

CAREER FRONT

ENROLLMENT NUMBER - 14503002, 14503009, 14503031

NAME OF STUDENTS - UMANG NYATI, SNEHA RAWAT, DEVIKA

NAME OF SUPERVISOR - MS. ADITI SHARMA



MAY 2018

MID SEMESTER REPORT

Submitted in partial fulfillment of the Degree of

5 Year Dual Degree Programme B. Tech

in

Computer Science Engineering

**DEPARTMENT OF COMPUTER SCIENCE ENGINEERING &
INFORMATION TECHNOLOGY**

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY (NOIDA)

TABLE OF CONTENTS

Chapter No.	Topics
	Acknowledgement
	Certificate
Chapter – 1	Introduction
	1.1 General Introduction
	1.2 Sources
	1.3 Summary of Relevant Papers
Chapter – 2	Results of Literature Survey
	2.1 Integrated Summary of the literature studied
	2.2 Problem Statement
	2.3 Solution Approach
Chapter – 3	Implementation and Testing
	3.1 Overall Description
	3.2 Implementation details and issues
	3.2.1 Description of Modules with respect to design
	3.3 Testing
	3.3.1 Testing Plan
	3.3.2 Component decomposition and type of testing required
	3.3.3 List of test cases
	3.3.4 Error and Exception Handling
	3.3.5 Limitations of the solutions
	3.4 Risk Analysis and Mitigation
References	

CHAPTER 1

INTRODUCTION

1.1 General Introduction

Resume screening is still the most time consuming part of recruiting. It is estimated that it may take up to 23 hours for just one hire. Career Front is a tool for resume screening and helping candidates by simplifying the process of creating a resume and increases their chances for job selection. Moreover, we aim to provide important information to candidates by predicting jobs for them considering their resume, what are the skills needs by the company when they post a job opening. It is also important to determine that which jobs are best match for the candidates according to their skills, which is time consuming for both candidates and company, therefore there is need to automate the process.

Moreover, with the increase in need of spreading word about the company, it is important for the company to provide relevant information at multiple platforms. We provide an in-built text summarizer that helps companies to summarize their detailed information in small amount of text.

Our tenets includes –

- a. Innovate – Dealing with science and bring modernized technology
- b. Deliver – Trying our best to deliver solutions to the problem with precision
- c. Impact – We will transform the process of hiring and transform the careers of many candidates

With the aim of simplifying the process, the tool employs methods such as skill matching, job prediction, skills required, text summarization, resume scorer, resume parser, resume builder, and personality detector. This also includes a novel approach used, which consists of a combination of a grammar checker and text simplifier which helps candidate in building resume that have higher chance of getting shortlisted.

1.2 Sources

The literature survey done helped in formulating and refining the problem statement. It was also helpful in identifying the key challenges that need to be faced for designing a practical approach that needs to be followed, and the priorities with which the tasks are required to be

done. The papers from renowned sources like IEEE and Springer were studied and referenced for the purposes of understanding the complexity of the problem being dealt with and coming up with an efficient solution.

1.3 Summary of Relevant papers

Table 1:

Title	RésuméMatcher: A personalized résumé-job matching system
Authors	Shiqiang Guo, Folami Alamudun, Tracy Hammond
Year of Publication	2016
Publication details	Expert Systems with Applications. 60. 10.1016/j.eswa.2016.04.013
Summary	<p>The paper describes a novel statistical similarity index. It determines the appropriateness of a job by calculating the similarity between the candidate model and the job model, generated from the resume and the job description respectively. This method changes the fundamental nature of job search from keyword-based search to model matching. Since search results are sorted by order of similarity score, the algorithm not only finds the most appropriate jobs, but also provides a ranking based on the similarity score. A job is determined to be relevant based on a novel similarity metric, which compares features from a resume object with features from job objects in the database. A list of jobs is returned to the user in sorted order of similarity to the user's resume.</p> <p>The paper also describes the finite-state machines serving to improve time and space efficiency. A finite state transducer (FST) is a finite state machine which, given an input string, is able to generate a unique output string. A lightweight FST framework which was developed is also described, which is capable of processing regular expression matching over labeled tokens.</p>
Web Link	https://www.sciencedirect.com/science/article/pii/S0957417416301798

Table 2:

Title	Quantifying Skill Relevance to Job Titles
Authors	Wenjun Zhou, Yun Zhu, Faizan Javed†, Mahmudur Rahman, Janani Balaji and Matt McNair
Year of Publication	2016
Publication details	IEEE International Conference on Big Data (Big Data), Washington, DC, 2016, pp. 1532-1541. doi: 10.1109/BigData.2016.7840761
Summary	<p>The objective of this investigation is to build up a systematic approach to access the relevance of different skill terms for each job title. Main use of this study is career profiling which includes identification of top skills for a given job title. First, starting with the identification of core skills which are essential, then suggestion and identification of standout skills to job seekers which are likely to help stand out from others. Focus should be on identifying skills, which are in high demand, but low in supply are generally very valuable. Quantifying skill relevance using the following methodology: TF-IDF based, global uniqueness and local uniqueness. TF-IDF based scores, aggregating skill counts under each job title and computation of skill-title pair. Quantifying Skill Dispersion using the three variants of dispersion measures: entropy-based, the dispersion purity based and the variation-based. Skills are labeled into one of four levels: 1. Essential 2. Important 3. OK 4. Wrong skills. Ranking list of relevant skills for a title using following evaluation metrics: Precision, Mean average precision and Normalized discounted cumulative gain.</p>

Web Link	http://ieeexplore.ieee.org/document/7840761/
-----------------	---

Table 3:

Title	Text Summarization using Sentence Scoring Method
Authors	T. Sri Rama Raju, Bhargav Allarpu
Year of Publication	2017
Publication details	International Research Journal of Engineering and Technology (IRJET), 2017, Vol. 04, Issue. 04.
Summary	<p>In this paper, an automated text summarization tool has been developed using Sentence Scoring Method, which involves finding the frequent terms, sentence ranking etc. Summary is extracted from the list of top ranked sentences. In Extractive summarization, the important sentence is picked from the text as a summary. The paper describes the implementation of document summarization using extractive method. This method consists of four phases they are: Pre-processing, Sentence scoring, Sentence ranking, Summary Extraction. The phase of pre-processing involves chopping the paragraph into words. After this phase, sentences are ranked based on features like frequency, sentence position, cie words, sentence length, etc. After each sentence is scored they are arranged in descending order of their score value i.e. the sentence whose score value is highest is in top position and the sentence whose score value is lowest is in bottom position. After ranking the sentences based on their total score the summary is produced selecting certain number of top ranked sentences where the number of sentences required is provided by the user.</p>

Web Link	https://irjet.net/archives/V4/i5/IRJET-V4I5493.pdf
-----------------	---

Table 4:

Title	Automatic Text Summarization of Wikipedia Articles
Authors	Dharmendra Hingu , Deep Shah, Sandeep S. Udmale
Year of Publication	2015
Publication details	International Conference on Communication, Information & Computing Technology (ICCICT), Mumbai, India. Jan. 1617, 2015
Summary	The main objective of this paper is that Wikipedia articles are given as input to system and extractive text summarization is presented by identifying text features and scoring the sentences accordingly. The text Sentences and perform stemming operations. We then score the sentences using the different text features. Two novel approaches implemented are using the citations present in the text and identifying synonyms. These features along with the traditional methods are used to score the sentences. The scores are used to classify the sentence to be in the summary text or not with the help of a neural network. The user can provide what percentage of the original text should be in the summary. It is found that scoring the sentences based on citations gives the best results.
Web Link	http://ieeexplore.ieee.org/document/7045732/

Table 5:

Title	A Survey on Abstractive Text Summarization
Authors	N. Moratanch, Dr. S. Chitrakala
Year of Publication	2016
Publication details	2016 International Conference on Circuit, Power and Computing Technologies [ICCPCT].
Summary	In this paper, an exhaustive survey on abstractive text summarization methods has been presented. The two broad abstractive summarization methods are structured based approach and semantic based approach. Structured based approach are basically categories is six sub categories, i.e Tree based method, graph based method and Semantic based methods are divided into 4 subcategories. In this survey all methods are proposed only, not implemented. This paper collectively summarizes and deciphers the various methodologies, challenges and issues of abstractive summarization. This survey has showcased various methods of abstractive summarization. Abstractive summarization methods produce highly cohesive.
Web Link	http://ieeexplore.ieee.org/document/7530193/

Table 6:

Title	A survey of job recommender systems
Authors	Shaha T. Al-Otaibi1 and Mourad Ykhlef
Year of Publication	2012
Publication details	International Journal of the Physical Sciences Vol. 7(29), pp. 5127-5142
Summary	Following are the various recommender system

	<p>techniques described in this paper:</p> <ol style="list-style-type: none"> 1. Collaborative filtering approach: It is further classified into two types: Memory-based and Model-based method. In memory-based method, user-item database is used where each user is a part of a group of users with similar interests and preferences of new items for user can be predicted through identifying neighbours. While, model-based model is produced from the historical rating and used to deduce predictions. 2. Content based filtering approach: Two main tasks related to CBF approach are: User profiling and the Item representation. Recommendation of items whose content is similar to the content that user has previously reviewed. 3. Knowledge based approach: This approach assists users in the determination of suitable solutions from complex product and service assortments. 4. Hybrid approach: Collaborative filtering is integrated with other techniques in an attempt to avoid the previous mentioned challenges which are classified as follows: Weighted hybrid recommender, Switching hybrid recommender, Mixed, Feature Combination, Cascade, Feature augmentation, Model. 5. A probabilistic hybrid approach. The
--	---

	<p>recommendation approach used both concepts: content-based filtering and collaborative filtering simultaneously. This assists partially to overcome the problem of data sparsity.</p>
Web Link	<p>http://www.academicjournals.org/IJPS</p>

CHAPTER 2

RESULTS OF LITERATURE SURVEY

2.1 Integrated Summary of Literature Review

Table 7

S.No.	Research Study	Objective & Analysis
1	Shiqiang Guo, Folami Alamudun, Tracy Hammond, 2016	Description of novel statistical similarity index and finite state machine to improve time and space efficiency
2	T. Sri Rama Raju, Bhargav Allarpu,	Objective to build a systematic approach to access the relevance of different skill terms for each job title
3	Laylavi, Farhad et. al; 2016	Paper describes an automatic tool text summarization tool with the help of sentence scoring method
4	Dharmendra Hingu , Deep Shah, Sandeep S. Udmale, 2016	The paper describes the summarization of Wikipedia articles with the help of citations present in the text and identifying synonyms
5	N. Moratanch, Dr. S. Chitrakala, 2016	Two methods of abstractive summarization are described that includes structured based approach and semantic based approach
6	Shaha T. Al-Otaibi1 and Mourad Ykhlef, 2012	Paper describes various recommender system techniques.
7	Alfirna Rizqi Lahitani, 2016	Paper describes the implementation of cosine

		similarity and TF-IDF
8	Linna Li, 2010	Paper describes the implementation of decision tree algorithm in the data mining model
9	Mashael S. Aldayel, 2012	The paper describes the application on KNN
10	Jiangtao Ren, 2009	The paper describes the naive Bayes model that could be applied on uncertain data

2.2 Problem Statement

The process of looking for a new job is itself a full time job. There is an overwhelming amount of job postings and many a times it is difficult for the candidate to determine which job best fit according to the skills he/she posses. A thoroughly background research has to be done about the job posting which is time consuming. Moreover, a particular job postings receives overwhelming amount of application to eliminate the candidates that are not good match for the job, therefore there is need to find an automated skill match between the resume of candidate and job description of the company. It is also important that a candidate should know what job they should be looking for, applying in a wrong industry is not suitable for both the candidate and the company. Apart from what job to apply, it is also important where to look for a job and what are skills much needed to find a job. Therefore, a system is needed that can match the skills of the candidate, predict a job that is relevant for them and recommend them the skills which are high in demand so that a candidate can secure a right job.

Today, companies need to spread a word about them therefore, it is important that relevant information is posted at various platforms, but re-writing the content every time consumes time and resources, so there is need of an automated text summarizer that can summarize the detailed description into smaller clusters.

2.3 Solution Approach

With the aim to help the candidate to find the right job in hassle free and less time consuming manner we recommend following approaches –

- a. Finding the similarity in the skills of candidates extracted from his/her resume with the job posting using cosine similarity, which will help both the candidate and company to find the best match by an automated process.
- b. Job Prediction, matching the resume with the job summary of the company by using various machine-learning algorithms: KNN, Decision Tree and Gaussian Naives Bayes Algorithm
- c. To make sure, candidates are well updated by the current skills trends needed in the industry, therefore, key words are extracted from job postings made in cities like Bangalore, Chennai, Delhi and percentage was calculated by counting the frequency of each skill divided by the total jobs found in the city to calculate percentage of skills required.
- d. Text Summarizer, with the increased number of companies, it has become important that company should spread a word and post job postings on every necessary platforms. The company should re-write their detailed description for every platform, therefore an automated text summarizer so that both time and resource can be utilized in more efficient way.

CHAPTER 3

IMPLEMENTATION AND TESTING

3.1 Overall Description

3.1.1 Introduction

Fetching a right job has become a major concern for the candidates. Finding a job that matches with the requirement of the company is very important. Candidates have to ensure that they are well updated with the skills required by the companies in the present time. Moreover, the human resource department dedicates a lot of time and resources in finding the right candidate for the company; therefore, there is a need of an automated process that helps both the company and the candidates. Career Front aims to help candidates by recommending them the skills set that are currently required by maximum number of companies. It also helps candidates predicting the job for the candidate by matching their skill set and job description of the company. It benefits both the company

and candidates by matching their skill set by the job description. With the increased competition in the industries, it is also important for the organizations to ensure that they have mentioned their details on multiple numbers of platforms. By becoming their helping hand we provide an in-built text summarizer that a company can use to extract important data from detailed information about the company and put them on various platforms.

3.1.2 Purpose

Right person for right job is the main purpose and main concern is laid on the following:

1. Identify the important skills required for a particular job by extracting job description of the company.
2. Predict the job by analyzing and comparing the job description and cv by various machine learning algorithms to predict the accurate results.
3. To provide best suggestions to the candidate as well as recruiter, use similarity measure to remove the bad match for the postings.
4. Provide the candidate with a brief and useful description of the company to have a general knowledge about the company in which they are applying.

3.1.3 Scope

1. Fetching a right job has itself became a individual task
2. Companies want to recruit the best fit
3. Companies want to minimize the time and resource consumed for recruiting a perfect candidate
4. Very less work done on making the recruitment process automated

3.1.4 Requirement Specification

Software Requirements

- i. CPU: 500MHz processor
- ii. Computer Processor: intel core i5
- iii. Computer Memory: 500gb or more
- iv. Graphics Hardware: Not required
- v. Good Internet connection for scrapping data

3.1.5 Functional and Non-functional Requirements

Functional Requirements:

- i. Upload resume with accurate job skills
- ii. Briefly provide qualification details in your resume
- iii. Enter job description clearly specifying the requirements
- iv. Description or About me blog of the company
- v. Codes must be syntactically free from errors
- vi. All essential libraries must be installed/ imported.
- vii. Short listing the best candidate available.

Non-Functional Requirements:

- i. Scalability- Our system should be able to work efficiently with larger dataset.
- ii. Performance Requirements- System should fetch results with minimum complexity.
- iii. Fault tolerance- It refers that operations can continue even in the presence of faults or missing data.
- iv. Correctness- It needs to be reliable and give correct and expected output.
- v. Openness- It refers to the ability of system to be re-implemented and get extended in various ways
- vi. Security- Our system should be protected from unauthorized uses. It shall not cause any harm to either the user or the system.

3.1.6 Product Perspective

The tool is independent software. It matches the skills of the candidates according to the job description given by any company. It works independent of any other job-matching website or resume screening website. The in-built text summarizer does not use any websites for extracting text, it takes into account the detailed description given by the company and presents it into a summary form.

a. System interface

Resume Entry
System

Job description
System

Company
Description System

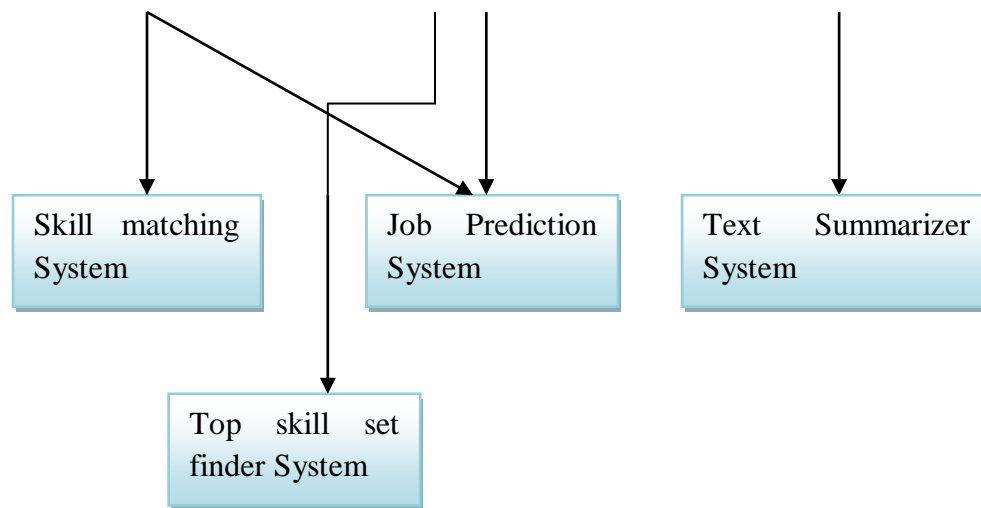


Figure 1

b. User Interface

The system comprises of two types of user namely, the candidate and member of human resource department. Candidate can use the system to avail job prediction and skill finder facility. Both the candidate and member of human resource department can use the skill matching system. The human resource department member would use text summarizer.

c. Hardware Interfaces

Career Front being a software does not consists of any additional hardware interfaces apart from a system which would is required to run the software.

d. Software Interface

Database is required to store the job descriptions of the company so as to match the resume of the candidate. Moreover, the database will be required to store the company detailed information.

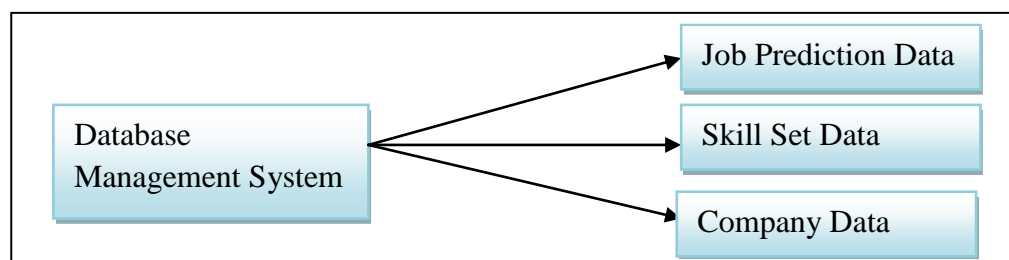


Figure 2

e. Operations

- Resume upload by the candidate
- Job description by the company
- Detailed Description upload by the company
- Data extraction from indeed.com
- Text summarization

f. Site adaption requirements

Data is required from indeed.com for visualizing that what skill sets are currently important and in demand by the companies.

3.2 Implementation Details and Issues

3.2.1 Description of Modules with respect to design:

With increasing demand of good employees, there is need to build an automated system to help the human resource department to find the best fit for their company. We provide a novel approach for doing so. Below mentioned are the modules that are currently integrated into the system –

- a. Skill finder – With the help of descriptions for various types of jobs posted on indeed.co.in, the system finds the current skill set trends that are required in the companies and thus help candidate to be updated and enhance their skills and grab a job. Key words are extracted from job postings made in cities like Bangalore, Chennai, Delhi and percentage was calculated by counting the frequency of each skill divided by the total jobs found in the city to calculate percentage of skills required.
- b. Skill Match – To ensure that candidates can determine that which jobs best matched the skills and preferences and ensuring that the companies get the best-fit candidate. the system aims to provide skill matcher which takes into account the resume of the candidate and match it with the keywords extracted from the job description helping

- to decide that if the job is suitable for the candidate or not. Cosine similarity is used to calculate the similarity between the skills and job description.
- c. Job Prediction – Most of the candidate does not know what type of job to look for, which are the jobs that they should apply for, thus to help the candidate, the system takes into account the resume of the candidate and matches it with the job summary of the company by using various machine learning algorithms like KNN and Decision tree to predicts that what are the best jobs suitable for the candidate
 - d. Text Summarizer – With the increased number of companies in the market, it has become necessary that a word about the company is spread on various platforms. Re-writing content for every platform is both resource and time-consuming process, they we provide and in-built text summarizer so that company can summarize the detailed description and post it in brief on several platforms.

The above system is novel and is independent of any other platform available in the market. There are currently no other systems that helps both the companies and candidates for fetching the job by the above mentioned process.

Algorithms Used:

1. Term Frequency-Inverse Document Frequency

The TF*IDF algorithm is used to weigh a keyword in any content and assign the importance to that keyword based on the number of times it appears in the document. More importantly, it checks how relevant the keyword is throughout the web, which is referred to as corpus.

For a term t in a document d , the weight $W(t,d)$ of term t in document d is given by:

$$W(t,d) = TF(t,d) \log (N/DF(t))$$

Where:

$TF(t,d)$ is the number of occurrences of t in document d .

$DF(t)$ is the number of documents containing the term t .

N is the total number of documents in the corpus.

2. Naive Bayes Classifier

Naive Bayes is a simple technique for constructing classifiers, models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable.

3. Cosine Similarity

Cosine similarity is a measure of the similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them. The cosine angle is the measure of overlap between the sentences in terms of their content.

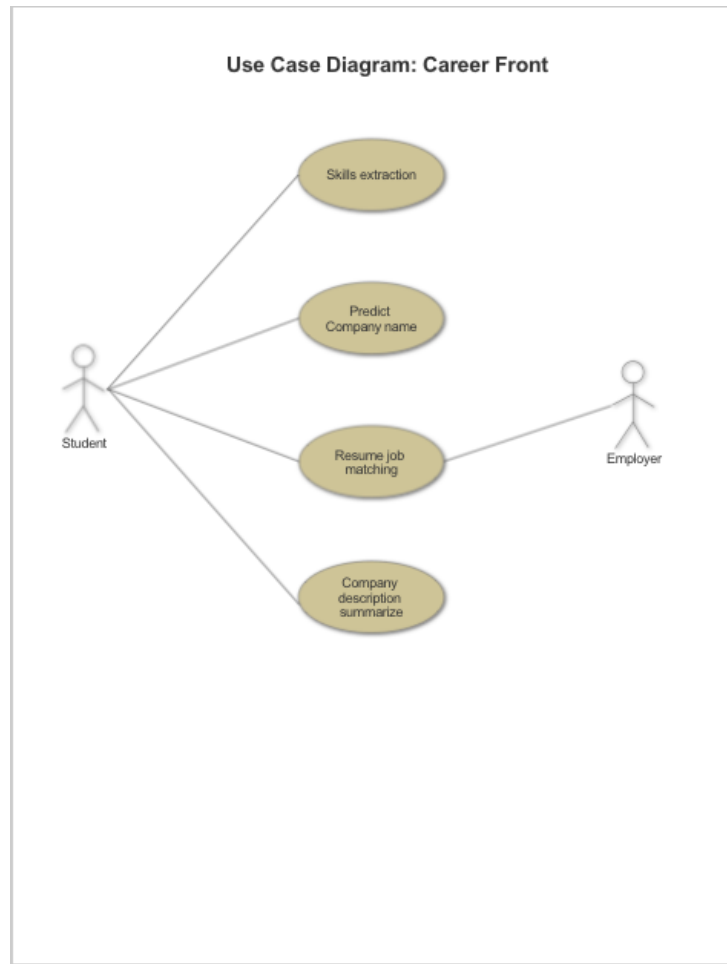
4. KNN Classifier

In k-NN classification, the output is a class membership. A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If $K = 1$, then the case is simply assigned to the class of its nearest neighbor. Choosing the optimal value for K is best done by first inspecting the data. In general, a large K value is more precise as it reduces the overall noise but there is no guarantee. Distance functions that can be used are Euclidean distance, Manhattan distance, Minkowski distance and in case of categorical data Hamming distance can be used.

5. Decision Tree

It is one of the predictive modeling approaches used in statistics, data mining and machine learning. Algorithms for constructing decision trees usually work top-down, by choosing a variable at each step that best splits the set of items. Different algorithms use different metrics for measuring "best". These generally measure the homogeneity of the target variable within the subsets. Some examples are given below. These metrics are applied to each candidate subset, and the resulting values are combined (e.g., averaged) to provide a measure of the quality of the split. Metrics used are Gini Index, Information gain, Entropy gain, etc.

Use Case Diagram



3.3 Testing

3.3.1 Testing Plans

a. Testing

Table 8

Types of Test	Will test be performed?	Comments/Explanations	Software Component
Requirement Testing	Yes	Input expected from user	To be tested on corpus implementation

Unit Testing	Yes	Python found compatible with algorithms	To be tested on Python
Integration	Yes	Successful software integration on all operating systems	The client program tested on different inputs
Performance	Yes	The software responds to any query in minimal time	To be tested on the software
Stress	Yes	Able to summarize sufficient amount of queries	To be tested on software
Compliance	Yes	In compliance with the existing competition needs	Suite to be tested and compared with existing technologies
Security	Yes	To ensure that candidates data are not leaked	Resume uploader
Load	Yes	Implementation allows multiple users to request at same time	To be tested on user's component
Stress	No	It is applied to check the response of the system when the system is given a load beyond its specified limits. This situation will not occur in our software, so it is not performed.	

Interface	Yes	Interactive interface for each user of the software	GUI interface to be tested on Python module (GTK).
Acceptance	No	This testing is carried out only when the software is developed for particular customer(s), but we have developed 'standardized' software for anonymous users.	

b. Test Team Details

Table 9

Role	Name	Specific Responsibilities
Test planner/Tester	Umang Nyati	<ol style="list-style-type: none"> 1. Ensure phase 1 is delivered to schedule and quality 2. Produce high level and detailed test conditions 3. Ensure test systems problems are reported immediately and followed up
Test Designer	Umang Nyati	<ol style="list-style-type: none"> 1. Identifying the appropriate techniques and guidelines to implement the required tests

		2. Analyzing the inputs, its effects and distinguishing clearly between the test cases
Tester	All members	1. Identifying the test data 2. Execute the test conditions 3. Produce high level and detailed test conditions

c. Test Scheduling

Table 10

Activity	Start Date	Completion Date	Hours
Study testing	18/03/2018	21/03/2018	6
Select Main Modules for testing	21/03/2018	23/03/2018	7
Obtain Input Data	23/03/2018	24/03/2018	4
Apply Various Testing Techniques	24/03/2018	26/03/2018	8
Analyze to resolve	26/03/2018	28/03/2018	4

d. Test Environment

Software items:

It refers to set of programs where the objective is to enhance the capabilities of the hardware machines. The minimum requirements are as follows -

- Operating System – Windows 7,10

3.3.2 Component decomposition and type of testing required

Table 11

Sr. No	Components that require testing	Type of testing Required	Technique for writing test cases
1	Running of algorithm	Unit, performance Stress	White box
2	Load Balancing	Unit, integration	Black box
3	Connectivity	Performance	White box
4	Results	Unit	Black box

3.3.3 List of test cases

Table 12

Test Case Id	Input	Expected Output	Status	Module
1	Self-driven	Self driven	Pass	Skills Match
2	Apache	Hard skill	Pass	Skills Match
3	Analytic	General skill	Pass	Skills Match
4	Algorithms	Hard skill	Pass	Skills Match
6	Job	General skill	Fail	Skills Match
7	Data scraped	Data in csv file	Pass	

	from indeed.com			
8	Csv file	Keywords formed using tf-idf and n-grams	Pass	Bag-of-words creation
9	Resume Parsing	Experience from resume	Fail	Job Prediction
10	Company Description	Summarized Data	Pass	Text Summarizer
11	Removing python library	Trace back error file not found	Pass	Text Summarizer

3.3.4 Error and Exception handling

Table 13

Test case id	Module	Test case for component	Debugging technique
1	Text Summarizer	Numerical Stability	Applying logarithmic probability

3.3.5 Limitations to the Solution

1. Text Summarizer needs to be more efficient and should produce results for every type of platform
2. Resume not accepted in every kind of format

3.4 Risk Analysis and Mitigation

Table 14

Risk Id	Classification	Description of Risk	Risk Area	Prob.	Impact	RE (P*I)
A1	Product Engineering	Design and selection of algorithms used	Functionality	Medium (3)	High (5)	15

A1	Product Engineering	Requirements of complete data to be worked on	Completeness	High (5)	Medium (3)	15
A3	Product Engineering	Inadequate systems properties, System properties do not match with tool's Requirements	Coding/Implementation	Low (1)	High (5)	5
A2	Product Engineering	Efficient Design Model runs for the test data in a go	Performance	Medium (3)	High (5)	15
B1	Development Environment	Development Process i.e Familiarity with languages, IDEs used	Familiarity	Low (1)	Medium (3)	3
B5	Development Environment	Work Environment should be stable and well cooperated	Cooperation	Low (1)	Medium (3)	3
B3	Development Environment	Management Process. Proper planning of phases	Planning	Low (1)	Medium (3)	3
B2	Development Environment	Development System. The csv files were too large that they caused system to hang multiple times	Capacity	Medium (3)	High (5)	15
C1	Program Constraints	Resources should be available at appropriate time	Schedule	Low (1)	Medium (3)	3

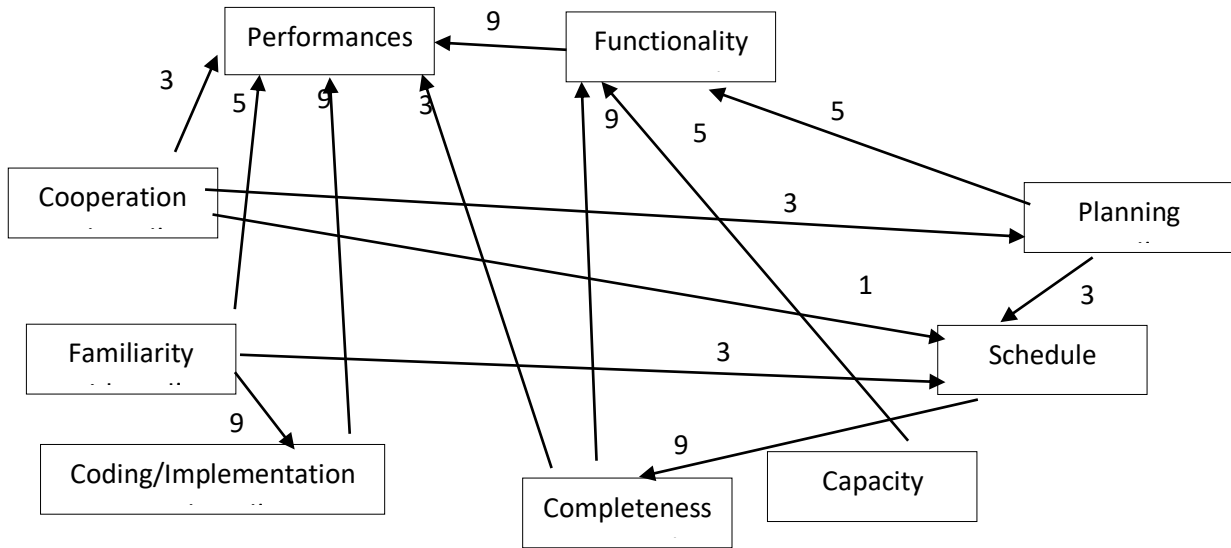


Fig 3: Interrelationship Graph

Table 15: Risk wise total weighting factor

S.No.	Risk Area	# of Risk statements	Weights (In+ Out)	Total Weight	Priority
1	Performance	6	3+5+9+3+9	29	1
2	Functionality	5	9+9+5+5	28	2
3	Completeness	2	9+9+3	21	3
4	Coding/Implementation	2	9+9	18	4
5	Familiarity	1	9+5+3	17	5
6	Schedule	2	1+3+3+9	16	6
7	Planning	4	3+5+3	11	8
8	Capacity	1	9	9	9
9	Cooperation	2	3+3+1	7	7

Table 16: Risk Statement

Risk Statement	Risk Area	Priority of Risk Area in IG
Risk of availability of data	Completeness	3

Mitigation:

Words and sentence tokenization, stop words removal using NLTK libraries

Effectiveness: Ease in breaking of sentences into words and removal of unnecessary words so that data could be read and processed correctly

Use of Pandas library to read and preprocess the csv data

Effectiveness: Easy and relevant in data processing

References

A. Jain, D. Bhatia and M. K. Thakur, "Extractive Text Summarization Using Word Vector Embedding," *2017 International Conference on Machine Learning and Data Science (MLDS)*, Noida, India, 2017, pp. 51-55

A. R. Lahitani, A. E. Permanasari and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," *2016 4th International Conference on Cyber and IT Service Management*, Bandung, 2016, pp. 1-6

Alotaibi, Shaha. (2012). A survey of job recommender systems. *International Journal of the Physical Sciences*. 7. 10.5897/IJPS12.482.

D. Hingu, D. Shah and S. S. Udmale, "Automatic text summarization of Wikipedia articles," *2015 International Conference on Communication, Information & Computing Technology (ICCICT)*, Mumbai, 2015, pp. 1-4

Guo, Shiqiang & Alamudun, Folami & Hammond, Tracy. (2016). Résumatcher: A Personalized Résumé-Job Matching System. *Expert Systems with Applications*. 60. 10.1016/j.eswa.2016.04.013.

J. Ren, S. D. Lee, X. Chen, B. Kao, R. Cheng and D. Cheung, "Naive Bayes Classification of Uncertain Data," *2009 Ninth IEEE International Conference on Data Mining*, Miami, FL, 2009, pp. 944-949.doi: 10.1109/ICDM.2009.90

Linna Li and Xuemin Zhang, "Study of data mining algorithm based on decision tree," *2010 International Conference On Computer Design and Applications*, Qinhuangdao, 2010, pp. V1-155-V1-158.doi: 10.1109/ICCDA.2010.5541172

M. S. Aldayel, "K-Nearest Neighbor classification for glass identification problem," *2012 International Conference on Computer Systems and Industrial Informatics*, Sharjah, 2012, pp. 1-5. doi: 10.1109/ICCSII.2012.6454522

N. Moratanch and S. Chitrakala, "A survey on extractive text summarization," *2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)*, Chennai, 2017, pp.1-6. doi: 10.1109/ICCCSP.2017.7944061

W. Zhou, Y. Zhu, F. Javed, M. Rahman, J. Balaji and M. McNair, "Quantifying skill relevance to job titles," *2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, 2016, pp1532-1541.doi: 10.1109/BigData.2016.7840761