

Recruitment System Using Psychometric Test and Text Mining

Rajavarman¹, Kalpana²

1,2 Assistant Professor

1,2 K.Ramakrishnan College of Technology, Trichy.

1.rajavarman.rr.r@gmail.com 2. kalpanabtech.raja@gmail.com

Abstract— Job recruitment has become a time-consuming process for the HR department in the present working scenario. The results of the existing recruitment systems are mostly unsatisfactory. As most of the candidates who were intended to move out of the company. Every wrong people the Hiring Resource department chooses costs the company heavily. We present a system that shortlists the eligible candidates in the early stage of the interview with the help of the Psychometric test. Text Mining techniques performed on the answers of the test which tokenizes the entire text, remove the stop words and then compares with the possible keywords and then provides a shortlist based on the scoring mechanism.

I. INTRODUCTION

Recruitment is one of the key processes for a company. A company seeks a smart candidate who can upgrade and advance the company to its competitors. Human Resource management looks for the finest candidates who are fit for the company. Human makes an error, a face-to-face interview is not enough to judge about a candidate. Sometimes it produces non-welcoming results which cost the company heavily. This model proposes an idea which involves psychometric test in the first place. The psychometric test pressurizes the candidates and evaluates the mental characteristics of them to ensure whether they are capable of the job. The Psychometric test contains ability and behavioural test. The ability test evaluates the logical thinking capacity of the candidate and the behavioural test is to ensure the mental characteristics of the candidate. The test has a time constraint on which a candidate must have to finish the test which gives the candidates much more stress. The test consists of text-based answers (short answers). The system needs to equip with text mining techniques using tools from python which tokenizes the entire text, remove the stop words, compares with the possible keywords and shortlists the candidates based on the scoring mechanisms.

II. LITERATURE SURVEY

[1] The recruitment system for candidates is a corporate management system. It easily helps employees recruit. HR is an important role in recruiting. It helps the manager appeal to the organization and selects the best candidates. Recruitment is a job connecting employers to job seekers. Thus, recruitment is a process where qualified applicants can find 19 jobs and attract them. The process starts with the search for new recruits and ends with application sub-missions. It results in a group of apps selecting new employees. The success of the recruitment program is largely determined internally by the company's participation in the human resources planning or not. Future employees may not attract enough staff with the required skills overnight. The study of alternative recruitment sources and more productive methods for recruitment require time. Recruitment plans can be developed once the best alternatives are identified. AngularJS has been developed by Google, initially published in 2010, as an open source framework that gives web pages and web apps a dynamically active character. AngularJS is programmed primarily in JavaScript. AngularJS uses HTML to expand its markup language syntax. AngularJS and HTML are more efficient and dynamic in developing web applications. The unit code of AngularJS can be tested. It uses injection of dependence and separates concerns. The reusable components are provided. Developers can

use AngularJS for more short code features. It was created by Google and first published in 2010. AngularJS programmed primarily in JavaScript. AngularJS uses HTML in addition to the bookmarking language. AngularJS and HTML are used more efficiently and dynamically in the development of web applications. You can check the AngularJS unit code. It uses injection of dependency and splits concerns. There are reusable components provided. Use AngularJS together for more short code functionality.

[2] Text mining is the process by which high volumes of unstructured text data are analyzed and explored. You can identify keywords, models, subjects and other attributes in the data concepts. Text mining is also referred to as text analysis.

One of the main causes of a business sector failure is the lack of a proper or inadequate risk analysis. Adopting a data-driven software driven by technology, like SAS Text Miner, to manage potential risks by keeping companies up-to-date with all current corporate market trends. Since text mining technology can collect relevant information from thousands of text data sources and link the collected information, it enables companies to access the right information at the right time, enhancing the entire risk management process.

Text mining technology, especially NLP, is becoming increasingly important in the area of customer care. Firms invest in text analysis software to increase their client experience by accessing texts from a number of sources, including surveys, customer feedback and customer calls, etc. The objective of the text analysis is to reduce the response time for companies and to assist clients in responding to their complaints quickly and efficiently. Business and organisations, as part of their company intelligence, have started using text mining. Text mining techniques also help companies analyze the strengths and weaknesses of their competitors and provide them with a competitive edge. Many text mining software packages are solely intended for analysis of the performance of social media platforms. This enables you to read and track news, blogs, e-mails, and so on on online texts. In order to enable people who interact with your brand and online content to understand its reactions, the text mining tools also analyze the numbers of posts, preferences and followers of your brand on social media. Phenotyping definition means a description of the criteria used in observational studies to detect a phenotype to promote the knowledge of a population's disease or adverse event. Different studies are very diverse and contradictory in the nature of "phenotypic definitions." In published studies and reports, there is no internationally accepted standard which supports "phenotyping definitions." The problem is when research studies are carried out or results are compared across studies. The method of text mining can identify various literature phenotype definitions. There are evidences that there are repeatable patterns in definitions of phenotyping. Records of Health (EHR). Text mining biomedical supports the discovery of biomedical knowledge. Therefore, we believe that literature definitions of mining phenotyping can support clinical research based on EHR. However,

the literature contains inconsistent, varied and unknown information on these definitions, especially when used in text mining. We look at "phenot models," therefore. Electronic recruitment is a term for every electronically based recruitment and management activity. Most HCM software providers do not use this term to describe their products. Generally, their systems are recruited with recruitment software or software. Sellers can stop recruitment on good reason. The name of e-recruitment is e-commerce. Online retail sales on the brick and mortar markets are clearly defined by e-commerce. However, this does not apply to advertising and hiring as almost every hiring is carried out online today. It does not cover any specific recruitment aspects and can be used with online recruitment and recruitment interchangeably. The wide and full concept This could include recruitment management systems, social recruitment, application tracking systems, recruitment marketing, applicant administration, talent acquisition and talent management software. The Open Online massive class is extended by MOOCs. It's an online course for unlimited involvement and open Internet access. MOOCs are used in our system to create a motivated talent system, mainly integrating recruitment processes. The proposed system actually allows firms to access a wide variety of social and geographical profiles. In contrast to previous ones, our system offers many businesses an opportunity to acquire MOOC data to get full visibility of the best candidates or to develop their own MOOC platforms. Recruiters can, indeed, find all the data saved on the MOOC platforms throughout

the period of recruitment via our system. These details concern the assessment test, the interest of applicants, the level of participation in training and other personal information such as their age, residency, etc. These data are not only for applicants with certificates. This section shows frameworks ordering the web pages into major classes. At first, two types of site page arrangements are characterized: object-based order and type-based order. This section shows frameworks which therefore order website pages in major classes. The first feature is two types of arrangement for the site page: subject - based and sort - based orders. These frameworks are basic Web Mining and Semantic Web instruments for the eventual destination. Some Internet site pages contain critical substances that are useful for eternity or for a long time. Once again, in a short period of time, certain pages are significant. Naturally, because of their substance, it is difficult to characterize this kind of website. This is a key task to improve the execution of motors recommending web crawlers and website sites. These site pages in this enterprise have been grouped into two AI calculations classifications. Therefore, characteristic pre-processing language and message digging systems were used for pre-processing content. The correct information was removed from the documents in this case and the AI calculations were used on the website pages. The vast majority of the focus in this project was on the content preparation stage and new procedures were shown to fill the hole, as opposed to different methodologies. The results show that the methodology proposed is useful to implement than various methods. Information mining provides the way to find concepts in major information indexes including AI, measurement and database convergence strategies. Data mining is an interdisciplinary software engineering subfield with the general aim of removing data (with smart techniques) in an information index and transforming the data into a smart structure for further use. Information mining is the "learning disclosure in databases" test venture or the KDD procedure. In addition to the harsh examination step, the report also provides information and data on executive views, pre - preparation information, model and derivative considerations, fascinating measurements of quality, intricate considerations, post - handling of the structure found, and representation. The distinction between information review and information mining is, for example, the condensation of history by breaking down the adequacy of a promotional effort, the use of explicit AI and factual models to help the information mining centers for anticipating the future and finding examples from the information.

[3] Bali is the most wonderful tourists destination. To increase the quality of the quality of the tourists industry it needs to devote more time to improve the quality of attention for the hotel which is the most integral part of the tourism. Most of the tourists chooses the best hotels based on the reviews from the online tourists websites. TripAdvisor.com is one of the world's largest tourist website. There are more than 1000 hotels seen in TripAdvisor website for Bali. As the hotels are the integral part of the tourism so the hotelers need to pay too much attention to the services and facilities they provide. With the growing development of the internet system online forum is the primary tool for a travel information. TripAdvisor is a freely accessible forum it contains about more than 25000 reviews for each of the hotels in the Bali. Considering the large amount of the reviews it needs to gain knowledge about the text review. Text Mining is used to handle a lot of unstructured data in the website and provide a knowledge opinions from the users reviews. To obtain the opinion of the reviews the system uses Text Mining techniques along with the semantic analysis technique to break down the text to obtain the opinion. It uses Recursive Neural Tensor Network for the sentiment analysis. The data were collected from the TripAdvisor.com with the help of an automated program. It uses web scraping technique to collect all the data from the online forum and analyse the data. The text Mining approach is applied to the data which contains many steps like tokenization, lemmatization and stemming. The tokenization process separates the words and the stemming process removes the words which offers less meaning than the other words. The main aspect of the system is to extract the hotel details from the reviews. The details like facilities, price and the services provided by the hotels were taken. And then sentimental analysis were taken place to extract opinion about the hotel and then categorises based on the reviews. The RNTN algorithm is a sentence based semantic analysis algorithm which

analysis each sentences and produce the output.

[4] Data Mining is a process of analytical investigation used to look for consistent, systematic or compatible relations among variables and to validate the findings by applying the detected patterns to new data subsets. Data collected from different applications and repositories required different data mining techniques, so that they provide precise results, to extract useful and new information from them. In educational data mining, the phenomenon of data about many daily transactions in academic establishments is a coming field in data mining. Educational data are collected from different sources, such as colleges, universities, schools and by keeping an eye on students and instructors' activities online. Online and offline data are available in two types. The data are collected from modern as well as traditional classrooms, which are behavioral data, information from educators, attendance of students, course information and student information. Online data are collected from geographically divided web-based educational stakeholders, online forums and networking websites. The WJ-48 algorithm is used to generate data in the form of IF-THEN statements that are a decision tree. K-means algorithm to classify data and the algorithm of linear regression into technique of predictive analyzes that find the link between independent and dependent variable. The main aim of the system is to identify the factors that affect the performance of students. The parents were informed on the basis of the questionnaire that affect the terms.

[5] The test is one of the most crucial tests to evaluate academic achievement, integrate ideas and recall skills. Tests are useful in evaluating students' learning outcomes as they offer a chance to demonstrate their skills and knowledge, including higher order thinking. While essay questions are beneficial for the learning and evaluation of students, several challenges were noticeable for the teachers. Manual grading of essays takes a considerable time for the teacher, as the gradient test takes time and is especially tedious for a large number of students. The perception of the subjectivity of the grading process can also be taken into consideration as the subjective nature of the assessment can lead to different results. Automated Essay Scoring (AES) is an instrument that allows teachers to save time and effort, make more objective assessments and avoid being subjective. Its main objective is to automatically predict student grade with different features. The proposed approach includes ontologically based information extraction, using basic algorithms for tokenisation, word tagging, counting, words and phrases, frequency distribution and semanticized text matching for automatic essay marking. In this study we aim to see whether teachers use ontology to compute students' essay degrees and to explore various natural language tools to extract the features for grading students' essays. The first phase involves the use of OntoGen for ontology generation and identification from a corpus of relevant essays in each field. The second phase consists of extracting features with NLTK tools to scoring the experiment. This will help teachers to identify conceptual fields and the similarity between the essays. It will start with the extraction of domain ontology from a text corpus. To identify keywords, OntoGen uses the SVM support algorithm. It proposes concepts based on the list of currently selected documents. In a root node it founded the concept hierarchy, and subconcepts in the ontological treasure are treated as children. Some datasets contain a source essay, which is based on a question. But the test type test is not only correct, so we compared the resemblance between the source. The amount of language or vocabulary that the students can display contributes to the essay score.

[6] Scholars or professionals who are experts in their fields write scholarly publications. Researchers involved in research often publish their findings. Through various publications such as journal articles, lectures, monographs, reviews these scholars communicate their research results. Some of the results of these research may still not be published and published in grey literature via open access platforms or through pay walls. For recovery purposes, all publications published in the medium mentioned above are indexed and abstracted in databases. Some of the most prominent databases that cover large-scale research. Only journals with high impact that communicate original research findings are highly indexed and

abstracting databases. As large quantities of research findings are conveyed in the form of scientific literature, it is essential to encode the knowledge in the findings to make best use of that knowledge. In this case, the role of text mining, which deals with text, comes into play with getting hidden information and text patterns. The highly quoted data mining articles have been identified using the Web of Science Database (WoS). Titles and abstracts from 50 Top articles were extracted from the WoS database. Since the WoS database contains only abstracts of the articles, the analysis of title and abstracts for selected articles is limited to this study. Each title and abstract is read to identify the principal and subdisciplines covered by considered articles by researchers. Several categories were assigned for the single article on the basis of the content of the article. Articles were then grouped according to categories assigned. The term-document matrix was calculated using assigned subcategories. In this study, the categories are columns and the articles are rows. The frequencies of the category (k) in question in the matrix represent the cell values (m_{kn}) for the number of articles (n). Social media and medical information systems were identified as less popular research fields relating to data mining by the results in this study. However, the above fields of research are highly public-specific. Social media is a popular tool for the general public and user based information on social media platforms can be used for the identifying of public behaviors and patterns. The results of social media research can thus be used to forecast general public behaviour.

[7] Modern recruitment process and people analysis as demonstrated by the millions of professional social network-ing users, work advertisements on employment portal web-sites and social networks. Data for peopleware are focused on recognizing required skills and professional personality the psycho - psychological characteristics in a successful team the creation of member and the allocation of work. The term peopleware therefore refers to how industry perceives the important role that human factors play in the process of development. In the process of analyzing peopleware, the distinction between hard and soft skills is clear. Skills and competencies are the key human analytics variables and can be drawn from job advertisements. The use of raw information on online job offers provides a rich source of analysis for people. Detecting the right skills and skills for a job from raw text data and associating them with a job seeker is a growing challenge. The detection and association of the skills and abilities for work from raw text data is increasingly difficult. StackOverflow is the selected professional web source for networking, and multivariate statistical data analysis has been used to test correlations between skills and competencies in the dataset of job offers. The present work falls within a new field in which researchers focus specifically on software development with the ability to extract data from human software.

III. EXISTING SYSTEM

The existing recruitment system uses psychometric test but the answers are in the form of Multiple Choice Questions which is easier for the machine to evaluate the answers.

DISADVANTAGE

- The Students can get a correct answer by lucky guess.

IV. PROPOSED SYSTEM

A. Candidate attend test

The test contains questions which are carefully selected from the databases. The questions are based on aptitude to check the logical thinking skills and the behavioural test to check the mental characteristics of the candidate. The results were used to shortlist the candidate.

Time constraints

To give more pressure the test contains a time limit for each question. These were used to check how well the candidate performs under stress.

B. Text Mining

The answers to the questions are in the form of text. The system uses text mining tools of python to analyse the text and provides the score based on the scoring mechanisms in the system.

C. Admin provider results

The goal of the system is to shortlist the candidates in the early stages of the interview, which reduces works on the HR department and they can conduct any other test if they want.

V. BENEFITS OF PROPOSED SYSTEM

- It reduces the cost and time for the hiring department of the company.
- It used to employ finest candidate which in turn helps for the growth of the organisation

VI. PSYCHOMETRIC TEST

Psychometric examinations are a standard scientific method for measuring the mental and behavioural capacities of individuals. Psychometric tests aim at measuring the suitability of candidates for a role based on the required personality and ability (or cognitive abilities). It helps to identify the personalities and cognitive skills of the candidates to what extent the skills match the functions required. The collected information from the psychometric test to identify the hidden aspects of the applicant that are hard to extract from a face to face interview. Psychometric tests are evaluated statistically and built on an objective and impartial basis. This is done by through standard methods of assessment, hence everyone who attends the test were given same questions and instruction. The psychometric tests are very reliable in predicting candidates' performance, and in most cases, the test report provides an accurate evaluation of the applicant.

VII. TYPES OF PSYCHOMETRIC TEST

A. Aptitude Test

Aptitude tests are intended to evaluate a certain or general set of skills, although this often depends on your job type. This group includes the following categories:

- **Numerical reasoning test** : Used for the purposes of identification of how data are interpreted often through a combination of written and statistical information in reports, charts and graphs. This may also be used to evaluate fundamental mathematical skills.
- **Verbal reasoning test** : Used as one way to evaluate your ability to make an informed decision by evaluating detailed written information.
- **Inductive reasoning test** : These require the identification of trends or patterns, usually using schemes. Try a free test below.
- **Diagrammatic reasoning test** : They evaluate your logical thinking capacity by using diagrams and flowcharts.
- **Logical reasoning test** : Its purpose is to evaluate your ability to conclude. You may receive some information and then be asked to decide what you received. These tests are often referred to as deductive reasoning tests.
- **Error checking test** : Assesses your ability to quickly identify any errors in complex data sets such as codes, or combinations of alpha-numeric characters.

B. Personality Test

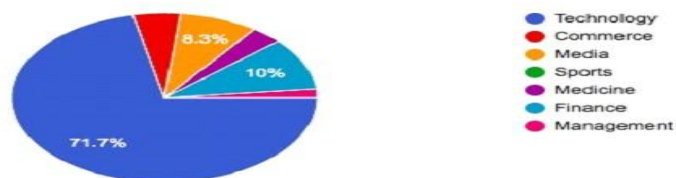
Personality assessments allow employers to assess your fitness based on the way you work and your behavior. This is used to determine how well you fit into the organization and company culture. Your answers are often linked to the ones of a high-performing employee or successful manager, which indicate if you have the same features. During the recruitment process, employers look at several factors from your knowledge and experience to your skills in such fields as decision making and teamwork. They are increasingly using personality tests to determine whether you have

therightattitudeandpersonalitytofitwiththecultureandvisionofyourcompany. OneofthemostwidelyusedpersonalitytestsisMyersBriggs.Afteryouhavemadesomeprogressabouthowyoufeeloractinacertainscenario,youareassignedoneofsevententypesofpersonalitiesbasedonyouranswers.Employersthenexaminethistypeofpersontoseeifyourcompanyvalueswouldfit.Therisseldomatimelimitbecauseitlookslke.

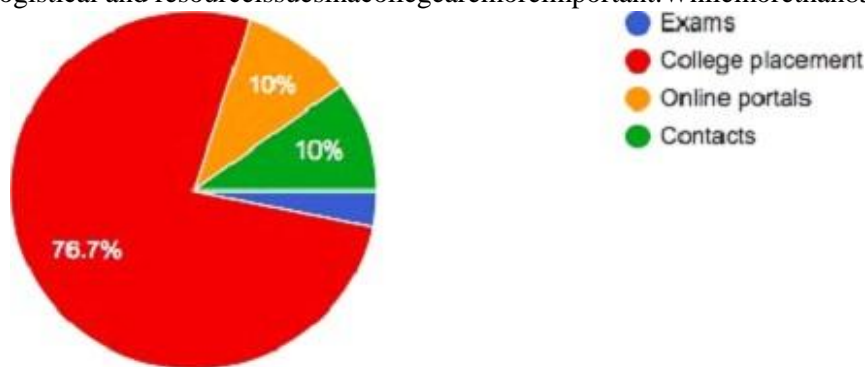
VIII. SURVEY

Theentirecomplexityoftheprevalingsystems,aspotentialemployees,struckusasaneextremeneedforanautomatedrevolution.Withsomepreliminaryteststhatwillbeusedtounderstandthecandidatemindset,theproposedmodelstreamlinestherecruitmentprocess.

Theaudienceconsistedmainlyofa20-35yearoldagegroup,whiconsciouslytriedtohavenewdatafromayoungaudience.Thedivisionoftheaudienceisshowninthefigurebelow.

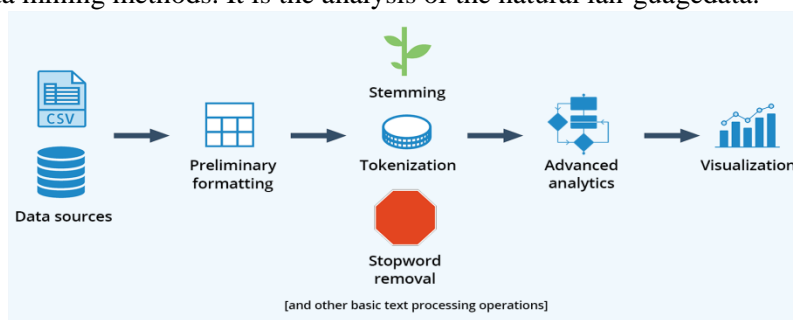


grouparethroughuniversityplacements.Theneedforamechanismliketheoneproposedisfurtherstressed,aslogisticalandresourceissuesinacollegearemoreimportant.Whilemorethan83



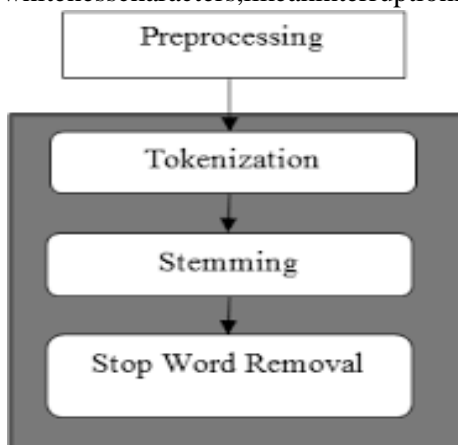
IX. TEXTMININGALGORITHM

In text mining the unstructured documents that come from different sources extract interesting information, knowledge or design. It transforms words and phrases into numerical values in unstructured information that can be linked to database structured information and analyzed using older data mining methods. It is the analysis of the natural language data.



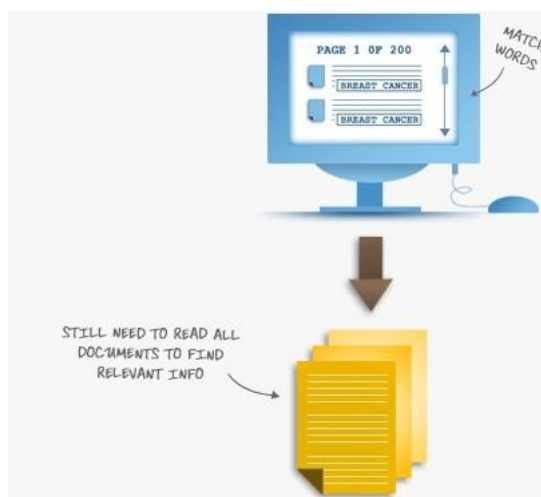
Different pre-processing operations are shown. Stemming is the process by which modulated words are reduced to their word base, root. Stop word is a word that is filtered out before or after the text has been processed. Tokenisation is the process of breaking a stream of textual content into words, terms, symbols or some other meaningful elements called token. In general, the tokenization process takes place on the word level. But sometimes it is difficult to define what a 'word' means. In simple heuristics, for example, a tokenizer regularly undertakes:

- The resulting list of tokens may include pointing and whitespace.
- All adjoining character strings are part of one token with numbers in the same way.
- The tokens are separated by whiteness characters, like an interruption in the space, line or punctuation.

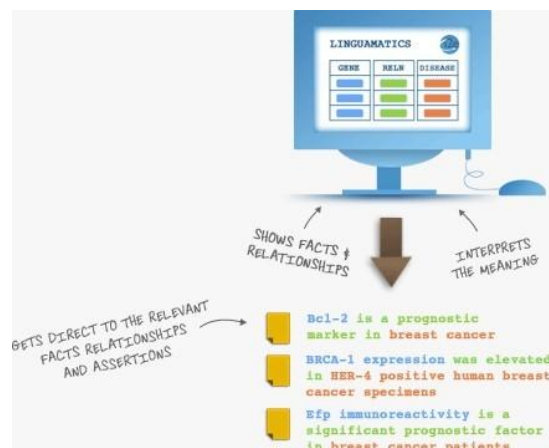


X. TEXT MINING VERSUS TRADITIONAL KEYWORD SEARCH

Traditional keyword search returns all the documents containing the keywords you specified. It's great, but still you have to read all these documents, in order to find out if they contain information relevant to your search. This information is relevant.



It can understand real meaning through sophisticated natural language processing (NLP) algorithms, which allow it to recognize similar concepts though they are expressed in very different manners and spellings and it can also read and analyze documents on your behalf. The text mining software can understand real meanings. A text mining search will identify facts, relationships and assertions which would be buried in an amount of free text or unstructured data otherwise.



XI. WORDTOKENIZATIONWITHPYTHONNLTK

A number of tokenizers are provided in the module byNLTK. ThetextshallbetokenizedfirstwiththePunkt-SentenceTokenizer into sentences. The following words areplayedwithfourtokenizersforeachsentence:

- TreebankWordTokenizer - This tokenizer uses regulartokenizingexpressionsliketheoneinTreebank.
- WordPunctTokenizer - This tokenizer divides a stringintosubstringsbydividingitinsubclassesonthespecificstring.
- Punct Word Tokenizer - This tokenizer divides the textinto a list of sentences ; uncontrolled algorithms areused.
- Whitespace Tokenizer - This tokenizer splits text intospace.

XII. METHODOLOGY

STEP1: SURVEY PHASE Inthisphasesurveyisconductedtoanalysethepresentdayscenarioandtherequirements needed for the job. Based on the survey thequestionsneededforthePsychometrictestweretaken.

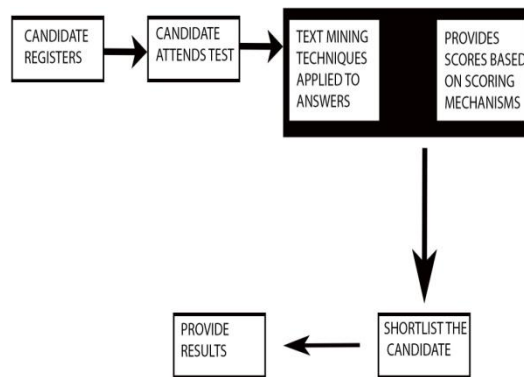
STEP2: BACKEND PROCESS Thequestiontakenwereloadedintothedatabaseandtheappropriateanswersforthemalsoloaded.

STEP3: TEST ENVIRONMENT The test environmentwas created which differs based on the requirement of everyjobs.

STEP 4: TEXT MINING The text mining algorithms aredeployed in the system. Which tokenizes the text, removesthestopwordandanalysethedatawiththepossiblekeywordsthatareloadedinthedatabase. Thescoringmechanismswereusedtoproducescoreforeveryquestions.

STEP 5: DISPLAY RESULTS The candiate shortlistedinformationisdisplayed.

XIII. ARCHITECTURALDESIGN



XIV. CONCLUSION

By this proposed model we can recruit the finest candidate from the market. This model saves time as well as the cost for the company. It increases the range of hiring the quality candidate.

REFERENCES

1. GauriKejkar, Amreen Khan, Richa Sharma, An enhancement for candidate recruitment system using Angularjs. 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1-4, 2017.
2. Anton Kanev, Stuart Cunningham, Terekhov Valery, Application of formal grammar in text mining and construction of an ontology. 2017 Internet Technologies and Applications (ITA), Pp. 53-37, 2017.
3. Samar Binkheder, Heng-Yi Wu, Sara Quinney, Lang Li, Analyzing Patterns of Literature-Based Phenotyping Definitions for Text Mining Applications. 2018 IEEE International Conference on Healthcare Informatics (ICHI), pp. 374-376, 2018.
4. Haddadi Lynda, Tali Mohammed El Amine, Bouarab-Dahmani Farida, BerkaneTassadit, E-recruitment support system based on MOOCs. 2017 International Conference on Mathematics and Information Technology (ICMIT), pp. 234-238, 2017.
5. SeyedMoeinBabapour, MeysamRoostae, Web pages classification: An effective approach based on text mining techniques. 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI), pp. 0320-0323, 2017.
6. PuteriPrameswari, IstiSurjandari, Enrico Laoh, Opinion mining from online reviews in Bali tourist area. 2017 3rd International Conference on Science in Information Technology (ICSITech), pp. 226-230, 2017.
7. Sheena Angra, Sachin Ahuja, Implementation of data mining algorithms on student's data using rapid miner. 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), pp. 387-391, 2017.
8. Jennifer O. Contreras, ShadiHilles, ZainabBintiAbubakar, Automated Essay Scoring with Ontology based on Text Mining and NLTK tools. 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), pp. 1-6, 2018.
9. P. K. Jayasekara, Abu K. S., Text Mining of Highly Cited Publications in Data Mining . 2018 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ET-TLIS), pp. 128-130, 2018.
10. Maria Papoutsoglou. Nikolaos Mittas, Lefteris Angelis, Mining People Analytics from StackOverflow Job Advertisements. 2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 108-115, 2017.