SMART RESUME
BI TOOL TO SHORTLIST CVs FOR A JOB VACANCY

Yureshani H.B.D

(IT15040404)

Bachelor of Science Special (Honors) Degree in Information Technology

Department of Software Engineering

Sri Lanka Institute Of Information Technology

Sri Lanka

September 2018

SMART RESUME -
BI TOOL TO SHORTLIST CVs FOR A JOB VACANCY

Yureshaani H.B.D

IT15040404

Final Report

The dissertation was submitted in partial fulfillment of the requirements for the B.Sc. Special (Honors) Degree in Information Technology.

Department of Software Engineering

Sri Lanka Institute Of Information Technology
Sri Lanka

September 2018

DECLARATION

I declare that this is my own work and this dissertation1 does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology the nonexclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The supervisor/s should certify the dissertation with the following declaration.

The above candidate has carried out research for the B.Sc Dissertation under my supervision.

Signature of the 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 have a collection of related CVs, and have the ability of generating the list of most qualified applicants among them, would be helpful. 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 about the applicant. There is a shortage of tools that support in selecting the best qualified set of employees to an employer. Smart Resume is a business intelligence tool for IT sector, which analyse and classify operational data with classification algorithms to present complex and competitive information to decision makers, in order to dynamically fulfil the business needs. It is built to satisfy the task of generating the list of most suitable candidates. 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.

ACKNOWLEDGEMENT

The work described in this document was carried out as our 4th-year research project for the subject Comprehensive Design Analysis Project. This becomes a success due to the result of combining all the hard work of the group members and the encouragement, support and guidance given by many others. Therefore, it is our duty to express our gratitude to all who gave us the support to complete this major task.

We are deeply indebted to our supervisor Mr.Lakmal Rupasinghe and all the other lecturers of Sri Lanka Institute of Information Technology who gave us their valuable suggestions and time to make this a success. And also their constant encouragement and support in the development of this research, was become a great help on the way to achieve this. We are also extremely grateful to Mr. Jayantha Amararachchi, Senior Lecturer/ Head-SLIIT Centre for Research who gave and confirmed the permission to carry out this research and for all the encouragement and guidance given.

We also would like to thank all our friend and colleagues for all their help, support, interest and valuable advices. Finally, we would like to thank all others whose names are not listed particularly but have given their support in many ways and encouraged us to make this a success.

TABLE OF CONTENT

Contents

DECLARATION i
ABSTRACT ii
ACKNOWLEDGEMENT iii
TABLE OF CONTENT iv
LIST OF TABLES v
LIST OF FIGURES vi
1 INTRODUCTION 1
1.1 Problem to be Addressed 1
1.2 Background Context 1
1.3 Research Gap 2
1.4 Research Questions 3
2 BODY 3
2.1 Literature Review 3
2.2 Methodology 3
2.2.1 Support Vector Machine 4
2.3 Research Findings 6
2.3.1 Tools 6
2.3.2 Technologies 6
2.4 Testing 6
3 RESULTS AND DISCUSSION 10
3.1 Results and Evidence 10
3.2 Discussion 10
4 CONCLUSION 11
5 REFERENCES 12

LIST OF TABLES

Table 1 Comparison with existing systems 3
Table 2 Test case 01 8
Table 3 Test case 02 8
Table 4 Test case 03 9

LIST OF FIGURES

Figure 1 SVM Model..5

1 INTRODUCTION

1.1 Problem to be Addressed

Data in business is very useful only if data is analyzed properly which will aid in strategic business decision making. In IT industry, workforce management is critical since there is huge workflow in the industry. Hiring individuals makes 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.

1.2 Background Context

With the globalization, Information and Communication Technology is playing a major role in economy of Sri Lanka today. IT industry has become wider and there is a large workforce in the sector. There are many companies under this industry and most of them are not second to each other. There is a large workforce working under each company and also there are many well qualified candidates who are willing to work in them. Therefore, selecting the right team for a project by satisfying time, Budget, Scope, and Quality goals can change the outcome of a project. Also It is not a secret that, there should be a great workforce or a dedicated and very well qualified team of employees in a company, in order to make the business success. Also it is a known fact that when there are large businesses, many other supportive businesses are starting around them.

In any industry, employee hiring process is common and traditional in Sri Lanka. The normal process is advertising the vacancy, calling Curriculum Vitae (CV), short listing them by referring the CVs and interviewing the short listed candidates. Every employer is looking for the best employee. In order to make this easy there are third party companies who combines the employer and employees by referring and matching the requirements of employer and

qualifications of the employee. These third party companies also should give a great effort to match the best suite candidates with the employers.

Hence, considering all these problems, there is a need to develop a comprehensive tool which in advance provides the valuable, fully data-driven and fact-based information and insight to the top level management in the companies so that they can make efficient, accurate and result-oriented strategic decision.

This Business Intelligence tool makes the prediction efficient by carrying out the statistical analysis and data mining technique to calculate the probability of how each candidate suits for a company requirement. Even though there are various kinds of business problems faced by ICT industry selecting the right employee set to make the company growth a success is majorly focused. This is because, currently employees are selected by a manual process, which takes a lot of time and effort since it is complex when a human judge another person ‘s skills. Therefore, there is a system to store all the knowledge in order to select the best candidate team for a company according to its requirements.

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 systems

1.4 Research Questions

2 BODY

2.1 Literature Review

In order to make our effort a success, we have done a literature survey on the specified sections. Here I have presented the survey based on predictive model building.

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]

2.2 Methodology

This section includes detailed descriptions of the techniques and mechanism employed to make the predictive model of Smart Resume reality. The descriptions include how software implementation of the 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.

When it comes to devising the best model for predicting the best team for IT industry, there is a need for identifying the most influencing attributes that drive an employee is suitable or not. For this, Smart Resume's prediction model has undergone lots of inspection and interviews with industry experts in the IT sectors which gave the clear and vivid idea on how the model should approach and devise. There can be hundreds of attributes of an employee and a job role that are required by an IT company. Not each and every attribute are responsible for identifying the suitability of an employee.

It has been found that only a few attributes are responsible for identifying the employee and his or her skill set for a job role. The identification of these has been done based on research, expert analysis and few statistical modeling techniques. The predictive model has followed the best, optimal and feasible approach in analyzing the most influencing attributes for statistical modeling.

When it comes to predicting the outcomes of any classification problem, either it can be specified prediction or the default prediction, correction manipulation of data is most important. Data can be in various formats with different impurity level. Hence, we focused our approach on developing best possible and efficient data extraction and transformation techniques. The other key findings we got is when we are cleansing the data, we shouldn‘t lose any valuable data. It is because data is most important for any company. So BI engineers in Sri Lanka based companies were consulted on how we can come up accurate and highly reliable ETL tool.

After data is loaded to ETL, it cleanses, transforms and process the telco data that are now ready to feed into the models. Models are trained by the imputed data. Once the models are trained, then real data set are fed to the prediction model using ETL tool again and then the result are predicted. Generally, training happens only one time if we are dealing with the same set of data. If the data is different, the model can be trained anytime with a diversely new set of data.

Predictive model is made with some predictors, and those predictors are variable factors which can make a change in behavior or result. As the first step the data is sampled by using proper sampling mechanism. As examples we can use randomized sampling or probability sampling. Here cluster- probability sampling is used. By using them system will produce Test Set, Training set and Holdout set.

Next suitable classification algorithm is being used to generate the most qualified candidate list. Support vector Machine has been chosen as the main classification algorithm to our approach.

2.2.1 Support Vector Machine

In SVM algorithm, the data will be mapped into higher dimensional input space and build a hyper plane in this space. The built hyper plane will divide the input data into classes according

to their behavior, by provide an optimal separation. In this algorithm, we can decide the correct class of an input data by evaluating the sign of,

$y(x) = wT\varphi(x) + b$ (1)

If $y(x) > 0$ we assign to class +1 and if $y(x) < 0$, we assign it to class -1. Here Ø (x) is a feature-space transformation, which can map to a space of higher, possibly infinite, dimensions. This is more effective in higher dimensional spaces. Below figure, illustrates the behavior of SVM algorithm.

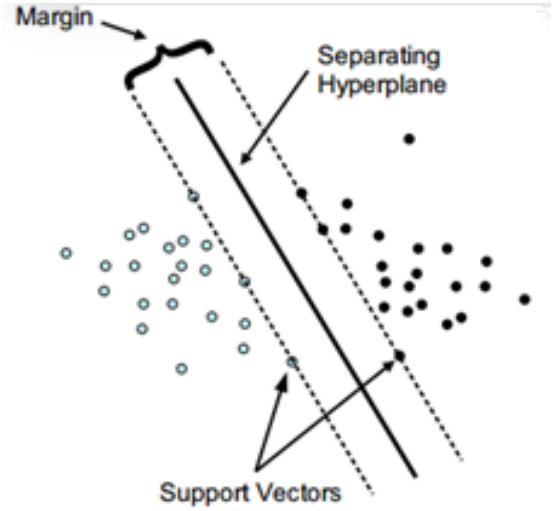

Figure 1 SVM Model

So in building the predictive model smart Resume will follow the following steps;

- Hypothesis Testing
- Data Sampling
- Algorithm building; Classification, Association

This will generate a team for a given job role of a given company depending on whether it is optimal or feasible. This optimality or feasibility will be selected during the prediction of the predictive model. With the help of this predictive model, the business performance can be shown as how it has been in the past, present and what will happen in the future, thus business predictions can be made wisely and easily.

2.3 Research Findings

Final Smart Resume system has provided a complete tool to classify and short list CV for IT industry. There in Predictive model building I have found a method to select the most suitable candidate list to a specific job role in a company.

In Predictive Model Builder and Visualizer Component for IT industry, in order to build the models, we have used scikit-learn a machine learning library in Python.

2.3.1 Tools

- Pycharm
- PHPMyAdmin
- MySQL Workbench
- Spyder

2.3.2 Technologies

- Python 3.6
- MySQL 8.0
- PHP 5.6(Laravel Framework 5.2)
- Jython
- HTML5

2.4 Testing

Software testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test. Software testing can also provide an objective, independent view of the software to allow the business to appreciate and understand the risks of software implementation. Test techniques include, but are not limited to the process of executing a program or application with the intent of finding software bugs (errors or other defects).

It involves the execution of a software component or system to evaluate one or more properties of interest.

Unit Testing

Unit testing, also known as a component or module testing, refers to tests that verify the functionality of a specific section of code, usually at the function level. We are planning to carry out testing individual modules at the latter stages of the research project.

Integration Testing

Integration testing works to expose defects in the interfaces and interaction between integrated components (modules). Progressively larger groups of tested software components corresponding to elements of the architectural design are integrated and tested until the software works as a system.

System Testing

System testing, or end-to-end testing, tests a completely integrated system to verify that it meets its requirements. We intend to perform a system testing to ensure that we have achieved all the objectives of our research up to the level of performance expected.

Test Cases for Predictive model building

Test Case ID	TC1
Use case description	Validate valid User Credentials
Pre –Condition	Login interface is loaded
Test Procedure	Type username and password Click login button
Test Input	User name = admin Password = admin123

Expected Output	User should be able to successfully log in to the system
Actual Output	User logs into the system

Table 2 Test case 01

Test Case ID	TC2
Use case description	Validate invalid User Credentials
Pre –Condition	Login interface is loaded
Test Procedure	Type username and password Click login button
Test Input	User name = abc Password = abc1
Expected Output	User should not be able to log in to the system and error message should be displayed
Actual Output	User cannot log in to the system and error message is displaying

Table 3 Test case 02

Test Case ID	TC3
Use case description	Insert job role requirements to the system
Pre –Condition	Requirement gathering page should be displayed
Test Procedure	Fill the form Click Submit button
Test Input	All the fields should be given a value

Expected Output	The data should be submitted
Actual Output	Data is submitting

Table 4 Test case 03

3 RESULTS AND DISCUSSION

3.1 Results and Evidence

Smart Resume takes data from different sources in different formats like from PDF format and CSV format. Those data can be unorganized and dirty. Hence Smart Resume has to take care of the Data Cleansing and Transformation. After that Smart Resume will make the predictions using the predictive models and the cleansed data saved in the db. Then the results will be saved again to the db. The visualizer will present those data using dashboard so that the non-technical users can easily understand the results.

3.2 Discussion

Smart Resume is a business intelligence tool that has been developed in order to fill the gap of a missing part from research world as well as in the business world. It consists with tools and technologies that have not been addressed so far in any other business intelligence tool. Smart Resume is one complete system with cooperated a combination of desktop application and web application.

4 CONCLUSION

Today for most companies like IT, receive huge number of CVs for a vacancy as there are lot of graduates coming out from a university within a year. The quality of the company depends with the capabilities of the recruiters. Therefore, getting the best qualified people from incoming applications is very difficult for them. Currently the selections of CVs is done manually. As the huge number of CVs are summarized by manually, qualified CVs can be skipped by CV errors and human errors. That is where Smart Resume come in to play by reducing all the difficult tasks of selecting optimal set of candidates in place of referring CVs. Smart Resume contains a Desktop application with ETL (Extract, Transform, Load) Tool to transform data into a meaningful order and Smart Resume has a Dashboard as a Web application in order to predict best candidates for a given job opportunity by analyzing the CV data. As the initial step Smart Resume analyze and extract the professional skills, personal skills, personal details and etc. from the CV according to the job vacancy. Finally, Smart Resume display the results in an attractive proper manner in a dashboard with simple graphs and charts which shows the user friendliness of the system.

5 REFERENCES

[1]Yang, J., Dai, C. and Ding, Z. (2017). *A Scheme of Terminal Mobility Prediction of Ultra Dense Network Based on SVM*. 1st ed. [ebook] Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8078755 [Accessed 3 Aug. 2018].

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