

**CAREER PATH IDENTIFICATION AND CAREER
OPPORTUNITY RECOMMENDATION BASED ON A
MACHINE LEARNING APPROACH**

2022 – 017

Hirimuthugoda UJ

(IT19138114)

B.Sc. (Hons) Degree in IT specializing in Software Engineering


Department of Computer Science & Software Engineering

Sri Lanka Institute of Information Technology Sri Lanka

September 2022

DECLARATION

I declare that this is my own work and this dissertation 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).

Name	Student ID	Signature
Hirimuthugoda UJ	IT19138114	

Signature of the supervisor:

Date:

Signature of the co- supervisor:

Date:

ACKNOWLEDGEMENT

The outcome of the effort is completely based on the effort and the dedication of the team members as well as the support, guidance and the encouragement provided by many others. I am grateful to all those who helped me in finishing the huge effort and achieving this result.

I am grateful to Prof. Samantha Thelijjagoda, Dr, Koliya Harshanath, Ms. Kugathasan Archchana, and the lecturers at Sri Lanka Institute of Information Technology for their recommendations, ongoing encouragement, and support in the creation of this research, especially for the variety of advanced and educational discussions.

I'd want to express my gratitude to Mahen Thammita, Akila Liyanage, and Nethsara Liyanage, members of my research team, for their support and encouragement in finishing this project. Also, I'd like to thank the Sri Lanka Institute of Information Technology (SLIIT) for identifying the resources that were available at the right time to groom and develop my skill.

Additionally, we would like to extend our thanks to all our colleagues and friends for their assistance, support, and invaluable advice. Finally, I'd like to thank everyone else who isn't expressly listed but has shown support in several ways and inspired me to make this a success.

ABSTRACT

Information technology is a massive industry consisting of a wide range of tech stacks and professional experts. Various career paths such as full stack developer, backend developer, front end developer, and many more careers are observed in the information technology industry. However, a lack of awareness of own talent may lead to poor decisions. On the other hand, Employees change their careers regularly, seeking for the most suitable and comfortable career for them, specially in IT industry. Therefore, identifying the capabilities and providing effective guidance to newcomers is important. This research focuses on implementing a career guidance system, which consists of the most suitable career path identification and career opportunities recommendations purposing to assist fresh IT undergraduates by providing effective career guidance. For this purpose, data were collected from IT industry-related job postings using some keyword extraction methods. Several models were implemented such as Support Vector Machine Classification, Logistic regression, Naive Bayes and their performance were compared using model accuracy and f1 score to select the accurate model. For the identification of most suitable career path, Support Vector Machine with RBF kernel resulted with highest accuracy of 83%. The ultimate goal of this research is to help IT undergraduates to identify their skills and provide a stable start for their career life.

TABLE OF CONTENTS

DECLARATION	2
ACKNOWLEDGEMENT	3
ABSTRACT	4
TABLE OF CONTENTS	5
LIST OF TABLES	8
LIST OF FIGURES	9
LIST OF ABBREVIATIONS	10
LIST OF APPENDICES	11
1. INTRODUCTION	12
1.1. Background Context	13
1.1.1. Requirements Gathering and Analysis	14
1.2. Research Gap	14
1.3. Research Problem	16
1.4. Research Objectives	17
1.4.1. Main Objective	17
1.4.2. System Objective	17
1.4.3. Specific Objectives	17
1. Student's Skill Identification	17
2. Career Path Identification	17
3. Job Opportunity Recommendation	17
2. 17	
2.1. Functional Requirements	18
2.2. 17	
3. METHODOLOGY	20
3.1. Student20	
3.2. Career Path Identification	22
3.3. 22	
3.4. Data Gathering	24
3.5. Model Selection	26
3.6. Training the Model	28

3.7.	Design the System	29
3.8.	Implementation and Testing	30
3.8.1.	Student Core Skill Identification	30
3.8.2.	Most Suitable Career Path Identification	30
3.8.3.	Job Opportunities Recommendation	30
3.8.4.	Tools and Technologies	31
3.8.5.	Front-end Implementation	34
3.8.6.	Testing	35
3.8.7.	Test Cases	37
3.9.	Commercialization Aspects of the Product	40
4.	RESULT	43
4.1.	Result	43
4.2.	Research Findings	44
4.3.	Discussion	44
5.	CONCLUSION	45
6.	REFERENCES	46

LIST OF TABLES

Table 1-1 - Research Gap	14
Table 2-1 - Skill Categorization	19
Table 2-2 - Example of Student Skill Profile	21
Table 2-3 - Filter by Experience Level	22
Table 2-4 - Sample Data Set	24
Table 2-5 - Test Case 1	35
Table 2-6 - Test Case 2	36
Table 2-7 - Test Case 3	36
Table 2-8 - Marketing Cost	38
Table 2-9 - Pricing	39
Table 2-10 - Budget Estimation	40

LIST OF FIGURES

Figure 2.1 High Level Architecture Diagram	19
Figure 2.2 Jaccard Similarity	23
Figure 2.3 Model Accuracy Comparison	26
Figure 2.4 Model Training Workflow	27
Figure 2.5 Career Profile UI	32

LIST OF ABBREVIATIONS

Abbreviations	Description
IT	Information Technology
SVM	Support Vector Machine

LIST OF APPENDICES

Appendix 1 - Model Accuracy Comparison	46
Appendix 2 - Keyword Extraction	46
Appendix 3 - Model Training	47
Appendix 4 - Project Gantt Chart	48

1. INTRODUCTION

The IT industry has shown rapid growth over the last few decades with a number of growing career opportunities. Since the IT industry provides a wide range of career paths, when stepping into the industry, freshers may get confused about which path to select as their career. According to research conducted by Peter Akosah [1], there is a chance of getting wrong decisions on career path selection due to some external factors such as family pressure, friend's career, and social pressure. Therefore, having a clear understanding about their own capabilities and personal interest is important when choosing a career path as a fresher. Many students start their first career after they completed the higher education or during the higher education. This stage can be considered as a key point of the person's career life since it's the first time they take the decision of their career path. Therefore, it is important to provide appropriate guidance to the students prior to their entry into the industry. A number of researches have been conducted focusing on career path identification and career guidance.

1.1. Background Context

A number of researches have been conducted focusing on providing more efficient career path identification and career guidance for university students. Those studies have focused on identification of student's skills and providing the most suitable career guidance by analysing the student profile following various machine learning approaches.

Vignesh S, Shivani Priyanka and Shree Mangju have come up with a career guidance system for the engineering students based on their skills[2]. They have conducted an assessment to evaluate the student skills, which includes psychological and the core-skill oriented questions. Students have been clustered into different departments (Computer Science Engineering, Electric and Electronic Engineering, Electronic and Communication Engineering, and Mechanical Engineering) based on the identified skills with the help of the K-means clustering algorithm.

In 2014, a team of researchers with Tajul Rosli, have implemented a career path recommendation system using fuzzy logic, [3] focusing on computer and mathematical students. In addition to the student's technical skills, they have considered the student's personality as well. Personality and skills data were collected through a series of interviews. Skills have been labelled with three linguistic variables such as "Good", "Medium", and "Weak". Considered careers are also labelled, respective to the student, with another three linguistic variables such as "Yes", "No", and "maybe".

Ashutosh Shankhdhar, Akash Agrawal and Deepak Sharma have come up with an intelligent decision support system using the decision tree algorithm. [4] They have collected student academic performance through a survey and some quizzes to collect data about student personality. In this study also, student personality was considered in addition to technical skills. Decision tree algorithm has been used to identify the most suited career for the student.

In terms of candidate profile classification, numerous researches have been conducted using text classification techniques.

Tere Gonzalez, Pano Santos, and Fernando Orozco have come up with an Adaptive Employee Profile Classification System that classifies employees according to their skills and experience by analysing their resume. [6]

Razkeen Shaikh, Nikita Phulkar, and Harsha Bhute have implemented a system to classify the candidate by analysing the profile using text categorization and semantic analysis. [7] Recommendations have been given by calculating the similarity between candidate skills and the required skill set.

Faizan Javed and Qinlong Luo have come up with a job title classification system for a large taxonomy of job categories [8]. Data set have been prepared normalizing the job titles and the job descriptions. Support Vector Machines and k-Nearest-Neighbour have been applied in their research.

1.1.1. Requirements Gathering and Analysis

During the requirements gathering and analysis phase, existing systems and studies have been reviewed that were utilized for career guidance for university students. The majority of studies have focused on a broad audience during career path identification and some of important key factors such as personal interest of the student were not considered. The goal of this research was to provide an efficient and reasonable career guidance for students using machine learning based techniques. Based on the above goal, all functional and non-functional requirements are determined, and system requirement specifications were created.

1.2. Research Gap

By examining existing career guidance systems, the student skill identification phase is usually performed by analysing student academic data or conducting the interviews and quizzes. When the student grades are evaluated module-wise, it won't be able to get accurate data on individual skills of the student. Students' career paths may be changed by the skills they gain through the learning process, and it is important to give them an

idea about how should they align their career path accordingly. None of the above-mentioned studies have attempted to evaluate student's skills continuously and identify the student core skills. To fill the aforementioned gaps, this study proposes a machine learning approach based on the identification of the student skills and the IT industry-specific requirements.

Study	Consider Student's Technical Skills	Consider Student's Personal Skills	Consider Student's Personal Interest	Continuous Student's Individual Skill Evaluation and Identification
Study 1 [2]	Yes	Yes	No	No
Study 2 [3]	Yes	Yes	Yes	No
Study 3 [4]	Yes	Yes	No	No

Table 1-1Research Gap

1.3. Research Problem

For a university student, Identification of suitable career path is a critical decision as a newcomer to the industry. Even the student has followed a specific subject stream targeting specific industry, since an industry provides multiple career paths focusing on specific skills and requirements, selecting the most suitable career path can be quite challenging. Without a proper understanding about industry requirements and their own capabilities, it may lead to incorrect decisions.

Addition to that, family pressure, friend's career and social pressure can directly make impact to the student's decision on career path selection. With these impacts, a student may be able to select a wrong career path and then he or she will have to face many difficulties in career life. Having to work with technologies and tools which are not interested in, not able to work with others in similar capacity kind of concerns will push the student to very difficult situations and it can be worst when they have to balance both work and academic life.

As a solution for this problem, a career guidance system is developed attaching to the machine learning based learning management system "AITor". It will be able to analyse the student skills continuously and identify the core skills and then, based on those identified skills and student personal interest, it will suggest the most reasonable career path. Addition to that system will suggest the most suitable job openings for the student according to the above-mentioned parameters.

1.4. Research Objectives

1.4.1. Main Objective

This research aims to identify the core skills of IT undergraduate students including both technical and personal skills, and then provides efficient career guidance for them considering their personal interest also. As the main output, this will suggest the most suitable career path for the student and the current job openings.

1.4.2. System Objective

This system is developed focusing on identification of student core skills through continuous assessment series and identify the most reasonable career path for a student using machine learning model based on above identified skills and then suggest the most suitable job openings for the student by matching the required skills and the student skills.

1.4.3. Specific Objectives

1. Student Skill Identification

As the first step of the career guidance system, student's core skills will be identified by the system. This will perform through a continuous assessment process and skills will be updated with the student progress. Both technical and personal skills will be analysed by the examination module.

2. Career Path Identification

Based on the identified skills and the student's personal interest, most suitable career path will be suggested to the student using the machine learning model.

3. Job Opportunity Recommendation

Based on identified skills and the suggested career path, student will be recommended current job openings.

2. PROJECT REQUIREMENTS

2.1.Functional Requirements

According to the objectives that are going to be achieved through this case study, the main requirement of the AITor platform would be, implementing an all-in-one solution that introduces a personalized experience into the traditional education system while providing detailed analysis on student education. In that case, career guidance module will perform individually for each and every student based on their skills and interest. Three main requirements of career guidance module can be listed as follows.

- Student core skills identification
- Most suitable career path identification
- Job opportunity recommendation

2.2.Non-Functional Requirements

Considering patterns and trends of users of existing E-Education platforms, it was decided to include three major non-functional needs in the proposed solution.

- Availability
- Scalability
- Accuracy
- Privacy

Learners are spending more time in E-Learning platforms as the means of education are vastly altered towards E-Learning rather than classroom education. Thus, the system's and its features' availability would be critical.

Because the change of traditional learning into E-Learning is currently underway, most learners, tutors, and institutes are still migrating to E-Education platforms such as AITor. As a result, the solution must be scalable as the number of stakeholders and resources grows.

The accuracy of career path recommendations is a one of important non-functional requirement for AITor, since it provides a support to the student to get a decision on their career life.

3. METHODOLOGY

High Level System Diagram

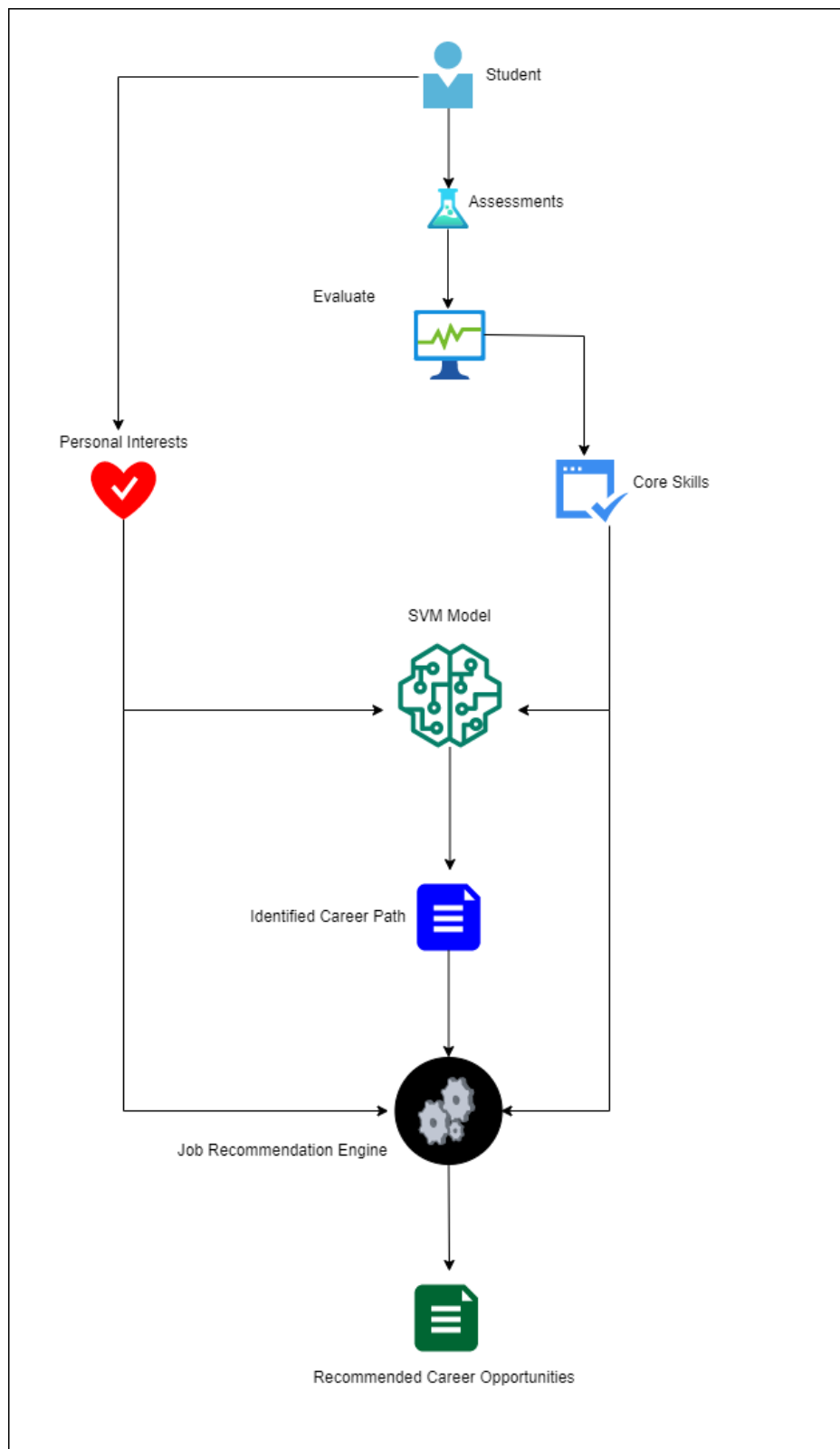


Figure 3.1 High Level Architecture Diagram

This section describes the implementation of the career guidance system which consists of three main modules.

- Student Skills and Personal Interest Identification
- Career Path Identification
- Career Opportunities Recommendation

3.1. Student Skills and Personal Interest Identification

In the initial stage, which is the preparation phase of the exam paper, all the questions in the exam paper will be tagged according to the required skills. Some existing studies have proposed this content tagging concept [9] for student skill identification, which can be used to address the above-mentioned problem, which is the difficulty of getting accurate data on student performance of individual skills. These skills can be technical skills, such as programming languages, frameworks, and databases, or soft skills, such as problem solving, analytical, and critical thinking skills. One question can be tagged with one or multiple tags. The tags are predefined and available in the preparation of the exam paper.

Skill	Category	Type
Java	Language	Technical Skill
React JS	Framework	Technical Skill
Problem Solving	-	Soft Skill

Table 2-1Skill Categorization

Second phase, examinations will be conducted through an online examination platform and obtained marks for each and every question will be recorded. In the 3rd

phase, examination results will be evaluated and average mark for each and every skill will be calculated. Based on the average mark, skills will be graded. If the skill has expected grade level or above, it will identify as a core skill of the student. Another important factor mentioned above is the student's personal interest. Even if the student showed better performance in different subject areas, it can't assume that the student is preferred in all of them. This system has implemented a feedback mechanism to collect the student's personal preferences in covered subject areas and any other related subjects. Student can rate their experience on covered learning areas or input any other personal interests related to the IT industry. Those ratings will be saved along with previously identified skills. Then, the intersection of the previously identified core skill set and the student's most preferred skill set will be extracted as the optimal skill set for the career recommendation.

3.2. Career Path Identification

Career path identification model has been implemented as a multi-class text classification model to take the student profile as the input and to recommend the most suitable career as the output. The student profile is maintained by the system in four aspects. Technical skills which are identified by the system, skills determined by the student, student career-related personal interests, and the personal skills of the student. Here, students are allowed to add skills to their profile since a student can have some skills in addition to what they learn in the academic subject stream. The following table showcases an example of a prepared student profile by the system.

Identified Technical Skills	JavaScript, Java, React, Node, MongoDB,
Personal Skills	Problem solving, Analytical
Personal Interest	AI, Machine Learning
Technical Skills Declared by the Student	Azure, Datagran

2-2Example of student skill profile

The model was deployed on a Django application by implementing the Representational State Transfer Protocol (REST) Application Programming Interface (API) which receives the input from the user and provides the most suitable career path as the output. In the initial phase, since student skills are not recorded in the system, it leads to a cold start. To avoid that, students will ask to provide their fields of interest or their own skills to the system. Later, they will be fine-tuned along with the learning and evaluation process.

3.3. Career Opportunities Recommendation

Career opportunities recommendation module was implemented that focuses on recommending career opportunities in the IT industry by matching the set of identified student skills with the skills extracted from job postings. Here the identified career path in the second step will be an optional parameter to recommend the career opportunities since career opportunities recommendation was focused on the recommendation of job opportunities based on student skills and experience level. Student can determine their experience level in the system. Job postings will be classified into three categories based on the required experience level.

Job Position/ Description	Experience Level
Intern/ Trainee	Entry level
Associate or 1-5 exp	Mid-level
Senior/ Lead or exp > 5 years	Senior level

Table 2-3 Filter by Experience level

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

A - Set 1
B - Set 2
J - Jaccard Distance

Figure 3.2 Jaccard Similarity

Student will have more choices based on the identified skills and if the student decided to go with the recommended career, the system will filter out the job posting related to the recommended career. Current job postings were accessed through an API and required skills were extracted. To extract skills from the job postings, resume parse [ref] library has been used here. The skills extracted were saved in each job posting as a JSON value. Then the similarity between the skills extracted from the job posting and the student skill set was calculated using the Jaccard similarity index. A student's skill set is defined by combining the identified skills and the defined skills by the student.

The Jaccard similarity index compares values for two sets to identify which values are shared and which are distinct. The similarity will be given in range of 0% to 100%. Based on the calculated similarity, most suitable job opportunities will recommend for the student. In this system, it was designed to recommend jobs that have similarities of 50% or above.

3.4. Data Gathering

In the data collection phase, it considered gathering data on various careers in the IT industry and their requirements. This research has been narrowed down to the below careers which are demanding in Computer Science and that can start as a fresher.

- Backend Developer
- Database Administrator
- Data Scientist
- Database Administrator
- Dev-ops Engineer
- Frontend Developer
- Full Stack Developer
- Mobile Application Developer
- Network Engineer
- Software Quality Assurance Engineer
- UI/UX Designer

The data set was prepared with existing data sets that contain job postings in the last few years. From the selected data sets, IT industry-related job postings were separated, and since some job postings contain entire job descriptions, skills were extracted using keyword extraction techniques. The data collected was organized as required skills along with the career path/position. Data points that contain null values or insufficient job descriptions have been removed. The extracted text data was converted to a matrix of token counts using CountVectorizer. Finally, the count matrix was transformed to a normalized tf-idf representation using TfidfTransformer. The table below describes the structure of the data set.

Career	Required Skills
Frontend Developer	JavaScript, React, HTML, CSS, SCSS, Designing, CI/CD, REST API, Git
Backend Developer	Java, MySql, Spring boot, AWS, Git, problem solving skills, Analytical skills
Mobile Application Developer	Flutter, Firebase, NodeJs, Swift, Cordova, Analytical skills
Backend Developer	Tomcat, Linux, JBoss, JSP, Eclipse, MySQL, Agile, JDBC
UI/UX Developer	Adobe Photoshop, UI/UX, Adobe XD, communication, graphic designs, wireframes

Table 2-4sample data set

3.5. Model Selection

Three machine learning algorithms have been considered in the model selecting phase.

- Naive Bayes
- Support Vector Machine
- Logistic Regression

The following figures show the comparison between different algorithms in terms of accuracy and f1 score.

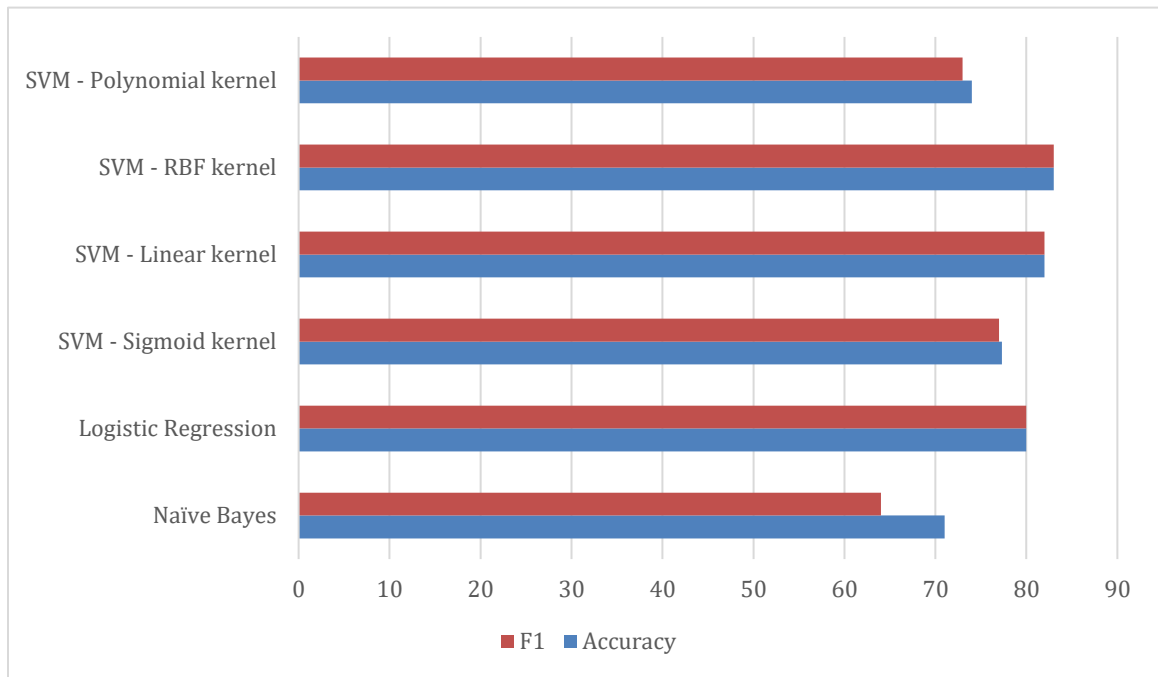


Figure 3.3 Model Accuracy Comparison

In the model selection phase, three main supervised machine learning algorithms were considered for the classification model, which are Naive Bayes, Logistic Regression, and Support Vector Machine. Support Vector Machine (SVM) algorithm was tested out with four kernels which are linear, polynomial, sigmoid, and RBF.

As described in the above figure, the Support Vector Machine algorithm with the RBF kernel has shown the highest accuracy, which is 83%. Support Vector Machine is a supervised machine learning algorithm used for both classification and regression problems.

SVM doesn't support multiclass classification natively and supports for binary classification. To apply SVM for a multiclass classification problem, here it has followed an approach that divides the data points in the particular class and rest, which called as one to rest approach. It can consecutively a certain class is distinguished from all other classes.

A particular type of Gaussian kernel, called a Radial basis function kernel (RBF kernel) projects high-dimensional data and searches for a linear separation for it. Based on the results of the evaluation phase, the SVM with an RBF kernel was chosen since it has shown the highest accuracy and, furthermore, has shown better performance in previous studies in multiclass text classification. [10]

3.6. Training the Model

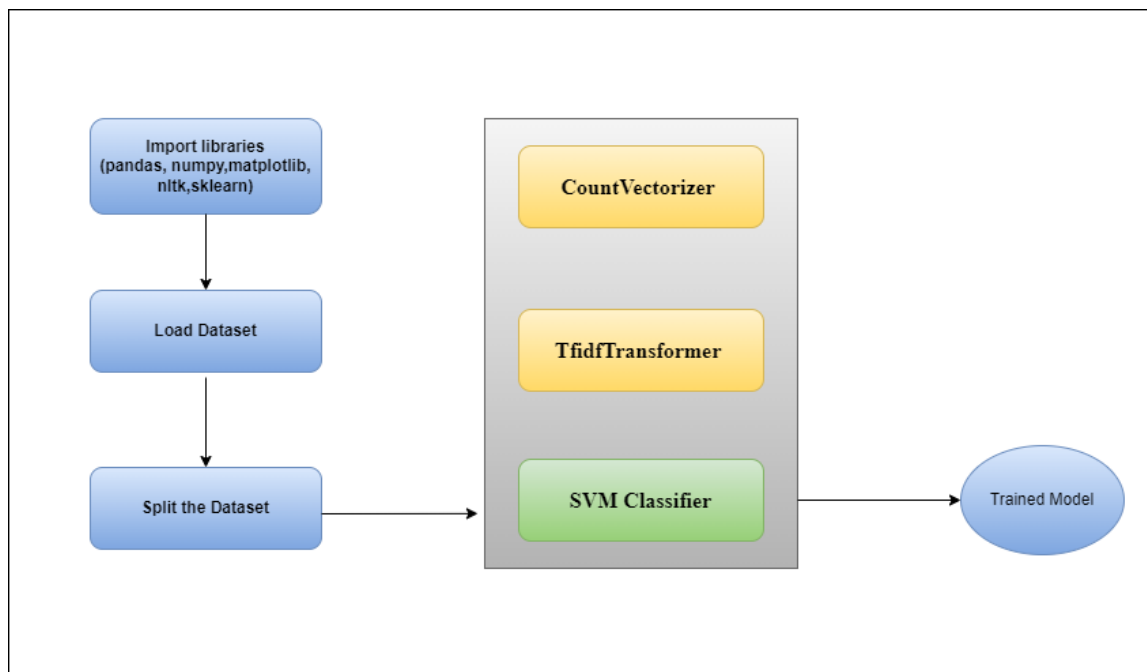


Figure 3.4 Model Training Workflow

The above resulted data set has been divided into the training and the testing data set, where 70% of the data set was used as the training data set to train the model, and 30% of the initial data set was used to test the trained model. While training the model, a pipeline has been used to work with the vectorizer, transformer, and classifier. CountVectorizer was used to convert the extracted keywords into numeric values. TfidfTransformer was used to convert the values into a tf-idf representation, which helps to determine how relevant each word is in the input phrase. Finally, model was trained with processed data to identify the most suitable career path using the SVM as the classifier.

3.7. Design the System

Before the implementations, entire system and the subsystems were planned and designed. The overall system and subsystem connections, inputs, and outputs were identified and designed. System design aids in the specification of hardware and system requirements, as well as the definition of overall system architecture.

Under the Designing phase, the proposed system is primarily focused on developing an SVM-based classification model, improving user experience, and system usability.

3.8. Implementation and Testing

3.8.1. Student Core Skill Identification

As the initial step of the system, student's technical and personal skills will be evaluated through a continuous exam series. Examinations will be tagged according to the skills which are going to evaluate, and skills will be graded at the end of the exam. These grades will be saved in the database and will update continuously with the learning progress of the student. Addition to that, students are allowed to update the system with their interested area related to the career.

3.8.2. Most Suitable Career Path Identification

Above identified skills and the student's interest will provide to the career path identification model through an API as an array of keywords. Model will identify the most suitable career path according to the identified core skill set and the personal interest of the student.

3.8.3. Job Opportunities Recommendation

In the job opportunities recommendation, initially it will perform through a similarity calculation process between the student skills and the required skills of the job postings. System will access to the current job posting through an external API or job postings given to the system by Institute. If the student decided to go with suggested career path, student can filter out recommended job postings by the suggested career path. Also, student can determine the experience level and it will provide to the job opportunity recommendation module to fine-tune the output. Suggested job postings will be exposed through an API.

3.8.4. Tools and Technologies

Frontend Technologies

React

React was created and is maintained by Facebook. React is a JavaScript library for creating user interfaces giving an overview of the MVC architecture. Suitable for large-scale web applications that can access and alter data in real time without having to reload the entire page

NPM

The Node Package Manager is a JavaScript package manager for the Node platform. It creates modules so that node can locate them and intelligently resolves dependency concerns. It can be designed to handle a wide range of scenarios. It is the most often used programming language for publishing, discovering, installing, and developing node applications.

Backend Technologies

Python and Libraries

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. It is a powerful programming language that is free to use and distribute. Python's standard library is large and includes functions that can be used for practically anything.

- **Pandas**

Pandas is an open-source data analysis and manipulation tool developed on top of the Python programming language that is fast, powerful, versatile, and simple to use.

- **NumPy**

NumPy is a Python library that adds support for huge, multi-dimensional arrays and matrices, as well as a vast number of high-level mathematical functions to operate on these arrays.

- **NLTK**

The Natural Language Toolkit, or more simply NLTK, is a collection of Python-coded tools and applications for symbolic and statistical natural language processing of English.

- **Django**

Django is a high-level Python web framework called supports quick development and straightforward, practical design. Machine learning models were deployed through Django application.

Spring Boot

Spring Boot is an open-source Java-based framework used to create a micro-Service.

Tools

- **Jupyter Notebook**

The Jupyter Notebook is a web application for creating and sharing computational documents. It provides a straightforward, efficient, document-focused experience.

Database

MySQL is a relational database management system (RDBMS) developed by Oracle that is based on structured query language (SQL).

3.8.5. Front-end implementation

Student Career Guidance Dashboard

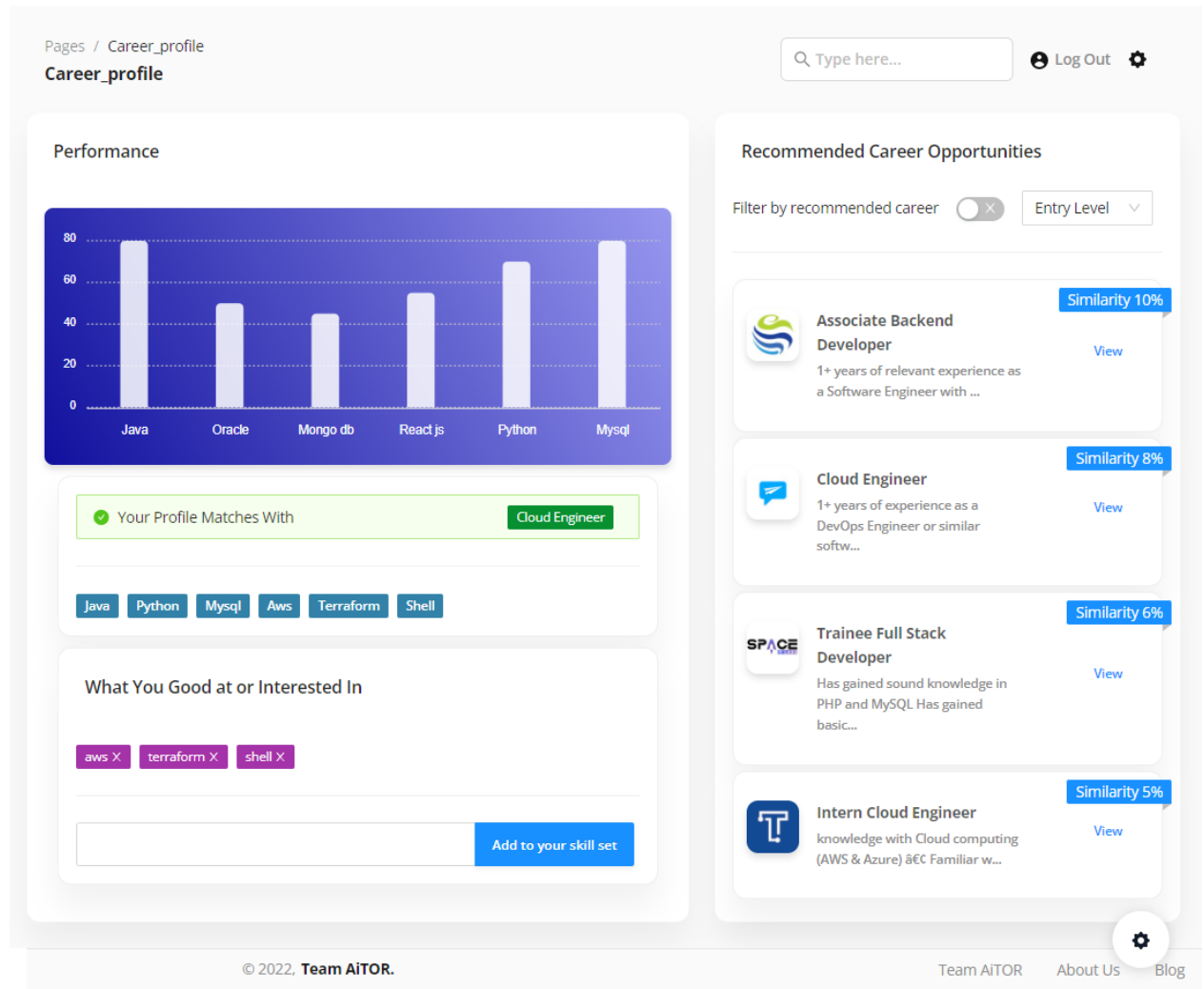


Figure 2.5 Career Profile UI

Student career guidance dashboard displays the overall analysis of the student performance for skills which are evaluated. Also, it will display the suggested job opportunities along with the suggested career path.

3.8.6. Testing

At this stage, a number of assessments were carried out to test the implemented system.

Unit Testing

Each component that has been implemented under relief optimization has been subjected to unit testing.

Component Testing

By integrating numerous units, component testing has been completed.

Integration Testing

To make sure that each component communicates with the others correctly, integration testing has been carried out.

System Testing

The entire system (AI for web application) is tested to ensure proper operation after the components have been connected.

After developing the career path identification model and job opportunities recommendation module, unit testing was done to test the accuracy of the system. Using data gathered after conducting interviews with industry professionals under different career paths, most suitable career path was identified and compared whether it matches the expected output.

3.8.7. Test Cases

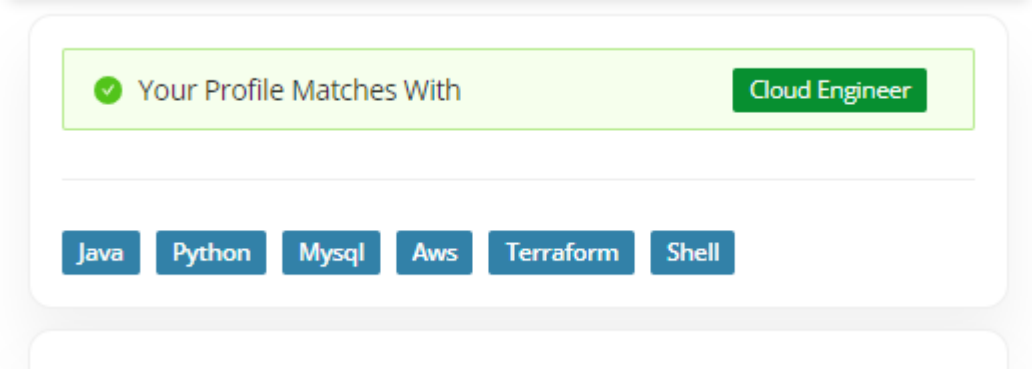
Test Id: 001	
Test Description: An Input of skill set of Cloud Engineer	
Test Data : “Java, Python, MySQL, AWS, Terraform, Shell”	
Expected Output: Cloud Engineer	<p>Actual Output</p>  <p>The screenshot displays a green notification bar at the top stating 'Your Profile Matches With' followed by a green button labeled 'Cloud Engineer'. Below this, a horizontal line separates the header from a list of skills. The skills are presented as blue buttons: 'Java', 'Python', 'Mysql', 'Aws', 'Terraform', and 'Shell'.</p>
Test Status: Pass	

Table 2-5 Test case 1

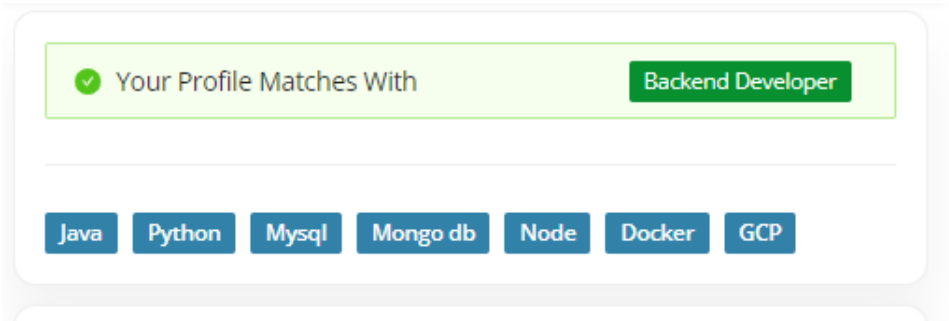
Test Id: 002	
Test Description: An Input of skill set of Backend Developer	
Test Data : “Java, Python, MySQL, Mongo DB, Node, Docker, GCP”	
Expected Output: Backend Developer	Actual Output: 
Test Status: Pass	

Table 2-6 Test case 2

Job Opportunities Recommendation based on Identified Career Path and Experience Level



<div><h3>Recommended Career Opportunities</h3><p>Filter by recommended career <input checked="" type="checkbox"/> Entry Level ▾</p><div><div><h4>Cloud Engineer</h4><p>1+ years of experience as a DevOps Engineer or similar softw...</p><p>Similarity 8%</p>View</div><div><h4>Intern Cloud Engineer</h4><p>knowledge with Cloud computing (AWS & Azure) â€C Familiar w...</p><p>Similarity 5%</p>View</div></div></div>	
Output: Success	

Table 2-7 Test case 3

3.9.Commercialization Aspects of the Product

Target Audience

AITor education platform is intended to provide most reliable and efficient support for the students and lecturers in any education institute. Primary target users will be university students and lecturers.

Demand for the system

With the situation raised with the covid-19 pandemic, online education became more popular all over the world. In Sri Lanka also most of the educational activities moved into the online approach. Comparing to the traditional classroom education concept, online education has some barriers, and this system is designed to address those problems and overcome them while providing most accurate and reliable output for the students and the lecturers.

Marketing Plan

Marketing Cost

Marketing strategy	Monthly	Currency
Product branding	50	USD
Content branding	20	USD
Email and newsletters	20	USD
Cost	90	USD

Table 2-8 - Marketing cost

Since product commercialization is a considerable fact in considering the entire system, it will take considerable cost to the marketing related activities.

1-month free trial

- With all functionalities of the system

Paid Version

- With all functionalities of the system
- 24/7 service

Free Version

- With limited functionalities

Pricing

Package	Features Included	Price
Pro starter	<ul style="list-style-type: none">• Student performance analyser• Target for educational institutes having student count of below 1000	USD 1500 per month 1.5 USD per Student

Prod Ultimate	<ul style="list-style-type: none"> • Student performance analyser • Carrier recommendation • Final grade forecasting and Tutor guide • Target for educational institutes having student count of over 1000 ~ 3000 	USD 3000 per month ~ 1.5 USD per Student
---------------	---	---

Table 2-9 Pricing

Budget

The budget that is needed for the entire AITor platform, can be classified into two categories. These categories will represent the cloud-based cost that needed for development and deployment and the other category represent the marketing cost for commercialize and market the AITor product

Cloud Based Cost

Service	Monthly	First months total	12Currency
AWS Fargate	36.04	432.48	USD
S3 Standard	1.16	13.92	USD
Data Transfer	0	0	USD
Amazon Simple Queue Service (SQS)	0	0	USD
Amazon Elastic Container Registry	0.5	6	USD
Amazon EC2		526.44	USD
Amazon Keyspace	2	64	USD
Other	20	240	USD
Cost	103.57	1282.84	USD

Table 2-10 - Budget estimation

In the cost calculation, it is assumed that 100 users are using the system. Since the entire system will be fully hosted in AWS cloud and will use different cloud features, the considerable portion of the cost will allocate acquiring cloud services. The “Other” category mentioned in the above table includes the costs for third party APIs that are used in getting learning materials into the system.

4. Result

4.1. Result

As the main objective of the career guidance module, the SVM multiclass classification model was built to identify the suitable career path for the student. Model has been trained with the dataset of recent job postings related to the IT industry and could be able to achieve 83% of model accuracy.

In the job opportunity recommendation, similarity calculation between student identified skill set and the required skills of job postings is performed with the Jaccard similarity index, and it shows a connection with the identified career by having a higher similarity index which are listed under identified career path.

4.2. Research findings

When building a multi-class classification model for the system using SVM models, accuracy becomes a critical factor. When attempting to improve the accuracy of an SVM-based multiclass classification model, it was identified that the kernel used in the model had a significant impact on the model's performance.

4.3. Discussion

In the career path identification scenario, the main responsibility is to identify the most suitable career path for the student based on the student skills and personal interest. For this purpose, SVM based multiclass classification model was built, trained, and tested with the valid data.

Initially model was built with Naive Bayes, SVM, Logistic regression and then based on the accuracy, SVM has been selected as the optimal model for the career path identification. SVM model accuracy also compared with different kernels including Polynomial kernel, RBF kernel, Linear kernel, and the Sigmoid kernel. Based on the results, SVM model with an RBF kernel has been selected since it has shown highest accuracy among all kernels which is 83%.

In the model training process, a dataset of IT industry related most recent job posting has been used and dataset has been processed and arranged along with the career path and the required skills. In the data processing process, 3rd party python library called resume parser has been used to extract the skills from the job postings.

The developed models used to identify the most suitable career path for the student. Model outcome was utilized through the job recommendation engine to recommend the current career opportunities to the student.

5. Conclusion

In this paper SVM based multiclass classification model was implemented, which can be used to identify the most suitable career path based on the student's skills and the personal interest. Career paths has been narrowed down to which can start as a fresher to the industry.

In the career opportunities recommendation, student core skills which are identified during the skill identification phase, student personal interests and the student experience level also considered.

There are some existing studies focused on resume data-based job recruitment which follows text extraction and classification techniques.[11] The same approach is followed in this study to build an effective career guidance system based on the university student's skills and the interests

This research is aimed to avoid the student getting wrong decisions when they are stepping into industry, due to impact of outside factors such as family pressure, friend's career, and social pressure.

Model outcomes can be used to give an efficient guidance for undergraduate students who are waiting to step into the IT industry. It will provide a reliable guidance for the student during the career path selection and addition to that finding a suitable career opportunity.

As for future directions, due to the lack of time and resources, various improvements, tests, and experiments are left for the future. Future works will include the recommendation of critical/ necessary skills for the student that need to be focused on based on the suggested career path.

6. REFERENCES

- [1] P. Akosah-Twumasi, T. I. Emeto, D. Lindsay, K. Tsey, and B. S. Malau-Aduli, “A Systematic Review of Factors That Influence Youths Career Choices—the Role of Culture,” *Front. Educ.*, vol. 3, 2018, Accessed: Jun. 11, 2022. [Online]. Available: <https://www.frontiersin.org/article/10.3389/educ.2018.00058>
- [2] S. Vignesh, C. Shivani Priyanka, H. Shree Manju, and K. Mythili, “An Intelligent Career Guidance System using Machine Learning,” in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, Mar. 2021, pp. 987–990. doi: 10.1109/ICACCS51430.2021.9441978.
- [3] T. R. Razak, M. A. Hashim, N. M. Noor, I. H. A. Halim, and N. F. F. Shamsul, “Career path recommendation system for UiTM Perlis students using fuzzy logic,” in *2014 5th International Conference on Intelligent and Advanced Systems (ICIAS)*, Kuala Lumpur, Malaysia, Jun. 2014, pp. 1–5. doi: 10.1109/ICIAS.2014.6869553.
- [4] A. Shankhdhar, A. Agrawal, D. Sharma, S. Chaturvedi, and M. Pushkarna, “Intelligent Decision Support System Using Decision Tree Method for Student Career,” in *2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC)*, Mathura, Uttar Pradesh, India, Feb. 2020, pp. 140–142. doi: 10.1109/PARC49193.2020.246974.
- [5] B. Patel, V. Kakuste, and M. Eirinaki, “CaPaR: A Career Path Recommendation Framework,” in *2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService)*, Redwood City, CA, USA, Apr. 2017, pp. 23–30. doi: 10.1109/BigDataService.2017.31.
- [6] T. Gonzalez *et al.*, “Adaptive Employee Profile Classification for Resource Planning Tool,” in *2012 Annual SRII Global Conference*, San Jose, CA, USA, Jul. 2012, pp. 544–553. doi: 10.1109/SRII.2012.67.
- [7] R. Shaikh, N. Phulkar, H. Bhute, S. K. Shaikh, and P. Bhapkar, “An Intelligent framework for E-Recruitment System Based on Text Categorization and Semantic Analysis,” in *2021 Third International Conference on Inventive Research in Computing*

Applications (ICIRCA), Coimbatore, India, Sep. 2021, pp. 1076–1080. doi: 10.1109/ICIRCA51532.2021.9544102.

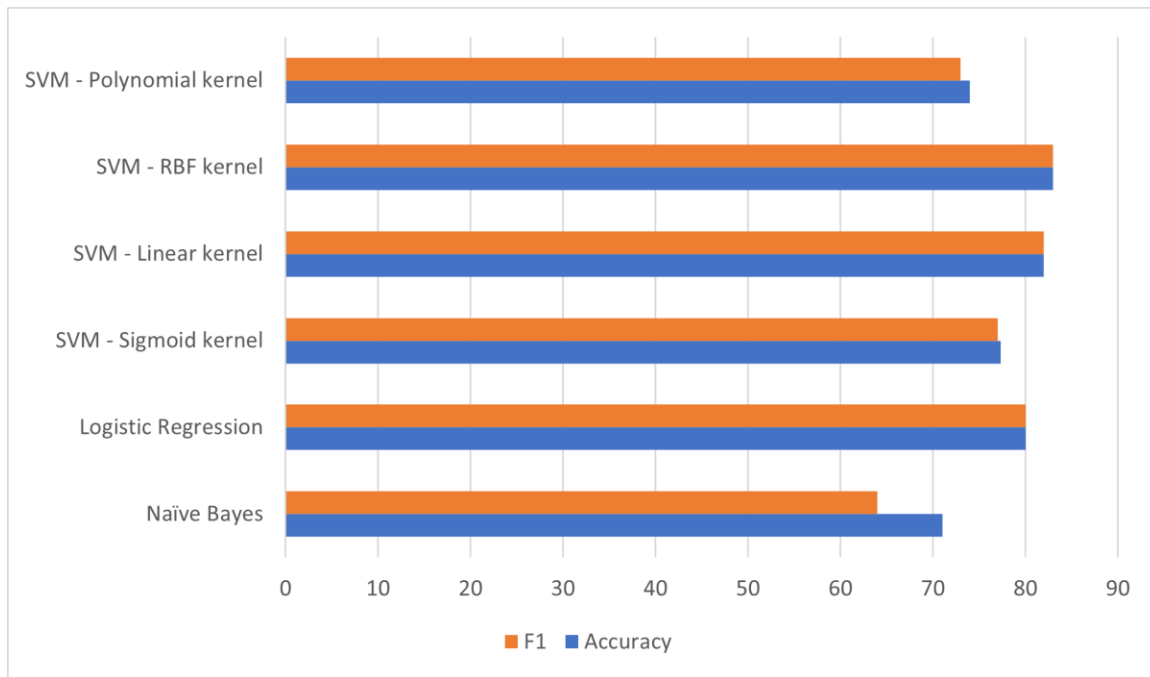
[8] F. Javed, Q. Luo, M. McNair, F. Jacob, M. Zhao, and T. S. Kang, “Carotene: A Job Title Classification System for the Online Recruitment Domain,” in *2015 IEEE First International Conference on Big Data Computing Service and Applications*, Redwood City, CA, Mar. 2015, pp. 286–293. doi: 10.1109/BigDataService.2015.61.

[9] A. Sharma, “A proposed e-learning system facilitating recommendation using content tagging and student learning styles,” in *2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH)*, Hyderabad, India, Aug. 2017, pp. 1–6. doi: 10.1109/ELELTECH.2017.8074989.

[10] baeldung, “Multiclass Classification Using Support Vector Machines | Baeldung on Computer Science,” Oct. 07, 2020. <https://www.baeldung.com/cs/svm-multiclass-classification> (accessed Jul. 02, 2022).

[11] R. Pal, S. Shaikh, S. Satpute, and S. Bhagwat, “Resume Classification using various Machine Learning Algorithms,” *ITM Web Conf.*, vol. 44, p. 03011, 2022, doi: 10.1051/itmconf/20224403011.

APPENDICES



Appendix 1 - Model Accuracy Comparison

```
def keywords_extract(self, text):  
    processed_text1 = text.replace(".", "")  
    processed_text2 = processed_text1.replace(",", "")  
    text = processed_text2.lower().split()  
    stops = set(stopwords.words("english"))  
    text = [w for w in text if not w in stops]  
    custom_stop_words = ["knowledge", "work", "ability"]  
    text = [w for w in text if not w in custom_stop_words]  
    text = list(set(text))  
    keywords = resumeparse.extract_skills(' '.join(text))  
    return keywords
```

Appendix 2 – Keyword Extraction

```

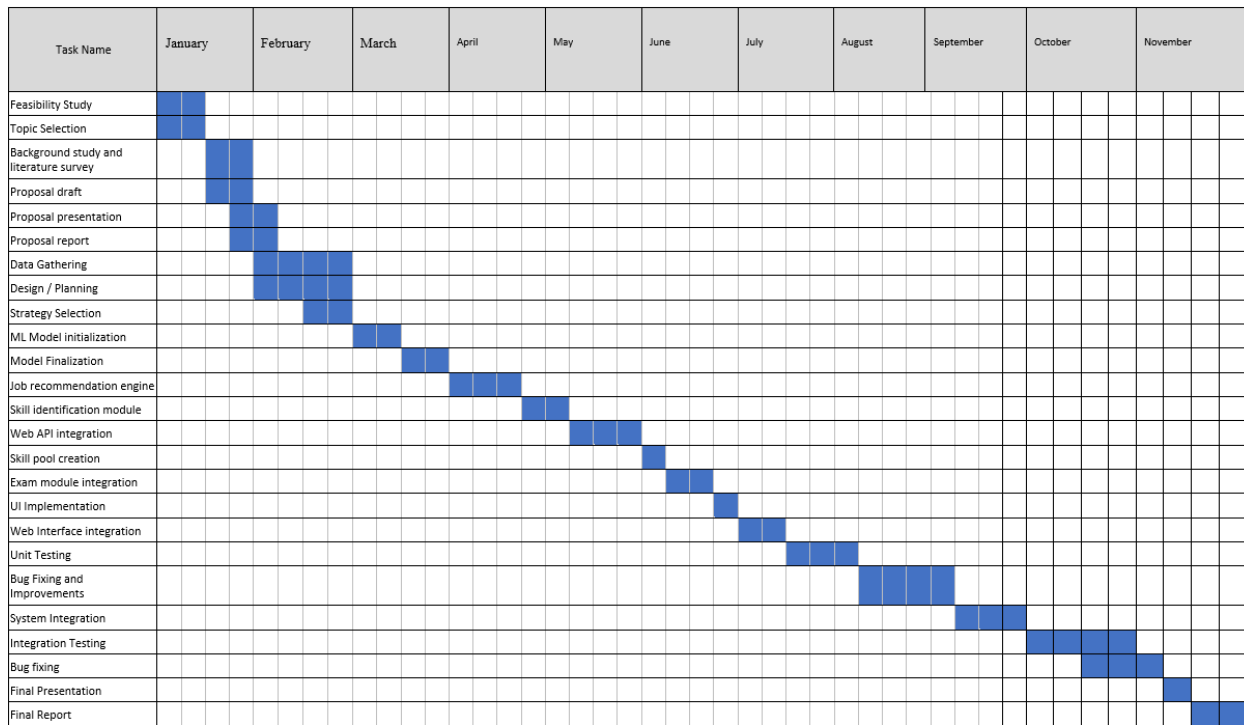
from sklearn.pipeline import Pipeline
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.metrics import classification_report
from sklearn import svm

sgd = Pipeline([('vect', CountVectorizer()),
                 ('tfidf', TfidfTransformer()),
                 ('clf', svm.SVC(kernel='rbf', gamma='scale', C=1)),
                 ])

sgd.fit(X_train, y_train)
y_pred = sgd.predict(X_test)
val2 = sgd.predict(skill)
tp = confusion_matrix(y_test, y_pred).ravel()
tn = tp
fp = tp
fn = tp
precision_score = tp / (tp + fp)
recall_score = tp / (tp + fn)

from sklearn.metrics import f1_score
rbf_f1 = f1_score(y_test, y_pred, average='weighted')
print('F1 (RBF Kernel): ', "%.2f" % (rbf_f1*100))
print("Support Vector Machine:", val2[0])
print('precision_score:', precision_score)
print('recall_score:', recall_score)
print('Confusion Matrix:', confusion_matrix(y_test, y_pred))
print('accuracy %s' % round(accuracy_score(y_pred, y_test), 3) + '\n')

```



Appendix 4- Project Gantt Chart

