CV ANALYSIS AND OPTIMIZING THE RECRUITMENT PROCESSIN THE IT INDUSTRY USING MACHINE LEARNING TECHNIQUES

Project ID: 2023-098

Project Proposal Report

De Silva S.R. – IT20216900

Bachelor of Science (Hons) Degree in Information Technology

Specializing in Data Science

Department of Information Technology

Faculty of Computing

Sri Lanka Institute of Information Technology

Sri Lanka

March 2023

CV ANALYSIS AND OPTIMIZING THE RECRUITMENT PROCESSIN THE IT INDUSTRY USING MACHINE LEARNING TECHNIQUES

Project ID: 2023-098

Project Proposal Report

Bachelor of Science (Hons) Degree in Information Technology

Specializing in Data Science

Department of Information Technology

Faculty of Computing

Sri Lanka Institute of Information Technology

Sri Lanka

March 2023

DECLARATION

I 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.

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).

Name	**Student ID**	**Signature**
De Silva S.R.	IT20216900	

The supervisor/s should certify the proposal report with the following declaration.

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

……..

Signature of the Supervisor:
(Dr. Anuradha Karunasena)

07/04/2023

…….

Date:

ABSTRACT

Choosing the right individuals is vital to an organization's success and expansion. However, the traditional recruitment approach, which involves manual procedures like screening resumes and academic qualifications, assessing technical and professional skills, is not only time-consuming but also ineffective. To meet the demands of employers, it is essential to adopt an efficient and reliable approach to assess the skills and abilities of candidates. To meet the demands of employers, it is essential to adopt an efficient and reliable approach to assess the skills and abilities of candidates. Our proposed solution aims to optimize the hiring process by implementing a system for the IT industry that can effectively identify the most suitable candidate/s for a given job role. To achieve this goal, this research component will analyze the data present in a candidate's academic transcript. This analysis will extract important features such as module names and grades, which will then be categorized into different types of skills using a graphical method. By doing this, we will be able to identify which candidate possesses the necessary technical skills required for the job role, thus making the recruitment process more efficient and beneficial for the company.

To accomplish this, we will employ Natural Language Processing and machine learning techniques to perform the necessary processing steps. This will enable our system to accurately evaluate a candidate's skills and abilities, providing employers with a more thorough understanding of their potential employees. Overall, our proposed solution offers a promising approach to enhancing the recruitment process and identifying the most effective candidates for a given job role.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

Abbreviation	Description
NLP	Natural Language Processing
CV	Curriculum Vitae
PDF	Portable Document Format
VPN	Virtual Private Network
CDN	Content Delivery Network
DNN	Deep Neural Network
CNN	Convolutional Neural Network

Table 1 : List of Abbreviations

INTRODUCTION

Background

The recruitment process in Sri Lanka typically goes through manual steps such as screening resumes, shortlisting candidates and conducting interviews and performing background checks. However, despite the vast evolving technology and advancements in the country, many companies still rely on the manual process in recruitment methods.

The IT industry is rapidly changing, and there is a great demand for fresh employees. Traditional recruiting processes, which include manual resume screening, phone screenings, and in-person interviews, may be time-consuming and resource-intensive, making it difficult for businesses to keep up with the industry's pace. Furthermore, there is a dearth of standardized techniques for assessing crucial areas such as technical skills and personality qualities, making it difficult to select the most suitable people for a certain position. As a result, a more streamlined and systematic approach to recruiting is required to account for these issues and enable more efficient and successful hiring methods.

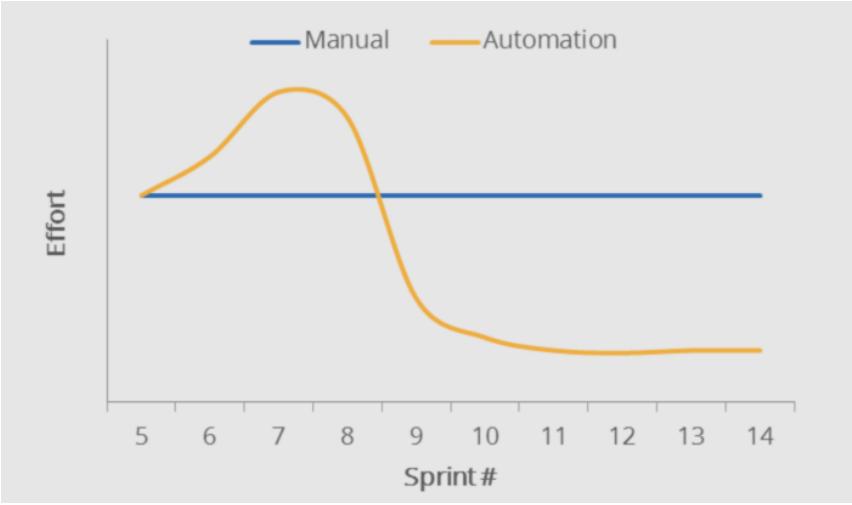
Figure 1: Time taken for the hiring process.

[1]

According to a recent ICTA study, the IT industry in Sri Lanka is developing at a 16% annual pace and is predicted to provide over 200,000 direct and indirect job possibilities by 2022. [2] Its expansion may be credited to the country's pool of highly skilled IT workers, who have obtained top-tier education and training and are competitive in the global IT market.

However, there are difficulties in satisfying the demand for IT employment and identifying the finest people for these positions. Despite the country's large talent pool, competent professionals are in limited supply, and competition for top talent is tough. Businesses may struggle to find the best individuals for these positions, and there may be a skills gap between what employers want and what candidates can provide.

Several research have been undertaken in order to enhance and streamline the recruiting process. The purpose of these research is to explore the elements that are taken into account when hiring new workers and to develop a completely automated system that can successfully evaluate candidates based on these characteristics in order to discover and recruit the best candidate for a job opening.



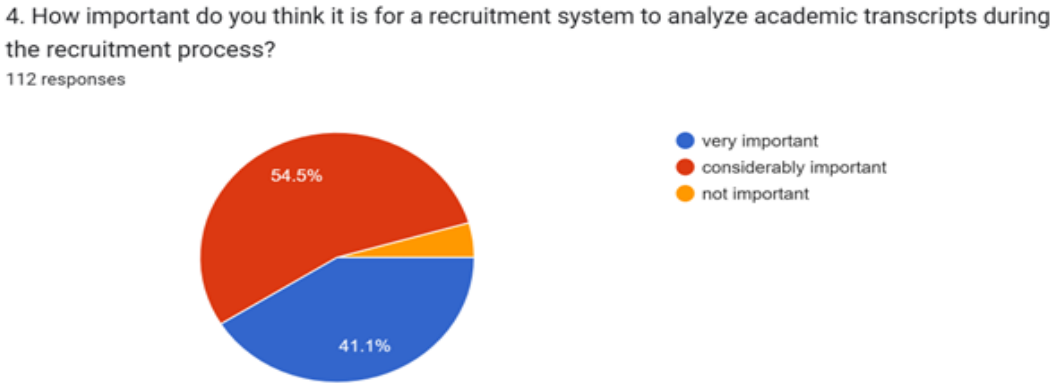
[3]

Figure 2: How automating the recruitment process reduce the effort

Component Overview

This research component specifically targets a particular university and aims to extract and analyze relevant information from academic transcripts and module outlines, which is then used to evaluate the suitability of fresh graduates for specific job roles. What sets this approach apart from others is that it analyzes the transcript and module outline together as a whole, rather than separately. By doing so, this component can accurately identify patterns and trends in a student's performance and provide valuable insights into their strengths and weaknesses.

Additionally, this component focuses on identifying the highest skilled areas of fresh graduates from the university. It helps organizations to make informed hiring decisions by providing a thorough and accurate picture of the candidate's academic background. By understanding a candidate's skills and knowledge, the organization can match them with the right job roles, increasing the chances of a successful hire. This, in turn, leads to higher productivity and efficiency, which is beneficial for the organization.



Overall, this academic transcript analysis component is an essential part of an online recruitment system. It helps organizations to simplify the hiring process, ensure accuracy and relevance of the analysis, and find the right fit for their job openings. It provides a competitive edge for the organization by increasing the likelihood of finding the right candidate and simplifying the hiring process, especially in IT companies in Sri Lanka.

Literature survey

The success and growth of an organization depend on hiring the right employees. The typical recruiting strategy, which relies on manual tasks like reviewing academic records and resumes and evaluating technical and professional abilities, is time-consuming and inefficient. It is crucial to use a reliable and efficient method to evaluate candidates' skills and talents in order to satisfy company criteria. There are some solutions that have been proposed and has been implemented to automate the recruitment process. The following are some of the solutions that has been proposed.

Academic transcripts have been extremely important when hiring a candidate for a certain job position. For the problems faced when analyzing an academic transcript in a manual recruitment process , some solutions have been proposed and have been implemented. Authors have developed systems based on analyzing the academic transcript when hiring a candidate. The following are some researches done to address those problems.

According to the study done by a group of students in 2020 which focuses on candidate selection for an interview for the position of Software engineers [4], This research conducts a resume analysis, a personality prediction and also The researchers examined applicants' GitHub accounts using text mining tools and discovered that the number of repositories, commits, and followers were strong indicators of job performance. Apart from that they have analyzed the academic transcripts and identified the modules and categorized them into technical skills and soft skills and then the candidate with the most technical skills will be selected and their respective grade for the module will be compared. The candidate who has grades above C+ will be selected and others will be rejected for the job role.

The research paper titled "Analyzing Competences in Software Testing: Combining Thematic Analysis with Natural Language Processing (NLP)" [5] addresses the lack of analysis of competences required in software testing education for fresh graduates in computer science. The authors use NLP techniques to analyze job descriptions and course

syllabi to identify the competences required in software testing. The findings suggest a gap between the competences required and the current software testing curriculum, as soft skills such as teamwork, communication, and leadership are not always taught in software testing courses. This paper provides valuable insights for designing software testing curricula that equip students with necessary competences to succeed in their careers.

Research Problem

Sri Lanka's IT sector is currently demonstrating the impressive natural intellect of its people through creative technological advancements. With the second-highest income earned among industries in the country, this sector contributes considerably to the nation's exports. Over 500 industries are associated with information technology in Sri Lanka, and the sector places a high value on employee labor and environmental standards.

Due to the availability of a highly qualified workforce and cost-effective operations, the IT industry has become extremely profitable. As a result, it has become a key driver of economic growth in Sri Lanka.

While there are many automated recruitment systems available, they often fail to cover all aspects of the recruitment process. For example, some systems only rank CVs without analyzing the academic transcript or social media profiles like GitHub and LinkedIn, leading to crucial information being missed. Additionally, most of these systems are developed and implemented in foreign countries and may not be suited for the Sri Lankan job market or may not comply with local laws and regulations. Therefore, there is a need for more comprehensive and localized recruitment systems that take into account all relevant aspects of a candidate's profile.

In Sri Lanka, a considerable number of IT companies do not evaluate the academic transcripts of potential candidates during the recruitment process. This omission may lead to the hiring of unqualified candidates, which can result in suboptimal job performance, leading to a waste of resources and time. In some cases, the lack of proper evaluation may also lead to the underutilization of qualified candidates. For fresh graduates, who do not have significant industry experience, analyzing their academic transcripts becomes crucial to assess their suitability for the role. This assessment can help in determining whether the candidate possesses the necessary skills and knowledge required for the position. Therefore, it is recommended that companies give due consideration to the academic transcripts of candidates while recruiting to ensure a well-informed hiring decision.

Overall, even though there are system that can be used to automate the recruitment process, they lack some important concepts that can be essential when recruiting a candidate.

Research Gap

Despite the fact that many research studies have concentrated on automating the hiring process by screening resumes, the suggested method seeks to optimize every step of the hiring process. . This is especially important for Sri Lanka's IT sector, which earns the second-highest income yet lacks a functional hiring process. A Chatbot is a feature of the suggested system that enables people to communicate with the computer and get any questions answered. The system stands apart from other systems due to this functionality, which is done using RASA.

Unlike previous systems which focused mainly on all types of candidates, this specific component mainly focuses on fresh graduates which can be very valuable when considering a candidate for an open job position. Apart from that, previous systems that evaluate academic transcripts have only analyzed the module name and categorized them as technical or soft skills based on their names, which is limited in scope. In contrast, the proposed system takes into account the module name, grade and the module description from the module outline to categorize skills into different areas, making it more comprehensive. Moreover, unlike previous systems that are tailored to specific job roles, the proposed system is designed to be utilized across the entire IT industry.

The proposed system's graphical representation of skill areas is also a key benefit as it enables companies to quickly and easily identify the areas in which a candidate excels. This is a significant improvement over past systems that did not offer a user-friendly visualization of skill areas. By providing a more detailed and accurate assessment of a candidate's skills, the proposed system can help organizations make informed hiring decisions and ultimately lead to more successful hires.

Research	Based In Sri Lanka	For all the job roles IT industry	Using ChatBot	Analyzing the academic transcript	Analyzing the module outlines	Graphical representation of expertise areas of a candidate
Research [4]	✓	✗	✗	✓	✗	✗
Research [5]	✗	✗	✗	✗	✓	✗
Proposed Component	✓	✓	✓	✓	✓	✓

Table 2: The Research Gap

RESEARCH OBJECTIVES

Main Objectives

The fundamental objective of this research is to automate the recruiting process in order to select the best applicant for a certain position based on technical abilities, academic credentials, and relevant personality attributes. Automating the recruiting process can assist to improve recruitment speed, productivity, accuracy, and cost. [7] Because of the company's delayed response, many prospects prefer alternative offers. As a result, the perfect applicant for the job may be overlooked. Apart from that, the automated system can help in reducing the biases in the recruitment process. This can lead to more diverse and inclusive candidate pool and workforce. [8] And therefore, Since the manual recruitment process is very time consuming and a lengthy procedure, it is more efficient to automate the recruitment process.

Specific Objectives

The research component aims to evaluate the academic performance of fresh graduates. By analyzing their academic transcript and module outline, the component seeks to identify the strengths and weaknesses of candidates, as well as the areas in which they are highly skilled. To accomplish this, the component uses a cutting-edge algorithm that accurately categorizes the skills and knowledge of candidates.

The result is a user-friendly graphical representation that can be easily understood by hiring managers. This representation provides a thorough and accurate picture of a candidate's academic background, helping recruiters to make informed hiring decisions. With this information, recruiters can better evaluate whether a candidate is a good fit for a particular position, based on their academic achievements and skills.

Overall, the academic transcript analysis component plays a critical role in the success of the online recruitment system. By providing valuable insights into the strengths and

weaknesses of candidates, this component helps to ensure that only the most suitable candidates are selected for job positions. This not only benefits the organization by ensuring that it hires the best talent available, but also benefits the candidates themselves by increasing their chances of being selected for positions that best match their skills and expertise.

METHODOLOGY

The objective of this component is to get insights about candidates' academic performance and potential and to assess whether they possess the necessary knowledge and skills for the job role. In this web application, the candidate's transcript will be uploaded and then a graph will be generated which will display the skills that the candidate possesses. The following steps are involved in analyzing academic transcripts and module outlines to extract and categorize skill areas. Firstly, academic transcripts and module outlines are gathered from a specific university. Then, the module outline data and transcript data are extracted and preprocessed. Necessary features are extracted from the module outline and transcript data. These extracted features are then mapped and categorized into different skill areas using a machine learning model. Finally, the skill areas are represented graphically to provide a comprehensive view of a candidate's skills and knowledge. These steps are crucial in simplifying the hiring process and increasing the likelihood of finding the right candidate for job openings in the IT industry in Sri Lanka.

Overall System Overview Diagram

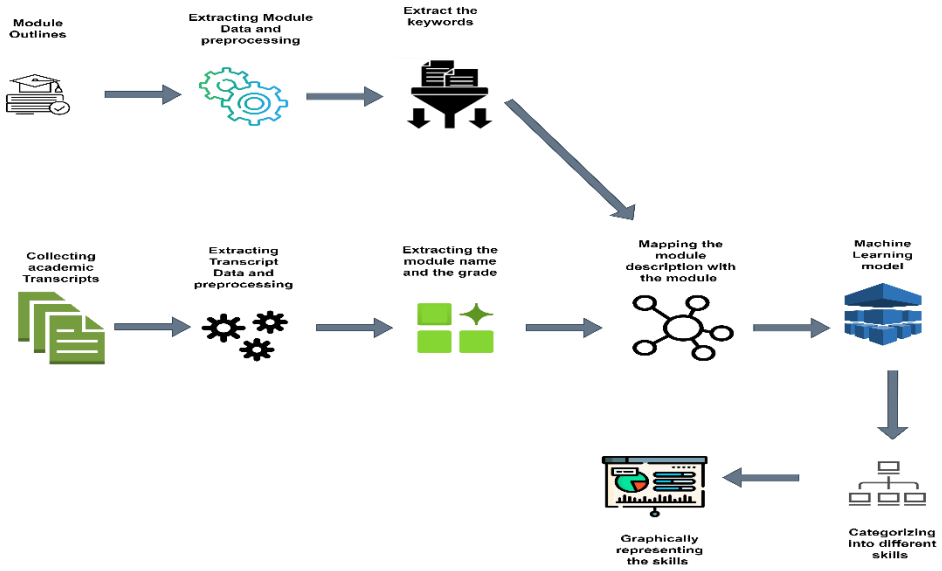


Figure 3: Overall System Overview Diagram

Software Solution

Agile is a popular approach to computer program development, which is designed to ensure that the final product is of high quality and meets the needs of clients. Agile methodology was chosen for this procedure because of its adaptability and flexibility. This is an iterative method in which the development process is divided into stages, each having its own set of deliverables at the conclusion.

[9]

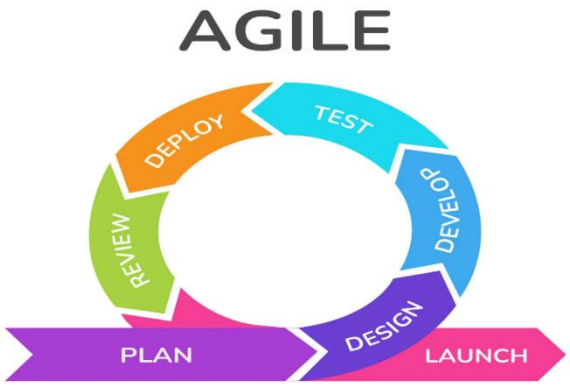


Figure 4: Agile Methodology

Requirements Gathering and Analysis

This phase focuses on gathering the requirements which is used to design this component. To gather information, a survey was conducted to gain people's opinions and ideas about the recruitment process.

User Requirements

The target audience for this particular component is the HR manager of the IT company. The component's requirements from the user's perspective are outlined below.

- The user will be able upload the candidate's academic transcript to the system.
- The system should be able to determine which areas candidate is highly skilled in by using the graphical representation generated by the system.

Functional Requirements

- A candidate should be able upload their academic transcript to the web application.
- Extracting the modules and the grade from a transcript
- Categorize the modules into different skills.
- Generate a graph by representing the skills of the candidates.

Non-functional Requirements

- **Reliability** - The system must be highly reliable without interruptions or breakdowns during the translation process.

- **Accuracy** - The system is intended to give reliable outputs since users rely on it for acquiring knowledge.

- **User Friendliness** – The system should be able to give users with a simple environment in which to perform what they desire.

- **Performance** – The system should be able to work properly and provide quick and accurate results.

- **Compliance** – All applicable laws and rules should be followed by the system. The candidate's private and secure information should be protected.

Feasibility Study

Technical Feasibility

While building a system to automate the recruiting process, technical feasibility is crucial. It is critical that the project team has competence in artificial intelligence, machine learning, and natural language processing to support the effective development of the proposed system. It is also necessary to ensure the security and privacy of candidate data in order to comply with data protection rules. Finally, assuming the project team has the necessary tools and skills, the suggested automated recruiting method is theoretically possible.

Operational Feasibility

The proposed system be able to analyze the transcripts and categorize the skills of candidates. The users should be able to understand and operate the system easily along with that it should be able to be integrated with the existing business processes and workflows.

Economic Feasibility

It is critical to consider economic feasibility while building a system that automates recruiting. The suggested approach intends to improve recruiting efficiency, save expenses, and streamline the process. The approach allows the organization to save money on recruiting expenditures such as advertising and hiring staff. The method can help accelerate the recruiting process by swiftly identifying competent individuals, lowering time-to-hire. Furthermore, the solution may free up HR teams to concentrate on vital responsibilities like employee development and retention, resulting in a more productive staff. In summary, the suggested automated recruiting system has strong economic viability since the system's advantages can outweigh the initial expenditures, resulting in cost savings and enhanced efficiency for the organization.

Design and Implementation

Once the requirements are gathered, we can start the designing phase. In here, a sketch of the proposed approach is done. Use case diagrams, sequence diagrams, use case scenarios, ER diagrams will be created.

During the implementation phase, the proposed system will be developed and constructed using various technologies, tools, and software solutions.

Tools and Technologies

IDE: **VS Code**

VS Code, a Microsoft-developed source code editor, is a versatile and lightweight tool that is suitable for various programming tasks, including coding, debugging, and version control. Its key benefits are its extensibility, allowing users to add new functionalities through its vast library of extensions, and its clean and user-friendly interface. Additionally, it has built-in support for Git, making it easier for developers to manage version control from within the editor.

Model Implementation : **Python using Google Colab**

Python is a user-friendly programming language that is particularly useful for quickly developing models and natural language processing. Google Colab is a cloud-based platform that offers high-performance computing resources and already-installed machine learning libraries, reducing the need for manual installation and ensuring compatibility with other frameworks.

Frontend: **Flask**

Flask is a Python web framework that enables the creation of a user interface for Python models, streamlining the interaction with them. By using Flask, a web application can be built that facilitates the input of data from users, which is then processed by the Python model. Finally, the output can be displayed back to the user via the Flask interface. This provides an accessible and user-friendly way to utilize Python models in various applications such as artificial intelligence, data analysis, and machine learning.

Database: **Azure Cosmos DB**

Azure Cosmos DB is a database service designed for modern app development that supports both NoSQL and relational data models. It was chosen as the database solution for storing unstructured data obtained from various sources such as CVs, academic transcripts, and social media profiles.

Deployment: **Azure App Service**

Libraries:

- Natural Language Toolkit (NLTK) - For removal of stop words, stemming.
- PyPDF2 or pdfplumber - Extract Text from PDF
- OCR (Optical Character Recognition) - Text Extraction from Images OpenCV
- PDF2Image - Convert PDF to Image
- NER(Name Entity Recognition) - SpaCy - Identification and Categorization
- Matplotlib, Seaborn, Plotly, Bokeh - Visualizations

Note: The tools, technologies, and algorithms proposed above are subject to change based on any challenges or problems that arise during the development process.

Integration and Testing

In order to ensure a high-quality product during the integration and testing phase, it is crucial to perform unit testing on each individual module of the system. Through system and integration testing, the methodology of the system can be evaluated. Ultimately, the result of these efforts will be the creation of a top-notch product.

Deployment

Azure – Platform as a service

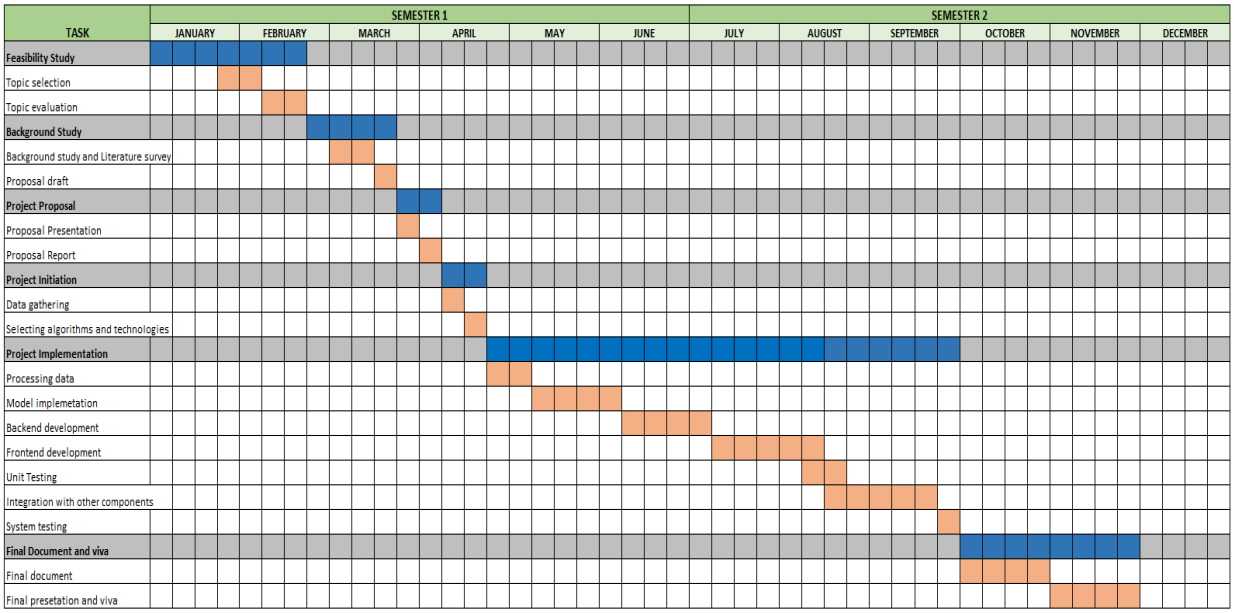
Azure is a cloud solution that makes it simple to migrate programs between your own servers and the public cloud. It provides a number of options to link your various systems, such as VPNs, caches, CDNs, and ExpressRoute connections. This makes it easier to operate and enhances its performance. You can rapidly deploy your app from anywhere and control which applications are loaded on which devices using Azure.

Review

The last stage of the Agile process entails providing the product to the client once it has been tested. If any revisions or alterations are required, a new iteration of the Agile process will begin to handle those changes. If no other revisions are necessary, the product will go on to the next stage of development. [10]

Work breakdown structure

Gannt Chart


TASK
SEMESTER 1
SEMESTER 2
JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
JULY
AUGUST
SEPTEMBER
OCTOBER
NOVEMBER
DECEMBER
Feasibility Study
Topic selection
Topic evaluation
Background Study
Background study and Literature survey
Proposal draft
Project Proposal
Proposal Presentation
Proposal Report
Project Initiation
Data gathering
Selecting algorithms and technologies
Project Implementation
Processing data
Model implemetation
Backend development
Frontend development
Unit Testing
Integration with other components
System testing
Final Document and viva
Final document
Final presetation and viva

COMMERCIALIZATION OF THE PRODUCT

When a candidate is hired, especially in Sri Lanka, the process takes a long time. The company examines resumes, shortlists them, interviews them, and conducts background checks. In the long run, this can be inefficient and may tend to lose the productivity of the company as well. As the information technology business is continually evolving, there is a great demand for new employees. And when the organizations take time to respond and undergo the manual recruiting procedure, they tend to lose applicants who can be highly beneficial to the organization.

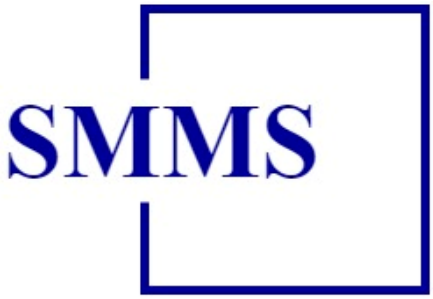


Figure 5: Logo - SMMS Software Solutions

Figure 6: Logo - Intellihire Software

SMMS, our team, has developed a cost-effective and automated software solution called “Intellihire” that addresses the challenges faced by companies during the recruitment process. Our solution incorporates advanced CV analysis and recruitment software, providing companies with a reliable tool to streamline their recruitment process and identify the most suitable candidates for a given job.

The analyzing academic transcript component is a cutting-edge software solution that analyses academic transcripts of fresh graduates, designed specifically for a leading university. With our advanced algorithm, you can easily and accurately evaluate a candidate's academic performance, strengths and weaknesses, and skill areas. This will enable you to identify the top talent from a pool of fresh graduates with little to no experience. If your company is looking to hire from another university, our software is

available for licensing at a reasonable cost. With our innovative solution, you can quickly and easily find the right candidates with the necessary skills and expertise to take your organization to the next level.

Our solution is scalable, meaning it can be used by companies of all sizes and across a range of industries. By automating the recruitment process, this system eliminates the need for manual and time-consuming tasks such as resume screening and candidate tracking. This saves companies both time and money while improving the accuracy of their recruitment process.

Furthermore, this system is designed with user-friendliness in mind, allowing even those without extensive technical knowledge to easily navigate the system. Additionally, our solution is customizable, so it can be tailored to meet the specific needs and requirements of each company.

Overall, this system is an efficient, reliable, and cost-effective solution that can help companies of all sizes streamline their recruitment process and identify the best candidates for the job. With its advanced CV analysis and recruitment software, our solution can be a valuable asset to any company looking to optimize their hiring process.

REFERENCES

[1] "harver.com," [Online]. Available: https://harver.com/blog/time-to-hire/.

[2] "www.icta.lk," [Online]. Available: https://www.icta.lk/news/sri-lanka-aiming-200000-ict-workforce-by-2022/.

[3] "www.browserstack.com," [Online]. Available: https://www.browserstack.com/guide/automation-testing-tutorial.

[4] M. P. E. S. R.G.U.S.Gajanayake, "Candidate Selection for the Interview using GitHub Profile and User Analysis for the Position of Software Engineer," p. 6, 2020.

[5] J. N. R. M. S. F. H. S. A. P. Tajmilur Rahman, "Analyzing Competences in Software Testing: Combining Thematic Analysis with Natural Language Processing (NLP)," p. 9, 2021.

[6] [Online]. Available: https://sapia.ai/resources/blog/recruitment-automation/.

[7] "www.testgorilla.com," [Online]. Available: https://www.testgorilla.com/.

[8] "www.krasamo.com," [Online]. Available: https://www.krasamo.com/agile-development-process/.

[9] "relevant.software," [Online]. Available: https://relevant.software/blog/agile-software-development-lifecycle-phases-explained/#Review.

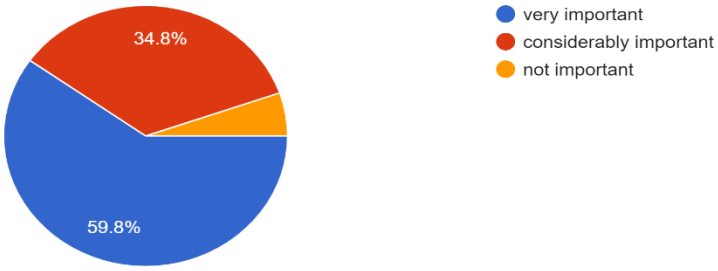
[10] "yello.co," [Online]. Available: https://yello.co/blog/interview-scheduling-statistics/.

[11] "Evaluating Automatic CV Shortlisting Tool For Job Recruitment Based On Machine Learning Techniques," 2022.

APPENDIX A

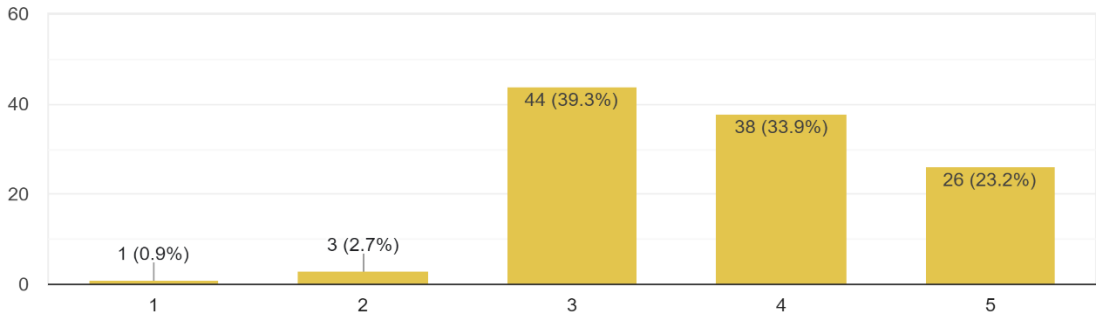
5. How important do you think it is for a recruitment system to be able to identify courses or programs that are particularly relevant to the position being recruited for?

112 responses

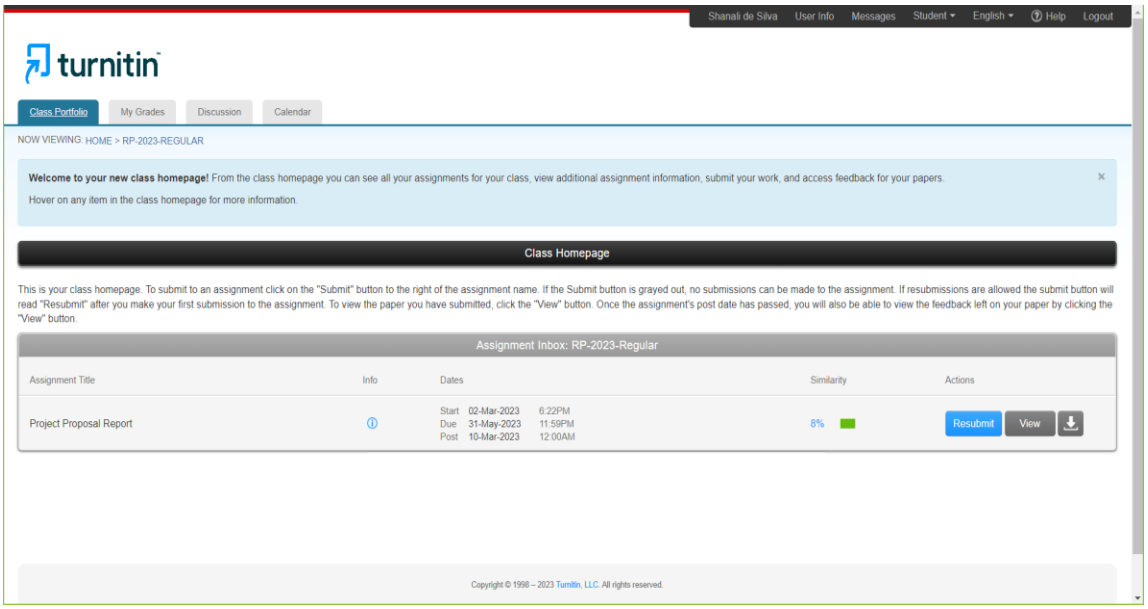


6. On a scale of 1 to 5 how important do you think academic performance is when considering a candidate for a job?

112 responses



APPENDIX B


Shanali de Silva
User Info
Messages
Student
English
Help
Logout
turnitin
Class Portfolio
My Grades
Discussion
Calendar
NOW VIEWING: HOME > RP-2023-REGULAR
Welcome to your new class homepage! From the class homepage you can see all your assignments for your class, view additional assignment information, submit your work, and access feedback for your papers.
Hover on any item in the class homepage for more information.
Class Homepage
This is your class homepage. To submit to an assignment click on the "Submit" button to the right of the assignment name. If the Submit button is grayed out, no submissions can be made to the assignment. If resubmissions are allowed the submit button will read "Resubmit" after you make your first submission to the assignment. To view the paper you have submitted, click the "View" button. Once the assignment's post date has passed, you will also be able to view the feedback left on your paper by clicking the "View" button.
Assignment Inbox: RP-2023-Regular
Assignment Title
Info
Dates
Similarity
Actions
Project Proposal Report
Start 02-Mar-2023 6:22PM
Due 31-May-2023 11:59PM
Post 10-Mar-2023 12:00AM
8%
Resubmit
View
Copyright © 1998 – 2023 Turnitin, LLC. All rights reserved.