

AI BASED STUDENT CAREER PREDICTION USING COGNITIVE BEHAVIOUR

THESIS REPORT

*Submitted in partial fulfillment of the requirements for the award of the
Degree of Master of Technology in Department of Electronics and
Communication Engineering with specialization in Communication Systems by
the A P J Abdul Kalam Technological University*

by

M G PRIYANKA

TKM20ECCS09



DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING

TKM COLLEGE OF ENGINEERING

KOLLAM 691 005

JULY 2022

AI BASED STUDENT CAREER PREDICTION USING COGNITIVE BEHAVIOUR

THESIS REPORT

*Submitted in partial fulfillment of the requirements for the award of the
Degree of Master of Technology in Department of Electronics and
Communication Engineering with specialization in Communication Systems by
the A P J Abdul Kalam Technological University*

by

M G PRIYANKA

TKM20ECCS09



DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING

TKM COLLEGE OF ENGINEERING

KOLLAM 691 005

JULY 2022

DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING

TKM COLLEGE OF ENGINEERING

KOLLAM 691 005



CERTIFICATE

*Certified that this thesis report titled “AI BASED STUDENT CAREER PREDICTION USING COGNITIVE BEHAVIOUR” is a bonafide record of the work done by **M G PRIYANKA** (Reg. No. TKM20ECCS09) under my supervision, in partial fulfillment of the requirements for the award of the Degree of Master of Technology in Electronics and Communication Engineering with specialization in Communication Systems by the A P J Abdul Kalam Technological University.*

Guide

Dr. ANZAR S M

ASSISTANT PROFESSOR

Dept. of ECE, TKMCE

Coordinator

Dr. NISHANTH N

ASSOCIATE PROFESSOR

Dept. of ECE, TKMCE

HoD

Prof. ABID HUSSAIN

Head, Dept. of ECE

TKMCE

Acknowledgement

At the outset, I find it my obligation to thank the Almighty God for giving me necessary wisdom to complete this Project First Phase presentation successfully.

I thank **Prof. ABID HUSSAIN MUHAMMED**, HOD, Department of Electronics and Communication, for his encouragement and support.

I express my sincere thanks to our PG coordinator, **Dr. NISHANTH N**, Associate Professor, Department of Electronics and Communication Engineering, for the support and encouragement during the course of this presentation.

I take this opportunity to express my sincere gratitude and profound thanks to my guide, **Dr. ANZAR S M**, Assistant Professor, Department of Electronics and Communication, for his advice, supervision and patience during the course of project preparation and presentation and for providing me guidance and critical inputs in the preparation and presentation of my project.

I also express my heartfelt thanks to all my teachers, friends and my parents for providing the much-needed support during the course of preparation and presentation of project.

M G PRIYANKA

TKM20ECCS09

ABSTRACT

One of the most important decisions we make in life is our career. This involves much more than what we do for a living. It cannot be stressed enough, so it is important to choose a career that we are happy with. Most of us are not as fortunate as others who simply know what they want to do and work in fulfilling jobs without thinking about it. Many people choose a profession without thinking about it or for the wrong reasons. They may choose professions that seem safe or lucrative. In this way they become dissatisfied. A well thought-out decision is the best approach to ensure that this does not happen to us.

The proposed work paves the way for students to choose their ideal career at school age. Bloom's Taxonomy forms the basis of the work. A hierarchical paradigm called Bloom's Taxonomy divides learning objectives into different levels of complexity, ranging from basic understanding and knowledge to complicated assessment and production. Bloom's Taxonomy includes three domains of learning - cognitive, emotional and psychomotor. Within each domain, learning can occur at different levels, from basic to advanced. The cognitive domain is primarily concerned with intellectual skills such as critical thinking, problem solving and knowledge building. The learner's attitudes, values, interests and appreciation are the core issues of the affective domain. The learners' ability to physically perform activities, execute movements and apply skills falls under the psychomotor domain. In this case, the cognitive domain is used in our work. Holland's hypothesis forms the basis of the proposed model. The input of the model is questions corresponding to RIASEC and the output is academic skills and occupations based on the results of academic skills. The model was developed in two phases. In the first phase, academic skills were predicted using Holland's system. Based on the academic skills acquired, the second stage involved the prediction of career. Since there was no pre-existing data set, we collected the necessary information ourselves. A Google form with 30 questions (5 questions each for Holland's code) was distributed to students and other participants. The Google form also included a field for the applicant's name and job title. 159 people answered our questions and we used the resulting dataset for our work. The required dataset was collected and then

set aside for preparation. Then the dataset was entered into Google Colab. After pre-processing, Decision Tree, Random Forest, K-Nearest Neighbour and Multioutput Regressor algorithms were used to build machine learning models. The model was trained on 80% of the collected data and then tested on the remaining 20%. When comparing the accuracy of each model, the multi-output regressor showed a maximum accuracy of 99.9%. Thus, the academic skills were successfully predicted. The next step was to determine the ideal occupation based on the academic skills acquired. Support Vector Machine, Gaussian Naive Bayes, Perceptron, Decision Tree and Random Forest were used to build machine learning models. Along with the academic skills acquired in the first phase of the work, the occupational information from the dataset was used to train and test the model. When comparing the model accuracy in this case, the SVM had the highest accuracy of 78.125 % and the career was effectively predicted. Python in Google Colab was used for all this. We also tried applying Synthetic Minority Oversampling Technique(SMOTE) to career prediction. For some of the models, the accuracy decreased, while for others it improved. Of the five models used, Gaussian Naive Bayes and Random Forest showed a decrease in accuracy, while Perceptron achieved the highest accuracy of 80%.

Contents

List of Figures	v
List of Tables	vi
1 Introduction	1
2 Literature Review	9
3 Proposed Methodology	15
4 Results and Future Scope	22
5 Conclusion	27
References	29

List of Figures

1.1	Six hierarchial levels of Cognitive domain.	4
1.2	Holland's RIASEC	6
3.1	Proposed model.	16
3.2	Example of RIASEC test	17
4.1	Pie chart of different models	23
4.2	Confusion matrix obtained for SVM model before SMOTE	24
4.3	Confusion matrix obtained for perceptron after SMOTE	25

List of Tables

4.1	Academic Skill Prediction - Model Accuracies	22
4.2	Accuracy of models before applying SMOTE	24
4.3	Accuracy of models after applying SMOTE	24

Chapter 1

Introduction

To choose a career is to choose the life we lead. All facets of our lives - personal, social and professional - are affected by our careers. Careers shape who we are. The type of career we have determines our everyday life. Our career choices affect everything from our social standing to our friends, lifestyle, neighbourhood and social circle. The time we spend with our loved ones, friends and even ourselves depends on our career choice. The amount of money we can spend on things like food, clothing, shelter, entertainment, etc. depends on how much money we earn from our jobs. The comforts and difficulties we experience at work depend on the industry or profession we choose. If we choose a job we like, we do it impeccably and receive praise and incentives that increase our self-esteem. If we choose a job that we do not find fulfilling, we will eventually become bored with our work, our colleagues and life in general. This leads to irritation and dissatisfaction with work and life. This leads to a range of emotional and physical problems such as anger and despair, as well as headaches, stomach aches, heartburn, ulcers, high blood pressure, nervousness and fatigue. A career is more than just a means to make money. A career must be pursued sincerely, passionately and with joy. The wrong career or a position that does not suit a person can occasionally lead to a number of major problems in life, including health complaints, psychological problems such as stress and depression, and a loss of interest in hobbies and social activities. Sometimes, it can even lead to drug or alcohol addiction. Therefore, it is crucial to find the ideal career for a person as early as possible in their academic career. With the right profession, success and joy in life can be achieved.

A collection of three hierarchical models known as Bloom's Taxonomy is used to categorise educational learning goals into levels of complexity and specificity. Learning objectives for the cognitive, affective, and psychomotor domains are covered by the three lists. The models bear the name of Benjamin Bloom, who served as chair of the committee of educators that created the taxonomy. In the 2001 revision of Bloom's taxonomy, the stages are named slightly differently and listed in a new order: Remembering, Understanding, Applying, Analysing, Evaluate, and Creating (rather than Synthesize).

1. Cognitive Domain

The cognitive domain includes learning and improving cognitive skills. It includes the ability to remember or recognise specific facts, logical processes, and conceptual frameworks that support the growth of intellectual skills and talents. The six levels of the cognitive domain are as follows:

- **Knowledge:** Knowing things without necessarily knowing what they mean involves recognising or remembering facts, terminology, basic ideas, or solutions. Some characteristics might be:
 - specific knowledge of vocabulary and facts.
 - understanding strategies for coping with specifics, including habits, patterns, and categories
 - understanding of the generalisations and abstractions of a field, including its guiding principles, theories, and organisational structures
- **Comprehension:** By organising, summarising, translating, generalising, describing, and expressing the essential thoughts, understanding means showing an understanding of facts and ideas.
- **Application:** Application is the process of using newly learned skills to solve problems in new circumstances. This requires the application of learned knowledge, facts, strategies, and rules. Students should be able to use existing knowledge to solve problems and recognise relationships in novel contexts.

- **Analysis:** Analyzing information means breaking it down into its component parts, figuring out how they relate to each other, determining motivations or reasons, drawing conclusions, and gathering data to support assumptions. It has the following characteristics:
 - Element analysis
 - Relationship analysis
 - Organizational analysis

- **Synthesis:** Synthesis is the process of combining different elements to create a new structure or pattern. It can also refer to the combination of parts into a whole or the combination of knowledge into a new meaning. It has the following characteristics:
 - the creation of a distinctive message
 - the creation of a plan or proposal for action
 - the creation of a collection of abstract relations.

- **Evaluation:** By judging facts, the reliability of theories, or the quality of a work according to certain standards, evaluation involves the presentation and defence of points of view. It has the following characteristics:
 - Decisions based on internal evidence
 - Decisions based on external standards

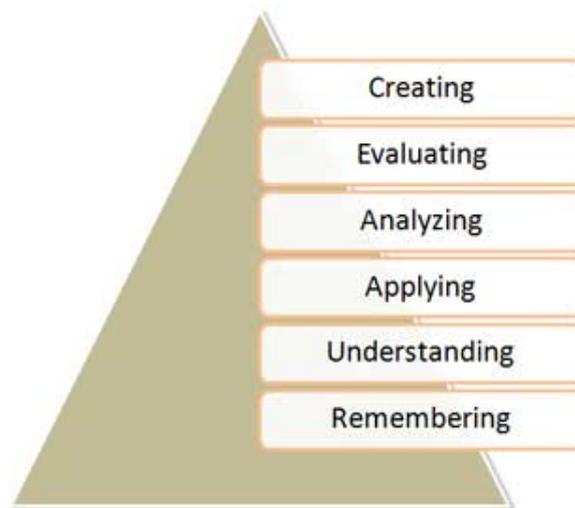


Figure 1.1: Six hierarchial levels of Cognitive domain.

2. Affective Domain

The ability to feel the joy or sorrow of other living beings is one of the abilities in the affective domain. Affective goals often focus on promoting awareness of and development of attitudes, emotions, and feelings. The affective domain includes five levels, from the lowest to the highest processes.

- **Receiving:** lowest level; the student observes only passively. Without this stage there can be no learning. Receiving also includes the student's ability to remember and recognise.
- **Responding:** The learner actively participates in the learning process by responding to and paying attention to the stimuli in some way.
- **Valuing:** The student gives value to anything, a phenomenon or a piece of knowledge. The knowledge learned is associated by the student with a value or set of values.
- **Organizing:** The learner compares, relates, and elaborates what he has learned; he is able to synthesise different values, knowledge, and ideas and incorporate them into his own scheme.
- **Characterizing:** In this phase, the learner strives to develop abstract knowledge.

3. Psychomotor Domain

The ability to physically use a tool or instrument, such as a hand or hammer, is called a psychomotor skill. Most psychomotor goals target behaviour change, skill improvement, or all three.

- **Perception:** The ability to use sensory cues to control motor activity includes all aspects of sensory stimulation, cue selection, and translation.
- **Set:** It contains sentences for the mind, the body, and the heart. These three groups of dispositions determine how a person will react in certain situations (sometimes called "mindsets"). This psychomotor subdivision is closely related to the "reacting to phenomena" subdivision of the affective domain.
- **Guided Response:** Early imitation and trial-and-error learning phases for a complex skill: practice is the key to adequate performance.
- **Mechanism:** The second phase of mastering a demanding skill: the actions learned have now become second nature, and they can be performed with great confidence and competence.
- **Complex Overt Response:** The expert execution of motor actions with complicated movement patterns: a fast, precise, and well-coordinated performance that uses as little energy as possible is a sign of skill. This category includes effortless execution and automatic execution.
- **Adaptation:** The person has well-developed skills and can adapt his or her movement patterns to the requirements at hand.
- **Origination:** Develop new movement patterns to address a specific problem or situation: Learning outcomes place a strong emphasis on creativity that builds on advanced skills.

The work done here is in Cognitive domain of Bloom's taxonomy. The proposed model is based on John L. Holland's code created. A system for classifying jobs into occupational categories, interest groups, or work environments is called Holland Codes. These categories correspond to work personalities in Holland's model.

Holland's theory of career choice states that choosing a job or educational institution whose culture matches or is similar to one's personality is most likely to lead to success and happiness. He established the RIASEC code, which lists the six personality types:



Figure 1.2: Holland's RIASEC

- **Realistic (R):**Realistic people like physical activity, stability, and manual labour-intensive activities such as construction, maintenance, operating machinery, and sports. They prefer to work with objects rather than concepts and people. They like to play sports and often spend time outdoors where they come in contact with animals and plants. People who belong to this group usually prefer "learning by doing" in a real, task-oriented environment rather than spending a lot of time in a classroom. Realistic people value material possessions and interact in an open, direct way. They believe they are good at mechanical and physical tasks, but may feel awkward or less skilled when it comes to interpersonal interactions. Realistic jobs are desirable because they foster technical skills and allow employees to produce results they can see. The military, veterinarians, engineers, and electricians are examples of realistic jobs.
- **Investigative (I):**People who like to study or are involved in mathematics or science are analytical, intelligent, and perceptive. They are attracted to obscure problems, and an overly organised environment can limit them. This group of people likes to apply logic to solve extremely difficult, abstract problems. Investigative types often work independently and do not seek leadership positions be-

cause they are introspective and motivated by innovative problem solving. They value education and science and consider themselves intelligent people with scientific or mathematical abilities, but lack leadership and persuasive skills. The ideal workplace for inquiry types fosters scientific skills, allows for individualised work, and focuses on finding new solutions to abstract, challenging problems. Medical technologists, biologists, chemists, and systems analysts are common investigative occupations.

- **Artistic (A):** Artistic people are innovative, perceptive, and inventive and enjoy pursuing creative activities such as writing, acting or directing, making music, or playing an instrument. They seek opportunities to express themselves through artistic creation. This group of people rejects convention and uniformity and prefers freedom and uncertainty. Artists are usually emotional and impulsive people who are very expressive and open in their communication. They respect aesthetics, consider themselves creative and non-conformist, and value or possess writing, dramatic, musical, or artistic skills, but do not possess bureaucratic or organisational qualities. The ideal workplace for creative types encourages innovation, the use of imagination, and the development of creative skills in a fluid, unstructured environment. Musicians, journalists, and interior designers are examples of common artistic professions.
- **Social (S):** Social people are humanistic, idealistic, responsible, and concerned about the welfare of others. They enjoy participating in group activities and educating, rehabilitating, guiding, or developing others. They usually care about interpersonal relationships, enjoy participating in social activities, and work well under pressure. Social personalities seek opportunities to work with others, solve problems through dialogue, and apply interpersonal skills, but may shy away from tasks that require repetitive use of machines or equipment. They speak kindly and diplomatically and have the ability to persuade others because they genuinely enjoy working with people. They see themselves as patient, kind, cheerful, and capable teachers, but lack mechanical aptitude. The social type prefers a workplace that encourages collaboration and allows for extensive social interactions. Teachers, counsellors, and social workers are examples of common

social occupations.

- **Enterprising (E):**Entrepreneurial people are enthusiastic, ambitious, daring, gregarious, and self-confident. They are attracted to leadership positions and love tasks that require them to influence others, such as sales. Although they dislike monotonous or methodical activities, they feel energised by using their interpersonal skills, leadership abilities, and persuasive powers to achieve organisational goals or financial benefits. Although they tend to be friendly and speak competently in public, they may come across as arrogant. Although lacking in scientific knowledge, they see themselves as strong, confident, and effective in leadership and public speaking. The entrepreneurial type prefers a work environment that rewards them with money, status, and power and encourages them to engage in tasks such as management, sales, and leadership. Salespeople, business leaders, and managers are common entrepreneurial occupations.
- **Conventional (C):**Efficiency, prudence, conformity, organisation, and conscientiousness are qualities that characterise conventional people. They are comfortable working under a set line of command and prefer to follow clear instructions rather than take leadership responsibility. They dislike ambiguity and prefer structured, orderly activities. They are adept at managing and manipulating data, setting up schedules, and using office technology, which they often enjoy. They are meticulous, persistent, and reliable, although they rarely aspire to leadership or a role in the limelight. Conventional types consider themselves capable of handling office, organisational, and numerical tasks and are responsible, efficient, and orderly. They may also believe themselves to lack imagination or originality. The conventional workplace is preferred because it promotes organisational skills such as data management and record keeping in a regulated environment, and reliability and compliance are highly valued. Secretaries, accountants, and bank clerks are examples of typical conventional jobs.

Chapter 2

Literature Review

To produce graduates for the digital workforce, Pongsaton *et al.* in [1] built the architecture of the intelligent career prediction system based on cognitive technology. The research techniques were divided into three phases:

1. Intelligent Career Prediction System Composition Synthesis.
2. Development of an intelligent career prediction system based on cognitive technologies to support university graduates in their career choice.
3. A determination of whether the design of the cognitive technology-based intelligent career prediction system is appropriate for creating graduates for the digital workforce.

Three components form the architecture of the intelligent career prediction system based on cognitive technology:

1. Experimental data were collected from students, teachers, supporting faculty, technical assistants and office administrators.
2. The user management, the forecast data management, the forecast management system and the forecast display system form the architecture of the intelligent occupational forecasting system.
3. Nine experts in intelligent career prediction systems and cognitive technology have evaluated the suitability of the design of the intelligent career prediction

system based on this technology for helping the graduates to chose the profession.

Mean and standard deviation are the statistics used in the study. The results of the evaluation show that the designed architecture was the most appropriate with a combined mean of 4.54 and standard deviation of 0.49.

The relationship between personality, expressed as HEXACO dimensions, and job selection was shown by Madhura *et al.* in [2] for 356,574 applicants from 10 work sectors. To find groups of applicants with comparable HEXACO dimension value combinations, they use Self-Organising Maps (SOM), a kind of unsupervised machine learning technique.

Three conclusions are drawn from the study:

- They discovered discrete personality domains associated with several career categories, including sales (customer facing), sales (call centre), graduate and administrative occupations, demonstrating a link between personality and career aspirations.
- Compared to positions like retail assistant and retail manager, roles such as sales assistant and graduate had clear regional demarcations.
- Across all applicants, they could not identify any particular number of macro-level personality groups.

This shows that there are no clear larger groups of HEXACO personality variables representing a specific number of personality types, although there are smaller proximal personality clusters. These results show how personality differences in individual traits influence career choice, a kind of bias in career choice.

This idea is also supported by existing theories such as Trait Activation Theory and the Attraction-Selection-Attrition model.

Giorgia *et al.* presented in [3] a study on the effectiveness of the b-test, an instrument used to detect individuals who misrepresent cognitive symptoms, in detecting

individuals who misrepresent mild cognitive impairment. They included three participant groups: healthy seniors (controls, $n = 21$), patients with mild cognitive disorder (malingers, $n = 21$), and patients with mild neurocognitive disorder (patients, $n = 21$). In addition to the b-test, they received two baseline neuropsychological assessments (MMSE, FAB). The malingers performed significantly worse than the patients and the control group on all error scores, and similarly to the patients on the time score. Malingers also scored significantly worse than the control group on all error scores. Patients scored significantly worse than controls in all categories, but both groups showed the same pattern of more errors of omission than errors of commission. The malingers, on the other hand, showed the opposite pattern, namely more committing than omitting errors. The machine learning models achieve an overall accuracy of more than 90% in distinguishing patients from malingers based on the b test results alone. The results suggest that the error scores of the b test effectively discriminate patients with mild neurocognitive disorders from malingering and can be used in conjunction with other tested methods such as the Medical Symptom Validity Test.

According to Graziella *et al.* in [4], incorporating machine learning-based analyses into the analytic workflow of psychological studies will maximise accuracy and reduce replicability problems. ML analysis of experimental data is model-independent and, unlike statistical inference, focuses largely on predictions. They also point out some potential dangers of using machine learning-based analysis of experiments. If not used properly, it can lead to optimistic predictions, as was the case when statistical inference was used. Solutions to these problems are also offered, such as the development of models with cross-validation and the use of ensemble models. ML Models are often considered "black boxes" and this article looks at methods that make predictions more transparent.

By incorporating machine learning methods for personality assessment into a more thorough construct validation framework, Wiebke *et al.* in [5] expanded the potential of these methods. They reviewed recent applications of machine learning for personality assessment and placed learning about research in the broader context of basic

construct validation and provide suggestions for how machine learning can be applied to advance knowledge about personality.

In [6], Jeromy *et al.* reviewed the major personality taxonomies and provided an overview of the considerations for measuring personality. They offer a comprehensive list of suggestions on how and when to use personality tests for academics and professionals. Various taxonomic representations of personality structure were discussed, focusing on hierarchical representations and five- and six-factor models such as the Big Five and the HEXACO model. They discussed the many advantages and disadvantages of each strategy. The article they read argues that the Big Five model is a useful descriptive framework for analysing personality in general, and discusses the main reasons for its popularity. They advise practitioners and researchers to consider additional taxonomic representations of personality such as the HEXACO. They suggest a number of cases where alternatives to the Big Five framework are more appropriate, and offer suggestions both for selecting measures in general and for conducting personality trait studies. While the Big Five is a good general personality framework that is acceptable in many contexts, researchers and practitioners should be aware of different measurement methods and use them where appropriate.

172 students participated in the study by Melissa *et al.* in [7] in which they took tests assessing their mathematical ability (the Math Computation Subtest of the Wide Range Achievement Test- IV), math anxiety (the Math Anxiety Rating Scale-Revised), general test anxiety (from the Adult Manifest Anxiety Scale-University version), and the three Working Memory Index tasks from the Wechsler Adult Intelligence Scale- IV Edition (WAIS-IV; Digit Span [DS Beyond the effects of gender, general test anxiety and mathematics knowledge, the results showed that mathematics anxiety predicted performance in arithmetic, but not DS or LNS. The results suggest that scores on the WAIS-IV working memory subtest may be negatively affected by arithmetic anxiety. The use of LNS in people who show significant arithmetic anxiety is a therapeutic intervention.

Andreas *et al.* [8] investigated whether creating more effective short tests us-

ing multilayer neural networks can improve personality assessment. The Big Five Inventory (BFI-10), the HEXACO Personality Inventory-60 and the International Personality Item Pool-300 item version (IPIP-300 or IPIP-NEO) were used to collect personality information from $N = 3,498$ participants from Germany, the USA and the UK (HEXACO-PI -R). Based on this data, they trained 40 multilayer neural networks to predict the Big 5 personality dimensions as well as the facet scores of a 30-item IPIP version. IPIP30-NNet, a neural network-based short test version, accurately predicted the Big 5 dimensions from IPIP-300 as well as their facets. Using a joint averaging approach and multiple regression, the correlations with the long test (IPIP-300) scores were much higher than with the short test scores.

The predictive validity of the IPIP30-NNet improved significantly, especially for aspect scores. In addition, they created Big 5 "superscores" as a combination of all three personality tests, which were very accurate predictors of IPIP30-NNet. The results show that neural network-based diagnostics can be used to create an extremely detailed individual personality profile from a small amount of data. They talk about the difficulties, possibilities and future prospects of using machine learning to improve conventional psychological tests.

One of the best-known career theories, John L. Holland's theory of career choice, is used by academics and professionals around the world. According to this theory, people should look for jobs that match their career interests if they want to be happy and successful. Implementing this in study and practise requires the calculation of coefficients that quantify the core ideas of this thesis, including the fit between person and environment. Holland, a freshly released R package, attempts to provide a comprehensive collection of hints, explanations and calculations of the most important coefficients. Florian G. Hartmann *et al.* in [9] describes the package in the current publication, which also addresses how it can be used in practise and research. For this purpose, the functions of the package are used and discussed. Suggestions are also made when there are many coefficients for the same theoretical concepts, and functions for upcoming versions are discussed. The Holland R package is a promising computational environment for Holland's main theoretical ideas, containing many coefficients.

The core of Holland's (1997) theory of career types is the idea of congruence between person and occupation, according to Donald E. Eggerth *et al.* in [10]. The C-index is probably the best instrument currently available for assessing the congruence of two Holland Code profiles. The postulated hexagonal structure of the Holland RIASEC model is reflected by the C-index, which is also easy to calculate and sensitive to code order. However, the C-index is only intended for comparing Holland code profiles. The Strong Interest Inventory (SII) and the Occupational Information Network (O* NET), both of which assign one- to three-letter Holland codes, do not match this, although they are compatible with the instruments and accompanying materials developed by Holland and his colleagues. Consequently, using the C-index with the SII or the O* NET is difficult. Furthermore, the authors claim that it is arbitrary to calculate congruence consistently with three-letter Holland codes and that congruence should only be calculated with Holland species that are evident in the profiles being compared. The modifications to the C-index proposed in this article retain the desirable properties of the original C-index, including reflection of hexagonal structure, sensitivity to order, and ease of calculation, and allow comparisons to be made between Holland code profiles of different lengths and/or of less than three letters in length.

Chapter 3

Proposed Methodology

As explained earlier, the proposed model is based on Holland's theorem. A taxonomy of interests [1] based on a theory of occupations and occupational choice, first developed by American psychologist John L. Holland, is known as Holland Codes or Holland Occupational Themes. The six types Holland originally identified were "motor, intellectual, aesthetic, supportive, persuasive, and adaptive." He later expanded them and changed their names to "Realistic (Doers), Investigative (Thinkers), Artistic (Creators), Social (Helpers), Enterprising (Persuaders), and Conventional (Organisers)." The six categories listed by Holland are somewhat related to each other. Since the initial letter of the region is identical to R-I-A-S-E-C when expressed as a circle connecting the regions with strong correlation, it is called the RIASEC model or hexagonal model.

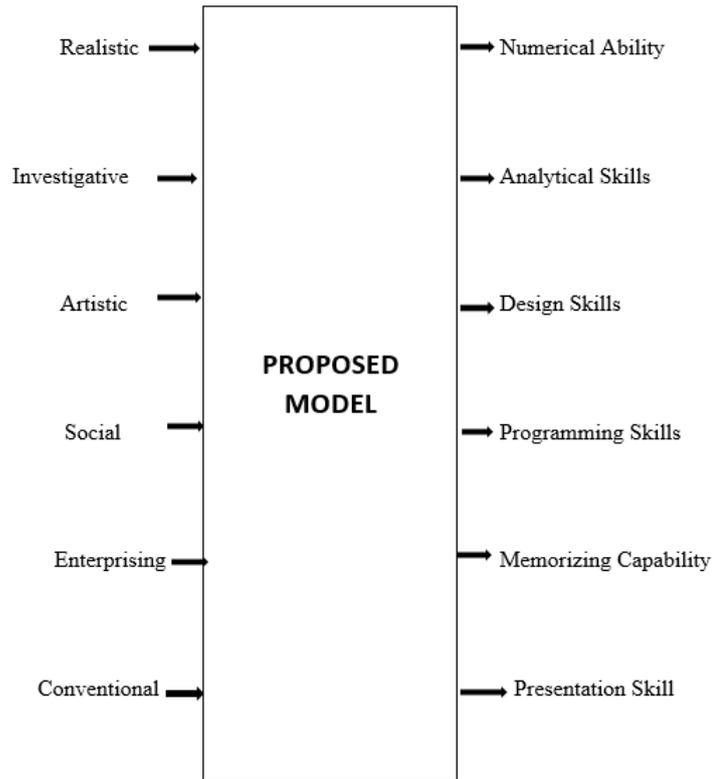


Figure 3.1: Proposed model.

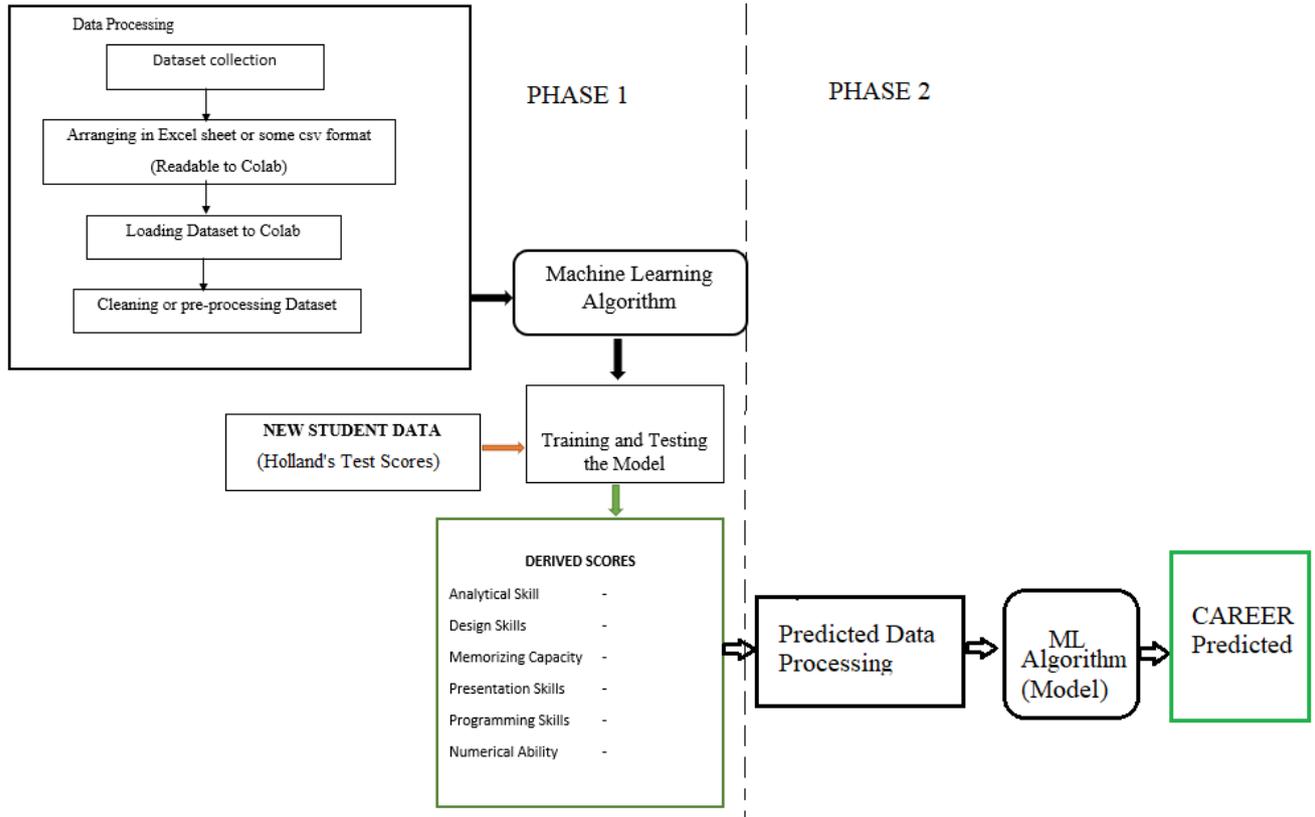
Holland Personality Styles Assessments provide a link between job families and career interests. A two- or three-letter Holland or RIASEC code is used for the assessments. Correlations between career personalities and essential characteristics, fields of study, interests, skills, and related occupations are determined by various tests. Because a data set was not yet available, a set of 30 questionnaires with 5 questions per trait, approved by a psychiatrist, was distributed via Google Forms. The RIASEC score was determined using the collected data set. Figure 3.2 shows an example of the RIASEC test that was discussed. We can determine the scores of each personality type - Realist, Researcher, Artist, Social, Entrepreneur, and Conventional - by comparing the test scores to Holland's score.

We chose to work with Google Colaboratory, a fully cloud-hosted, free Jupyter notebook environment in which to create and run Python programmes. The following Python libraries are needed for this machine learning project:

- **NumPy:** The Python package NumPy is used to manipulate arrays. It also has matrices, Fourier transform and functions for linear algebra work. NumPy was developed in 2005 by Travis Oliphant. You can use it for free, since it is an open-source project.
- **TensorFlow:** TensorFlow provides a choice of methods for creating and refining models in JavaScript or Python. It is a machine learning and artificial intelligence software library that is open source and free.
- **Pandas:** For processing and analysing data, the Python programming language has a software package called pandas. It contains specific data structures and procedures for working with time series and mathematical tables.
- **Matplotlib:** For 2D representations of arrays, Matplotlib is a fantastic Python visualisation library. A cross-platform data visualisation package called Matplotlib was created to deal with the larger SciPy stack, and is based on NumPy arrays.
- **Seaborn:** A Python library for data visualisation based on Matplotlib is called Seaborn. It provides a sophisticated drawing tool for creating eye-catching and instructive statistical graphs.

The machine learning model receives the output of Holland test for testing and training. From this, the academic skills score is determined. Academic skills acquired include numerical skills, analytical skills, design skills, programming skills, memory skills, and presentation skills. Depending on how well each skill is developed, career progression is predicted.

Two phases of construction were completed. The first phase involved data processing, building a machine learning algorithm, testing and training the model, and finally collecting the results for academic capabilities. Data collection, uploading the



data to Google Collaboratory, and preprocessing the data were part of the data processing procedure.

The second phase of the project began with the analysis of data on academic skills projected. After processing, a machine learning model will be created, trained and tested to predict the best career.

The first step was to collect the required data set. Since no dataset was available, a 30-question survey was distributed via Google Forms to obtain RIASEC values. There were a total of five questions, five for each attribute. The questions asked were:

1. I like to work on cars.
2. I like to build things.
3. I like to take care of animals.
4. I like putting things together or assembling things.

5. I like to cook.
6. I like to do puzzles.
7. I like to do experiments.
8. I enjoy science.
9. I enjoy trying to figure out how things work.
10. I like to analyze things (problems/situations).
11. I am good at working independently.
12. I like to read about art and music.
13. I enjoy creative writing.
14. I am a creative person.
15. I like to play instruments or sing.
16. I like to work in teams.
17. I like to teach or train people.
18. I like trying to help people solve their problems.
19. I enjoy learning about other cultures.
20. I like to get into discussions about issues.
21. I am an ambitious person, I set goals for myself.
22. I like to try to influence or persuade people.
23. I like selling things.
24. I am quick to take on new responsibilities.
25. I would like to start my own business.
26. I like to organize things (files, desks/offices).

27. I like to have clear instructions to follow.
28. I wouldn't mind working 8 hours per day in an office.
29. I pay attention to details.
30. I would like to work in an office.

Ratings were to be on a linear scale of 0 to 10, from extremely disliking to extremely liking. The RIASEC score was calculated based on these questions. The form also asked about the respondent's current employment. 160 people responded to our survey. Thus, the data we needed were collected.

The data set collected does not directly contain RIASEC values. Therefore, the data were processed to obtain RIASEC values directly, and further work was done based on these results.

To determine whether machine learning algorithms are better at predicting academic ability, several were tested. Another machine learning model was created to predict careers after predicting academic ability. Using the calculated values for the academic skills and the collected career data from the dataset, it was tested and trained. The optimal algorithm is determined here using different algorithms. In addition, the Synthetic Minority Over-Sampling Technique (SMOTE) was tried.

Chapter 4

Results and Future Scope

After testing and training, the model is presented with a set of random values that simulate the responses to the Holland test. The different algorithms used are decision tree, Random Forest, Multi Output Regressor and K-Nearest Neighbor. The results of the different algorithms are presented here as pie charts. The results given are for academic ability prediction. The accuracy of each model is also shown below.

Decision Tree	Random Forest	KNN	Multi Output Regressor
16.15%	81.52%	84.46%	99.99%

Table 4.1: Academic Skill Prediction - Model Accuracies

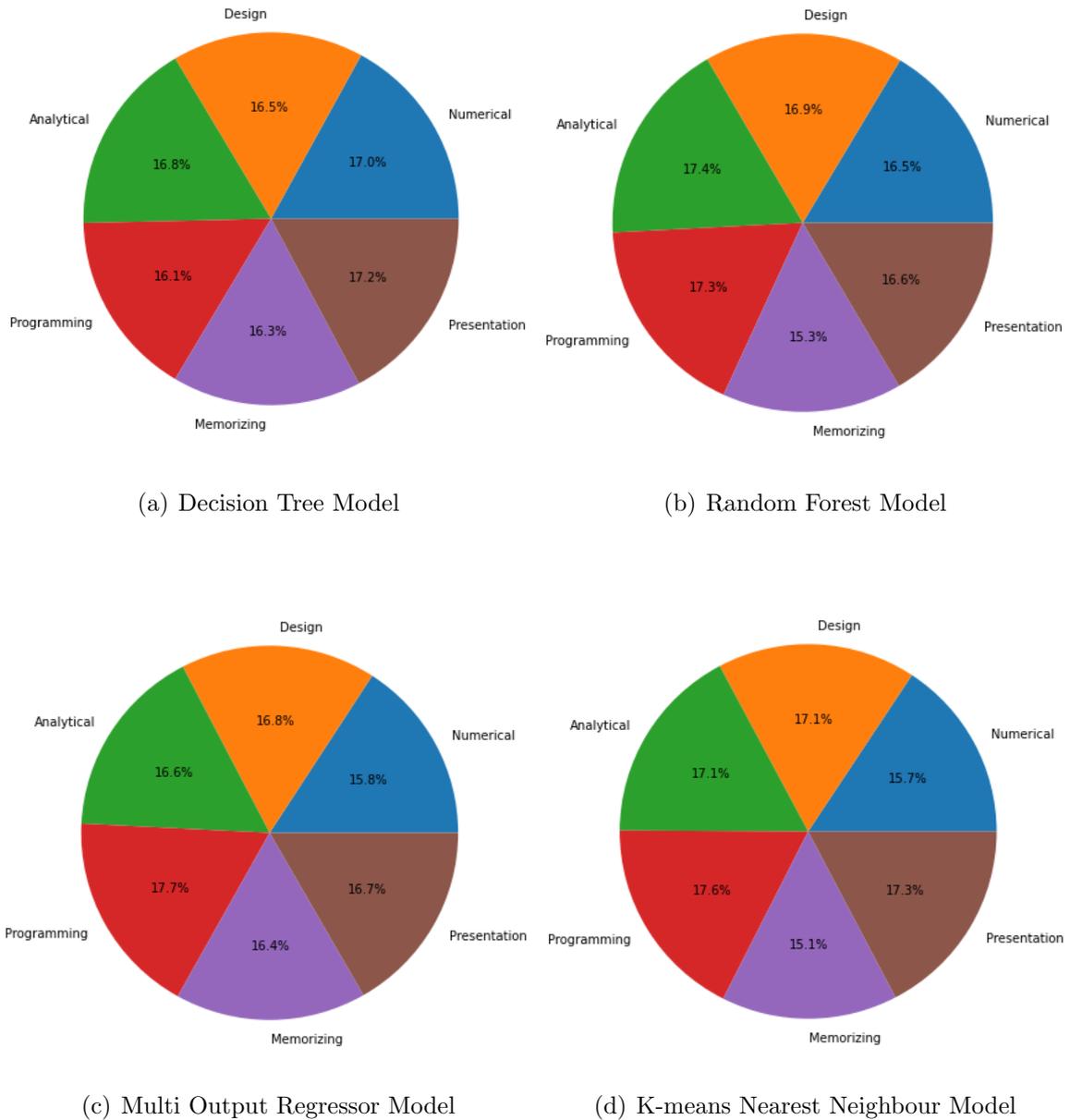


Figure 4.1: Pie chart of different models

The ideal career is proposed by analysing the acquired academic competences. For this purpose, machine learning algorithms are developed and a variety of algorithms are tested for accuracy. A dataset of occupations is used for testing and training. The accuracy of each model is given here. Support Vector Machine, Gaussian Naive Bayes, Perceptron, Decision Tree and Random Forest are among the various techniques used.

SVM	Gaussian Naive Bayes	Perceptron	Decision Tree	Random Forest
78.125%	46.875%	37.5%	50%	71.875%

Table 4.2: Accuracy of models before applying SMOTE



Figure 4.2: Confusion matrix obtained for SVM model before SMOTE

We also tried using SMOTE and the results are shown.

SVM	Gaussian Naive Bayes	Perceptron	Decision Tree	Random Forest
78.6%	40%	80%	60%	67.5%

Table 4.3: Accuracy of models after applying SMOTE

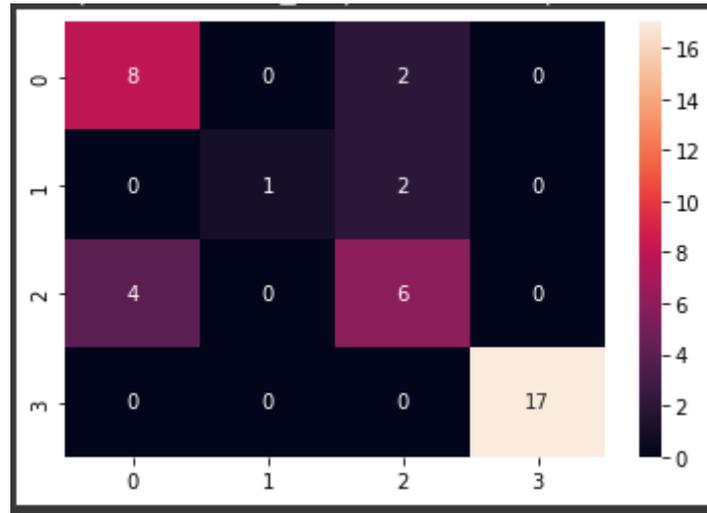


Figure 4.3: Confusion matrix obtained for perceptron after SMOTE

The results showed that the accuracy of SVM, Perceptron and Decision Tree models increased when SMOTE was applied, while Gaussian Naive Bayes and Random Forest decreased.

The process of model development includes a step called model evaluation. It is helpful to find the model that best represents our data and predicts how well the model will perform in the future. To improve the accuracy of the model, we can optimise its hyperparameters. At the same time, we can look at the confusion matrix to see if we can increase the proportion of true positives and true negatives. To determine the accuracy of each model in this study, we use the k-fold cross-validation method.

Making the right career choice early in an academic career is very important because it helps the person prepare to achieve the goal. Our primary goal was to evaluate an individual's personal preferences in different work contexts and sectors and to predict his or her skill level and a preferred career area. To this end, we used the Holland test, and we concluded that by correlating personality traits with various academic skills, the values of these skills can be predicted using various machine learning approaches, and from there, an occupational domain can also be predicted using ML models. The crucial and difficult step in this process is the preprocessing of the data. It is possible to create a well-designed graphical user interface (GUI) for predicting future data. Holland's test results can be used to predict student

performance and ability, which is valuable primarily for improving teacher and student teaching and learning. In addition, by identifying their brilliant abilities using the proposed model, students can make better career choices in the future.

Chapter 5

Conclusion

Choosing the right profession is as important as making money. One of the most important decisions one makes in life is choosing a profession. It involves much more than just what we want to do for a living. Career planning is a lifelong process that involves the following:

- considering one's choices, values, talents, and interests
- investigating the available possibilities for life, employment, and learning
- making sure the assignment matches our unique situation
- we are always adjusting our work and learning plans to better support us in adjusting to changes in our lives and the workplace.

This study proposes to use machine learning to identify the ideal occupation based on Bloom's cognitive behavioural taxonomy. Mental behaviour and cognitive behaviour are interrelated. The work was conducted in two phases. Phase 1 consisted of data collection and prediction of academic ability. Phase 2 was devoted to career prediction.

Since no dataset was available, data was collected through a Google form that included name, profession and 30 questions (5 questions per trait). We received responses from 160 people. The dataset was then preprocessed and loaded into Google Colab. There, a machine learning model was developed using Python, using algorithms such as Decision Tree, Random Forest, K-Nearest Neighbour and Multi-Output Regressor algorithms. 80% of the collected data was used to train and 20% to test

the model. The accuracy of each model was then compared and of the four models, the Multi-Output Regressor showed a maximum accuracy of 99.9%. The proposed approach first builds a ML model to predict academic ability using the data from the obtained datasets. Based on the cognitive behaviour and the predicted academic skills, the appropriate career is predicted.

In the second phase, machine learning models were developed using algorithms such as SVM, Gaussian Naive Bayes, Decision Tree, Perceptron and Random Forest algorithms. The model was trained and tested using the academic skills from the first phase of the work and the career obtained from the collected data. Again, the accuracy of the model was compared and SVM showed the highest accuracy of 78.125%. We then attempted to apply the Synthetic Minority Over-Sampling Technique (SMOTE) to career prediction. Here Gaussian Naive Bayes and Random Forest showed a decrease in accuracy, while Perceptron showed an improved accuracy of 80%.

The proposed work is intended to assist students make the best career choices during their school years so that they can experience pleasure and satisfaction in what they do.

References

- [1] Palee, Pongsaton and Wannapiroon, Panita and Nilsook, Prachyanun. (2020). The Architecture of Intelligent Career Prediction System based on the Cognitive Technology for Producing Graduates to the Digital Manpower. *International Journal of Advanced Computer Science and Applications*. 11. 10.14569/IJACSA.2020.0111214.
- [2] Jayaratne, Madhura & Jayatilleke, Buddhi & Zhang, Yiqing. (2021). Using Unsupervised Machine Learning to Uncover Associations between Personality and Job Preference. 10.13140/RG.2.2.23558.91201.
- [3] Pace, Giorgia & Orrù, Graziella & Monaro, Merylin & Gnoato, Francesca & Vitaliani, Roberta & Boone, Kyle & Gemignani, Angelo & Sartori, Giuseppe. (2019). Malingering Detection of Cognitive Impairment With the b Test Is Boosted Using Machine Learning. *Frontiers in Psychology*. 10. 1650. 10.3389/fpsyg.2019.01650.
- [4] Orrù G, Monaro M, Conversano C, Gemignani A and Sartori G (2020) Machine Learning in Psychometrics and Psychological Research. *Front. Psychol.* 10:2970. doi: 10.3389/fpsyg.2019.02970
- [5] Bleidorn, Wiebke & Hopwood, Christopher. (2018). Using Machine Learning to Advance Personality Assessment and Theory. 10.31234/osf.io/ctr5g.
- [6] Jeromy Anglim & Peter O’connor (2019) Measurement and research using the Big Five, HEXACO, and narrow traits: A primer for researchers and practitioners, *Australian Journal of Psychology*, 71:1, 16-25, DOI: 10.1111/ajpy.12202
- [7] Buelow, Melissa & Frakey, Laura. (2013). Math Anxiety Differentially Affects WAIS-IV Arithmetic Performance in Undergraduates. *Archives of clinical neuropsychology*.

- chology : the official journal of the National Academy of Neuropsychologists. 28. 10.1093/arclin/act006.
- [8] Glöckner, Andreas & Michels, Moritz & Giersch, Daniel. (2020). Predicting Personality Test Scores with Machine Learning Methodology: Investigation of a New Approach to Psychological Assessment. 10.31234/osf.io/ysd3f.
- [9] Hartmann, F. G.; Heine, J.-H.; Ertl, B. Concepts and Coefficients Based on John L. Holland's Theory of Vocational Choice—Examining the R Package *holland*. *Psych* 2021, 3, 728–750. <https://doi.org/10.3390/psych3040047>
- [10] Donald E. Eggerth and Michael E. Andrew. Modifying the C Index for Use With Holland Codes of Unequal Length *Journal of Career Assessment* 2006 14: 267 DOI: 10.1177/1069072705283976
- [11] S. U. Monrad, N. L. B. Zaidi, K. L. Grob, J. B. Kurtz, A. W. Tai, M. Hortsch, L. D. Gruppen, and S. A. Santen, “What faculty write versus what students see? Perspectives on multiple-choice questions using Bloom’s taxonomy,” *Med. Teacher*, vol. 43, pp. 1–12, Jan. 2021.
- [12] S. Das, S. K. Das Mandal, and A. Basu, “Classification of action verbs of Bloom’s taxonomy cognitive domain: An empirical study,” *J. Educ.*, vol. 201, Apr. 2021, Art. no. 002205742110021.
- [13] J. Zhang, C. Wong, N. Giacaman, and A. Luxton-Reilly, “Automated classification of computing education questions using Bloom’s taxonomy,” in *Proc. Australas. Comput. Educ. Conf.*, Feb. 2021, pp. 58–65.
- [14] A. N. Rahmatih, D. Indraswati, G. Gunawan, A. Widodo, M. A. Maulyda, and M. Erfan, “An analysis of questioning skill in elementary school pre-service teachers based on Bloom’s taxonomy,” *J. Phys., Conf. Ser.*, vol. 1779, no. 1, Feb. 2021, Art. no. 012073.
- [15] B. D. Wijanarko, Y. Heryadi, H. Toba, and W. Budiharto, “Question generation model based on key-phrase, context-free grammar, and Bloom’s taxonomy,” *Educ. Inf. Technol.*, vol. 26, no. 2, pp. 2207–2223, Mar. 2021.

- [16] C. Servin, C. Tang, M. Geissler, M. Stange, and C. Tucker, “Enhanced verbs for Bloom’s taxonomy with focus on computing and technical areas,” in *Proc. 52nd ACM Tech. Symp. Comput. Sci. Educ.*, Mar. 2021, p. 1270.
- [17] M. D. Laddha, V. T. Lokare, A. W. Kiwelekar, and L. D. Netak, “Classifications of the sum mative assessment for revised Bloom’s taxonomy by using deep learning,” 2021, *arXiv:2104.08819*. [Online]. Available: <http://arxiv.org/abs/2104.08819>
- [18] G. N. R. Prasad, “Evaluating student performance based on Bloom’s taxonomy levels,” *J. Phys., Conf. Ser.*, vol. 1797, no. 1, Feb. 2021, Art. no. 012063.
- [19] M. Mohammed and N. Omar, “Question classification based on Bloom’s taxonomy cognitive domain using modified TF-IDF and word2vec,” *PLoS ONE*, vol. 15, no. 3, Mar. 2020, Art. no. e0230442.
- [20] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, “Text classification algorithms: A survey,” *Information*, vol. 10, no. 4, p. 150, Apr. 2019.
- [21] A. J. Swart and M. Daneti, “Analyzing learning outcomes for electronic fundamentals using Bloom’s taxonomy,” in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2019, pp. 39–44.
- [22] K. Atiullah, S. W. Fitriati, and D. Rukmini, “Using revised Bloom’s taxonomy to evaluate higher order thinking skills (hots) in reading comprehension questions of english textbook for year X of high school,” *English Educ. J.*, vol. 9, no. 4, pp. 428–436, 2019.