Sri Lanka Institute of Information Technology

SMART RESUME
BI TOOL TO SELECT THE CVs APPLIED FOR A JOB VACANCY

Project ID: 18-005

Software Requirement Specification

Author: Senarath S.P.

B.Sc. Special (Honors) Degree in Information Technology

Specialized in Interactive Media

Submitted on: 12/05/2018

SMART RESUME
BI TOOL TO SELECT THE CVs APPLIED FOR A JOB VACANCY

Project ID: 18-005

Author: Senarath S.P.

Supervisor : Mr. Lakmal Rupasinghe

Submitted on: 12/05/2018

DECLARATION

I declare this Software Requirement Specification report entitled “Smart resume - bi tool to select the cvs applied for a job vacancy” submitted to the Sri Lanka Institute of Information Technology is not a copy of a document done by any organization, university or any other institute or a previous student project group at Sri Lanka Institute of Information Technology and it is a record of original work done by me under the guidance of Mr. Lakmal Rupasinghe(Supervisor). The diagrams, research results and all other documented components were developed by me and I have put the references that I have used.

……

Senarath S.P.

Contents

List of Figures

List of Tables

1. INTRODUCTION

1.1 Purpose

The purpose of this SRS document is to outline the requirements and present a detailed description of the process needed for **Personalized hangouts planning and Business analytics**. The document contains the necessary requirements of the system, as well as the process to create and discover them. It will explain the functional and non-functional requirements, purpose and features of the component, the interfaces of the component, design constraints, project approach, what the component will do, the constraints under which it must operate and how the component will interact with other external applications. The information is organized in such a way that the developers will not only understand the boundaries within which they need to work, but also what functionality needs to be developed and in what order. This document is intended to be proposed to a customer for approval and also this document is targeting the, designers, developers and other stakeholders as its audience.

1.2 Scope

The component to be produced is called "**BOT Creation".** This component has four major sub components called Download the CVs in the email automatically, to read the downloaded emails, Classify the data in to relevant columns and Save the classified data in CSV format. Smart resume consists with a desktop application. "Downloading the CVs in the email automatically" provides the facility of downloading the relevant CV automatically for a given email address.

"To read the downloaded emails" provide the capability of read the downloaded cv. "Classify the data in to relevant columns" provide the capability of identify the terms in cv and save the data in relevant columns and "Save the classified data in CSV format" provides the facility of save the data in csv file format in internal storage.

Smart Resume CV Selector is a Business intelligent tool, which analyzes and classify operational data with algorithms and presents a web application that facilitates task automating to select most suitable and most qualified candidates depending on attributes given by the user. An ETL (Extract Transform and Load) tool will analyze the data and load them into a data warehouse then prediction

model will design the optimal and feasible solution in IT industry. Depending on the attributes, Smart Resume will dynamically visualize the candidate list.

The document is organized into several sections in order to give a clear understanding of project requirements. And it gives the detailed description of both functional and non-functional requirements for the above-mentioned system. Functional behaviors are associated with the specific functions, tasks or behaviors of the system. Nonfunctional requirements are the constraints on various attributes of the function, which need to be considered. This SRS covers the requirements for the Data warehousing using ETL tool. If the requirements changes in the future, it is possible to change the specification accordingly. Also, covers the hardware and software requirements need to implement the ETL tool and covers the limitations.

The scope of the **predictive model builder** is having the following functional areas. There are three primary components that need to be handled in order to complete Smart Resume predictive model builder component.

1. Model Building and Prediction

2. Predictive Model Evaluation

3. Dashboard Simulation

Model Building and Prediction is the most important part of the whole project. Scope of Model building and Prediction will include the following steps.

1. Hypothesis Testing

2. Data Sampling

3. Building Algorithms for Classification

This document also describes Model Evaluation and Dashboard Simulation. The scope of the predictive model evaluation and dashboard simulation is limited to implementing the CV Management System having the following functional areas. Two primary components need to be handled in order to complete Model evaluation and Dashboard simulation component.

1. Predictive Model Evaluation
2. Dashboard Simulation

If the requirement changes in future, it is possible to change the specification accordingly. It also covers the details of hardware and software requirements need to implement the predictive model and gives a detailed description of the externally visible behaviour of the predictive model and covers the areas that contain limitations while completing the predictive model for IT Industry.

The proposed Smart Resume aims at describing processes of building cost effective, flexible, ease of use and productive enterprise focused Business Intelligence tool.
Driven by fact –Data to decision, Smart Resume Business Intelligence (BI) tool make it possible to reach unambiguous decisions in the right time, minimize costs and provide the opportunity to verify and correlate the decisions with the entire enterprise's strategy.
In the dashboard, visualizer will give the detailed view of data for the user to identify the candidate information with graphs. This will help the user to recruit the best employees for the company.

Other Objectives

- To make the Business intelligent tool easy to understand and easy to use for non-techy people
- Develop a system with high accuracy, efficiency, flexibility, and support for other non-functional requirements
- Build solutions to select optimal or feasible CVs for IT industry

1.2.2 Benefits

- Smart Resume is a comprehensive product with ETL tool and visualization features so there is no need to buy separate ETL tool

- Small or large all IT industry firms can use this product
- Reduce time to check CV and analyze them
- Reduce time which use to insert data one by one.
- No need of expert knowledge in order to deal with this component.
- It helps to get faster answer to business operations.

1.3 Definitions, Acronyms, and Abbreviations

Term	**Definition**
SRS	Software Requirement Specification
BOT	A bot (short for "robot") is an automated program that runs over the Internet
OS	Operating System
User	Desktop app user
HTTPS	Hypertext Transfer Protocol Secure
Wi –Fi	A facility allowing computers or other devices to connect to the Internet or communicate with one another wirelessly.
Front End	Program that is responsible for interaction with the users. Users interact directly with it. It contains interfaces.
Back End	Application responsible for do all the processing. It is a service. Users don't interact directly with it. It doesn't contain interface.
ETL	Extract Transform Load Tool
DW	Data Warehouse
MySQL	MySQL database
CSV	CSV

CV	Curriculum vitae
GUI	Graphical user interface
IT	Information Technology
BI	Business Intelligence
HR	Human Resource
ROC	Receiver Operating Characteristics
TP	True Positive – correctly predicted that suitable for job vacancy
TN	True Negative – correctly predicted that not suitable for job vacancy
FP	False Positive – Incorrectly predicted that suitable for job vacancy
FN	False Negative – Incorrectly predicted that not suitable for job vacancy
AUC	Area under curve

Table 1: Definitions and Abbreviations

1.4 Overview

The remainder of this SRS document includes three sections and appendixes. The second section provides an overall view of the component functionality and interaction with other components. This section also discusses the specific requirements such as functional and nonfunctional requirements, design constraints and various approaches. Furthermore, this section also mentions the system constraints, User characteristics and assumptions about the product.

The third section provides the requirements specification in detail and a description of the different interfaces. Different specification techniques are used in order to specify the requirements more clearly for different audiences.

The rest of the sections that organized in this document are Project perspective and descriptions, Different interfaces that the system consists, Requirements of the system, Summary of major

functionality, Users and characteristics of the system and the background of the general factors affect the system.

1.4.1 Goals and Tasks

- Identify the Relevant and most suitable attributes to store in the data warehouse(Extract)
- Validate the data in CSV rows
- Remove redundant data, cleaned unorganized data (Data Cleansing)
- Transform data into a normalized format.
- Map data into a data warehouse
- Create interfaces for the user
- Get user requirements regarding the candidates.
- Create predictive model to select the best cv list.
- Show the results.
- Evaluate the prediction models to predict the optimal and feasible CVs.
- Increase model accuracy and performance.
- Develop informative dynamic dashboard.

1.4.2 Users

- HR workers in IT firms
- Top Level Management
- DB Administrator

2. OVERALL DESCRIPTIONS

“Downloading the CVs in the email automatically” provides the facility of downloading the relevant CV automatically for a given email address. If the user enters email address, bot go to inbox and download the cv.

User will interact simple interface to enter their email address.

After the download cv, bot will read all the details in cv line by line. Then the identify the terms like,

Name: - Kamal, Nisal, Harsha

Telephone No: - 071-*******, 077-*******, +94-*******

Age: - 20,30,52

Email Address :- asaa@gmail.com

After the identify terms, input the details in xl file, then the xl file save csv format in repository.

Smart Resume ETL Tool involve to store data from CSV files and store them in a data warehouse for further analysis and decision-making processes. To get data set for the predictive model's data should be stored in a data warehouse. To store those set of data will create an ETL tool to get the Dataset of CSV files.

In order to save data in data warehouse, data should be in a normalized format and fit into the data warehouse. For this purpose, data must be Extracted, Transform and d cleansing the data by removing redundant data and validating data sets. Lastly, dataset map into a data warehouse for further processing.

The predictive model will be built using a sample from the data set which will be created by the ETL tool of Smart Resume. Statistical inference is used to draw conclusions about the data and processes to identify the relationships in data. With the use of attributes which will be dynamically selected by the user final team formation should be predict. Hence, using statistical methods most important variables should be identified.

Classification is used to construct the predictive model using the historical data that accurately predicts the teams. Different algorithms are used to find relationships between predictor attributes. This process is done using test data to test the predictive model as it measures the predictive model accuracy. This model evaluation method is used to identify the best predictive model [7]. For further evaluation optimization techniques can be applied to increase the model accuracy.

Smart Resume - model evaluation and dashboard simulation component involve the predictive model evaluation, analysis, implementation and dashboard simulation of CV insights in regards to the data set available. The predictive model is an “expression of relationships between variables in the form of an equation”.

The evaluation of the predictive model is carried out using confusion Metrix, ROC curve, and Model accuracy methods. Classification table describes the performance of a classification model.

	Predicted: 0	**Predicted: 1**
Actual: 0	TN	FP
Actual: 1	FN	TP

Table 2- Classification Table

Every observation in the testing set is represented in exactly one box. It is a 2x2 metrix because there are 2 response classes. Classification Accuracy metric need to be calculated using classification metrix.

Classification Accuracy = (TP+TN)/ (TP+TN+FP+FN)

Receiver Operating Characteristic (ROC) Curves provide a graphical representation of the range of possible cut points with their associated sensitivity vs. 1-specificity. This illustrates the merit of the particular predictor/predictive model, making it possible to identify different cut-points for specific applications – depending on the 'cost' of misclassification. Estimates of the area under the curve provide an indication of the utility of the predictor and a means of comparing (testing) two or more predictive model. ROC Curves plot the true positive rate (sensitivity) against the false positive rate (1-specificity) for the different possible cut points of a diagnostic test. Each point on the ROC curve represents a sensitivity/specificity pair.

Finally, after processing all data set based on predefined predictive models, it generates comprehensive, self-descriptive and exploratory dashboard, which provide fact – driven visualization of the data. The result will be displayed as graphs, charts, hierarchy with selected employees for the job interview. Further analyzed details of a particular employee also can be viewed as well as descriptive reports also available in the dashboard.

2.1 Product perspective

There are several bot available the world. This bot provides to chat, game and manage the send email. But there is no any available bot to download cv and read the cv. There are some researches like,

SQL Server Integration Service (SSIS)

SSIS[4] is a component of a SQL server Database which performs integration on windows environment. It is quite expensive ETL tool in the market. Main disadvantage of this tool is it only operate on windows platforms only.

Jobscan Recruitment Solution Tool

- Here we have to submit one CV at a time.
- No BOT to save or download CVs.

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 3- Existing systems Vs Smart Resume

Web Page Downloading and Classification

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, 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 the paper an artificial neural network and genetic algorithm is 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 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]

2.1.1 User interfaces

The user interface (UI) comprises the logical face between software product and its users. This component deals with the web application.

This interface is used to Login the web application.

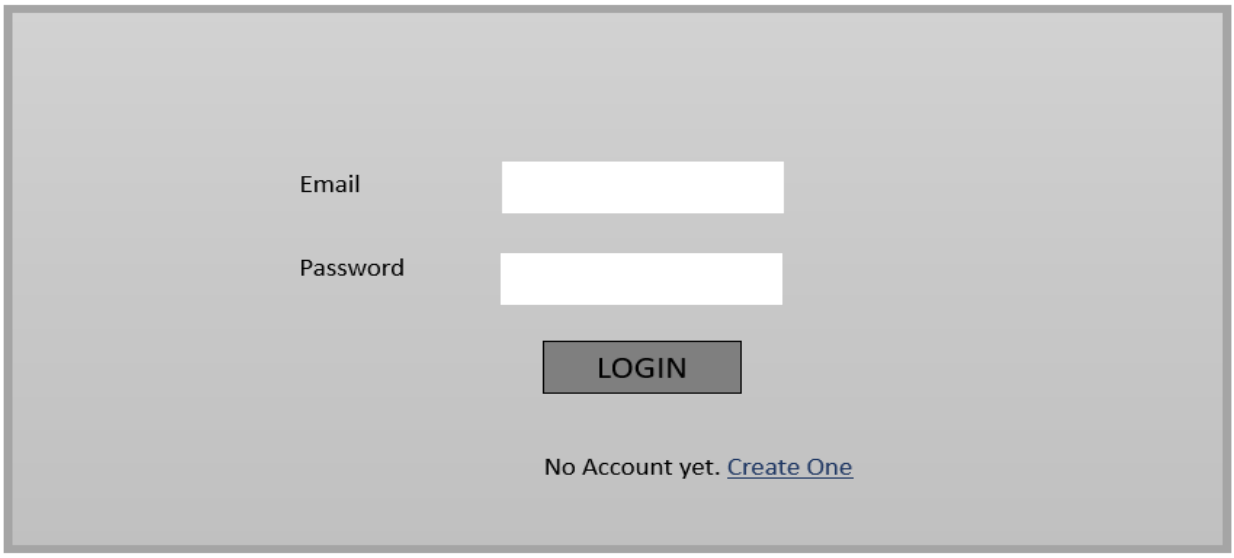


Figure 1: Login UI

This interface is used to Sign up or Login the web application.

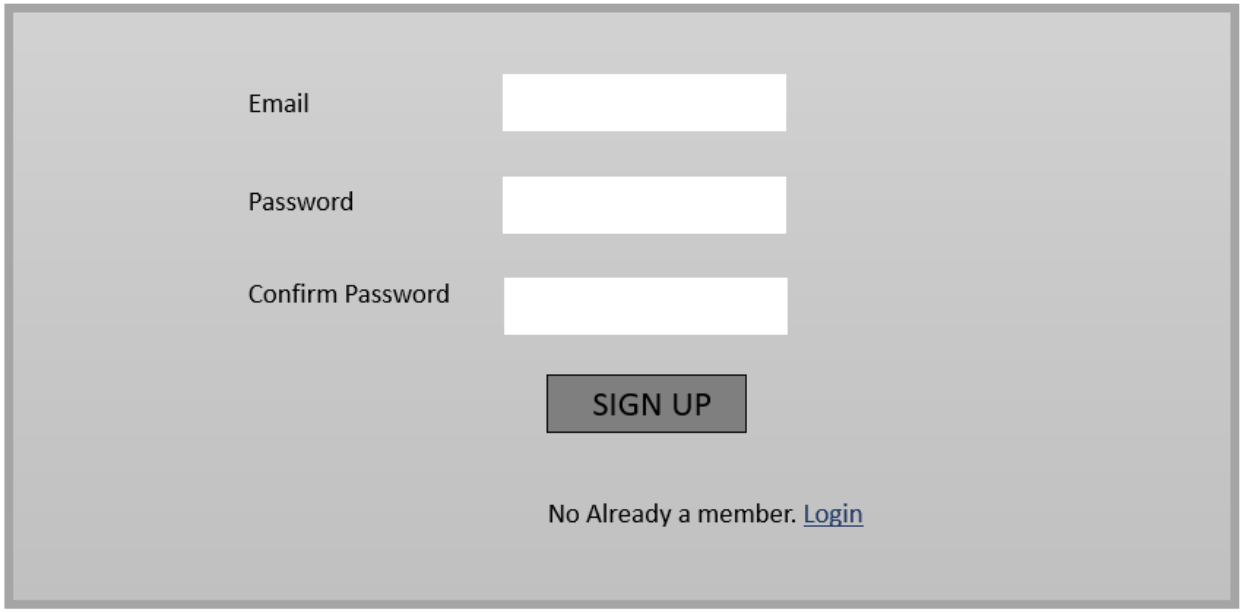


Figure 2: Sign up UI

This interface is used to enter users email address and it shows the how many cvs are downloaded.

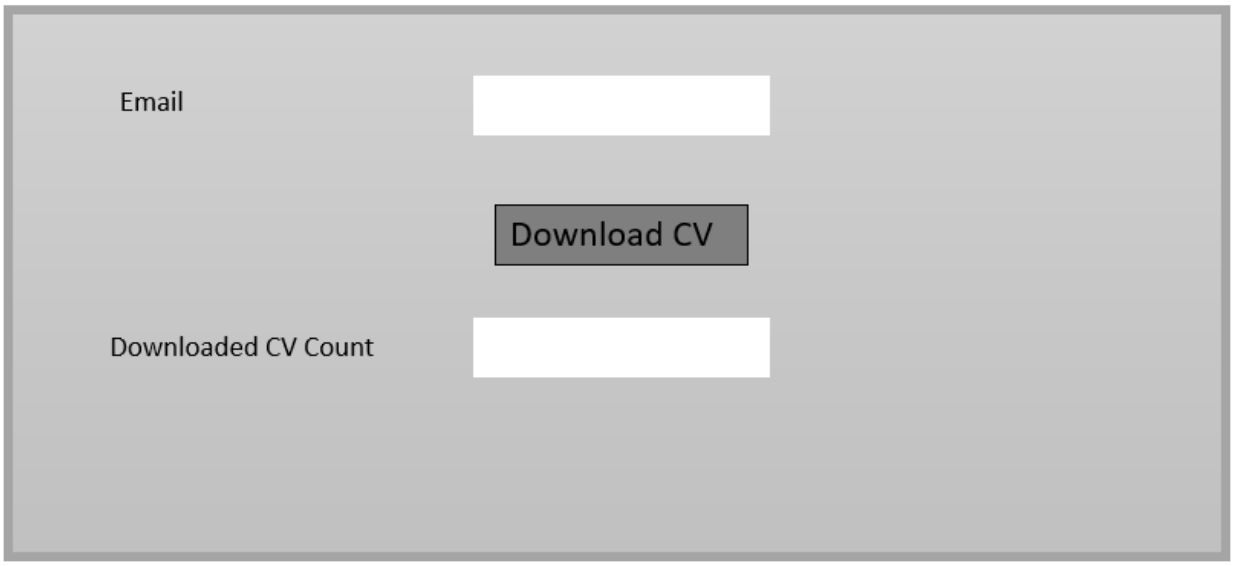


Figure 3: Show Email Count UI

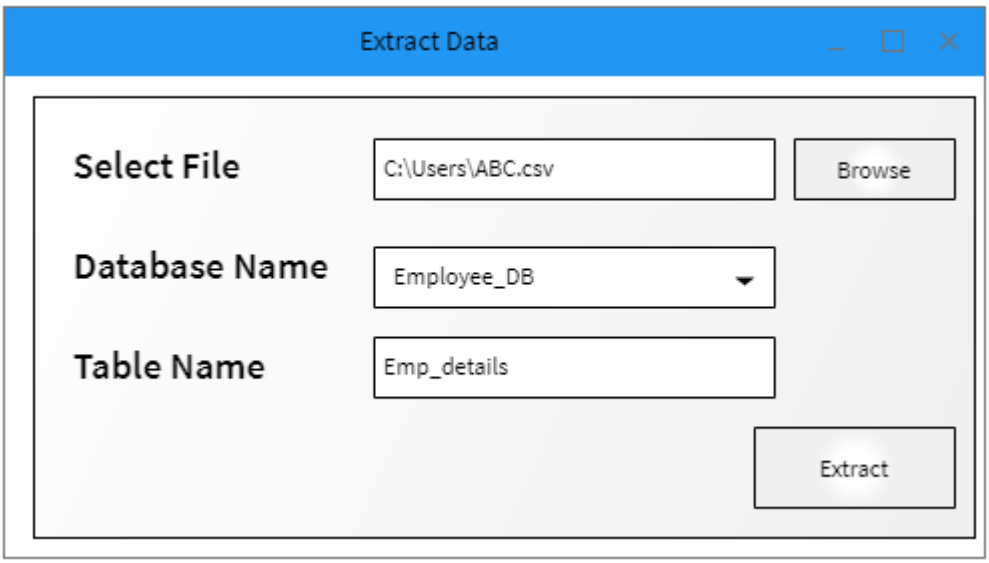


Figure 4- Data Extraction Interface

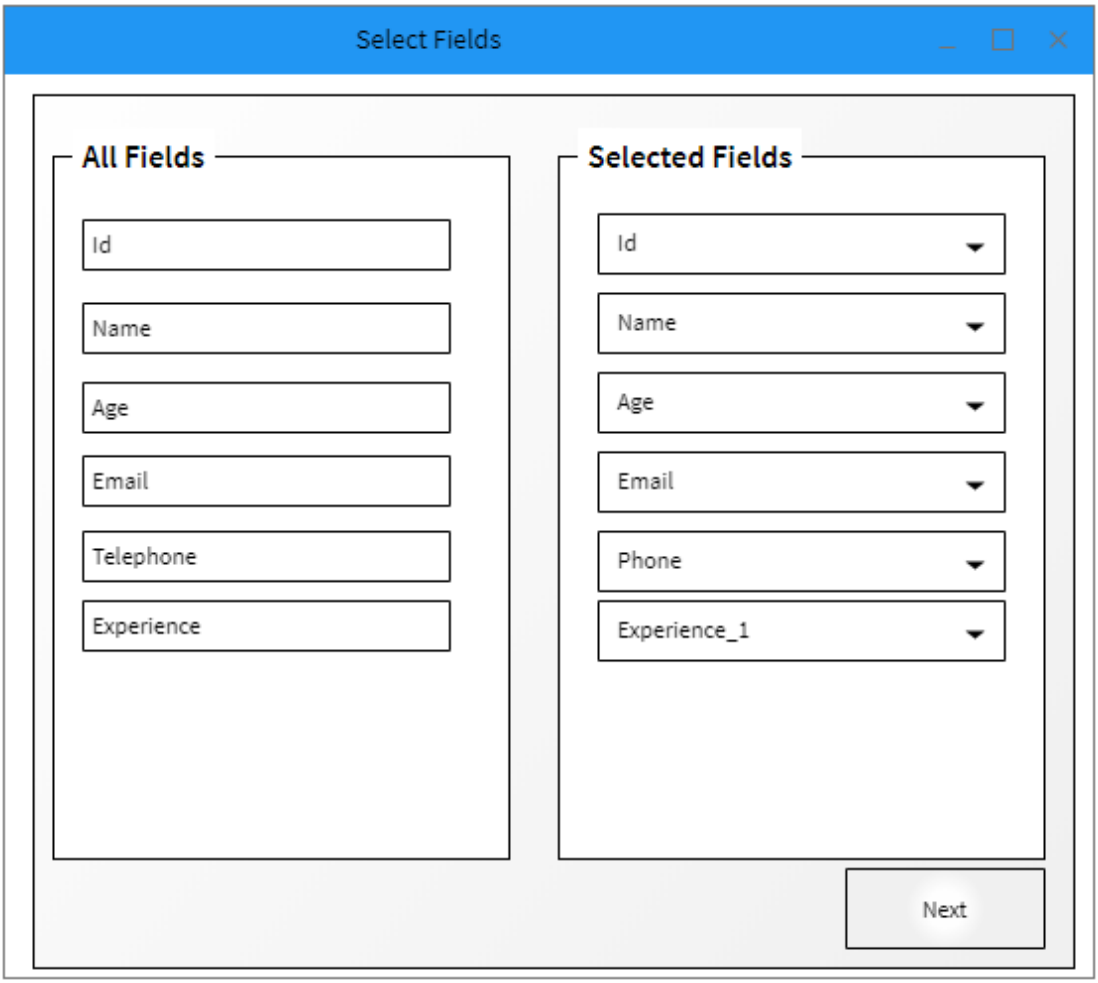


Figure 5- Field Selection Interface

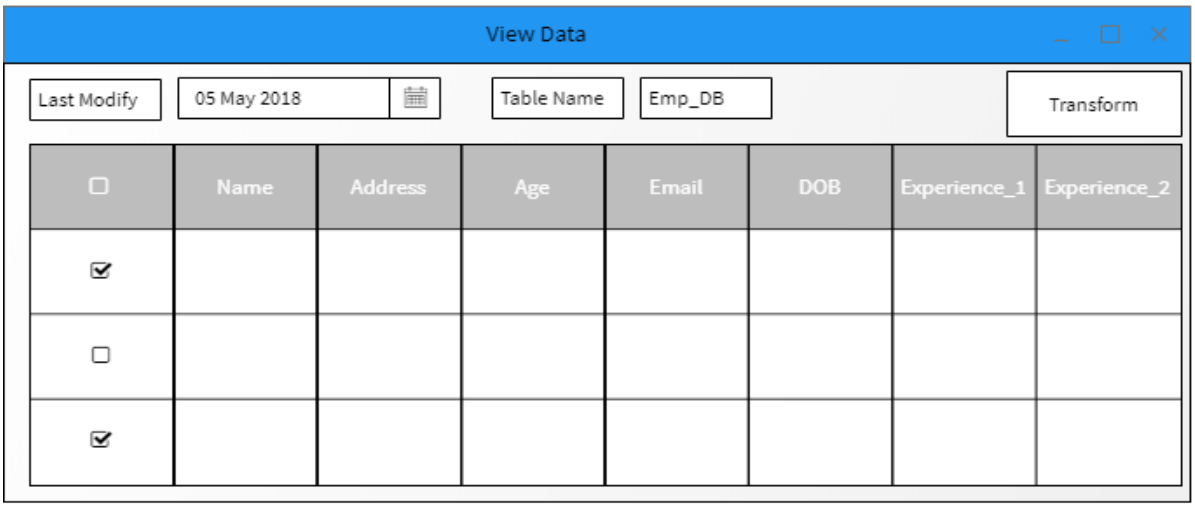


Figure 6- View Data Interface

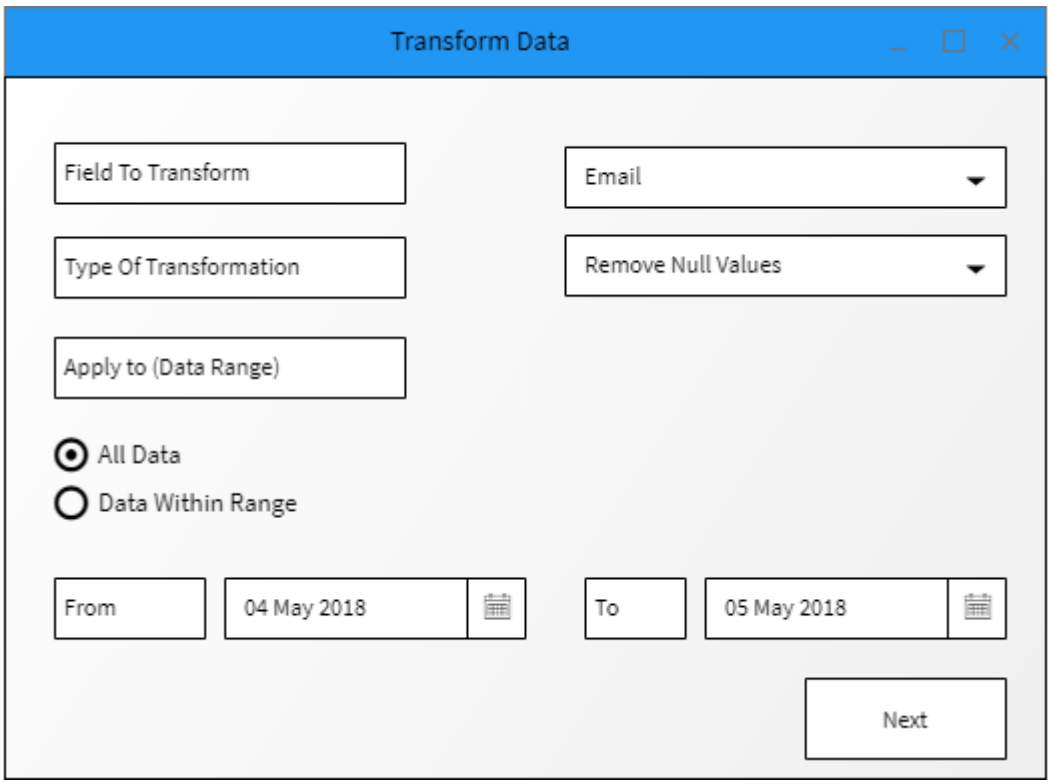


Figure 7- Data Transform Interface

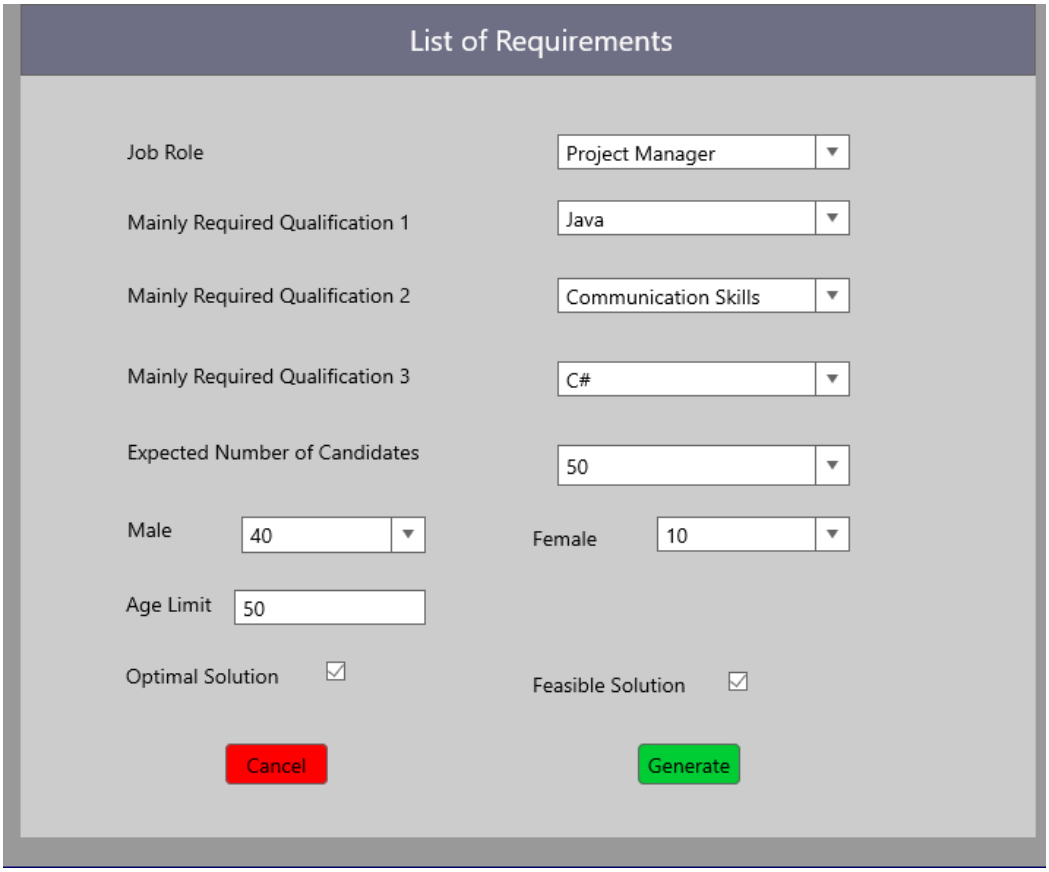


Figure 8- User Requirements Gathering Interface

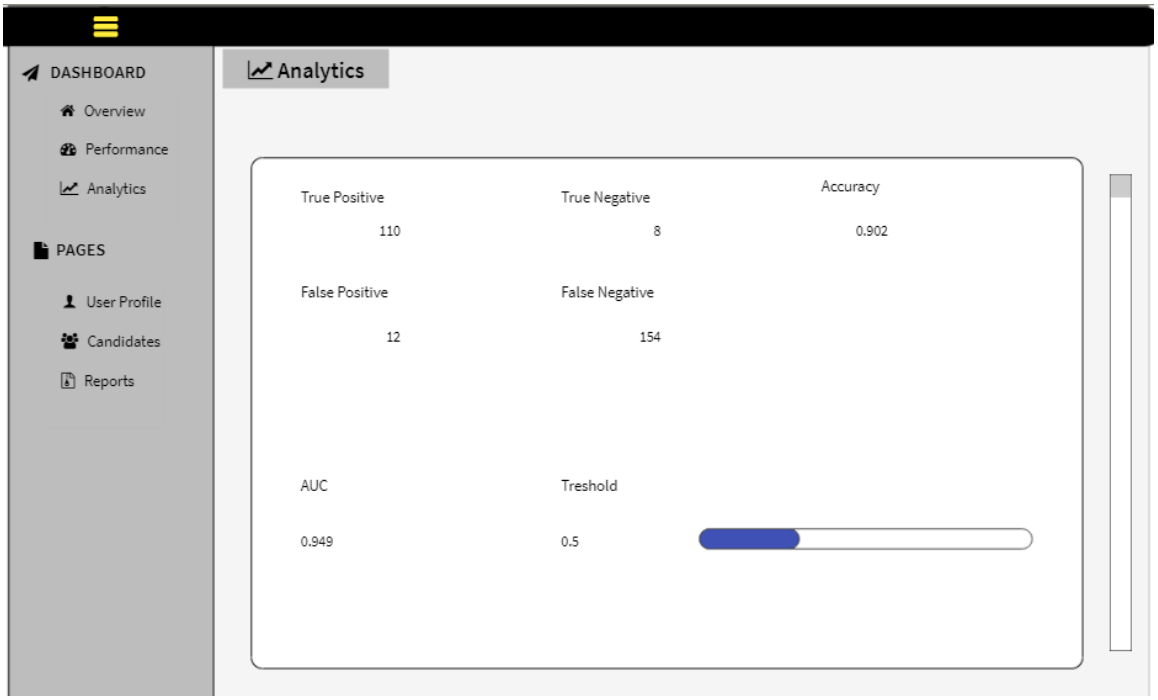


Figure 9- Statistics of the analyzed data

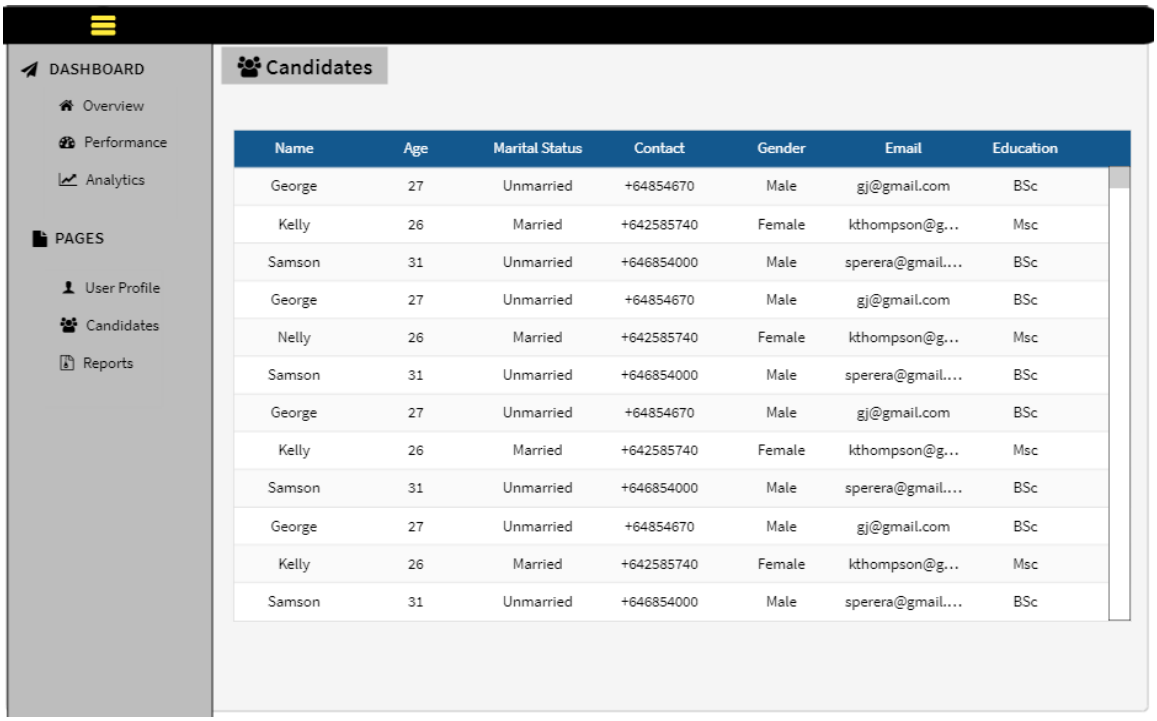


Figure 10- View Selected candidates

2.1.2 Hardware interfaces

As this will be a web application there isn't any specific hardware, it does not have any direct hardware interfaces. Basically, there are few requirements like,

- Desktop or Laptop with Windows operating system
- Wi-Fi router

Smart Resume ETL Tool does not require any special Hardware interfaces apart from a computer.

2.1.3 Software interfaces

- In order to use web application users should install a web Browser in the computer to access the web application.
- MySQL
- Jython Library

- Apache Server

2.1.4 Memory constraints

- Web server should have minimum of 4GB RAM since some functions should be perform at real-time.
- Internet
- Database Connection Interface

2.1.5 Operations

Prior to use, the web application would require a sign up before perform the functionalities. Here are some basic operations done by the user,

- Sign up or Login
- Enter the email address
- RAM of 2GB or higher
- Interpret results with the dashboard.
- Viewing reports

2.1.6 Site adaptation requirements

In order to run the desktop application user must install the application on their computer. The System will display the information only in English Language. Basically this a desktop application and the development team is planning to provide a web application in later stages.

Operations of Database Administrator:

- Extract Data from Given Source.
- Cleansing of data with necessary attributes and get it transformed.
- Load the data set to the data mart.

Operations of Top Level Management

- Extract, Transform and Load Dataset to the Data warehouse
- Select suitable attributes and generate optimal or feasible team.
- Interpret results with the dashboard.
- The server must have MySQL server installed on it.

- The user machine should have Java Virtual Machine installed.

2.2 Product functions

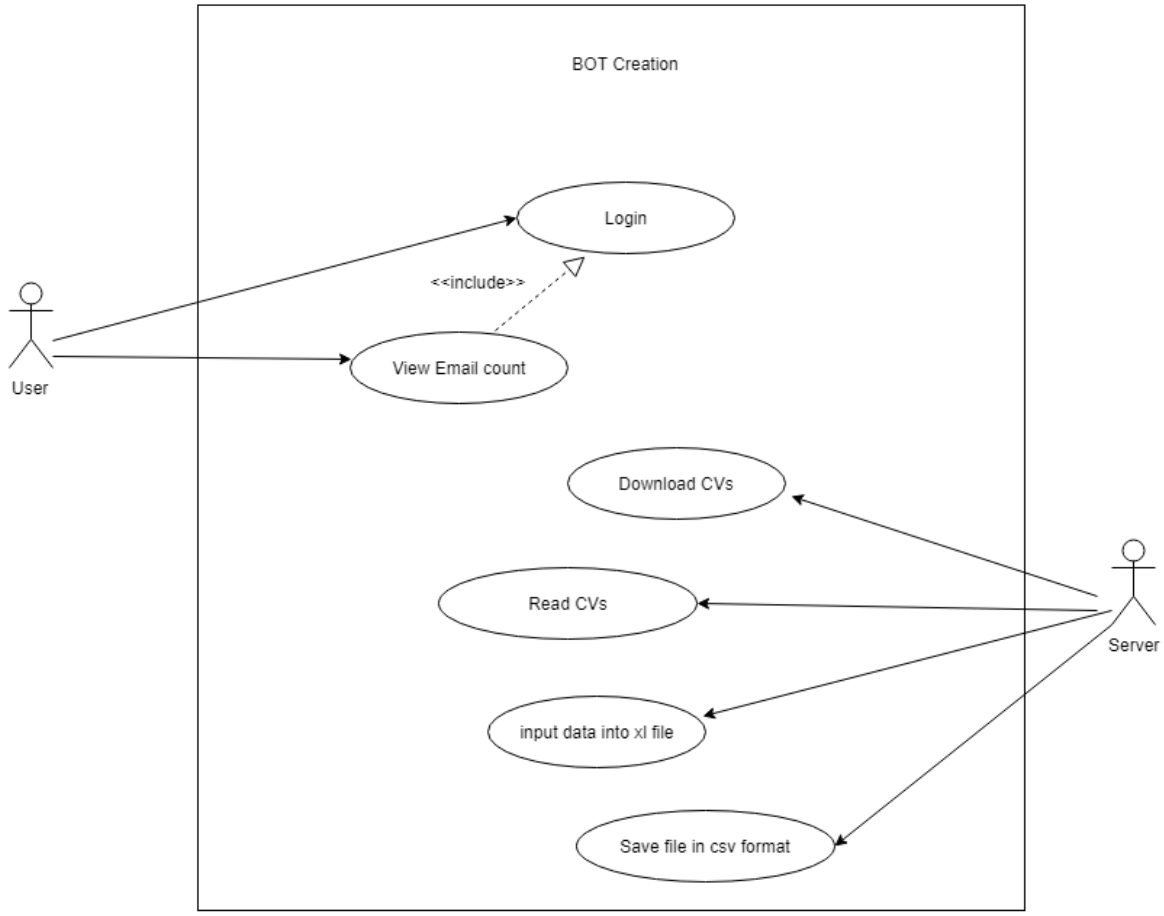


Figure 11: Use case Diagram

Login

Use Case No	01
Use Case	Login
Actors	User
Pre-Condition	User should be a registered user
Flow of Event	1. Enter the user name and password. 2. Click login
Post Conditions	Allow the user to use the web application
Alternatives	Display an error message to inform that invalid username or password are provided

Table 4- Use Case Scenario For Login

View Email

Use Case No	02
Use Case	View Email
Actors	User
Pre-Condition	User should be a registered user
Flow of Event	1. Enter the email address. 2. Click download
Post Conditions	Allow the user to view downloaded cv count

Table 5- Use Case Scenario For View Email

Download CV

Use Case No	03
Use Case	Download cv
Actors	Server
Pre-Condition	User should be a enter the email address
Flow of Event	1. Enter the email address. 2. Click download
Post Conditions	Allow the user to view downloaded cv count

Table 6- Use Case Scenario For Download CV

Read CV

Use Case No	04
Use Case	Read CV
Actors	Server
Pre-Condition	downloaded the cv
Flow of Event	1. Enter the email address. 2. Click download
Post Conditions	

Table 7- Use Case Scenario For Read CV

Write the details

Use Case No	05
Use Case	Write the details
Actors	Server
Pre-Condition	read the cv
Flow of Event	1. Enter the email address. 2. Click download
Post Conditions	

Table 8- Use Case Scenario For Write the details

Save files

Use Case No	06
Use Case	Save files
Actors	Server
Pre-Condition	Write the details
Flow of Event	1. Enter the email address. 2. Click download
Post Conditions	

Table 9- Use Case Scenario For Save Files

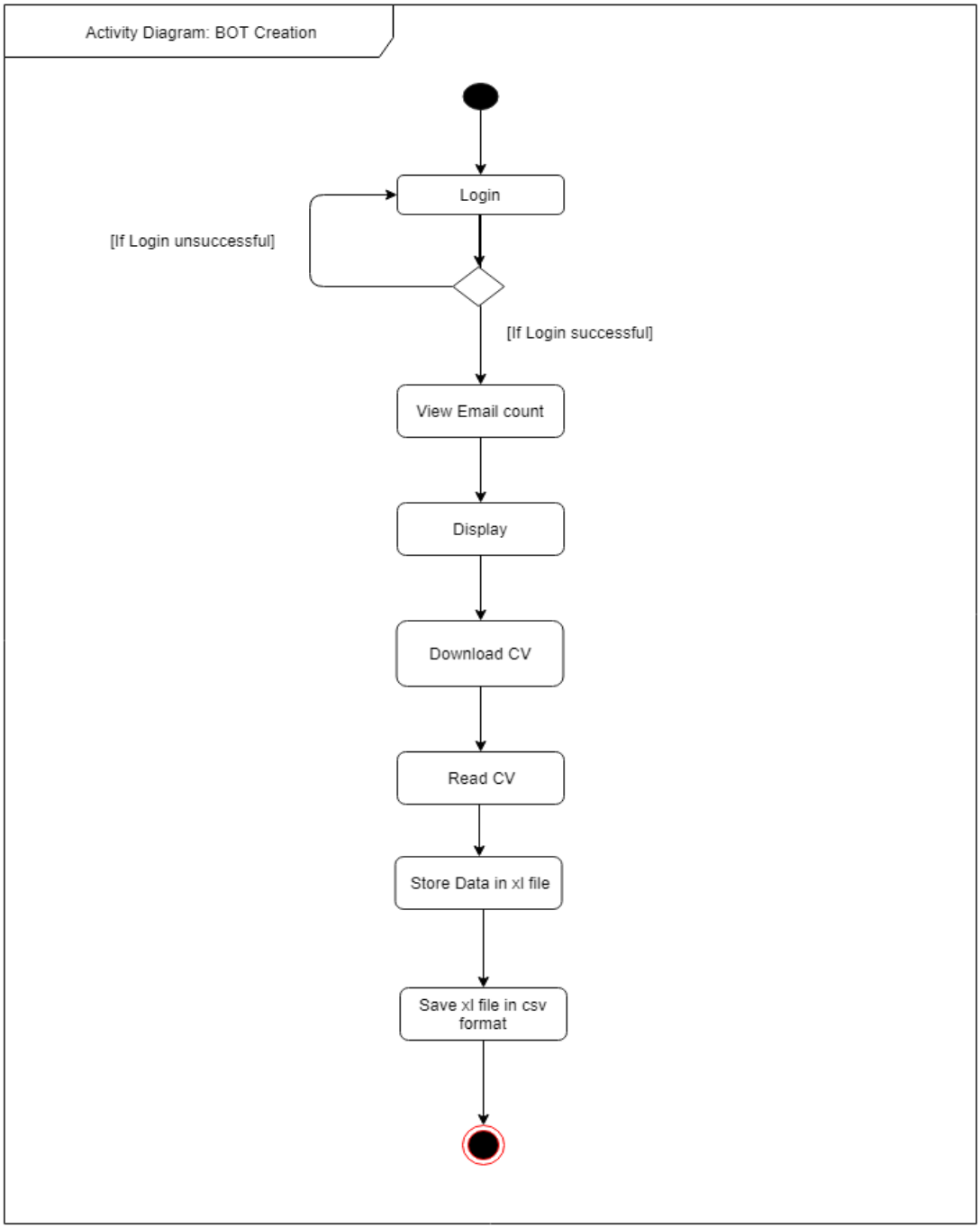


Figure 12: Activity Diagram

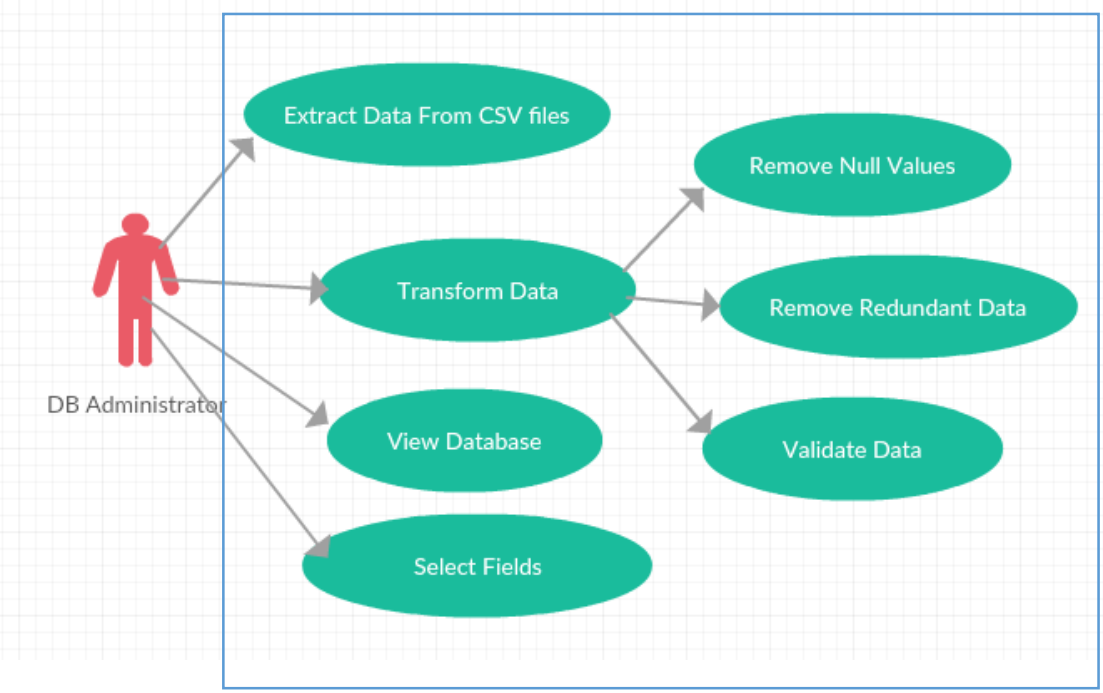


Figure 13 – Ues case diagram for ETL Tool

Use Case No	01
Use Case Name	Extract Data
Pre-Conditions	Database connection is active
Actors	Database Administrator
Main Success Scenario	1. Select the source csv file 2. Select or check the target database 3. Check the database table name 4. Click "Extract" button
Extensions	1a. Invalid Files 1b. Select Valid csv file 4a. Invalid source files 4b. Select a valid file and click "Extract"

Table 10- Use case scenario for extracting Data

Use Case No	02
Use Case Name	Select Fields
Pre-Conditions	Csv file is selected Database connection active
Actors	Database Administrator
Main Success Scenario	1. check necessary fields selected 2.Click “Next”
Extensions	1a. Select relevant field from dropdown 1b.Click “Next”

Table 11- Use case scenario for Selecting Fields

Use Case No	03
Use Case Name	View Database Table
Pre-Conditions	Database connection is active
Actors	Database Administrator
Main Success Scenario	1. User select the Database table 2. Click “Show Data” 3. System display Database table with data
Extensions	

Table 12- Use case scenario for viewing data

Use Case No	04
Use Case Name	Transform Data(Remove null values)
Pre-Conditions	Database connection active
Actors	Database administrator
Main Success Scenario	1. Select the database table 2.Select Field to Transform 3. select type as “Remove null values” from dropdown list 4.select the data range 5.Click “Next” 6.Specify the default value to be replace 7.Click “OK”
Extensions	7a. No data within the range 7b. specify a different range and Click “OK” 8a. user Click “Cancel” 8b.cancel the process.

Table 13- Use case scenario for Transforming Data by Removing Null values

Use Case No	05
Use Case Name	Transform Data – Remove Redundant Data
Pre-Conditions	Database Connection is active
Actors	Database administrator
Main Success Scenario	1.Select the Database table 2.Select type as “Remove redundant Data” 3.Click “OK”
Extensions	3a. No data within the given data range 3b. user click “Cancel” 3c.Cancel the process

Table 14-Use case scenario for Transforming Data by Removing Redundant Data

Use Case No	06
Use Case Name	Validating Fields
Pre-Conditions	Database Connection is active
Actors	Database Administrator
Main Success Scenario	1.Select the Field that want to validate 2. select type as “Validating” 3.Specify the Default value to be replace with 4.Click ok
Extensions	1a No data within the field 4a. user Click “Cancel” 4b. Cancel the validation process.

Table 15-Use case scenario for Transforming Data by Validating Data

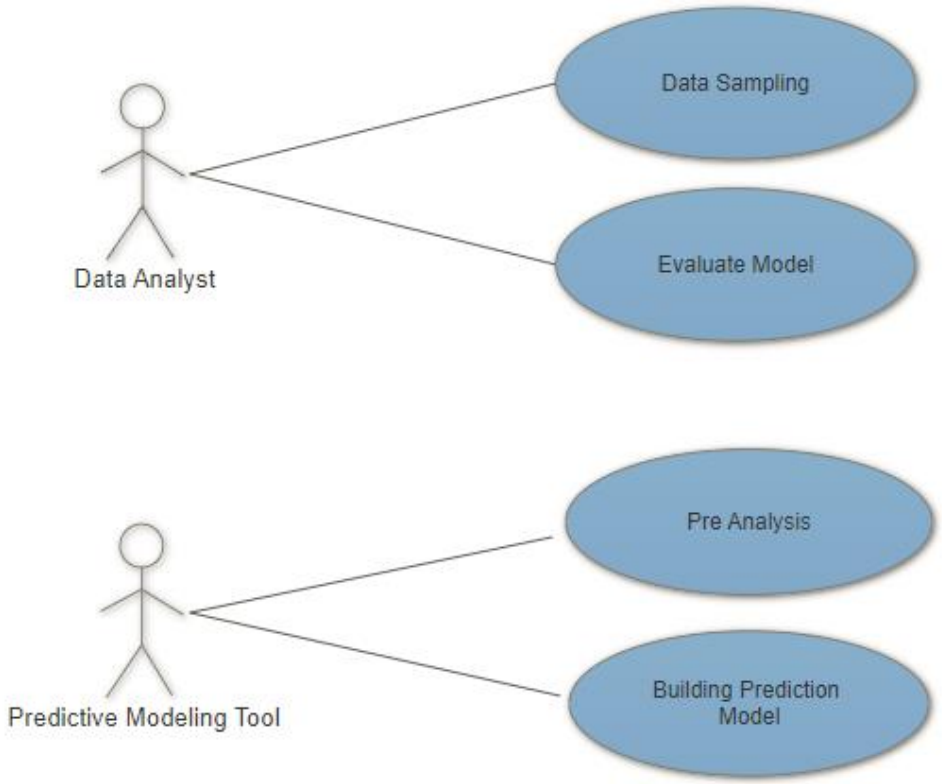


Figure 11- Usecase Diagram for predictive Model

Use Case No	01
Use case Name	Insert attributes and generate team
Pre –Condition	Database connection is active. System should be available.
Actor	Logged User (Top Level Management).
Main Success Scenarios	1. Insert necessary requirements 2. Click ‘Generate’ button
Extension	

Table 16- Insert user requirements

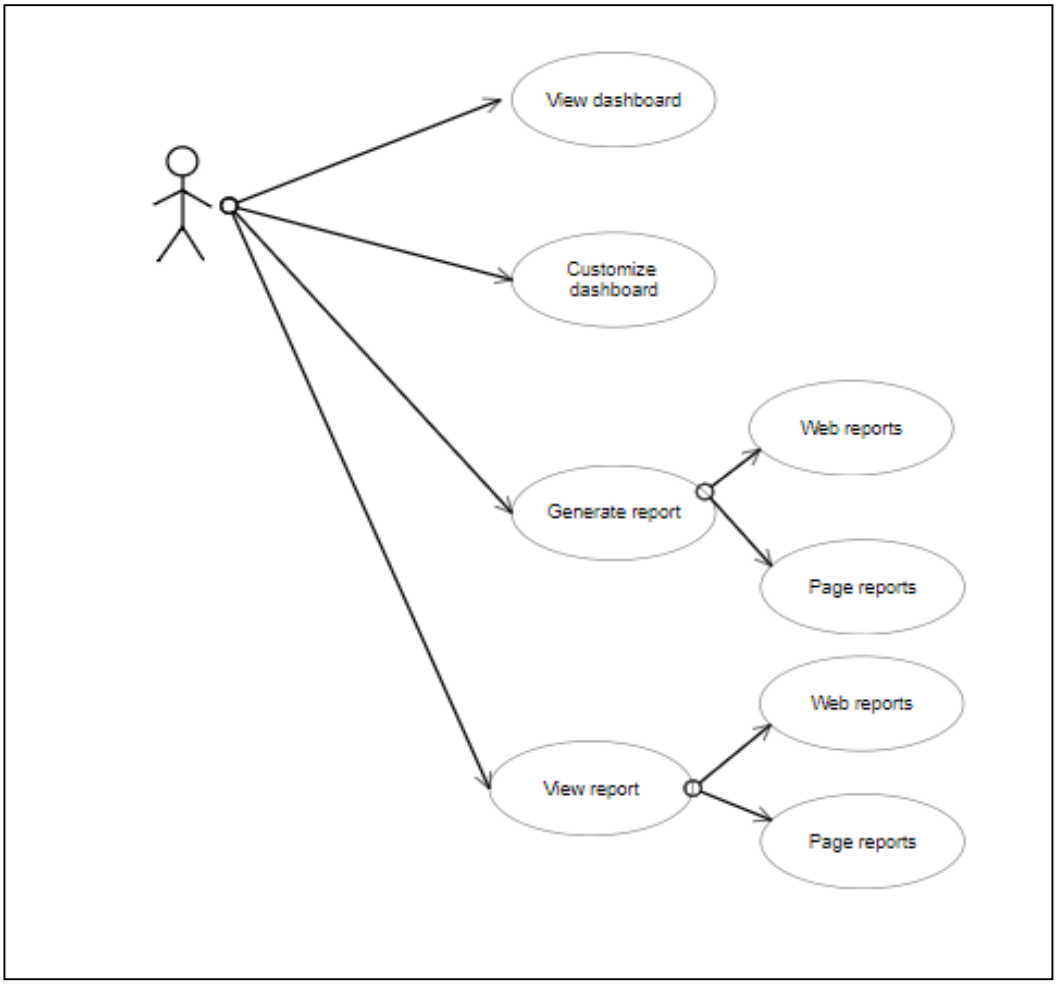


Figure 12- Use case Diagram - View dashboard

Use Case No	01
Use case name	View evaluated statistical details.
Pre – condition	Database connection is active. Web application should be running correctly
Actors	User
Main success scenario	1. Navigate to application 2. Click on dashboard 3. Click Statistics
Post – condition	View all the metrics calculated
Extensions	

Table 17- View evaluated statistical details

Use case No	02
Use case name	View candidates analysis details
Pre – condition	Database connection is active. Web application should be running correctly
Actor	User
Main success scenario	4. Navigate to application 5. Click on dashboard 6. Click selected candidate details
Post – condition	View a table of all the attributes with details of each and every candidates get selected
Extensions	

Table 18- View candidates analysis details

2.3 User characteristics

There is only a desktop application user in here. This system is mainly focused for HR staff. In there also only the authorized parties such as HR Manager have permission to access login the system. This application is mainly developed for the Admin.

Administrators

Administrators must have some knowledge to handle the component. Basically, administrators of the system must have knowledge about the desktop applications and computer handling.

User	Privileges	Activities
Database Administrator	Full Access to the Database system	• Extracting data from CSV files. • Transform Data into normalize format • Load the Data into Data warehouse

Table 19: User DB Admin

User	Privilege	Activities
Decision Maker (Top Level Management)	Full access to the system. Mainly focuses on Visualization layer /Dashboard	1. Get the Optimal or Feasible Solution 2. Predictive results with dashboard

Table 20: User Top Level Management

User must not be very techy person and not an intelligence professional about business intelligence. But user must have an idea about Database Management.

This system can be used by several users such as

- Database Administrators
- Top Level Management
- HR Managers

2.4 Constraints

- User login and the user authentication is must for both applications
- Server should be able to handle multiple requests at a time

- The response should be generated within minimum time constraint
- The server should be able to perform its normal operations without exceeding 2GB of RAM
- Having basic English knowledge also will be a constraint for the users
- Time for loading the web pages containing images should take no longer than 10 seconds.
- Major constraint will be limitation of available time. ETL Tool expected to complete within 5 month of time period
- All the tools and technologies should be open source.

2.5 Assumptions and dependencies

- No data will be lost/corrupted during the communication between server and desktop application
- Users can not use application without internet connection
- All the users have basic knowledge of using computer and internet.
- There is an active internet connection
- Servers up and running 24x7 hours
- Sufficient memory and processing powers in all computes.

2.6 Apportioning of requirements

Essential Requirements

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

Desirable Requirements

- Allow user mobile phone

Example: The requirements described in sections 1 and 2 of this document are referred to as primary specifications; those in section 3 are referred to as requirements (or functional) specifications. The two levels of requirements are intended to be consistent. Inconsistencies are to be logged as defects. In the event that a requirement is stated within both primary and functional specifications, the application will be built from functional specification since it is more detailed.

3. SPECIFIC REQUIREMENTS

3.1 External interface requirements

3.1.1 User interfaces

Name of item	Login Profile
Description of purpose	Maintain the user login
Source of input or destination of output	Basic user details
Valid range, accuracy and/or tolerance	100%
Units of measure	-
Timing	-
Window formats/organization	-
Relationships to other inputs/outputs	

Table 21: Login Profile User Interface

Name of item	View cv count
Description of purpose	Enter the email address to download cv
Source of input or destination of output	Basic get the cv details
Valid range, accuracy and/or tolerance	100%
Units of measure	-

Timing	-
Window formats/organization	-
Relationships to other inputs/outputs	Converted cv details to csv format

Table 22: View cv count User Interface

- **Data Extraction Interface:**

  This is the first step of extracting data. Data shroud be loaded from CSV files. So data should store in CSV files in order to extract data to the data warehouse. Loaded Database and the table can be seen by user and if they want they can provide Database name in necessary time. [6]

- **Field Selection Interface**

  In this interface user can see what the attributes map into data warehouse are and what are the selected table headings of database table. The can select from dropdown list if there are any changes to apply. Left side list have extracted attributes and right side has mapping attributes in the database table. User should click next to proceed to the next step.

- **View Data Interface**

  In this interface user can see what are the available database tables and its data. On top of the interface user can see last modified data and table name. If they want further transformation they can transform from there.

- **Data Transformation Interface**

  This interface uses to transform data if user want to transform. When mapping normally cleaning data is done, but if user want further transformation user can transform by selecting relevant transform type from the dropdown box and also giving relevant attribute and data ranges that they want to transform.

- **User Requirements Gathering Interface**

  This is where the user inputs the user requirements in order to generate the most qualified candidate CV list. This should be filled by the responsible parties of shortlisting CVs like top

level management. They have to enter the below details in order to get the best output as they want.

Job Role, Mainly Required Qualifications, Expected Number of Candidates, expected number of Male candidates, expected number of female candidates, Age Limit and Finally, they have to select whether they want either the Optimal solution or Feasible solution or both.

And then click the 'Generate' button.

- **Visualization Interface**

  This is the interface where all the visualization is done. This will display graphical graphs to indicate a number of candidates, who have selected and who have not selected. After selecting attributes and click on "Generate" button the best candidates for the job will be displayed in the dashboard as a table, with relevant details. The user can click on particular candidate's name and view analyzed statistical details of that candidate.

3.1.2 Hardware interfaces

- Development team must have at least a desktop/ laptop with minimum 1.8GHz processing power and 256GB hard drive space.

3.1.3 Software interfaces

Name	Version	Purpose
Java	8	Main developing environment
php	7.2	Run python scripts and web application development
Python	3.6.5	Develop dashboard functionalities
Tomcat	8.5.30	Development server
AngularJS	1.3	Used for the web application front end development

Web Browser	Chrome / Firefox	Used for the web application development and testing

Table 23: Software Interfaces

- **MySQL**

  MySQL is used for database management system for the system. For data transformation and data mapping use MySQL. Also to create data marts.

- **Jython Library**

  Use to connect interfaces between java and python. As all interfaces and validations develop by java and use python for calculations and algorithm developments.

- **Apache Server:**

  As the visualization dashboard component of the system is planned to be developed using PHP, Apache server will be used as the web server. Dashboard and reports will be available through the web application, which is developed on top of apache server.

3.2 Performance requirements

- CV should be downloaded the minimum time.
- System should read cv in minimum time.
- System should display error messages in less than 1 minute
- System should display the downloaded cv count in minimum time.
- Database can handle 50000 records.
- Based on the derived predictive lists, reports must be generated within 10 seconds.

3.3 Design Constraints

In designing this system, we have to consider about desktop application designing. Administrators of a particular company are the primary users of the desktop application. Other than that, there are no design constraints.

- The system based on IT industry only.
- Limitation of available time to develop the system.
- Performance, accuracy, reliability, security should achieve.

3.4 Software system attributes

In this section, we are explaining the attribute that we are going to offer through the system. There are some explanations in follow.

3.4.1 Reliability

Reliability of a system is the ability to perform its normal operations with minimum failures over a minimum time in a given environment.

- The required information by the users should be provide within minimum time without any failure
- Data must not be corrupted

Reliability is the probability that an application will accurately perform its task under stated environmental conditions. It means that system support to allow user to work normally in many environmental conditions. Application must be tested unit vise and integrated vise in order to eliminate errors and give user a proper reliability. All technologies and configuration is used to provide better reliability to the system.

Reliability is the ability of the predictive model run with a minimum number of failures. The predictive model has to go through a testing i.e. application must test by fixing each possible bug. Each sub-component of the model evaluation and dashboard simulation component will be tested and finally integrated system also be tested and to make sure the desired output is obtained. Predictive model output also has to be tested to make sure the output is meaningful and optimal. Since the model evaluation and dashboard simulation component is a component to give the optimal and feasible prediction result. Hence, the reliability of the component is much expected.

3.4.2 Availability

- Server must be available 98% of the time

Database servers are running 24x7 hours. So that ETL tool can be performed in every time. [5] Immediate server's aggregation to data warehouse can increase high availability

Availability is a key performance measure in the predictive analysis. Model evaluation and dashboard simulation components are in working condition at a given time and that should be able to deliver the required services for intended users that range from top –level management to other decision makers. All and necessary information for user requested can be viewed and available at any given time.

3.4.3 Security

- Information such as password will not be saved as clear text
- Only authorized users are given permission to access the android application
- It indicates how to protect application from unauthorized people. Database has usernames and passwords in order to grant security. Only authorized people can access to the database and work with ETL Tool.
- All critical data is secured during storage and transmission through proper access controls. Along with this, expected to implement VPN for Visualization tier or this can be high security can be achieved through authenticating and authorizing users.

3.4.4 Maintainability

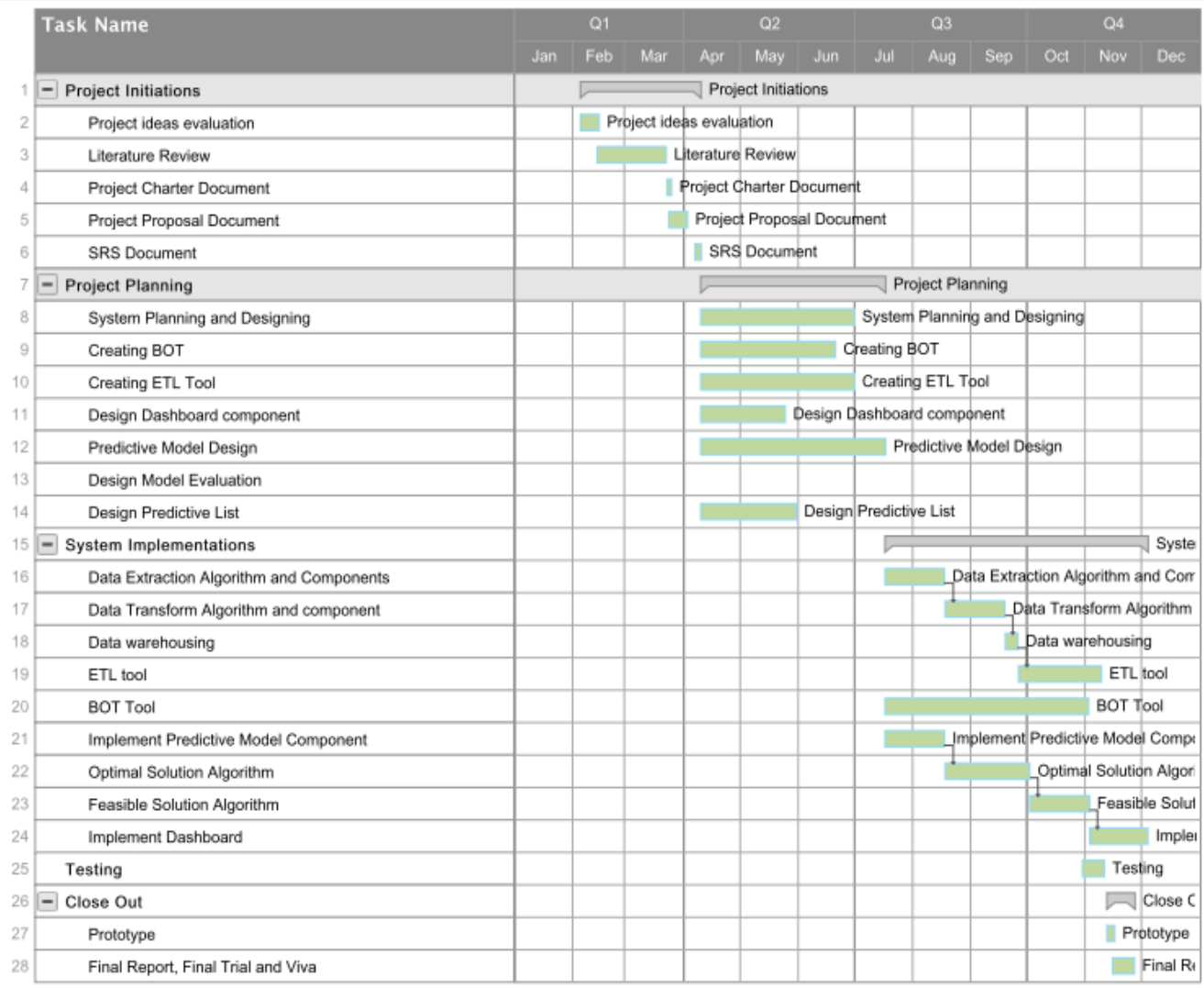
- Proper code commenting and indentation will be implemented
- Proper coding standards and naming conventions will be used at development time.
- Each attributes are validated using proper transformation functions.
- Creating user friendly interfaces with some standards.

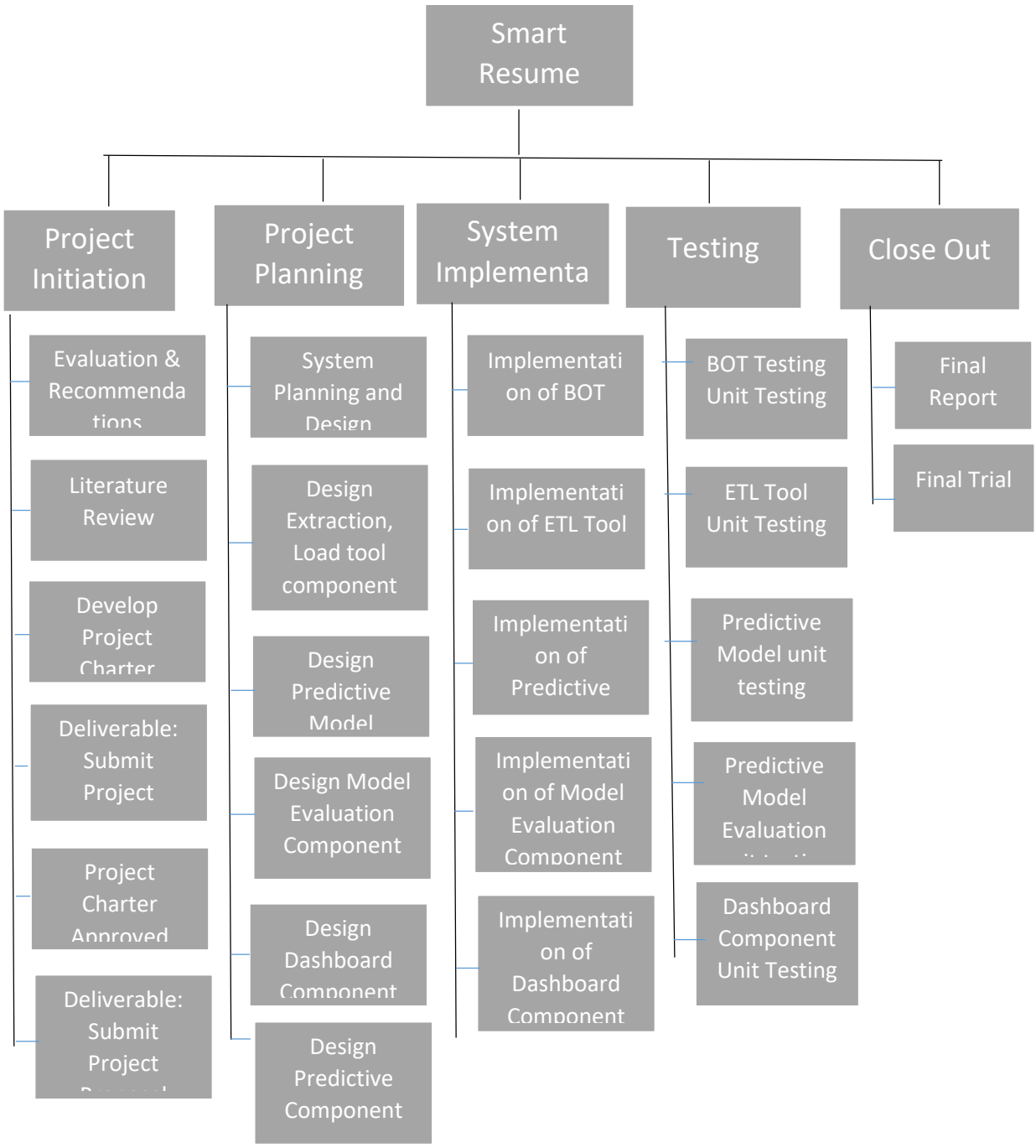
3.6 Other Requirements

- Data that Map into Database are acquire from valid sources for further processing
- Use open source technologies.

4. SUPPORTING INFORMATION

4.1 Appendices




Smart Resume
Project Initiation
Project Planning
System Implementa
Testing
Close Out
Evaluation & Recommendations
Literature Review
Develop Project Charter
Deliverable: Submit Project
Project Charter Approved
Deliverable: Submit Project
System Planning and Design
Design Extraction, Load tool component
Design Predictive Model
Design Model Evaluation Component
Design Dashboard Component
Design Predictive Component
Implementation of BOT
Implementation of ETL Tool
Implementation of Predictive
Implementation of Model Evaluation Component
Implementation of Dashboard Component
BOT Testing Unit Testing
ETL Tool Unit Testing
Predictive Model unit testing
Predictive Model Evaluation
Dashboard Component Unit Testing
Final Report
Final Trial

4.2 References

[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.

[1] Ephraim Turban, Decision Support and Business Intelligence Systems, 10th edition, 2014, pp.6-15.

[2] [Online]. Available: http://www.bitool.net/software/mysql.html

[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]R. Mukherjee, "A Comparative Review Of Data Warehousing ETL Tools With New Trends And Industry Insight," *2017 IEEE 7th Int. Adv. Comput. Conf.*, pp. 944–949, 2017.

[5]A. Sabtu *et al.*, "The challenges of Extract, Transform and Loading (ETL) system implementation for near real-time environment," *Int. Conf. Res. Innov. Inf. Syst. ICRIIS*, pp. 3–7, 2017.

[6]P. Figueiras, R. Costa, G. Guerreiro, H. Antunes, A. Rosa, and R. Jardim-Goncalves, "User interface support for a big ETL data processing pipeline an application scenario on highway toll charging models," *2017 Int. Conf. Eng. Technol. Innov.*, pp. 1437–1444, 2017.

[7] [Online]. Available: https://docs.oracle.com/cd/B13789_01/datamine.101/b10698/3predict.htm[Accessed: April 29, 2018]