SMART RESUME
BI TOOL TO SHORTLIST CVs FOR A JOB VACANCY

Project Id: 18-005

Project Proposal Report

Bachelor of Science Special (Hons) Degree in Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

02nd April 2018

SMART RESUME

BI TOOL TO SHORTLIST CVs FOR A JOB VACANCY

Project Id: 18-005

Project Proposal Report

Senarath S.P. – IT15099778
Madhushika K.A – IT15097798
Yureshani H.B.D– IT15040404
Y.I.Kodithuwakku – IT14115776

Supervisor: Mr. Lakmal Rupasinghe

Bachelor of Science Special (Hons) Degree in Information Technology

Department of Software Engineering

Sri Lanka Institute of Information Technology
Sri Lanka

02nd April 2018

DECLARATION

We declare that this is our own work and this proposal 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 our 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.

	STUDENT NAME	STUDENT NO.	SIGNATURE
1	Senarath S.P. (GROUP LEADER)	IT15099778	
2	Madhushika K.A	IT15097798	
3	Yureshani H.B.D	IT15040404	
4	Y.I.Kodithuwakku	IT14115776	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Name of supervisor: Mr. Lakmal Rupasinghe

Signature of supervisor: Date:

ABSTRACT

Today in Sri Lanka, most industries follow up one traditional process in hiring new employees. The normal process includes, advertising the vacancy, calling Curriculum Vitaes (CV), short listing them by referring the CVs and interviewing the short listed candidates. Having the right set of CVs is vital since a CV is the representation of the qualifications of an applicant. Also, when it comes to an emergency project, the employer should be able to hire the best employee set within a minimum time period. In this case help of a third party CV storage which already has a collection of related CVs, and has the ability to generate the list of most qualified applicants among them, would be helpful. Already there are many CV storages where people can submit their CVs and afterward they assign the applicants to relevant jobs. But still, there is a shortage of tools that support in selecting the best-qualified set of employees to an employer. The submitted CVs should be read properly and check several attributes such as skills, experiences and some personal information in order to select the best. It is much time consuming for a human to read and draw a mind image of the applicant. Smart Resume is a business intelligence tool for the IT sector, which analyze and classify operational data with classification algorithms to present complex and competitive information to decision makers, in order to dynamically fulfill the business needs. It is built to satisfy the task of generating the list of most suitable candidates. Third party CV storages can use this tool and provide the best solution for any client company. In this paper, we present a combination of desktop and web application that facilitates the task of automating the selection of the most suitable and qualified candidates depending on the attributes given by the user like Age, Gender, Work Experience, Soft skills and Education Qualifications. Depending on the relationship of the attributes (Internal and External) **Smart Resume** will dynamically visualize the most optimal or feasible candidate list.

TABLE OF CONTENTS

ABSTRACT	i
TABLE OF CONTENT	ii
1. INTRODUCTION	**1**
1.1 Background	1
1.2 Literature Survey	3
1.3 Research Gap	6
1.4 Research Problem	6
2. OBJECTIVES	**9**
2.1 Main Objective	9
2.2 Specific Objectives	9
3. METHODOLOGY	**11**
3.1 System Overview	11
3.2 Work Break Down Structure	17
3.3 Software Development Life Cycle	18
3.4 Gantt chart	19
4. DESCRIPTION OF PERSONAL AND FACILITIES	**20**
5. REFERENCES LIST	**22**

LIST OF FIGURES

FIGURE 1: SYSTEM OVERVIEW 11
FIGURE 2: PROCESS OF ETL TOOL 14
FIGURE 3: MODEL BUILDING PREDICTION 15
FIGURE 4: WORK BREAK DOWN STRUCTURE 17
FIGURE 5: SOFTWARE DEVELOPMENT LIFE CYCLE 18
FIGURE 6: GANTT CHART 19

LIST OF TABLES

TABLE 1: COMPARISON WITH EXISTING SYSTEM 6
TABLE 2: DESCRIPTION OF PERSONAL AND FACILITIES 21

1. INTRODUCTION

1.1 Background

Data in business is very useful only if data is analyzed properly which will aid in strategic business decision making. In the IT industry, workforce management is critical since there is a huge workflow in the industry. Hiring individuals make a big impact on the productivity of the company. To hire best-qualified employees there must be a system, as it is time-consuming to read and study all the curriculum vitae. Standardly curriculum vitae have 2-3 pages sometimes it may have more than 3 pages. So it may take a long time to go through all the details on it. And also, there is a probability of missing some CVs, as there is no proper method to store them. Therefore, we must find solutions to automatically analyze data, classify and summarize it along with discovering and characterizing trends and flagging anomalies in order to ease the decision-making process effectively in a company. **Smart Resume** targets at developing a set of tools, technologies and programmed products that are used to collect, integrate and make data available for better, faster decision making.

In **Smart Resume**, initially, all the CVs that are receiving by the email server in PDF form will be downloaded automatically. After downloading, each and every CV will be fully read from the system itself automatically. Then the downloaded CVs will be converted into CSV (Comma Separated Values) file format and saved. Here the internet BOT technology will be used to implement this functionality. As this is a BOT technology-based system, this will be an end to end system. The system will be worked under one training data set. An Internet BOT simply means a web robot. It is basically a software application which is used to run the automated tasks over the internet. A request to the server will send and the data will be fetched.

There are various attributes help to identify the best candidates for a given vacancy. After the data saved in CSV format, an ETL (Extract, Transform, Load) tool will extract relevant

attributes from the Source Files, clean them and save them in the data warehouse for further analyzes.

As the next step, client requirement for the job role is gathered. Then the prediction will be made by considering both client requirements and candidate data. It always depends on the nature of the job role, tasks, and required qualifications. **Smart Resume** will provide a shortlisted CV collection of candidates according to the user requirements.

Finally, by processing the entire data set based on predefined predictive models, it will generate comprehensive, descriptive and operational dashboards based on fact view data for those responsible for the job vacancies. The optimal or feasible result will be represented as a solution to the task in a user-friendly way. Graphs, charts, hierarchies, and tables will be used to represent those data.

1.2 Literature Survey

According to the research paper, the system downloads the Web pages using basically using Microsoft's Windows Internet API Tool (Winlnet). Also, to go through the links, PDFs, to identify texts, generate the successors of the downloading Web pages, Breadth-First search algorithm, and the Constraint Satisfaction method is used. There are two processes happening in the system; Downloading and Classification. Winlnet is there to connect to web servers when downloading data in different formats; HTML, images, and PDF. Several tasks are done by Winlnet such as requests to the web server for downloading the pages, determining a transfer mode (ASCII or binary) based on the relevant Web page's header. For controlling data flow and to track the downloading progress Breadth-First search algorithm is used. Furthermore, to make sure that the downloading moves smoothly without any error, the same Web page is not downloaded twice, and to revisit the unsuccessfully downloaded Web pages again. Breadth-First search algorithm is used because the links among Web pages are similar to a tree structure. Under the classification, contents of the hyperlinks will be categorized into texts, abstract PDF files and etc. For this purpose the Constraint Satisfaction method is used. As the

Classification is happening when downloading, only the necessary pages will be downloaded and placed in the relevant directories according to their formats. [1]

According to the paper an artificial neural network and genetic algorithm are used to solve effective text recognition problem. In order to do that a hetero-associative neural network is used to train the system for deciphering digits from pdf or jpeg images which are not readable. For the purpose of analyzing texts from handwritten or text file a crossover based genetic algorithm used. The algorithm solves the problem of deciphering digits and characters from the image. It's done by parsing image and converting it to a pixel array. The algorithm selects digits and characters and performs crossover with trained patterns with variable heights. [2]

The Objective of this research is to migrate Historical Data of Risk Management Product To a Data Warehouse. And Get Previous Day Reports for Further Analysis and Creating Reports. And They Use Large No of Data Such as Risk Management data, Ledger Data and Financial Application Data. In Methodology, they considered about migration Data from ETL tool instead of PL/SQL based transformation, Manually Transformation when Mapping Data. In Mapping Data- First, they Identify the Source Tables then What are the Entities That They want to Map into the Oracle Database. They Use Agile Methodology For their Approach in Reseach. [3]

The Objective of this research is to create a Model to active ETL tool. Use SQL Queries to Approach the Process for Mapping Data.SQL Queries use for All construction Process like data extraction, cleaning Procedures, direct Storage and for front-end information delivery. Try to Reduce the complexity of the process by applying SQL queries. [4]

This survey provides a comprehensive overview of various aspects of feature selection. They have introduced two architectures - a categorizing framework and a unifying platform. They categorize the large body of feature selection algorithms, reveal future directions for developing new algorithms, and guide the selection of algorithms for intelligent feature selection. A new feature selection algorithm can be incorporated into the framework according to the three dimensions. As data mining develops and expands

to new application areas, feature selection also faces new challenges. They have represented here some challenges in research and development of feature selection. [5]

In this paper, they have proposed a novel concept of predominant correlation, introduce an efficient way of analyzing feature redundancy, and design a fast correlation-based filter approach. The feature selection results are further verified by applying two different classification algorithms to data with and without feature selection. Their approach demonstrates its efficiency and effectiveness in dealing with high dimensional data for classification. [6]

This research paper focused on the possibility of creating a classification model to predict employee performance. When working on performance, many attributes have been tested, and some of them are effective in predicting performance. The Job Title was the strongest attribute, then the type of university, with little influence of degree and degree. The age attribute did not show any clear effect, while the marital status and gender showed an effect on some of the experiments to predict performance.
For the departments of management of companies and departments of Human Resources, this model or enhanced model can be used to predict the effectiveness of a new candidate. In this case, management can take several actions to avoid the risk associated with the recruitment of poorly performed employees. In this research as future work, it is recommended to collect more correct data from several companies.
Databases can be used for current and previous employees to have the correct performance rate for each of them. When a suitable model is created, the software can be designed for use by the HR, including the rules created to predict employee performance. [7]

This Research paper presents that selecting the appropriate metric to determine the optimal solution for obtaining an optimized classifier is a decisive step. The correct choice of the metric ensures that generative type classification training classifier is optimal. In this article, it is expected that the reviews of some metrics to recognize the optimum solution will sensitize data mining Researchers on this topic and encourage them to think carefully before choosing and applying metrically suitable to optimize the classification of the training.

In addition, this article also suggests several important aspects in constructing a better metric for recognizing the optimal solution for the generative type of classification algorithms. [8]

1.3 Research Gap

Even though there are existing proposed products in the market area, they do not address most of the problems that the proposed system is going to address. The following table shows a comparison of features between the existing products or applications and the proposed solution "**Smart Resume**".

Features	**Oracle BI**	**Birst**	**Jobscan**	**Smart Recruit**	**Smart Resume**
BOT(Automated)					✓
ETL Tool					✓
Optimal Solution			✓		✓
Feasible Solution					✓
All the user does not have to interact with the system	✓	✓	✓	✓	✓
Visualization of data in an abstract way		✓	✓	✓	✓
Predictive Analytics	✓	✓	✓		✓

Table 1: Comparison with existing System

1.4 Research Problem

The world is a chain of businesses. As the businesses get bigger day by day, the complexity and the competition are highly increasing. New trends are being incorporated into business ecosystems. New technologies are evolving rapidly, and it has significant influence in the organizational processes. So, day by day small to large all companies have

to update themselves in terms of resources, manpower, and infrastructures in order to maintain a competent and business system.

It is a known and documented fact that information-driven culture is needed in a company to meet customer needs in today's world. To achieve this, they need a tool or service to discover and prioritize business challenges across their organization with these new one's assessment methods and bring information closer to them, so they can make a smarter decision.

When hiring new employees, a company will have to spend much time, effort and cost on finding suitable candidates among thousands of the educated and qualified ones. As of the recruiting process using in the industry nowadays, companies have to spend a huge cost and time on selecting the perfect ones for the vacant position.

1. Advertise the vacancy
2. Call Curriculum Vitae of the interested candidates.
3. Shortlisting the applied candidate list by referring their Curriculum Vitaes.
4. Interview the shortlisted candidates and recruit the most suitable ones for the position.

But practically, it takes a lot of time and effort for a human to judge an employee's skill and talent just by reading their Curriculum Vitaes. Normally, a CV should contain 2 to 3 pages and all the relevant qualifications should be listed there. Because, according to normal policy, the time dedicated to reading one CV is 6 to 7 seconds. The reader should be able to grab the relevant information within that time period.

But, practically, there may be well qualified, talented candidates, who have a large skill set and a CV extended from 7 to 8 pages since it has to hold each and every qualification they achieved. Sometimes, the required qualifications for the specific position they applied, would be included in the last pages of the CV. In this kind of scenario, the reader would miss the important skills or points because they cannot waste much time on one CV. It is much time consuming for a human reader to read one CV end to end. And also, the most qualified candidates would not be called to the interview just because their CV

is too long or not well formatted. It is a huge disadvantage not only to the candidate but also to the company, since the company may lose the best employee to their vacancy.

On the other hand, there may be hundreds of applicants for a vacancy of a large IT industry. In such a scenario, it is very hard and time-consuming to download each and every CV and read them one by one in order to shortlist in human hands and send them to relevant companies who need suitable candidates. And also they need to save those candidates details in their databases for further uses such that another company also want the same set of candidates.

There is no any similar tool to satisfy the exact issue, but there are some similar commercially available BI tools but below are the drawbacks.

1. They are far too expensive, which are being developed by big vendors and often target the big clients.
2. Small and medium scale companies cannot afford a big cost or a time to find the suitable candidates because many of the employees have to engage in multi-tasks inside the company.
3. Cost for advertising for each candidate is very high.

Considering all the above facts, we can come to a point that there is a real need of cost-effective Business Intelligent (BI) tools that can cater the need of recruiting the best employees to a company by sending a set of selected candidates for relevant companies who require them. Therefore, the purpose of this research is to address such mentioned issues.

2. OBJECTIVES

2.1 Main Objective

Introduce an intelligent system to select the CVs based on the characteristics/attributes of the job vacancy. Depending on that given characteristics an optimal or feasible CVs will be filtered. The optimal CV selection is the one with the lowest probability of unfavorable outcomes an optimal solution is a theoretically proven solution. But this might not be a correct logical solution and we may have to come up with a feasible team. Hence the tool has the option of providing the most feasible (possible and practical) solution as well. The **Smart Resume** enables users to select the most optimal or feasible CVs according to the given attributes for the job vacancy of IT company or industry. The current market of the BI tools has a very complex interface, which requires professional knowledge to perform tasks. Smart Resume will not require technical knowledge or professional expertise to interact and it will be developed in a simple way with fewer controls to increase adaptability and comfort for the user.

2.2 Specific Objectives

- Download the CVs in the automatically
- To read the downloaded CVs
- Classify the data into relevant columns
- Save the classified data in CSV format
- Mapping CSV data into data warehouse by Extracting most suitable attributes, then cleaning the extracted data into more analytical manner by removing redundant data omitting null values and apply appropriate values. Then map the cleaned and

transformed data into the data warehouse created using MySQL for further Processing and analysis.

- ✓ Identify the Attributes with the use of Relevant data
- ✓ Extract the Relevant Data from CSV
- ✓ Cleaning Unorganized, Redundant Data (Data Cleansing)
- ✓ Transform Cleaned Data into Different Format
- ✓ Creating data warehouse using MySQL
- ✓ Load (Map) Cleaned and Transformed Data into data warehouse

- Build a solution to select the most optimal and feasible candidate list.

  - ✓ The main objective of this research part is to generate the most suitable candidate list in a faster and more accurate way considering attributes given by the user, related to the IT sector. The prediction is expected to be highly accurate and the final decision is sent to create a graphical view.

- After building the Predictive Model, the predicted details will be stored in a MySQL database. Using that data will find out that that result is the optimal or feasible CV selection according to the given attributes. ROC and classification Metrics will be used for those things. Also, the predicted result will be representing as Graphs, charts, hierarchies, and tables. It will provide a better user interaction through the interactive dashboard.

3. METHODOLOGY

This section includes detailed descriptions about the techniques and mechanism employed to make **Smart Resume** a reality. The descriptions include how software implementation of our project is carried out, what are the materials and data needed, and how they will be collected. It also includes time frames and schedules that are required in achieving its

objectives. In addition to them, the research areas that we have identified in order to carry out this project are explained rationally.

3.1 System Overview

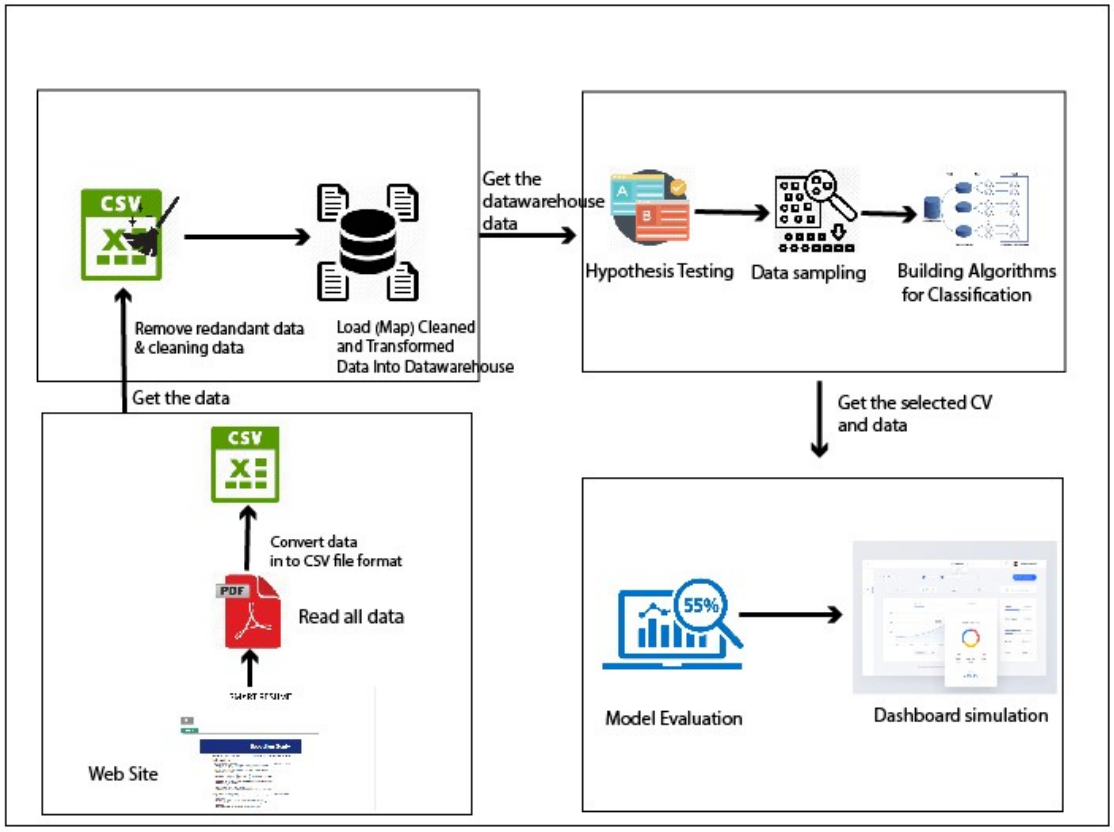

Figure 1: System Overview

a) BOT Creation

An Internet bot will be implemented to do the following tasks.

- Download the CVs in the email automatically
- To read the downloaded emails
- Classify the data into relevant columns
- Save the classified data in CSV format

To implement the internet bot Neural Networks will be used over NLP (Natural Language Processing).

b) ETL Tool

Data in a business is very useful if data analyzed properly to take proper decisions. There are various formats of data. For the research, we use **CSV** (comma separated values) data as row data to use in **ELT tools**. The user can select required attributes to analyze the data hence system consume selected attributes for predictive analysis. Additionally, the system detects odd data and removes it from processing. So we can get accuracy and best quality output.

ETL tool consists of 3 main parts.

- ✓ Extract
- ✓ Transform
- ✓ Load

Extract

The extract step covers the data extraction from the source system and makes it accessible for further processing. The main objective of the extracting step is to retrieve all the required data from the source system with as little resources as possible.

In most of the organizations, there are different types of data sources. CSV, XML, pdf, XLS are some examples. As we convert cv data into CSV format, necessary attributes have to be extracted for further processing.

Clean

After that data needs to be transformed into more organized steps. The cleaning step is one of the most important as it ensures the quality of the data in the data warehouse. Cleaning should perform basic data unification rules, such as:

- ✓ Remove unwanted data, incomplete data or incorrectly formatted data.

- ✓ Making identifiers unique (sex categories male/female/unknown, m/f/null, man/woman/not available are translated to standard male/female/unknown).
- ✓ Convert null values into standardized not available/not provided value Experience filter into year wise, Education(filter degree, a/l), Working category(software engineer/QA/tech lead/developers)

Likewise, we clean the data and come up with fit to the model.

Transform

In this step, data need to merged, aggregate, summarize or filter on the nature of the integration scenario.

Load

The final step is the loading. Transformed and cleaned data are stored in data warehouse for further analysis.

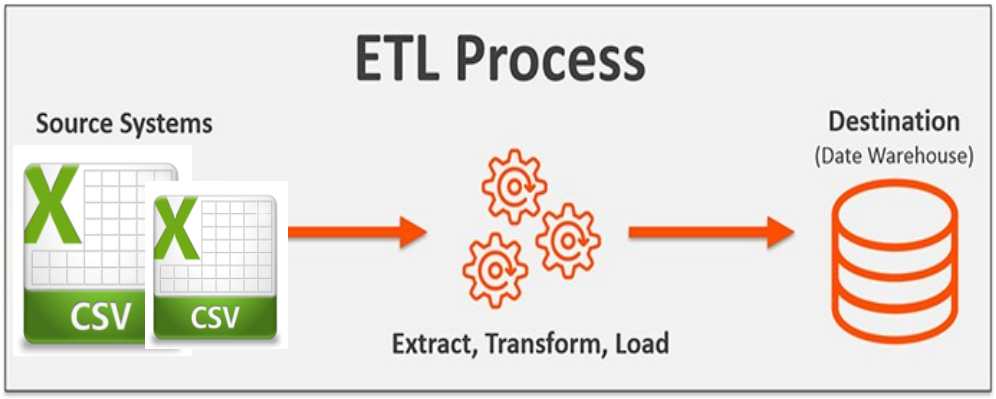

Figure 2: Process of ETL Tool

c) Model Building and Prediction

Smart Resume mainly works with data mining, classification, and forecasting. Once the relevant data is collected from the data warehouse, **Smart Resume** will make the statistical classification model which will help to make predictions.

So, in building this classification and predictive model, **Smart Resume** will follow the following steps;

1. Hypothesis Testing
2. Data Sampling
3. Building Algorithms for Classification

This will generate a list or a collection of most suitable candidates for a given qualification set, depending on whether their requirement is optimal or feasible. This optimality or feasibility will be selected by the user and also the required qualification list will also be provided by the user. The classification algorithms will sort out and filter the most suitable CV list by using the user given attribute list. With the help of this classification model, the filtering process will be done with more accuracy, thus predictions can be made wisely and easily.

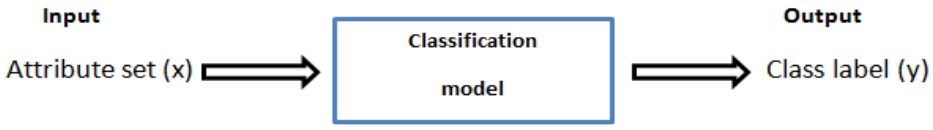

Figure 3: Model Building Prediction

The classification model should be made up of a number of predictors, which are variable factors that are likely to influence future behavior or results. The first step is to sample the main dataset to obtain a representative, and statistically valid sample of the whole using a suitable sampling technique. A proper sampling technique such as randomized statistical sampling or probability samplings like cluster sampling and stratified sampling must be used for this purpose. Using these techniques, the system has to produce Test Set, Training Set, and Holdout set.

Then, the suitable classification algorithm will be selected, based on the type of the required prediction. Selection of the proper algorithm has to be done with proper justifications.

Finally, the relevant data set will be added to the selected algorithm in order to find the list of the best candidates and produce the final prediction.

d) Model Evaluation & Prediction

The evaluation of the predictive model is carried out using three methods.

- ✓ Classification Table (Confusion Metrix)
- ✓ ROC (Receiver Operating Characteristics)
- ✓ Model Accuracy

The classification table shows the number of correct and incorrect forecasts made in comparison with the actual results (objective value) to the data.

The ROC curve is a graphical mark that illustrates the operation of a binary classifier system, such as it's the threshold of discrimination is different. A curve is created by tracking the true positive rate against false positive speed in various threshold parameters.

The precision of the model is measured by the proportion of correct predictions to the total number of cases evaluated.

e) Dashboard Simulation

After analyzing the filtered result data will present it in a format that will make user perfectly understands the difference between raw data and predictable data that would be able to represent the number of candidates got selected for the interview.

These data representations will help the Evaluators/Interviewers take the decisions on selecting most suitable professional candidates according to the company requirement.

The dashboard will represent the data in an interactive way using graphs, charts, hierarchies, and tables. The dashboard will provide you Summary of results and key points of analyzed data.

This approach enables the user to identify optimal or feasible CVs according to the specified attributes (requirement) given by the company. So that, the applicants who haven't submitted a good quality CV, also got the opportunity. It is not an only a static panel, where the user presents a set of predefined data each time they are loaded. This

system is expected that it will provide a better user interaction through the interactive dashboard.

3.2 Work Break Down Structure

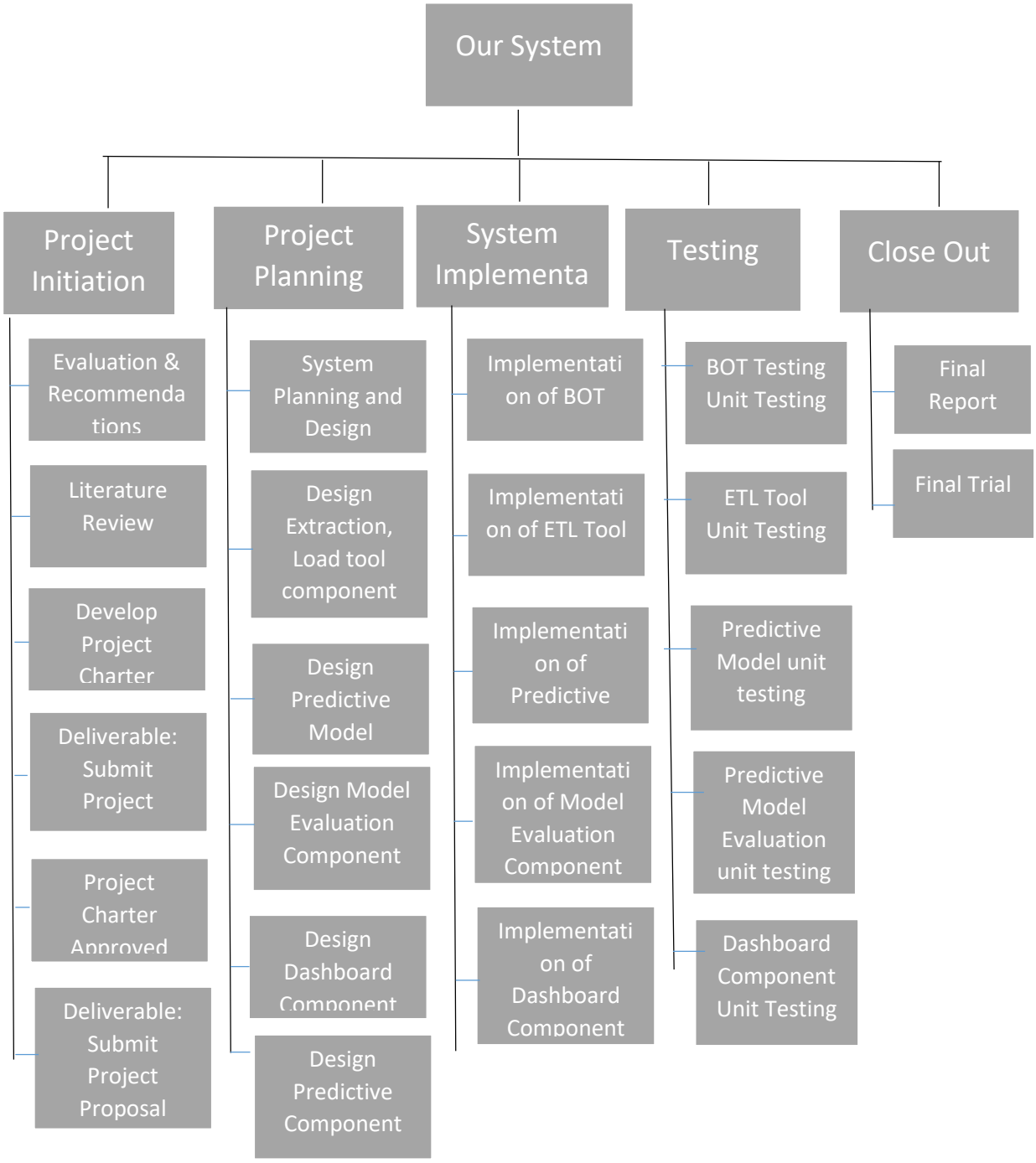

Figure 4: Work Break down Structure

3.3 Software Development Life Cycle

The software development methodology which is going to be used is the Agile Scrum process. Agile Scrum process facilitates lots of features which developers can use to develop a system in a flexible way. This process allows for requirement changes so easily as developers want to fulfill the customer's new requirements. Since the research project are associated with a lot of changes, it is hard to follow a set sequence of developments such as the waterfall model. Scrum concentrates on task management within a team based environment. As this project is done by a group of four members this methodology will help a team of performance all the activities and solve the problem faced by individual members. As well as that it facilitates a common understanding of each area of the project by conducting a daily scrum meeting.

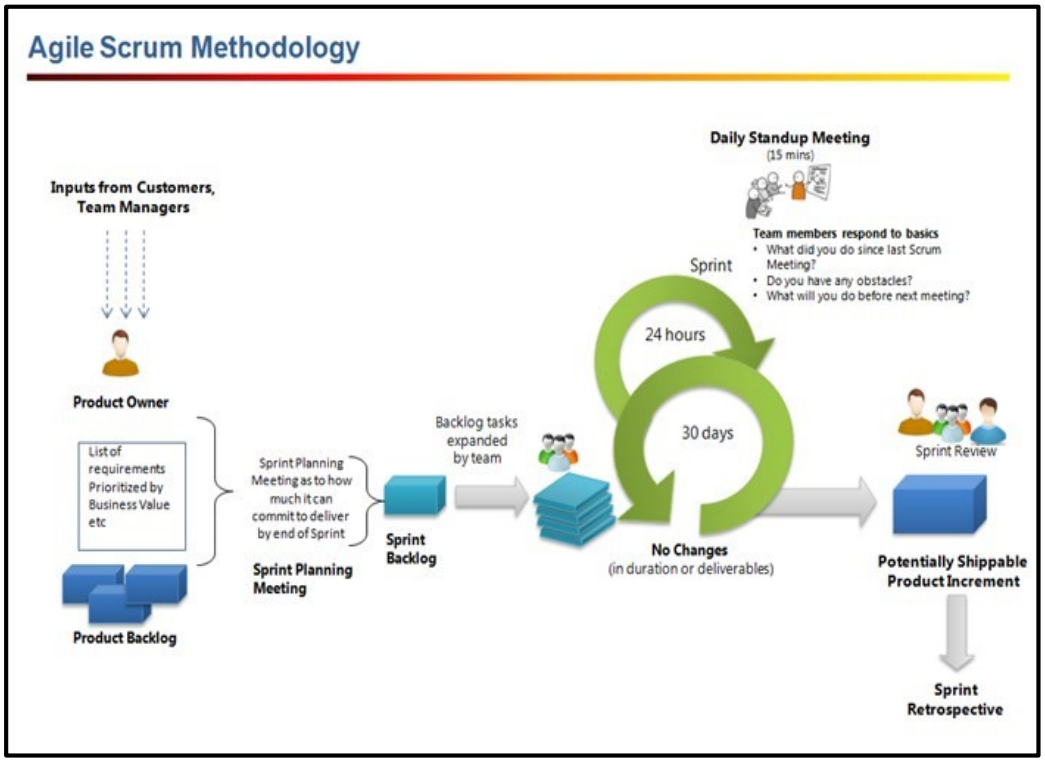

Figure 5: Software Development Life Cycle

3.4 Gantt Chart

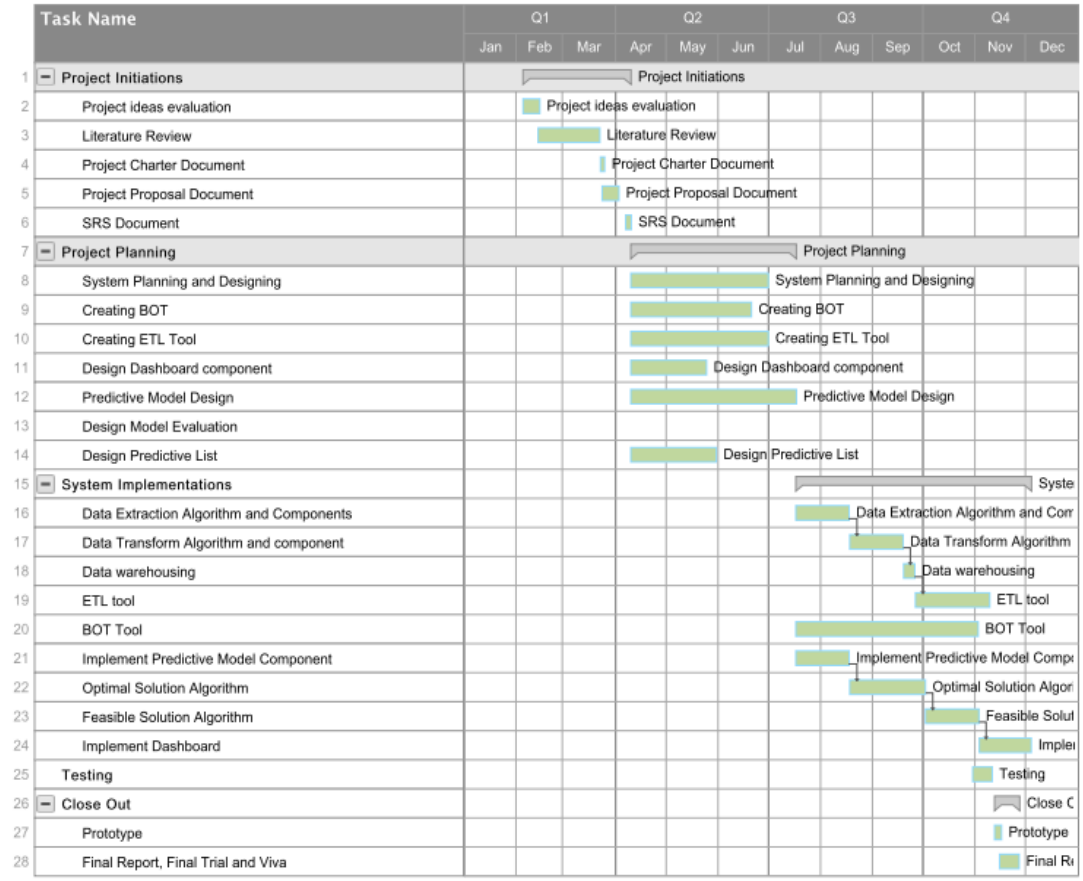

Figure 6: Gantt chart

4. DESCRIPTION OF PERSONAL AND FACILITIES

Member	Component	Task
Senarath S.P IT15099778	Creating BOT	Download the CVs in the email automatically Read Downloaded emails Classify the Data into relevant columns Save the classified data in CSV format Documenting Testing
Madhushika K.A IT15097798	Data Warehousing using ETL Tool	Identify the Attributes with the use of Relevant data Extract the Relevant Data From CSV Cleaning Unorganized, Redundant Data (Data Cleansing) Transform Extracted and Cleaned Data into Different Format Creating Dataware house using MySQL Load (Map) Cleaned and Transformed Data into Dataware house Data quality management Documenting Testing
Yureshani H.B.D IT15040404	Model Building and Prediction	Hypothesis Testing Data Sampling Building Algorithm for Classification Documenting Testing

Y.I Kodithuwakku IT14115776	Model Evaluation & Prediction	Evaluate the predicted model if it is the optimal or feasible result Documenting Testing
	Dashboard Simulation	Monitor the accuracy of the prediction models Documenting Testing

Table 2: Description of Personal and Facilities

5. REFERENCES LIST

[1]L. Q. Tran, C. W. Moon, D. X. Le, and G. R. Thoma, “Web Page Downloading and Classification,” *Proc. IEEE Symp. Comput. Med. Syst.*, pp. 321–326, 2001.

[2]M. Agarwal, “Text recognition from image using Artificial Neural Network and Genetic Algorithm,” pp. 1610–1617, 2015.

[3]C. Shrinivasan, “Data migration from a product to a data warehouse using ETL tool,” *Proc. Eur. Conf. Softw. Maint. Reengineering, CSMR*, pp. 63–65, 2011.

[4]Z. Chen and T. Zhao, “A new tool for ETL process,” *Proc. 2012 Int. Conf. Image Anal. Signal Process. IASP 2012*, pp. 269–273, 2012.

[5]H. Liu and L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering", *Pdfs.semanticscholar.org*. [Online]. Available:https://pdfs.semanticscholar.org/185e/7e2d397478d0d81b89c6f712b1fa7d62c979.pdf. [Accessed: 01- Apr- 2018].

[6] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution", *Aaai.org*, 2003. [Online]. Available: https://www.aaai.org/Papers/ICML/2003/ICML03-111.pdf. [Accessed: 01- Apr- 2018].

[7] Q. A. Al-Radaideh and E. Al Nagi, "Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance", *Thesai.org*, 2012. [Online]. Available: https://thesai.org/Downloads/Volume3No2/Paper%2025%20-%20Using%20Data%20Mining%20Techniques%20to%20Build%20a%20Classification%20Model%20for%20Predicting%20Employees%20Performance.pdf. [Accessed: 01- Apr- 2018].

[8] H. M. and S. M.N., "A REVIEW ON EVALUATION METRICS FOR DATA CLASSIFICATION EVALUATIONS", *Pdfs.semanticscholar.org*, 2015. [Online]. Available:https://pdfs.semanticscholar.org/6174/3124c2a4b4e550731ac39508c7d18e520979.pdf. [Accessed: 01- Apr- 2018].