant job in their system and again he/she might be searching in their system most probably after few years. Therefore, during that time, their job preferences might get changed. For example, if a job seeker looked for a software engineer position few years back and now he/she might be looking for a senior software engineer position or may be looking for a quality assurance engineer position in different companies. However, the system has already collected particular preference few years back and even though his/her preference changes in the future, system might suggests most of the jobs highly related to prior preference. Therefore has a high chance of missing out attractive job vacancies.

Other approaches were proposed based on agent technology for employer centered e-recruitment systems [23], [24], [25]. Unlike Pasquale De Meo et al approach, these proposed systems interact with users to construct their profile or preferences, which are then used by the system. Each of these proposed employer centric e-recruitment systems provide additional features beyond classical key word search. For example, this system [24] not only considers user preferences but also exploits economic factors. This approach [23] enables employers to conduct online interviews for initial screening. Problem with this approach is that companies may need to conduct several short interviews if the number of applicants are huge. Therefore, it is time consuming and costly process. This approach [25] tries to filter job applicants based on specific characteristics required for a sales person position by means on questionnaires. One of the problems with this proposed system is that job applicants might be answering favorable to these questions. For example, since system focuses only on recruiting sale persons, almost all job applicants mark as "strongly agree" for "Are you a talkative person?" type of questions.

2.4 Personality based e-recruitment systems

This proposed system [27] exploits personality traits of job seekers as means of special criteria for job searching beyond general criteria. The search results are displayed as weighted result of both special and general criteria and sorted in descending order. So based on the weighted result, most relevant job vacancies will be listed above. This system uses Five Factors model as a framework. In order to evaluate job seekers personality traits, system uses three different personality assessment forms to capture “Thinking Style”, “Occupational Interest” and “Behavioral Traits”.

Scale Name (Scale Description)	1	2	3	4	5
1) Energy Level (Tendency to display endurance and capacity for a fast pace.)					
2) Assertiveness (Tendency to take charge of people and situations. Leads more than follows.)					
3) Sociability (Tendency to be outgoing, people-oriented and participate with others.)					
4) Manageability (Tendency to follow policies, accept external controls and supervision and work within the rules.)					
5) Attitude (Tendency to have a positive attitude regarding people and outcomes.)					
6) Decisiveness (Uses available information to make decisions quickly.)					
7) Accommodating (Tendency to be friendly, cooperative, agreeable. To be a team person.)					
8) Independence (Tendency to be self-reliant, self-directed, to take independent action and make own decisions.)					
9) Objective Judgement (The ability to think clearly and be objective in decision-making.)					

Figure 2.1: Profile for Personality Traits used in [27]

Source: [27]

This proposed system [27] provides search functionality for both job seekers and employers. For example, job seekers can search suitable jobs based on their personal information, educational qualifications and based on their personality traits. On the other hand, employers also search for suitable candidates not only based on general criteria such as educational qualifications, experiences and salary but also based on the degree of match between candidate's personalities with personalities specified by employer.

One of the problems with this approach is that personality assessment is not comprehensive. Each of these three profiles for assessment only having few very broad sections. Profile for personality traits having nine sections and occupation interest profile and thinking style profile having six sections each. Since it assesses very broad topics, job seeker might not be clear on which scale score they might fall in to. Another problem is that in order to best fit with certain job positions, applicants can manipulate their profiles. For example, an applicant who is looking for good sale representative position might put high scores for "Energy Level", "Sociability", "Manageability" and "Decisiveness". So they will be ranked at the top of the list whenever an employer searches for best suitable candidates with "Matching on personality characteristic specification" option enabled in this proposed system. Further weights calculations for general criteria and special criteria are not described in this paper. In this proposed system, evaluation on personality traits are based on nonstandard methods because there are well defined standards [1], [2], [3], [4] for evaluating "Five Factor Model".

Another approach [5] was proposed to exploit personality characteristics using linguistic analysis on candidates' blog posts and objective criteria from his/her "LinkedIn" profile. Authors' objective of this proposed system was to limit interviewing and background analysis of candidates. One of the main challenges of this type of approach is that candidates must own a personal blog as well as a "LinkedIn" account. Authors used many different techniques to find correlation between job positions and personalities extracted from linguistic analysis. They used "Linear Regression", "M5' model tree", "REPTree decision tree" and "Support Vector Regression". Authors concluded that "Tree" models and "Support Vector Regression" with PUK universal kernel show best results.

2.5 Linguistic markers for personality traits

These studies [28], [29], [30], [31], [32] focused on extracting linguistic markers mainly for extraversion and each of other four personality traits in Big Five. This includes linguistic markers for different levels of language production such as written and speech. Some of the personality traits are more reflected via language use than others. For example, Dewaele and Furnham, Furnham, Mehl et al, Oberlander and Gill and Pennebaker and King found several linguistic markers for extraversion since extraversion is highly correlated with language use than other four personality traits in Big Five. Linguistic markers for extraversion, neuroticism, conscientiousness, agreeableness and openness to experience from previous studies [28], [29], [30], [31], [32] are as follows.

List of linguistic markers of extraversion

- Talk more, louder and repetitively with less number of unfilled pauses and hesitations
- Talks on many different topics
- Includes more positive emotional words than negative emotional words
- Shows more agreements than negations
- Use of many self-references
- Use of fewer articles (e.g. “a”, “an”, “the”)
- higher speech rates, shorter silences, higher verbal output
- Informal language use than introverts. For example, extravert more often greet with “hi” whereas introverts often use “hello”
- Fewer words per sentence or clause
- Use of more social language
- Total word count is relatively higher than introvert
- Use of more verbs, adverbs, pronouns and few tentative words
- Poor lexical choice
- More backchannel behavior

List of linguistic markers of introversion

- Use of broader vocabulary
- Talk less with many unfilled pauses
- Talks often express dissatisfaction and contain more negations
- Prefer strict topic selection such as on single topic
- Fewer self-references and many articles
- Many negative emotional words and fewer positive emotional words compare to extrovert
- Use of formal language
- Higher number of words per sentence or clause
- Fewer social words and increase use of tentative words
- Less backchannel behavior

Linguistic markers of high neuroticism

- Usage of many 1^{st} person singular pronouns
- Usage of many negative emotional words
- Use of less positive emotional words
- Use of fewer articles

Linguistic markers of conscientiousness

- Avoid negations and negative emotional words
- Avoid words which express discrepancies such as “should” and “would”

Linguistic markers of agreeableness

- Express more positive emotions
- Express less negative emotions

Linguistic markers of openness to experience

- Preference for longer words
- Tentative words such as “perhaps” and “maybe”

- Avoid 1st person singular pronouns and present tense forms

2.6 Predicting personality traits

Few studies [33], [34], [35], [36] were focused on exacting author's personality traits from written text as well as spoken language. One [34] of these studies was concentrated on exacting author's personality from personal weblogs or blogs. These weblogs mainly contain various studies carried out by each participant. In order to identify each participant's personality traits, revised NEO personality inventory [2] was used. Because in this study all participants were bloggers, some of these five personality traits such as extraversion and openness to experience might not be applicable. This is because bloggers write online blogs in their field of expertise to mass internet users and it implies that they are highly extravert and open type people. In order to measure the bloggers' personality traits, each participant was asked to take personality test from revised NEO personality inventory [2]. Surprisingly, scores for extraversion shows normal distribution whereas openness to experience was not normally distributed. Therefore, their study was restricted to only four personality dimensions excluding openness to experience. This study [34] was conducted as binary classification task as well as multiple classification task separately. This means that in binary classification either extremes of each personality traits will be classified (for example, whether a personal weblog is either having high agreeableness or low agreeableness). Whereas in multiple classification task, each personality traits will be grouped into three or five categories. For example, in multiple classification, a personal weblog can be classified as either highest neuroticism or relatively high neuroticism or medium level or relatively low neuroticism (relatively high emotional stability) or lowest neuroticism (highest emotional stability).

Authors of this research used "Support Vector Machines" and "Naïve Bayes" for binary classification task and used only "Naïve Bayes" for multiple classification task. This is because "Naïve Bayes" outperformed "Support Vector Machines" on binary classification task.

As for feature selection, authors used word-based bi- and tri-grams. Further, these features were selected either by “automatic” approach (using information gain) or by “manual” approach. In this study, authors have used different groups of participants (called “Tasks”) and varying levels of restriction on feature selection. For example, based on the scores from revised NEO personality inventory [2] instrument, participants were grouped as follows (below shows only for the binary classification tasks).

1. Those who fall above 1 standard deviation above the mean and those who fall below 1 standard deviation below the mean (around 50% of the participants were excluded in the task)
2. Those who fall above 0.5 standard deviation above the mean and those who fall below 0.5 standard deviation below the mean
3. Divide by exactly half (50% of participants are on one extreme and rest are on other side)

Five different levels of restriction on feature selection were as follows (below shows few of them).

1. n-grams most commonly occurring in the corpus
2. n-grams (equal to or more than five times occurring in each blog) distinctive for two extremes of each of four personality traits

Finally, results were compared with baseline where baseline is the majority classification. Further authors have reported that they achieved high accuracy for each of the four personality traits when compared to baseline.

Similar natural language processing technique has also been exploited in this study [33]. Their argument on using bigram analysis is that bigrams contain information about the interconnection between words, thus it exploits the contextual information of language the use. In this [33] study, authors only focused on one personality trait which was extraversion. They used the corpus of email texts. Conducted Eysenck’s EPQ-R (Eysenck Personality Questionnaire Revised) [37] personality test to identify personality traits of individual participants.

Based on their results, authors of this study [33] grouped bi-grams into three categories. These categories are bi-grams only used by extroverts, bi-grams only used by introverts and bi-grams used by both parties. Further authors have grouped features into eight main categories solely based on different use of the language by extroverts and introverts. They are shown with some example in the Table 2.1.

Table 2.1: Shows the main eight feature categories used by [33]

Main Category	Extravert	Introvert
Surface Realization Feature	Starts with “hi”. Multiple exclamation marks/full stops. Informal style and ‘loose’ use of language	Starts with “hello”
Quantification	Use of looser and less specific words such as “a bit”, “couple of”	Exaggeration. Greater use of quantifiers such as “a lot”, “a few”, “all the”, “one of”, “lots of”, “loads of”
Social Devices	Relax and informal. Stylistic expressions such as “catch up”, “take care”	No stylistic expressions
Self/Other Reference	First person bigram such as “i’ll be”, “i was”	Focus on self and more first person singular pronoun such as “I don’t”, “i went”
Valence	Positive affect or Positive disposition such as “a good”, “looking forward”	More negations such as “i don’t”, “don’t know”
Ability	Ability to do words such as “want to”, “need to”, “able to”	More timidly words such as “trying to”, “going to”
Modality	Use of stronger predictive words such as “i’ll be”	Use of weaker words such as “should be”
Message Planning/Expression	Use of connecting words such as “which”	Coordinating conjunction words such as “and”, “but”

In this approach [35] personality traits were extracted by analyzing the text’s style of writing. There are researches on exploring full meaning of a given text rather than tries to interpret the meaning of that text from its topic. This is a separate research area in computational stylistics. One of the interesting features is to explore the personality of the author from his/her text. Therefore, it [35] exploits lexical stylistics features to extract personality traits of authors. These lexical stylistics features are as follow.

1. Function words
2. Systemic Functional Grammar
3. Cohesion
4. Assessment
5. Appraisal

Argamon et al constructed a model to predict extraversion and neuroticism from both stream-of-consciousness essays and deep self-analysis essays. This is a binary classification model which means that only the two extremes of each of these two personality traits can be classified (e.g. extraversion or introversion and high neuroticism or low neuroticism) using Sequential Minimal Optimization (SMO) learning algorithm.

Results show that their accuracy rates for two extremes of each of the personality traits (neuroticism and extraversion) were not quite good. For neuroticism, overall accuracy rate was around 53%, 52% for stream-of-consciousness and deep self-analysis writing tasks respectively. Appraisal was the most useful feature set with more that 57% of accuracy rate for both writing tasks. Argamon et al argued that even though the accuracy rates were low, but it was quite significant for a single short text. Further authors of this study mentioned that in order to achieve high accuracy rate, either focused questions such as personality instruments (see section 2.8) or extended interaction, which means that continuously analyzing text written by same author over a period of time. For extraversion, over all accuracy rates were approximately 58%, 52% for stream-of-consciousness and deep self-analysis writing tasks respectively. Again, these accuracy rates are low.

Results clearly emphasized that negative appraisal is correlated with high neuroticism and positive appraisal is correlated with low neuroticism. On the other hand, “Function words” are most significant predictor for extraversion. Argamon et al also has listed top sixteen function words for each of two extremes of extraversion as in Figure 2.2.

Stream-of-consciousness (SoC)		Deep self-analysis (DSA)	
Low	High	Low	High
perhaps	may	seven	second
outside	immediate	try	comes
nobody	anyways	self	inner
fifth	yourself	except	normally
particular	mean	getting	get
uses	am	during	enough
second	so	hardly	very
their	being	sensible	particular

Figure 2.2: Top sixteen function words for low and high extraversion in stream-of-consciousness and deep self-analysis essays

This [36] research study extracts personality traits from two difference data sources. First one is a corpus of essays written by psychology students and it contains 2479 essays (over 1.9 million words). These students were asked to write whatever comes into the minds for 20 minutes and this is the same corpus used by Argamon et al. Finally, students' personalities were assessed using Five Factor Inventory [8] questionnaire.

Second source is voice conversations captured using Electronically Activated Recorded (EAR) from 96 participants' daily life over 2 days. Then these recorded conversations were converted to transcripts. Then individual participant's transcripts (utterances) were annotated with subjective information (such as type of interaction, location, activity, mood and language use) as well as participant's personality. In this data source, participants' personalities were assessed using battery of questionnaires (self-reports) as well as rating from 18 independent observers.

For feature extraction, authors used Linguistic Inquiry and Word Count (LIWC) tool for content and syntax analysis, 14 features from MRC Psycholinguistic database, to obtain psycholinguistic statistic of words, four utterance type features to tag whether an utterance type is command or prompt or question or assertion, and finally prosodic features. Out of these four features selection types, utterance type features and prosody features were more applicable for dialogues. Thus, these two features were used only in EAR corpus not in essay corpus.

Linguistic Inquiry and Word Count (LIWC) contains 88 word categories of linguistic features such as social words, cognitive words, positive emotional words, negative emotional words, etc. This content analysis tool assists to capture syntactic as well as semantic features from text. Syntactic features include standard counts such as word count, words with more than 6 letters, different pronouns count, etc. Whereas semantic features include positive emotional words, negative emotional words, anger words, sadness words, social words, certainty words, etc.

Authors constructed various regression models using “Linear regression”, “M5’ regression tree” with linear models, “M5’ decision tree” with regular leaves and “REP-Tree decision tree” since the output values vary continuously alone each personality dimension. Results were compared with baseline model, which is the mean personality score for each personality dimension from the whole training set.

All of these regression models clearly showed an improvement over baseline for essay corpus but showed poor improvement over baseline (higher than baseline error) for EAR data source. Few significant improvements were achieved for extraversion, emotional stability and conscientiousness for EAR corpus with observer ratings but not for EAR corpus with self-reports.

Authors also developed models for personality classification using “J48 decision tree”, “Nearest Neighbor”, “Naïve Bayes”, “JRip rules set”, “AdaboostM1” and “SMO Support Vector Machines (SMO)”. These classification models were compared with baseline model, which returns majority class for each personality dimension from the entire training set. For classification task, only EAR corpus with self-reports and observer ratings were used. Out of these classification algorithms, “Naïve Bayes” showed high accuracy rates for extraversion, emotional stability and conscientiousness. All the classification algorithms except “Nearest Neighbor” produce high accuracy rates for extraversion compare to other four personality dimension. This Results reproduced the previous finding that extraversion can be easily predictable personality dimension in spoken language than others.

Feature set	Acts	LIWC	MRC	Prosody
Set size	4	88	14	11
Extraversion	49.11	70.97•	67.96•	69.62•
Emotional stability	59.00	68.82•	61.04	61.88
Agreeableness	57.14	53.26	56.60	48.82
Conscientiousness	56.92	60.18•	65.76•	50.48
Openness	54.02	57.96	53.96	62.20•

• statistically significant improvement over the majority class baseline (two-tailed paired t-test, $p < 0.05$)

Figure 2.3: Break down of accuracy rates achieved by each of the four feature sets for "Naïve Bayes" classifier [36]

Figure 2.3 shows that effectiveness of each feature set. It is clear that Linguistic Inquiry Word Count (LIWC) utility with 88 word categories is a good predictor for extraversion, emotional stability and conscientiousness. Whereas MRC psycholinguistic dataset with 14 features produces better results for extraversion and conscientiousness. Prosody is a good indicator for extraversion and openness to experience. Speed acts with four utterance types features perform poorly on all the personality traits as its accuracy rates were not statistically significant improvement over baseline.

Finding shows new markers such as voice's pitch and variation of intensity are correlated with extraversion. Conscientiousness related with less use of swear words, content related to sexuality and preference for longer words.

This research [38] exploits same set of features used in the previous study [36] but with only EAR corpus and focused on personality recognition as a ranking problem. So that authors used "RankBoost" algorithm to rank participants based on each personality traits in Big Five. Their findings include new prosodic such as extraverts speak at high pitch and low conscientiousness people speak loudly. New markers such as low agreeableness correlated with high use of swear words and high agreeableness correlated with longer words with shorter sentences. In 2007, same set of authors from [36], [38] grouped with Mehl et al to publish another research work [33] with more granularity, using the same set of feature selection methods and the dataset.

This study [33] discussed few limitations of using LIWC utility to distinguish extrovert/introvert. Most of the feature sets in LIWC are more relevant for predicting neuroticism. This is because that this dictionary was originally constructed using texts written by distressed patients for therapeutic purposes. Thus, these texts often express variation in neuroticism than variation in extraversion. Another limitation is that, LIWC as well as MRC psycholinguistic dataset relies on simple word stem match and the context of the word is not considered. Again, this property may be suitable for neuroticism but not the extraversion.

2.7 Big five personality traits and job performance

The Five Factor Model evolved over many decades of research studies [21]. Many researchers have studied the relationship between personality traits and job performance. These studies [40], [41] have shown that relationship between personality traits and job performance is quite low. However these [6], [8] researches revised the conclusions made by those studies on the relationship and questioned the validity of the methods they have used. Further, findings of these [6], [8] researches have emphasized the relationship between Big Five and job performance.

The finding from this study [7] has also concluded that combination of low emotional stability, extraversion, openness to experience and conscientiousness having around 15% of the variance in task performance and creativity. This research sampled about 159 employees at corporate pharmacy group, which includes pharmacists as well as non-pharmacists. In order to evaluate their personality traits, they used NEO Personality Inventory Revised (NEO-PI-R) [2] that includes 240 questionnaires. For performance evaluation, they used Performance Appraisal Questionnaire (PAQ).

2.8 Standard set of questionnaires

Over past several decades, research on personality traits has been studied extensively and researchers have developed standard techniques for extracting individual's personality traits. All of them are questionnaires-centric. These questions and the number of item counts have been revised over the years.

The well-recognized measures are listed below:

1. NEO Personality Inventory (NEO-PI) [1]
2. Revised NEO Personality Inventory (NEO-PI-R) [2]
3. NEO Five-Factor Inventory versions (NEO-FFI) [2]
4. Goldberg's Big Five markers [3]
5. Hogan Personality Inventory (HPI) [4]

Number of item counts in each of these questionnaires seems to be an issue for many researchers. This is because these personality assessments consume lot of time due to large number of item counts. For example, Revised NEO Personality Inventory (NEO-PI-R) contains 240-item, NEO-FFI contains 60-item and Goldberg's Big Five markers contains a set of 100 unipolar terms. Therefore, to reduce the item counts, alternative approaches [42], [43] were proposed based on the above inventories. Big Five Inventory [42] having 44-item with short phrases with relatively accessible vocabulary. This approach [43] proposed Ten-Item Personality Inventory (TIPI), which contains only 10 items but with serious limitations [43] such as validity of this instrument since it uses short measures.

3. METHODOLOGY

3.1 Introduction

This thesis focuses on predicting the candidate personality traits using scenario-based analysis with machine learning techniques. Here candidate will be asked to participate in a written conversation where questions are not straightforward as in standard questionnaires. Based on the answers given by the candidates, their personality trait will be extracted using machine learning techniques. These written conversations will be related to a natural dialogue type and not necessarily restrict to a particular subject. For example, candidate will be asked to compare two restaurants where he/she has been to and asked to recommend a restaurant.

This thesis exploits the use of natural languages to predict personality traits using Big Five model and the evolution of Big Five model was based on lexical hypothesis [21].

3.2 Focused personality traits

This study focuses on two of the most important personality traits, which are extraversion and neuroticism (emotional stability). Previous studies (see section 2.7) showed that significant correlation between these two personality traits and job performance. For example, recruiting an extrovert type of candidates helps to build strong team players, which is highly expected soft skill in the modern business world. On the other hand, high emotional stable people able to do critical tasks while handling stress.

3.3 Data collection

3.3.1 Personality dataset

Personality Dataset was used for the purpose of training and testing the proposed model in this thesis. It contains 580 generated utterances [44] and annotated with personality ratings by independent human judges. These personality rating were given using Ten-Item Personality Inventory [43]. This means that human judges first read an utterance and then try to answer the ten questions in that inventory. Answers were based on each judge's independent imagination on what type of personalities of a person would have most likely produced this utterance. Based on the answers given by judge, final rating for each personality traits was calculated. Two or three independent judges gave final ratings. So that, each utterance was annotated with the final scores (for each personality traits in Big Five) from two or three independent judges as well as an average score for each personality traits. Further, all 580 utterances were annotated with extraversion as well as naturalness whereas only 320 utterances annotated with other four personality dimension which are neuroticism, agreeableness, conscientiousness and openness to experience.

Sample utterance – only annotated for extraversion and naturalness

```
<realization>I am sure you would like Daniel and Aureole.
Daniel just has wonderful servers and the ambience is lovely.
The food is kind of brilliant, even if it's expensive.
Aureole features great service and the atmosphere is beautiful.
The food is excellent, even if it's costly.</realization>
```

Figure 3.1: Sample utterance from the data source

```
<ratings>
    <rating type="avg.extra" value="6.33333333333333"/>
    <rating type="avg.naturalness" value="5.66666666666667"/>
    <rating type="userA.extra" value="6.5"/>
    <rating type="userA.naturalness" value="6"/>
    <rating type="userB.extra" value="6"/>
    <rating type="userB.naturalness" value="6"/>
    <rating type="userC.extra" value="6.5"/>
    <rating type="userC.naturalness" value="5"/>
</ratings>
```

Figure 3.2: Personality scores for extraversion and naturalness by three individual human judges (namely userA, userB and userC) and average scores as well

Sample utterance – annotated with all five personality traits in Big Five

```
<realization>Err... I am not sure. Mmhm... I mean,
Mavalli Palace is low-cost
and Once Upon A Tart is a cafe and sandwich place,
it's a cafe and sandwich place,
also it's located in TriBeCa/SoHo.</realization>
```

Figure 3.3: Another sample utterance from the data source

```
<ratings>
    <rating type="avg.agree" value="4.75"/>
    <rating type="avg.consc" value="5"/>
    <rating type="avg.ems" value="4.5"/>
    <rating type="avg.extra" value="1.75"/>
    <rating type="avg.naturalness" value="4"/>
    <rating type="avg.open" value="2.5"/>
    <rating type="userC.agree" value="3.5"/>
    <rating type="userC.consc" value="5"/>
    <rating type="userC.ems" value="3"/>
    <rating type="userC.extra" value="1.5"/>
    <rating type="userC.naturalness" value="3"/>
    <rating type="userC.open" value="2.5"/>
    <rating type="userD.agree" value="6"/>
    <rating type="userD.consc" value="5"/>
    <rating type="userD.ems" value="6"/>
    <rating type="userD.extra" value="2"/>
    <rating type="userD.naturalness" value="5"/>
    <rating type="userD.open" value="2.5"/>
</ratings>
```

Figure 3.4: Two independent human judges' (namely userC and userD) personality scores and the average scores for each of the five personality traits in Big Five and naturalness

3.3.2 Yelp open dataset

This Yelp Open Dataset is from yelp. Yelp is a commercial website, which facilitates to find restaurants, nightlife and home services such as electricians, home cleaners, mover, dentists, etc. This dataset totally contains 5,200,000 reviews for 174,000 businesses from several different geographical locations. This dataset can be downloadable in JSON or SQL format. JSON dataset is structured as in the following way.

Business – this section contains all the attributes pertain to that particular business. For example, address, overall rating, business categories such as Mexican, burgers opening hours, etc.

Review – full review text, number of stars given by reviewer, date of review, number of votes received for “useful”, “funny” and “cool” categories, etc. are included in this review section.

User – this part carries user related attributes such as number of review counts, joined date, number of “useful”, “funny” and “cool” votes received from this user, number of fans, average number of stars given for all reviews, number of compliments such as hot compliments, profile compliments, cute compliments received by other registered users, etc.

Check-in - count of check-ins on hourly basis throughout week.

Tip - it contains tip text (shorter than reviews) by users regarding the business, date of tip published, number of ‘likes’ received for the tips, etc.

Photos – attributes regarding any photos such as caption, label, etc.

Randomly, 1000 restaurant reviews were extracted from Yelp Open Dataset and these reviews were used to produce most predictable feature sets for this research work. In each review set, only review texts which having one (low rate) and five (high rate) score ratings were considered. Table 3.1 shows some samples extracted from this database.

Table 3.1: Shows four sample review texts extracted from Yelp Open Dataset with user given review scores

Review Star	Review Text
1	This place is horrible, we were so excited to try it since I got a gift card for my birthday. We went in an ordered are whole meal and they did not except are gift card, because their system was

	down. Unacceptable, this would have been so helpful if we would have known this prior!!"
1	I wish I could give 1.5 stars. Nothing special. Lack of flavor. The entrees were either sweet or spicy. The crab Rangoon were.....different. The filling had a mealy consistency. Friend rice was bland. Plenty of other places to spend 50$ on takeout. Save your money.
5	I love this place i'd recommend it to anyone ! We always order it togo and it never disappoints! The food always taste fresh and is always ready on time! Definitely our favorite lunch spot !
5	Super clean restaurant and friendly staff. FRESH food. Hasn't been sitting under heat lamps. NO MSG, this is the good stuff. I have to have the Kung Pao Chicken weekly.

3.4 Scenario based questions

As mentioned in section 3.1, questions are not straightforward as standard set of questionnaires used in varies inventories (see section 2.8). These questions are based on recommendations and comparisons of a place. This thesis particularly concentrates on recommendation and comparisons of restaurants. Thus, the system knows the context upon which candidates' answers will be based on. This means that, in this scenario, candidates' answers will be mostly based on food, price, service, cuisine and atmosphere/decoration. Therefore, candidates' answers can be categories into five different contents. Thus, it makes easier to understand and analyze the content of the candidates' answers such as some of its semantic properties.

Most of the previous studies were focused on extracting personality traits from essays corpus such as stream-of-consciousness and deep self-analysis [35], [36], [39], email corpus [33], weblogs [34], EAR data source [36], [38], [39] and standard questionnaire. Understanding the semantic properties from these data sources have been a difficult task, thus most of the time authors were relied on LIWC utility and MRC psycholinguistic database to identify syntactic as well as semantic information.

In this thesis, questions are in static form related to recommendation as well as comparison of restaurants candidates have been to.

Sample questions are as follows

- What are the restaurants you have been to?
- From those restaurants (you mentioned above), how do you compare them?
- Out of those restaurants, which one do you recommend most?

It is clear that candidates' answers to the above static questions will be around restaurant domain. Major advantage is that the context of the text as a whole is known. Further, these static questions can be easily extended to other kind of places and products. For example, other than restaurant comparisons and recommendations it can be also extended to movies, tourist destination spots, hotels, online store products such as books, electronic items, etc.

3.5 Feature extraction

3.5.1 LIWC word category

LIWC utility contains 88 word categories and only 1 word category was used in this thesis. LIWC utility was constructed by carefully choosing the words and categorizing them over many years of research [32]. Table 3.2 shows the word category used in this thesis.

Table 3.2: Shows the word category used from LIWC utility

Word Category	No. of Words	Sample words
Negations	31	NOPE, NOTHING, NEITHER, UHUH, NOT, etc.

3.5.2 Natural language processing technique

Yelp Open Dataset was used to extract important predictor feature set. This dataset having user reviews for many types of businesses. Out of those reviews, only restaurant related user reviews were carefully chosen for this study. In order to do that, “categories” tag was used to filter out user reviews, which are not mainly related to restaurants. Excluded categories were “Salon”, “Gyms” and “Hair” and included category was “Food”. For example, user reviews which are tagged with “Restaurants” and “Gyms”, mostly talk about the quality of services provided by the gymnasium and rest of the review on cafeteria attached to that gymnasium. So those reviews were not been considered for the research study.

Total of 1000 restaurants reviews were randomly selected for this study. Out of those reviews, half of them were selected with the score of 1 and another half with the score of 5. Review score 1 means that user has given very low score since he/she had worst experience with the restaurant whereas review score 5 means that user has given very high score since the experience was great with that restaurant. So that, corresponding review texts reflex the kind of experiences user had with the restaurant (see Table 3.1).

Bi-grams were extracted from these randomly selected 1000 review texts (with score either 1 or 5 only). Review texts, which were marked as one by the reviewers convey negative emotion and on the other hand review texts which were marked as 5 by the reviewers convey positive emotion. So, two sets of bigram were constructed. One is positive emotional bi-grams, which only occur in five star review text, and other is negative emotional bi-grams that only occur in one star review text.

Further, each review texts were preprocessed before extracting bi-grams. Following were the steps carried out in preprocessing stage.

- Removed all special characters such as “\n” and “\””
- Changed all words to lowercase characters
- Removed words such proper noun singular, proper noun plural and symbols
- Any numbers were converted to tag called “::NUM”
- Any percentages format such as “3/4” were converted to tag called “::PERCENTAGE”
- Any dates format were converted to tag called “::DATE”
- Any time format were converted to tag called “::TIME”

- Any “/” symbol in the middle of a single word were split into two separate words (both sides of “/”)

Each review needs to be tagged with Part-Of-Speech (POS) in order to remove some of the words such as proper noun singular, proper noun plural and symbol. So to do that, each review text was processed through Stanford Log-linear POS Tagger. This tool was developed and maintains by The Stanford Natural Language Processing Group, which includes members from linguistic department as well as computer science department. Output from the POS Tagger is the tokens assigned with POS information.

Table 3.3: Sample output from Stanford Log-linear POS Tagger

Sample Review text	Output from Stanford Log-linear POS Tagger
The food is great and customer service is the best! The Dan Dan noodles are dynamite but they come standard pretty spicy... I placed a pick up order by phone and asked for them to be mild, but they ended up being spicy! When I got home and tried to eat them they were too spicy. I called in and spoke to the manager and he took care of it completely and replaced the order for me at no charge!	The/DT, food/NN, is/VBZ, great/JJ, and/CC, customer/NN, service/NN, is/VBZ, the/DT, best/JJS, !/., The/DT, Dan/NNP, Dan/NNP, noodles/NNS, are/VBP, dynamite/NN, but/CC, they/PRP, come/VBP, standard/JJ, pretty/RB, spicy/JJ, .../:, I/PRP, placed/VBD, a/DT, pick/NN, up/RP, order/NN, by/IN, phone/NN, and/CC, asked/VBD, for/IN, them/PRP, to/TO, be/VB, mild/JJ, ,/,, but/CC, they/PRP, ended/VBD, up/RP, being/VBG, spicy/NN, !/., When/WRB, I/PRP, got/VBD, home/NN, and/CC, tried/VBD, to/TO, eat/VB, them/PRP, they/PRP, were/VBD, too/RB, spicy/JJ, ./., I/PRP, called/VBD, in/RP, and/CC, spoke/VBD, to/TO, the/DT, manager/NN, and/CC, he/PRP, took/VBD, care/NN, of/IN, it/PRP, completely/RB, and/CC, replaced/VBD, the/DT, order/NN, for/IN, me/PRP, at/IN, no/DT, charge/NN, !/.

Table 3.3 shows a sample review text extracted from Yelp Open Dataset and the corresponding output of Stanford Log-linear POS Tagger for that review text. For example, proper noun singular words were tagged this “NNP” and symbols were marked with either “.” or ‘:’ in the Table 3.3.

Figure 3.5 describes the meaning of each tag from Stanford Log-linear POS Tagger.

Tag	Description	Example	Tag	Description	Example
CC	coordin. conjunction	*and, but, or*	SYM	symbol	+,%, &
CD	cardinal number	*one, two*	TO	“to”	*to*
DT	determiner	*a, the*	UH	interjection	*ah, oops*
EX	existential ‘there’	*there*	VB	verb base form	*eat*
FW	foreign word	*mea culpa*	VBD	verb past tense	*ate*
IN	preposition/sub-conj	*of, in, by*	VBG	verb gerund	*eating*
JJ	adjective	*yellow*	VBN	verb past participle	*eaten*
JJR	adj., comparative	*bigger*	VBP	verb non-3sg pres	*eat*
JJS	adj., superlative	*wildest*	VBZ	verb 3sg pres	*eats*
LS	list item marker	*1, 2, One*	WDT	wh-determiner	*which, that*
MD	modal	*can, should*	WP	wh-pronoun	*what, who*
NN	noun, sing. or mass	*llama*	WP$	possessive wh-	*whose*
NNS	noun, plural	*llamas*	WRB	wh-adverb	*how, where*
NNP	proper noun, sing.	*IBM*	$	dollar sign	*$*
NNPS	proper noun, plural	*Carolinas*	#	pound sign	*#*
PDT	predeterminer	*all, both*	“	left quote	‘ or “
POS	possessive ending	*’s*	”	right quote	’ or ”
PRP	personal pronoun	*I, you, he*	(	left parenthesis	[, (, {, <
PRP$	possessive pronoun	*your, one’s*	)	right parenthesis	],), }, >
RB	adverb	*quickly, never*	,	comma	,
RBR	adverb, comparative	*faster*	.	sentence-final punc	. ! ?
RBS	adverb, superlative	*fastest*	:	mid-sentence punc	: ; ... – -
RP	particle	*up, off*			

Figure 3.5: Shows Penn Treebank POS tags with examples (punctuation mark was also included)

These bi-gram feature sets are most appropriate for extracting positive and negative emotions. This is because candidates need to answer restaurant related scenario-based questions and these all bi-gram feature sets were drawn from review texts, which were related to restaurants domain.

Below list shows top 20 most frequent positive emotional bi-grams only occurs in review text with score 5.

1. the tour
2. tour and
3. very friendly
4. was amazing
5. vegan pizza
6. love this
7. and friendly
8. fresh and
9. a tour
10. the distillery
11. good as
12. atmosphere is
13. of tea
14. is awesome
15. a pot
16. the brewery
17. the tasting
18. wish they
19. was delicious
20. 'll definitely

Below list shows 20 most frequent negative emotional bi-grams only occurs in review text with score 1.

1. the worst
2. i asked
3. tasted like
4. will never
5. at me
6. she was
7. iced coffee
8. was ok
9. your money

10. not worth
11. the girl
12. the bottom
13. ever been
14. took ::num
15. when she
16. ::num mins
17. does not
18. have no
19. the movie
20. and no

Unigrams can be also extracted from “Yelp Open Dataset”. This can be done using this Stanford Log-linear POS Tagger and select only adjective words, which are marked as “JJ”. This is because adjective words describe the noun and most likely to express positive or negative emotions.

However in this [45] study, authors extracted a comprehensive list of positive and negative emotional words. List contains 2006 positive opinion words as well as 4783 negative opinion words. Alone with negation word category from LIWC utility, this comprehensive list of positive and negative emotional words were also used.

3.5.3 Process of assigning topics

Firstly, each candidate’s answers were processed through Stanford Log-linear POS Tagger in order to tag with POS information. At this point, answers were broken into separate smaller parts using words such as coordinating conjunction (words tagged as ‘CC’) and symbols such as sentence-final punctuations mark as well as comma.

Sentence-final punctuation marks were used to separate each sentence and these sentence-final punctuation marks are “.”, “!”, “?”. Besides sentence-final punctuation marks, comma was also used to further break down the sentence into smaller parts.

Reason behind using coordinating conjunction words as delimiter is that conjunction words are used to join two phrases, clauses or sentences. So that each of these phrase, clause or sentence might have used to convey different contents. Thus, exploiting these conjunction words such as “and”, “but”, “or”, “as”, “if” and “when” are very useful to break down a sentence into meaningful phrases, clauses or small sentences.

Second major step is to tag each of these phrases, clauses or small sentences with labels from seven broad categories as shown in Table 3.4.

Table 3.4: Shows the seven content labels which are used to tag

Labels	Keywords
Name (of the restaurant)	Name of the restaurants (candidates been to) will be asked at the first place. So that name are used as the keyword to identify the “Name” label. Beside that general words such as “restaurant”, “place”/”this place” etc.
Food	General word such as “food” and other keywords such as “rice”, “sandwich”, “curry”, “chicken”, “prawn”, “drink”, etc.
Price	Keywords such as “price”, “expensive”, “inexpensive” and “dollar” are used.
Service	General word such as “service” and many other words such as “server”, “staff”, “waiter” and “waitress” are used to tag “Service” label.
Cuisine	Exact keywords such as “Thai”, “French”, “Chinese”, “Indian”, “Italian”, “Korean”, “German”, and “Bistro” are good predictors for “Cuisine”.
Atmosphere/Decoration	The words such as “decoration”, “atmosphere”, “ambience”, etc.
NoContent	This is the default label used when none of the above label matches.

This should be noted that there might be more than one label (except ‘NOCONTENT’) can be applied to one phrase or clause or small sentence. For analyzing purpose, utterance number and sub part number are also used. For example, when first phrase or clause or small sentence in the second utterance speaks about food and price, then it will be tagged as “food_price_2_1”. Table 3.5 shows how a review text from Yelp Open Dataset is tagged with the labels shown in above Table 3.4.

Review text:

Super clean restaurant and friendly staff. FRESH food. Hasn't been sitting under heat lamps. NO MSG, this is the good stuff. I have to have the Kung Pao Chicken weekly.

Table 3.5: Shows how the text is tagged with content labels

Tagged label		Part of review sentences
name_1_1	⟹	[Super/NNP, clean/JJ, restaurant/NN]
service_1_2	⟹	[friendly/JJ, staff/NN]
food_2_1	⟹	[FRESH/JJ, food/NN]
NOCONTENT_3_1	⟹	[Has/VBZ, n't/RB, been/VBN, sitting/VBG, under/IN, heat/NN, lamps/NNS]
NOCONTENT_4_1	⟹	[NO/DT, MSG/NNP]
NOCONTENT_4_2	⟹	[this/DT, is/VBZ, the/DT, good/JJ, stuff/NN]
food_5_1	⟹	[I/PRP, have/VBP, to/TO, have/VB, the/DT, Kung/NNP, Pao/NNP, Chicken/NNP, weekly/JJ]

As shown in Table 3.5, sentences further broken down into smaller parts of sentence. This technique allows to preciously tagging each of them under seven content labels shown in Table 3.4. As discussed in section 3.1, a candidate’s answer can be broadly categories under five different contents such as food, price, service, cuisine and atmosphere/decoration. This means that most of the time a candidate’s answers around on these broad categories.

This technique allows counting the number of topics a candidate’s text talks about. This is one of most predictable feature for extraversion since extroverts speak on many topics whereas introverts speak on fewer topics (see section 2.5).

Table 3.6: Describes five broad categories (or contents) a typical candidate's answer may contain

Categories (Contents)	Description
Food	In this category, candidates talk about the quality of the food such as its taste, neatness, decoration, etc.
Price	Candidates speak about whether the prices of food are high, reasonable, or inexpensive.
Service	How well the staffs treated candidates and the quality of services received when they were at the restaurants.
Cuisine	This section explains the type of cuisine such as whether it's a Thai, Chinese, Italian, German or Indian, etc.
Atmosphere/Decoration	Here candidates express their opinion on appearance and environmental conditions of the restaurant's as a whole.

Next major step is to measure the degree of positive emotions and negative emotions in each candidate's answers. In order to identify the degree of positive emotions and negative emotions, two main sources were used (see section 3.5). Number of matches was counted separately for positive emotions and negative emotions. In each attempt to find a match, steam of each word was derived using Stanford Log-linear POS Tagger. Finally, calculated total counts were used to measure the degree of positive emotions and negative emotions and fed into artificial neural network.

Measuring the degree of positive emotions and negative emotions in candidates' answers help to measure the extraversion as well as neuroticism. This is because extraversion and neuroticism mainly depend on the usage of positive and negative emotions in candidates' text (see section 2.5).

3.6 Construction of the prediction model

3.6.1 Artificial neural networks

In this thesis, artificial neural network was used to produce scores for personality traits. To my knowledge, this technique has not been exploited so far in this research domain. Since the output is a continuous numerical value (between scales of one to seven), artificial neural network with regression was used. Stochastic gradient descent optimization algorithm has been used and Mean Squared Error (MSE) has been used as the loss function. In order to implement a neural network, “DeepLearning4j” [46] library has been used. This is an open-source Java-based library for deep-learning neural networks. It has modules for preprocess dataset, construct neural networks, provides options to process matrix data on CPU or GPU, etc.

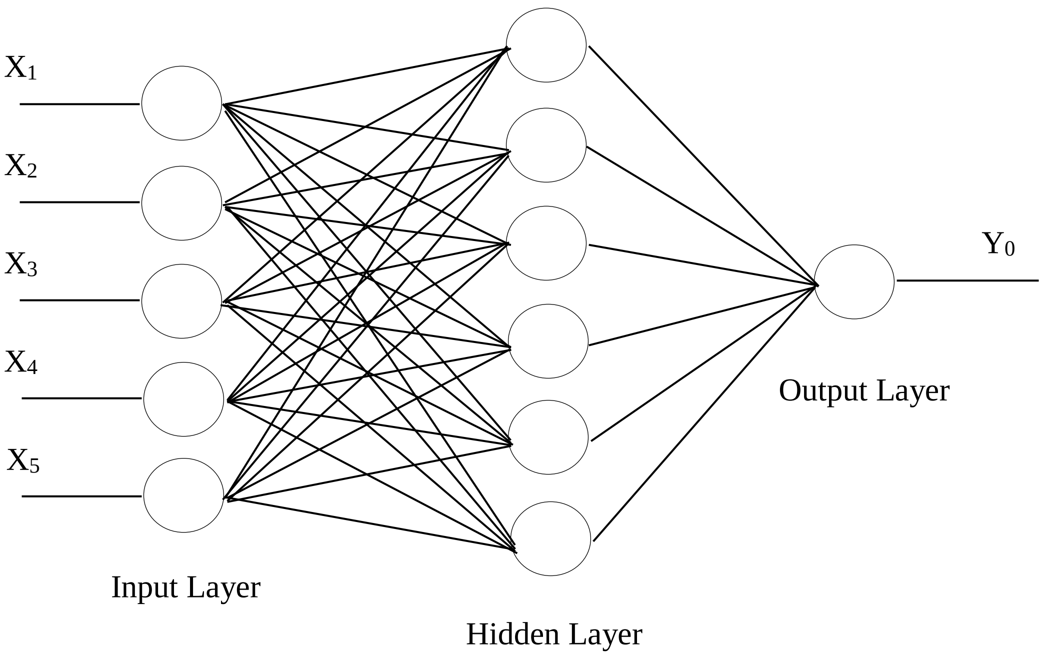

Figure 3.6: Shows the structure of Artificial Neural Network

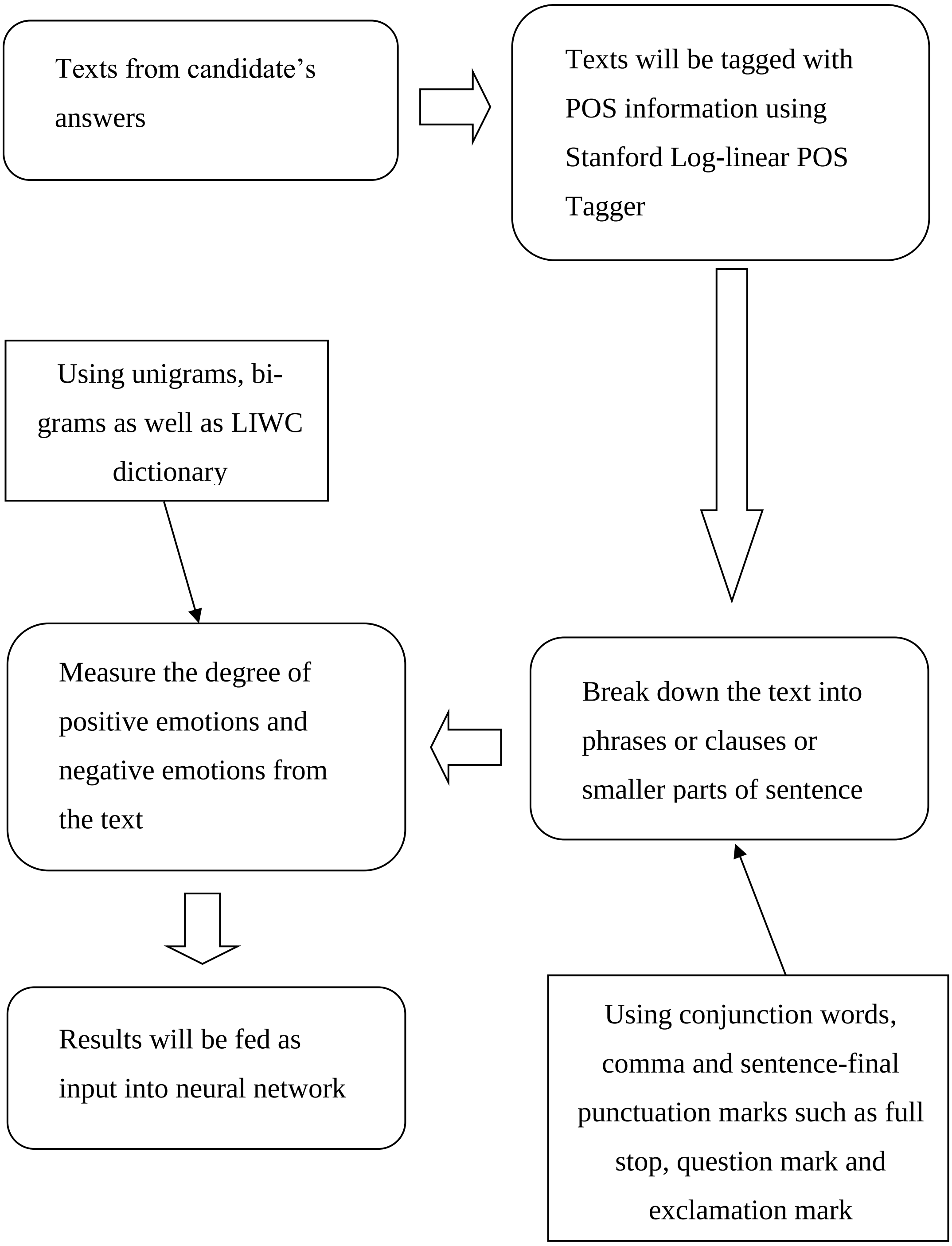

Figure 3.7: Shows all the main processes involved in measuring degree of positive emotions and negative emotions in a candidate's answer

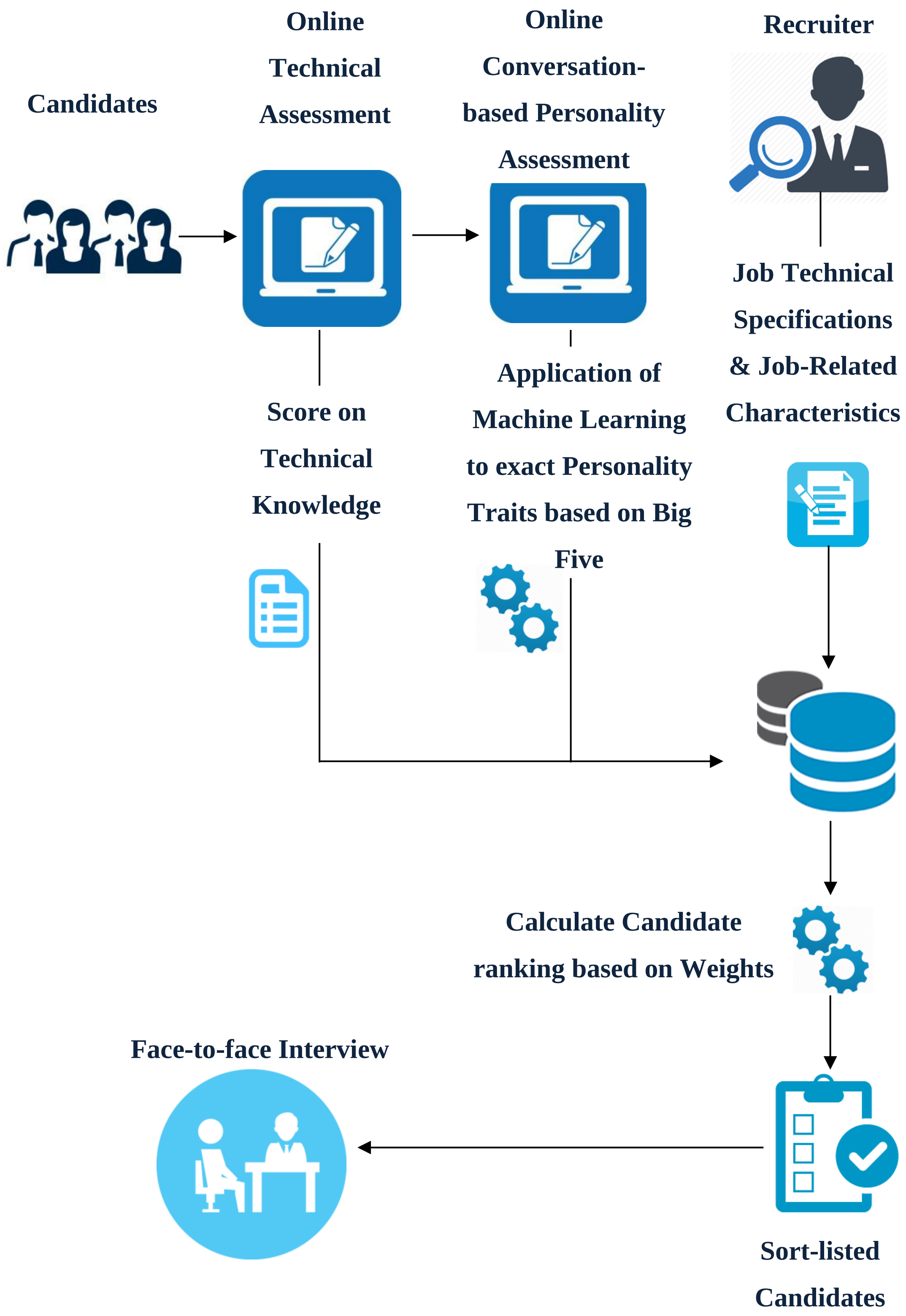

Figure 3.8: Shows how candidates' personality traits evaluations can be incorporated into existing e-recruitment system

4. EXPERIMENTAL EVALUATION

4.1 Training and testing

Input features were extracted from the given candidate answer. Then these features were fed into artificial neural network and corresponding output value was compared against with the actual scores. For the purpose of evaluation, annotated personality dataset (see section 3.3.1) was used.

Extraversion and neuroticism

For extraversion, 300 annotated utterances from personality dataset were used for training while 30 records were held out for testing. These 30 records had a naturalness score of more than six. Besides that, 100 records, which had a naturalness score of five and above used for testing. For neuroticism, 320 records were taken for evaluation. From these records, for training 290 records and for testing 30 records with a naturalness score of five and above were used. Besides that, 10-fold cross validation also performed.

Initially, 15 input features were used to construct the model based on the linguistic markers for personality traits from previous literature studies (see section 2.5). However, the error rate was significantly high when evaluating the model with all 15 input features. Error rate was calculated as Root Mean Squared Error (RMSE), which measures the difference between model prediction and actual measurement. Following shows the initially used input features.

1. Number of total words and symbols
2. Number of sentences
3. Number of words per sentence
4. Number of topics
5. Number of self-reference words
6. Number of articles
7. Number of negations words

8. Number of fillers words
9. Number of tentative words
10. Number of social words
11. Number of positive emotional words
12. Number of negative emotional words
13. Number of positive bi-grams match
14. Number of negative bi-grams match
15. Number of common bi-grams match

Beside positive bi-grams and negative bi-grams, common bi-grams were also produced using the same "Yelp Open Dataset". These common bi-grams occur both in 1 star review texts and 5 star review texts. Reducing some of these input feature sets reduce error rate. Table 4.1 shows the best predictive features which give low error rate for extraversion and neuroticism.

Table 4.1: Table shows the predictive input feature sets for extraversion and neuroticism

Input Features	Extraversion	Neuroticism
Total number of words and symbols	✗	✗
Total number of topic contents (see section 3.5.3)	✗	
Number of positive emotional words match (see section 3.5.2)	✗	✗
Number of negative emotional words match (see section 3.5.2)	✗	✗
Number of positive bi-grams match (see section 3.5.2)	✗	✗
Number of negative bi-grams match (see section 3.5.2)	✗	✗

Two different artificial neural networks were constructed for each of the two prediction tasks (extraversion and neuroticism). This is because different combinations of parameters produce best prediction result in each of these two prediction tasks. Six hyper-parameters of artificial neural network are presented in the Table 4.2. All these six parameters were manipulated to find the optimized combination of parameters, which produces lowest RMSE.

Table 4.2: Table shows the combinations of hyper-parameter values used to construct artificial neural networks for extraversion and neuroticism

Hyper parameters	**Extroversion**	**Neuroticism**
Number of epochs	400	100
Number of iterations	1	1
Batch size	35	75
Number of hidden layers	1	1
Number of hidden layer nodes	6	7
Learning rate	0. 017	0. 011

4.2 Evaluation measures

Following are the evaluation measures used to evaluate different regression models.

1. Mean Squared Error (MSE)
2. Root Mean Squared Error (RMSE)
3. Mean Absolute Error (MAE)
4. Relative Squared Error (RSE)
5. Relative Absolute Error (RAE)

RSE and RAE used to measure error relative to a simple predictor. For this study, MSE, RMSE and MAE were used for model evaluations as well as model comparisons with previous literature. These three evaluation methods are calculated as follows.

$$\text{Mean Squared Error (MSE)} = \frac{\sum_{i=1}^{n}(p_i - a_i)^2}{n}$$

$$\text{Root Mean Squared Error (RMSE)} = \sqrt{\frac{\sum_{i=1}^{n}(p_i - a_i)^2}{n}}$$

$$\text{Mean Absolute Error (MAE)} = \frac{\sum_{i=1}^{n} |p_i - a_i|}{n}$$

This [39] study was used as a benchmark for personality recognition from the text (see section 2.6). Results from artificial neural network model were compared with the model from this [39] research.

It has two types of models to choose. They are

1. Observed personality from spoken language
2. Self-assessed personality from written language

First model is for identifying personality traits from natural conversation. This means that, any transcripts of natural conversations can be used. Other model is used for predicting personality traits from written text. Since candidates' answers are in written form, second type of model was chosen. It also includes four trained models for different algorithms. They are as follows.

1. Linear Regression
2. M5' Model Tree
3. M5' Regression Tree
4. Support Vector Machine with Linear Kernel

Out of four, the best result producing trained model was selected which was support vector machine with linear kernel.

4.3 Overall performance

4.3.1 Neural network model

Result produced by the neural network model is shown in Table 4.3.

Table 4.3: Shows the results for extraversion and neuroticism from neural network model

Evaluation Measures	Extraversion	Neuroticism
Mean squared error (MSE)	1.47	1.59
Root mean squared error (RMSE)	1.21	1.26
Mean absolute error (MAE)	1.00	0.98

When evaluating the extraversion model with 100 test dataset results are shown in Table 4.4.

Table 4.4: Shows the results for extraversion using 100 test set

Evaluation Measures	Extraversion
Mean squared error (MSE)	1.51
Root mean squared error (RMSE)	1.23
Mean absolute error (MAE)	0.98

Cross validation results for neuroticism model are shown in Table 4.5.

Table 4.5: Evaluation results for neuroticism when using 10-fold cross validation

Evaluation Measures	Neuroticism
Mean squared error (MSE)	1. 40
Root mean squared error (RMSE)	1. 18
Mean absolute error (MAE)	1. 00

4.3.2 Baseline method

Neural network results were compared with baseline. Baseline is the mean personality score from the training set.

Table 4.6: Table shows the baseline results for extraversion and neuroticism

Evaluation Measures	Extraversion	Neuroticism
Mean squared error (MSE)	2. 65	2. 30
Root mean squared error (RMSE)	1. 63	1. 52
Mean absolute error (MAE)	1. 53	1. 38

4.3.3 Benchmark model

Results produced by the model from this [39] study showed in Table 4.7.

Table 4.7: Shows the result for extraversion and neuroticism using model from [39]

Evaluation Measures	Extraversion	Neuroticism
Mean squared error (MSE)	2.58	3. 56
Root mean squared error (RMSE)	1.61	1. 89
Mean absolute error (MAE)	1.36	1. 72

4.4 Overall comparison

This section compares the overall performances of baseline method and benchmark model with the model discussed in this thesis. RMSE (see section 4.2) was used as the evaluation measure.

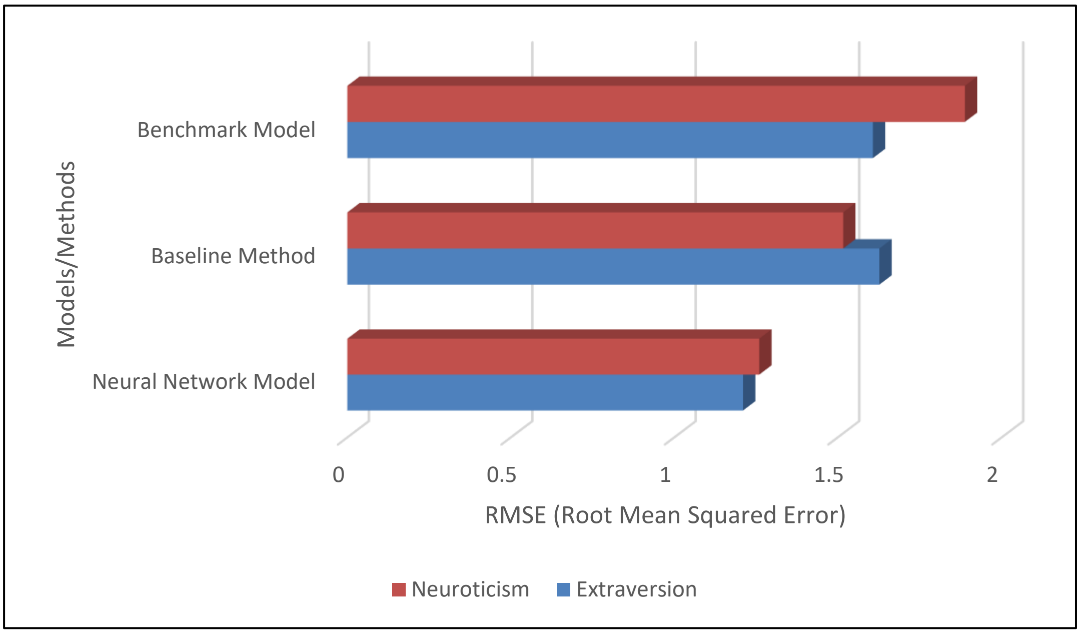

Figure 4.1: Bar chart shows the performance of the models using RMSE

As illustrated in the bar chart above, artificial neural network model discussed in this thesis outperforms baseline method as well as benchmark model.

4.5 Discussion

Firstly, baseline evaluation method was carried out. The output results for RMSE for extraversion is 1.63 and for neuroticism is 1.52. When comparing benchmark model from this [39] study produced RMSE of 1.61 and 1.89 for extraversion and neuroticism respectively. Meanwhile model discussed in this study output RMSE of 1.21 and 1.26 for extraversion and neuroticism respectively. Thus, it clearly shows that model discussed in this thesis outperforms baseline as well as existing model.

This [39] study was focused on developing a model to predict personality traits for written texts as well as spoken language. This model depends on dictionary technique, which includes 88 word categories from LIWC utility as well as 14 additional features from MRC psycholinguistic database. However, model discussed in this thesis depends only 1 word category from LIWC utility and produces good results compare to [39].

This is important to note that predicting personality traits from short text is a challenging task since short text might not well reflex a candidate's personality. However, using the techniques (see section 3.5) discussed in this thesis provides the capability to predict candidate's personality with low error rate.

Further, an output result (which is the continuous numeric value) can be converted into binary classification task or multiple classification task. For example, extraversion scores within 1.00-4.00 can be marked as introvert and extraversion scores within the range of 4.01-7.00 can be marked as extrovert. This threshold value is up to the recruiting companies to decide based on whom they would consider as extrovert and introvert.

5. CONCLUSION AND FUTURE WORK

5.1 Conclusion

As a conclusion, study on personality traits has been conducted over several decades and findings show that the personality traits of an individual have an impact on his/her job performance. Based on these studies, there should be effective mechanism to capture candidates' personality traits and compare with job related characteristics required for a particular job position. Thus, this technique helps to recruit right candidates to the right position. There are several e-recruitment services are available and they are either job seeker centric or employer centric or both. Further, there were few systems proposed by researchers, which utilize personality traits of the candidates in recruitment process.

Employer centric e-recruitment systems mainly depend on keyword search on general criteria to recruit suitable candidates. Beside that there are other functionalities provided by various e-recruitment systems such as facilitating short initial on-line interviews with candidates [23] for initial screening, evaluating candidates using psychometric tools [25] such as questionnaires. Further, candidates' thinking style, occupational interest and personality traits were predicted using short broad questionnaires [27].

Different techniques have been used to exact personality traits of the candidates. Questionnaires are fundamental standard instrument [1], [2], [47], [3], [4] used to evaluate personality traits. However, using questionnaires have significant drawbacks in this context. One of the issues is candidate might answer favorable to questions which he/she might think that associated with the required characteristics for a particular job position. Besides that, studies were focused on predict personality traits of candidates using linguistic analysis based on their posts or comments on social media [3]. This approach requires candidates to own a personal blog.

In recent decades, few studies were conducted to capture the author's personality traits in written text using different techniques (see section 2.6). These techniques can be broadly categorized into two. One being dictionary technique and other being natural language processing technique. In dictionary technique, authors used word categories from LIWC utility and features from MRC psycholinguistic database. For natural language processing technique, n-grams were extracted from annotated text and used in predicting personality traits.

This research work focuses on predicting candidates' personality traits based on scenario based analysis using machine learning techniques. Candidates will be asked to participate in scenario-based written conversation and finally their personality traits will be extracted based on it. One of the significant advantages over other systems proposed so far in the domain of e-recruitment is that candidates will be more interested to participate on conversational type of questions rather than answering large number of straightforward questions. Another major advantage of this approach is that candidates are not aware about the personality assessment thus would lead to more accurate and effective personality trait prediction. This thesis uses famous Big Five model as a general framework for identifies individual differences and studies on Big Five model primarily emerged based on lexicon research.

5.2 Future works

This research can be further extended by finding more valuable predictable features to improve the accuracy. This research focused only on extraversion and neuroticism in Big Five personality traits, which are the most influential factors for job performance. Research further can be extended to extract other three personality traits, which are agreeableness, conscientiousness and openness to experience. The challenge here is that to find best predictable feature sets for other three personality traits.

In this thesis, particular domain related questions were asked and personality traits were extracted from the context-known short candidate answers. As a future work, this research can be further extended by extracted personality traits from general context short answers.

REFERENCES

[1] R. McCrae and P. Costa, "The NEO personality inventory manual", *Psychological Assessment Resources, Odessa*, 1985.

[2] P. Costa and R. MacCrae, "Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO FFI): Professional manual", *Psychological Assessment Resources*, 1992.

[3] L. Goldberg, "The development of markers for the Big-Five factor structure", *Psychological Assessment*, vol. 4, no. 1, pp. 26-42, 1992.

[4] R. Hogan and J. Hogan, *Hogan Personality Inventory manual*. Tulsa, OK: Hogan Assessment Systems, 1995.

[5] E. Faliagka, K. Ramantas, A. Tsakalidis and G. Tzimas, "Application of machine learning algorithms to an online recruitment system", in *Proc. of the International Conference on Internet and Web Applications and Services*, 2012.

[6] G. Hurtz and J. Donovan, "Personality and job performance: The Big Five revisited.", *Journal of Applied Psychology*, vol. 85, no. 6, pp. 869-879, 2000.

[7] S. Rothmann and E. Coetzer, "The big five personality dimensions and job performance", *SA Journal of Industrial Psychology*, vol. 29, no. 1, 2003.

[8] M. Barrick and M. Mount, "THE BIG FIVE PERSONALITY DIMENSIONS AND JOB PERFORMANCE: A META-ANALYSIS", *Personnel Psychology*, vol. 44, no. 1, pp. 1-26, 1991.

[9] M. Barrick and M. Mount, "Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance.", *Journal of Applied Psychology*, vol. 78, no. 1, pp. 111-118, 1993.

[10] A. Holm, "E-recruitment: towards an ubiquitous recruitment process and candidate relationship management", *German Journal of Human Resource Management*, vol. 26, no. 3, pp. 241-259, 2012.

[11] E. Galanaki, "The decision to recruit online: a descriptive study", *Career Development International*, vol. 7, no. 4, pp. 243-251, 2002.

[12] P. De Meo, G. Quattrone, G. Terracina and D. Ursino, "An XML-Based Multiagent System for Supporting Online Recruitment Services", *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 37, no. 4, pp. 464-480, 2007.

[13] "Monster Jobs - Job Search, Career Advice & Hiring Resources | Monster.com", *Monster.com*, n.d. [Online]. Available: http://www.monster.com. [Accessed: 23- Oct- 2016].

[14] "My Job Search - Turbo charge your career with tips and guides", *My Job Search*, n.d. [Online]. Available: https://myjobsearch.com. [Accessed: 23- Oct- 2016].

[15] "Job Finder - Jobs Matched with the Best Employers", *Jobfinder.com*, n.d. [Online]. Available: http://www.jobfinder.com. [Accessed: 23- Oct- 2016].

[16] "jobsDB - Jobs in Hong Kong, Indonesia, Philippines, Singapore, Thailand", *Jobsdb.com*, n.d. [Online]. Available: http://www.jobsdb.com. [Accessed: 23- Oct- 2016].

[17] C. Deillon, "The advantages and challenges of turning towards e-recruitment for small and medium sized businesses (SME)", Ph.D, University of Fribourg, 2014.

[18] I. Myers, M. McCaulley and R. Most, *Manual, a guide to the development and use of the Myers-Briggs type indicator*. Palo Alto, Ca.: Consulting Psychologists Press, 1985.

[19] "The Myers & Briggs Foundation", *Myersbriggs.org*, n.d. [Online]. Available: http://www.myersbriggs.org. [Accessed: 07- Nov- 2016].

[20] J. Holland, *Making vocational choices: A theory of vocational personalities and work environments*, 3rd ed. Psychological Assessment Resources, 1997.

[21] L. Goldberg, "The structure of phenotypic personality traits.", *American Psychologist*, vol. 48, no. 1, pp. 26-34, 1993.

[22] "The Myers & Briggs Foundation - MBTI® Basics", *Myersbriggs.org*, n.d. [Online]. Available: http://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/. [Accessed: 07- Nov- 2016].

[23] G. Dafoulas, A. Nikolaou and M. Turega, "E-services in the Internet job market", in *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, 2003, p. 9.

[24] W. Gates and M. Nissen, "Designing agent-based electronic employment markets", *Electronic Commerce Research*, vol. 1, no. 3, pp. 239--263, 2001.

[25] R. Khosla, "An online multi-agent e-sales recruitment system", in *Proceedings of the IEEE/WIC International Conference on Web Intelligence*, 2003, pp. 111--117.

[26] H. Lieberman, "Letizia: An agent that assists Web browsing", *IJCAI (1)*, pp. 924--929, 1995.

[27] S. Supjarerndee, Y. Temtanapat and U. Phalavonk, "Recruitment filtering with personality-job fit model", in *Proceedings. International Conference on Information Technology: Coding and Computing*, 2002, pp. 46-51.

[28] J. Dewaele and A. Furnham, "Extraversion: The Unloved Variable in Applied Linguistic Research", *Language Learning*, vol. 49, no. 3, pp. 509-544, 1999.

[29] H. Giles and W. Robinson, *Handbook of language and social psychology*. Chichester [etc.]: J. Wiley & Sons, 1990.

[30] M. Mehl, S. Gosling and J. Pennebaker, "Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life.", *Journal of Personality and Social Psychology*, vol. 90, no. 5, pp. 862-877, 2006.

[31] J. Oberlander and A. Gill, "Language With Character: A Stratified Corpus Comparison of Individual Differences in E-Mail Communication", *Discourse Processes*, vol. 42, no. 3, pp. 239-270, 2006.

[32] J. Pennebaker and L. King, "Linguistic styles: Language use as an individual difference.", *Journal of Personality and Social Psychology*, vol. 77, no. 6, pp. 1296-1312, 1999.

[33] A. Gill and J. Oberlander, "Taking Care of the Linguistic Features of Extraversion", in *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2002.

[34] J. Oberlander and S. Nowson, "Whose thumb is it anyway? Classifying author personality from weblog text", in *Proceedings of the COLING/ACL on Main conference poster sessions*, 2006, pp. 627-634.

[35] S. Argamon, S. Dhawle, M. Koppel, and J. Pennebaker, “Lexical predictors of personality type”, 2005.

[36] F. Mairesse and M. Walker, "Words Mark the Nerds: Computational Models of Personality Recognition through Language", in *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2006.

[37] S. Eysenck, H. Eysenck and P. Barrett, "A revised version of the psychoticism scale", *Personality and Individual Differences*, vol. 6, no. 1, pp. 21-29, 1985.

[38] F. Mairesse and M. Walker, "Automatic Recognition of Personality in Conversation", in *Proceedings of the Human Language Technology Conference of the NAACL*, 2006, pp. 85–88.

[39] F. Mairesse, M. Walker, M. Mehl and R. Moore, "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text", *Journal of Artificial Intelligence Research*, vol. 30, pp. 457-500, 2007.

[40] R. Reilly and G. Chao, "VALIDITY AND FAIRNESS OF SOME ALTERNATIVE EMPLOYEE SELECTION PROCEDURES", *Personnel Psychology*, vol. 35, no. 1, pp. 1-62, 1982.

[41] N. Schmitt, R. Gooding, R. Noe and M. Kirsch, "METAANALYSES OF VALIDITY STUDIES PUBLISHED BETWEEN 1964 AND 1982 AND THE INVESTIGATION OF STUDY CHARACTERISTICS", *Personnel Psychology*, vol. 37, no. 3, pp. 407-422, 1984.

[42] O. John and S. Srivastava, "The Big Five trait taxonomy: History, measurement, and theoretical perspectives", *Handbook of personality: Theory and research*, vol. 2, pp. 102-138, 1999.

[43] S. Gosling, P. Rentfrow and W. Swann, "A very brief measure of the Big-Five personality domains", *Journal of Research in Personality*, vol. 37, no. 6, pp. 504-528, 2003.

[44] F. Mairesse and M. Walker, "PERSONAGE: Personality Generation for Dialogue", in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 2007, pp. 496–503.

[45] M. Hu and B. Liu, "Mining and summarizing customer reviews", in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168-177.

[46] Deeplearning4j.org. n.d. What Is DeepLearning4j? - Deeplearning4j: Open-source, Distributed Deep Learning for the JVM. [online] Available at: https://deeplearning4j.org/overview [Accessed 08 Nov. 2017].

[47] R. McCrae, P. Costa, Jr. and T. Martin, "The NEO–PI–3: A More Readable Revised NEO Personality Inventory", *Journal of Personality Assessment*, vol. 84, no. 3, pp. 261-270, 2005.