The International Journal of Social Sciences World
TIJOSSW is Available Online at:
https://www.growingscholar.org/journal/index.php/TIJOSSW
Vol. 5 No. 2, July-December 2023, pages: 159~172
DOI: https://doi.org/10.5281/zenodo.8255829
ISSN **2690-5167**
Growingscholar Publisher

Design of Occupational Recommendations in the Field of ICT Based on the BNSP Occupational Scheme Using Machine Learning Approaches

Usanto Usanto [1]

Article history:
Received May 01, 2023; Accepted: July 22. 2023; Displayed Online: August 17, 2023; Published: December 30, 2023

Keywords

Occupational;

BNSP;

Classification;

Technology;

RIASEC;

Abstract

Selecting a suitable profession is crucial for individuals to build their careers and futures. Not everyone possesses sufficient knowledge and understanding about professions that align with their abilities and behavioral patterns. Indonesia has the National Professional Certification Agency (BNSP), which aims to enhance human resource quality by developing standardized professional certification systems. The research methodology combines data analysis, exploration, and machine learning model development. It involves observing individuals who have completed their studies and are employed. Survey forms and observations of students who haven't finished their studies are provided for them to fill out. The system's design offers relevant professional recommendations based on interests and aptitudes. This application's information needs and expected outcomes are to serve as a substitute for experts, particularly for teenagers and high school students. It assists them in identifying their early interests and provides insights and recommendations for careers that align with their interests. In cases where users seek more detailed insights and recommendations, consulting directly with psychologists is advised. The third phase involves analysis, encompassing the following steps: (1)Data Parameter Processing. (2). Calculation of Personality Probability for Interest Determination. (3). Expert Evaluation of Values.

[1] *Institut Teknologi dan Bisnis Swadharma, Jakarta, Indonesia, Email : usanto.s@swadharma.ac.id; usanto1006@gmail.com*

1. Introduction

Information and Communication Technology (ICT) advancement has reshaped the global job landscape, including in Indonesia. More individuals are becoming interested in careers in ICT due to the wide range of potential opportunities and evolving innovations. However, when making decisions about the right profession or field of study, many people face challenges finding accurate and relevant information that aligns with their interests, talents, and potential. This challenge becomes more complex because the ICT field encompasses various specializations and has undergone significant fragmentation into specific roles, including software developers, data scientists, UI/UX designers, and cybersecurity experts (Cappelli et al., 2020). Moreover, job trends in the ICT field can change rapidly following new technological developments. Therefore, an approach is needed to provide accurate, relevant, and adaptable professional recommendations to follow changing job market trends in the ICT field.

The National Professional Certification Agency (BNSP) plays a key role in determining competency standards for the workforce in Indonesia, including in the ICT field. The competency standards issued by BNSP serve as references for industry practitioners, educational institutions, and individuals seeking to understand the qualifications and competencies required in a profession.(Badan Nasional Sertifikasi Profesi., 2020) Therefore, utilizing the BNSP occupation scheme as a foundation for designing a professional recommendation system in the ICT field can potentially address the challenge of matching individuals' abilities with job market demands. In this context, a machine-learning approach can be an effective solution. This technology enables large-scale professional data analysis, job market trends, and individual profiles. By utilizing machine learning algorithms, the recommendation system can identify complex patterns and alignments between individual characteristics and the professional requirements in the ICT field. As a result, individuals will receive more detailed information aligned with their interests and qualifications, while industries will gain a more suitable workforce to meet their needs.

To create an innovative and adaptive solution to the dynamic changes in the ICT world, integrating the BNSP occupation scheme and the machine learning approach becomes the main focus in designing the ICT profession recommendation system. By combining these two approaches, the recommendation system is expected to provide more accurate and relevant information to assist individuals in making wiser career decisions in the ICT field. The need for such professional recommendation systems becomes increasingly crucial due to the growing interest in ICT careers (Smith et al., 2021). Many often struggle to find accurate and relevant information about professions that match their skills and interests. Moreover, many students also face confusion when deciding which major to choose in the field of ICT. The National Professional Certification Agency (BNSP) is responsible for setting competency standards for the workforce in Indonesia, including in the ICT field (Badan et al.., 2020). Therefore, using the BNSP occupation scheme as a reference in the professional recommendation system in the ICT field is expected to provide more accurate recommendations that align with the job market needs in Indonesia.

This research utilizes a machine learning approach to build an ICT profession recommendation system based on the BNSP occupation scheme to address this issue. The machine learning method is chosen due to its ability to process data quickly and accurately, enabling the system to generate accurate professional recommendations based on user needs and interests. This research will collect data about BNSP occupations in the ICT field and user profiles. These data will be processed using machine learning algorithms such as clustering and classification to generate professional recommendations that match users' skills and interests. The classification algorithm is a set of steps or methods to classify data into predetermined classes or categories. The classification algorithm utilizes training data to train the classification model and uses the model to make predictions on

new data. By using discovered rules or patterns, the classification algorithm aids in grouping data and making predictions based on the built model. Classification algorithms are methods or approaches to classify data into predetermined categories or classes based on existing rules or patterns. Classification algorithms build a classification model based on training data and use this model to predict new data (Hastie et al., 2017).

Classification is the approach or algorithm to group or classify data into predetermined classes or categories. The classification method uses training data to learn existing patterns and build a model that can be used to classify new data into appropriate classes. Thus, the classification method assists in grouping and classifying data based on observed characteristics or attributes. Classification methods are approaches or algorithms that classify data into predetermined classes or categories. Classification methods learn patterns from training data to generate a model that can be used to classify new data (Witten & Frank, 2016).

Supervised learning is a machine learning paradigm in which a model or algorithm learns from labeled training data. In Supervised Learning, training data consists of known input-output pairs, and the main goal is to learn a function or rule that can map inputs to correct outputs. The model or algorithm uses the training data as guidance to train itself and discover patterns or rules in the data. With the learned patterns or rules, the model can be used to make predictions or classifications on new unlabeled data. Supervised learning is a machine learning paradigm where a model or algorithm learns to map input to output using known training data. Training data serves as a guide to train the model and discover patterns or rules in the data to make predictions or classifications on new data (Alpaydin, 2020).

System design consists of identification, analysis, planning, and implementation stages used to create a structured system to meet needs and achieve set objectives (Sarawak et al., 2014). System elements encompass interconnected and interacting components within a unit to achieve system objectives (Bertalanffy, 2015). The developed recommendation system is expected to assist prospective ICT professionals in finding professions that align with their skills and interests while facilitating their preparation to meet the competency standards set by BNSP. A Recommendation System is a computer system that provides suggestions or recommendations to users about items they might be interested in based on data analysis, item characteristics, and decision-making methods. A Recommendation System aims to enhance user experience and aid in discovering relevant and valuable content (Davenport & Prusak, 2017). A recommendation system is a system that leverages user data, item information, and information processing techniques to generate suggestions or recommendations about the most relevant, interesting, or suitable items based on user preferences. This system assists users in navigating and finding content, products, or services that might interest them(Burke, 2017).

2. Materials and Methods

The research method combines data analysis, exploration, and machine learning model development. This approach is well-suited for situations that solve concrete problems in practical environments, such as software development, information systems, or UI/UX design. Here is a more detailed breakdown of the approach:

1. Data Analysis: Analisis Data, Conduct data analysis related to the BNSP occupational scheme and professions in the field of IT. This involves deeply understanding the existing occupational scheme, including profession classification and related competencies.

Understanding trends and job market needs in IT is necessary to identify opportunities and challenges.

2. Data Exploration: Eksplorasi Data, Collect and analyze data about IT-related professions covered by the BNSP occupational scheme. This could include information about job descriptions, required competencies, expected education, and work experience.
3. Machine Learning Model Development: Pengembangan Model Machine Learning Utilize machine learning techniques, such as classification or recommendation algorithms, to develop a model that can recommend IT professions based on provided criteria. Prepare an appropriate dataset with features that encompass job descriptions, competencies, and other relevant attributes.
   3.1 Data Preprocessing: Perform data preprocessing steps like filtering, cleaning, merging, and vectorization (if needed) on the collected data.
   3.2 Algorithm Selection: Choose a suitable machine learning algorithm for the recommendation task. Depending on the data and recommendation goal, algorithms like multiclass classification or collaborative filtering could be options.
   3.3 Model Evaluation: Evaluate the model's performance using relevant metrics such as accuracy, precision, recall, or other metrics appropriate for the recommendation goal.
4. Validation, Iteration, and Result Interpretation: After model development, validate it using previously unseen data. If the model's performance needs improvement, iterate the model development process, including feature selection, parameter tuning, and data preprocessing. Once the model is deemed satisfactory, interpret the results and generate recommendations based on the provided input.

3. Results and Discussions

The RIASEC Test is an evaluation tool used to assist individuals in determining career interests based on the RIASEC theory. This test provides information about an individual's preferences for specific jobs and activities, thereby aiding in choosing a suitable career path (Liao et al., 2016). The RIASEC Test can help individuals identify and understand their career interests and provide information about work environments that align with their personality. This test helps individuals determine the appropriate educational and training paths to achieve their career goals. The RIASEC Test is one of the tools used in career counseling to help individuals explore their career interests and choose suitable career paths. This test assists individuals in understanding the connection between their personalities and the jobs that match them (Sampson Jr et al., 2020). The RIASEC Test is widely used in career counseling and human resources to aid individuals in exploring their career interests and selecting suitable career paths. This test provides a clear picture of job types that align with an individual's personality and interests, thus aiding in career decision-making (Swanson & Fouad, 2014).

Interests in RIASEC refer to preferred areas such as college majors, job fields, and professions. Not everyone knows their interests, which often leads to reluctance to identify them due to the cost and time involved in consultations with psychologists. Interests and talents are crucial in determining college majors, professions, and jobs. An expert system for interest and talent detection is an alternative approach to solving this problem. One of the models used for interest and talent detection is the RIASEC model. In a study using RIASEC for career detection in adolescents in Bulgaria and Serbia, the results were considered valid, with almost 90% of the outcomes being deemed good. Saadullah et al. discuss using the RIASEC model for detecting the interests of accounting students to enhance their performance.

Holland conducted research to map an individual's interests into 6 criteria present in the RIASEC model—realistic, investigative, by filling out a questionnaire containing characteristics from the RIASEC model. Each characteristic is grouped according to interests and talents. The result displays the highest values from the selected criteria, and based on these results, recommendations are made for professions, jobs, or educational majors that users can considerUntuk menangani permasalahan sistem, penting untuk mengidentifikasi akar masalahnya dan menerapkan langkah-langkah perbaikan yang tepat. Ini mungkin melibatkan analisis menyeluruh, pemeliharaan rutin, pengujian dan pembaruan, serta memastikan adanya dukungan yang memadai untuk sistem tersebut. Dalam menganalisa rekomendasi profesi berdasarkan skema okupasi BSNP digunakan metode SWOT (Strength,Weakness,Opportunity,Threat) sebagai berikut :

Table. 1 SWOT Analysis

Internal / Eksternal	*Strengths*	*Weaknesses*
	• Support from the National Professional Certification Agency (BNSP) provides legitimacy and confidence in the recommendation system. • The existence of competency standards from BNSP facilitates the determination of profession recommendation criteria. • There are already professional certification institutions in Indonesia. • Focusing on the rapidly growing ICT field adds value to the thesis's relevance to market needs. • The graduation rate in the ICT field is increasing every year..	• The absence of a profession recommendation system in the ICT field through behavioral pattern testing. • Many young generations are unable to make choices for their further studies. • Many young generations struggle to determine a profession that suits their ICT-related qualifications.
Opportunities • Creating a mechanism of a system to predict relevant professions accordingly. • Identifying behavioral patterns of professions in the field of ICT that will be chosen	Strategy SO • Enhancing Deeper Data Collection for the Development of Superior Algorithms & Partnerships with Relevant Institutions • Utilizing structured behavioral pattern data to develop a mechanism for predicting relevant professions. • Leveraging the focus on the rapidly growing ICT industry.	Strategy WO • Identifying behavioral patterns of the chosen ICT field. • Providing a profession recommendation system based on behavioral patterns .
Threats • ncongruity of behavioral patterns with professions. • Rapid changes in job market demands can lead to recommendations becoming less relevant over time .	Strategy ST • Developing a superior and competitive recommendation system compared to existing or other researchers' profession recommendation systems. • Building a robust data security and privacy mechanism to safeguard the integrity and confidentiality of user data.	Strategy WT • Maintaining close collaboration with educational institutions or the ICT industry to access more relevant and high-quality data. • Following the latest trends and developments in the ICT field and adapting the used machine learning models to remain relevant and effective.

The system starts with the user opening a web page, then inputting data in the form of a name, after which the data will be automatically saved. Next, the user inputs their raised test value, and the raised test data is processed in the Google Cloud and continued to the data processing stage to clean, transform, and normalize the data so that it is ready to be used. Furthermore, the results of professional recommendations are displayed as output. The system starts by selecting data withdrawal from the database that has been input by the user in the google cloud. The data is processed in the data processing stage to clean, transform, and normalize the data so that it is ready for use. Next, the system builds a recommendation model using predefined methods. The built model is evaluated to measure the performance and accuracy of the model. After the evaluation is complete, the system can generate appropriate professional recommendations based on the input data and the model that has been built. The results of the professional recommendation are displayed as output.

In this research, Profession Recommendation System Based on Big Five Personality Dimensions Using Fuzzy Inference System Tsukamoto." Based on the Test Results, an Accuracy Value of 63% and an Error Value of 37% Were Obtained. This Error Value is Due to the Suboptimal Determination

of Membership Functions. Therefore, for Further Research, the Evolutionary Algorithm Method Can Be Used to Determine Optimal Membership Functions to Achieve Better Accuracy Values (Mar'i et al., 2019). This Approach was Also Used by (Parewe et al., 2018), Who Employed a Hybrid of Fuzzy Logic and Genetic Algorithm, Demonstrating Improved Accuracy Levels. In a Previous Study Conducted by Made Hanindia Prami Swari, Rahel Widya Arianti, Faisal Muttaqin in 2020, in a Journal Titled "Case-Based Reasoning for Providing Profession Recommendations Based on Students' Interests and Talents Using Simple Matching Coefficient Similarity." Based on the Testing of the Evaluation Data, the Recommendation System Achieved an Accuracy Result of 0.8333 for Real-Life Cases

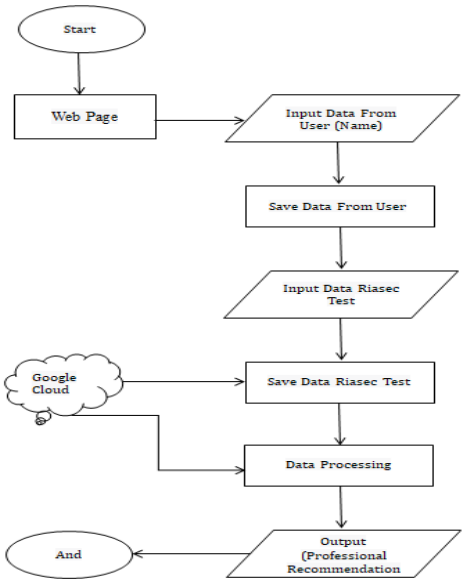


Figure 1. Flowchart User

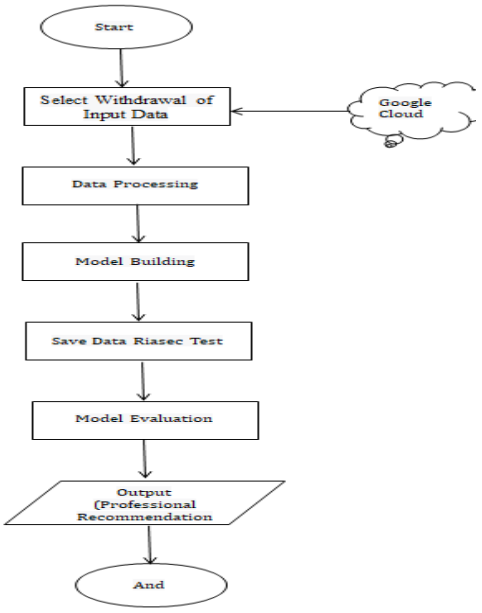


Figure 2. Flowchart System

Model Creation

The creation of the model involves two types of modeling, the first being the clustering model and the second being the classification model. These models are implemented here using the Python language and the Training Data obtained from observations. (Explaining where the form's content comes from and how to increase the number of forms. The training data contains Name, Profession, and RIASEC Test from 14 fields, where aliases have been provided to facilitate the input process in the system, as follows.

Table2. Training Data Description

No.	Aliases	Description	Represent
1	Name	User	User Name
2	P1	Profession	Profession
3	P2	Enjoy working outdoors/fieldwork	Adventurer
4	P3	80% Work with laptop/PC	Office Worker

5	P4	Do you think the weight of 1kg of rice is not the same as that of 1kg of cotton?	Analytical
6	P5	Do you think buying 100g packaged soap is more cost-effective than buying 1kg packaged soap?	Analytical
7	P6	Sometimes you feel comfortable working alone to solve a problem	Independent
8	P7	You enjoy a work routine indoors/outdoors with an 8-hour daily rhythm every day. You have many friends	Routine
9	P8	Anda memiliki banyak teman	Interaktif
10	P9	You are someone with good communication skills	Communication
11	P10	You are someone with high self-confidence and easy adaptability	Adaptable
12	P11	You are ambitious	Ambitious
13	P12	You often are not on time for your work	Discipline
14	P13	You often depend on others in your work	Confidence

While data on IT professions that were collected consist of 31 professions, as follows :

Table 3. Profesions Data

No.	Profesi
1	Ahli Desain Grafis
2	Analis Program
3	Asisten Pemrogram
4	Associate Data Analyst
5	Big Data Scientist
6	Business Intelligence Analyst
7	Cloud Engineer
8	Data Arcitecht
9	Data Entry Coordinator
10	Data Entry Supervisor
11	Data Management Supervisor
12	Data Quality Engineer
13	Database Administrator
14	Desainer Grafis Senior
15	Expert Programmer
16	Implementation Project Manager
17	Junior/Associate Data Engineer
18	Junior/Associate Data Scientist
19	Mobile Programmer
20	Motion Graphic Artist (Intermediate Animator)
21	Network Control Analyst
22	Network Specialist
23	Pemrogram Junior
24	Pengembang Web
25	Program Analyst
26	Provide Client IT Support Services
27	Senior Mobile Programmer
28	Senior Programmer
29	Senior System Analist
30	Software Development Expert
31	Specialist Annotator

Clustering

Data Clustering is the process of placing similar data into groups. The partition clustering algorithm arranges a set of data into several groups so that the similarity within a group is greater than between groups. The main purpose of this application is to replace experts for teenagers, especially high school students, in identifying their interests and providing knowledge and job recommendations that match their interests from an early age. Direct consultation with a psychologist is recommended if the user seeks more detailed information about their interests and comprehensive suggestions. The third stage is the analysis phase, including:

1. Processing parameter data (personality traits for determining interests) used for determining interests, including insert, update, and delete on all data.

2. Based on Holland's theory, calculating the probability values of personality traits to determine interests that align with the RIASEC model.
3. Verification of values by experts.

In principle, probability is utilized to determine interests and talents using the RIASEC method. Probability is used to obtain outcomes by comparing one probability with another. The Probability Theorem used in this case is as shown in the formula below:

$P(\alpha) = n/N$

Where:

- P(α) = probability of variable α (interest type)
- α = Interest type (Realistic, Investigative, Artistic, Social, Enterprising, Conventional)
- n = Result/sample (selected user personality traits per variable)
- N = Population (total personality traits per variable)
- Probability Theorem and Comparison

$P(\alpha 1) Termadapap\ P(\alpha 2) = p(\alpha 1)/p(\alpha 2)$

Example Case:

The obtained interest test results are as follows:

a) R (Realistic) = 8
b) I (Investigative) = 7
c) A (Artistic) = 8
d) S (Social) = 4
e) E (Enterprising) = 8
f) C (Conventional) = 4

Solution:

Calculation using the probability theorem for each variable

a) Realistic (R) Given n value = 8 and N value = 10
   $P(R) = n/N = 8/10 = 0.8$
b) Investigative (I) Given n value = 7 and N value = 10
   $P(I) = n/N = 7/10 = 0.7$
c) Artistic (A) Given n value = 8 and N value = 10
   $P(A) = n/N = 8/10 = 0.8$
d) Social (S) Given n value = 4 and N value = 10
   $P(A) = n/N = 4/10 = 0.4$
e) Enterprising (E) Given n value = 8 and N value = 10
   $P(E) = n/N = 8/10 = 0.8$
f) Conventional (C) Given n value = 4 and N value = 10
   $P(C) = n/N = 4/10 = 0.4$

From the above data, it is concluded that the highest values are P(R), P(A), and P(E), which represent Realistic, Artistic, and Enterprising. Thus, User Latif tends toward Realistic, Artistic, and Enterprising interests. The Probability Theorem and Comparison:

a) Comparison of Probability of R to I $P(\alpha 1) Termadapap\ P(\alpha 2) = p(R)/p(I) = 0.8/0.7 = 8/7$
b) Comparison of Probability of R to A $P(\alpha 1) Termadapap\ P(\alpha 2) = p(R)/p(A) = 0.8/0.8 = 8/8 = 1$
c) Comparison of Probability of R to S $P(\alpha 1) Termadapap\ P(\alpha 2) = p(R)/p(S) = 0.8/0.4 = 8/4 = 2$
d) Comparison of Probability of R to E $P(\alpha 1) Termadapap\ P(\alpha 2) = p(R)/p(E) = 0.8/0.8 = 8/8 = 1$
e) Comparison of Probability of R to C $P(\alpha 1) Termadapap\ P(\alpha 2) = p(R)/p(C) = 0.8/0.4 = 8/4 = 2$
f) Comparison of Probability of I to A $P(\alpha 1) Termadapap\ P(\alpha 2) = p(I)/p(A) = 0.7/0.8 = 7/8$
g) Comparison of Probability of I to S $P(\alpha 1) Termadapap\ P(\alpha 2) = p(I)/p(S) = 0.7/0.4 = 7/4$

h) Comparison of Probability of I to E $P(\alpha 1)Termadapap\ P(\alpha 2) = p(I)/p(E) = 0.7/0.8 = 7/8$
i) Comparison of Probability of I to C $P(\alpha 1)Termadapap\ P(\alpha 2) = p(I)/p(C) = 0.7/0.4 = 7/4$
j) Comparison of Probability of A to S $P(\alpha 1)Termadapap\ P(\alpha 2) = p(A)/p(S) = 0.8/0.4 = 8/4 = 2$
k) Comparison of Probability of A to E $P(\alpha 1)Termadapap\ P(\alpha 2) = p(A)/p(S) = 0.8/0.8 = 8/8 = 1$
l) Comparison of Probability of A to C $P(\alpha 1)Termadapap\ P(\alpha 2) = p(A)\ p(C) = 0.8\ 0.4 = 8\ 4 = 2$
m) Comparison of Probability of S to E $P(\alpha 1)Termadapap\ P(\alpha 2) = p(S)\ p(E) = 0.4\ 0.8 = 4\ 8 = 0.5$
n) Comparison of Probability of S to C $P(\alpha 1)Termadapap\ P(\alpha 2) = p(S)\ p(C) = 0.4\ 0.4 = 4\ 4 = 1$
o) Comparison of Probability of E to C $P(\alpha 1)Termadapap\ P(\alpha 2) = p(E)\ p(C) = 0.8\ 0.4 = 8\ 4 = 2$

The above results conclude that the Probability Theorem and probability are used to validate the comparison of variable values. A comparison of variables can be considered valid if the ratio of one variable to another matches the ratio of their probabilities. Based on the stated problem analysis, the best solution to solve the problem is creating an expert system to diagnose teenagers' interests, especially high school students, using the RIASEC model through a website-based platform. The website platform is chosen due to its easy accessibility and no need for installation; it can simply be accessed through a browser on both desktop and smartphone. The RIASEC model is used to determine a person's interests to obtain the best outcomes (Hidayat & Wahyuni, 2019).

Data clustering is the process of grouping similar data into clusters. Partitioning clustering algorithms divide a dataset into several clusters so that the similarity within a cluster is greater than between clusters (Hammouda & Karray, 2010). Clustering is conducted in two ways. The first way involves manual clustering or grouping of professions based on similar characteristics. Out of the 31 available professions, a clustering process was applied to group them into 3 clusters.

Table 4. Cluster Data

Cluster	Profession
(1) Grouping Data and Analysis:	a. Associate Data Analyst b. Junior/Associate Data Engineer c. Junior/Associate Data Scientist d. Big Data Scientist e. Data Quality Engineer f. Data Management Supervisor g. Data Entry Coordinator h. Specialist Annonator
(2) Group of Software Development and Programming:	a. Junior programmer b. Mobile Programmer c. Senior Mobile Programmer d. Expert Programmer e. Program Analyst f. Software Development Expert g. Senior Programmer h. Asisten Programmer i. Analyst Program j. Web Developer
(3) Group of Infrastructure and Networking:	a. Network Specialist b. Cloud Engineer c. Network Control Analyst d. Provider Client IT Support Services

Then, the manually created clusters were input into the training data as the field c3, based on the data provided above. Next, the second clustering was performed in Google Colab using Python.

```
x=df_input[['P2','P3','P4','P5','P6','P7','P8','P9','P10','P11','P12','P13']]
```

Figure 3. Input Variable

After determining the input variables to use the Elbow Method, which is one of the methods to determine the appropriate number of clusters through the percentage of the comparison results between the number of clusters that will form an elbow at a certain point, the next step is to formulate the elbow method.

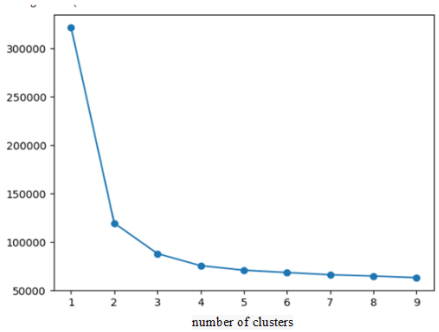


Figure 3. Elbow Method

It can be seen from the above elbow method plot that the number of clusters that can be chosen is 3 (three). Subsequently, a CSV file containing the 3 clusters is obtained, automatically creating a new field.

Table 5. Field Cluster

	Unnamed: 0	Timestamp	Nama		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	cluster
0	0	3/16/2023 19:23:35	Fajri Ardiansyah	Network Specialist	1	5	1	4	5	3	4	4	4	3	3	1	0	
1	1	3/14/2023 15:32:36	Encang Taufik	Data Entry Operator	3	2	4	1	3	4	5	4	4	4	2	1	1	
2	2	3/14/2023 16:40:07	Ridho Alamsyah	Pemrogram Junior	3	5	1	1	3	3	4	4	5	4	4	3	2	
3	3	3/14/2023 16:40:07	Lia	Data Entry Operator	3	5	1	4	4	3	4	4	4	4	3	3	0	
4	4	3/14/2023 15:09:22	Alex	Network Specialist	4	5	1	2	3	3	4	4	5	3	1	2	2	

It can be observed in the table below that within cluster 0, which is the Data and Analysis cluster, the dominant profession is "Associate Data Analyst." Therefore, this profession is considered a reference for cluster 0

Table 5. Data and Analysis Cluster

Cluster Data dan Analisis	
Associate Data Analyst	6
Big Data Scientist	2
Data Entry Coordinator	3
Data Management Supervisor	1
Data Quality Engineer	2
Junior/Associate Data Engineer	2
Junior/Associate Data Scientist	3
Specialist Annotator	3

It can be observed from the table below that within Cluster 1, which corresponds to the Software Development and Programming Cluster, the dominant profession is the programmer. Therefore, this profession is taken as a reference for Cluster 1.

Table 6. Software and Programming Cluster Development

Cluster Pengembangan Perangkat Lunak dan Pemograman	
Programmer	17
Senior Programmer	5
Mobile Programmer	5
Program Analyst	3
Senior Mobile Programmer	2
Pengembang Web	2
Asisten Pemrogram	2
Analis Program	2
Software Development Expert	1
Pemrogram Junior	1

The table below shows that within cluster 1, or the Development of Software and Programming cluster, the dominant profession is the programmer. Therefore, this profession is taken as a reference for cluster 2.

Table 7. Cluster Infrastructure and Network

Cluster Infrakstruktur dan Jaringan	
Network Specialist	15
Provide Client IT Support Services	3
Cloud Engineer	3
Network Control Analyst	2

Below is the cluster analysis graph of the RIASEC test that has been conducted for the analysis



Figure 4. Cluster Analysis

The conclusion drawn from the above image is that the Data Analysis and Data Cluster exhibit the following characteristics: Average Analytical 6, Average Office Worker 5, Average Adaptable 4, Average Confidence 4, Average Independent 4, Average Communication 4, Average Ambitious 4, Average Adventurous 3, Average Routine 3, Average Discipline 3, and Average Interactive 2. The Programming Cluster displays the following characteristics: Average Analytical 4, Average Office Worker 5, Average Adaptable 4, Average Confidence 4, Average Independent 4, Average Communication 4, Average Ambitious 1, Average Adventurous 3, Average Routine 3, Average Discipline 3, and Average Interactive 2.

The Network Cluster showcases the following characteristics: Average Analytical 3, Average Office Worker 5, Average Adaptable 4, Average Confidence 4, Average Independent 4, Average Communication 4, Average Ambitious 4, Average Adventurous 3, Average Routine 3, Average Discipline 3, and Average Interactive 2

Classification

After identifying the clusters obtained from the elbow method, the next step is to proceed with the second modeling, the Classification Model. In this case, the Random Forest classification model is utilized, which demonstrates the following accuracy.

```
print(accuracy_score(y_test,predict_rf))

0.996
```

Figure 5. Accuracy Random Forest

Display Structure

The following is a view of the career recommendation system as follows:



Figure 6. Input Riasec Test and Recommendation Results Display

4. Conclusion

Information Needs: This application's information needs and expected outcomes are to serve as a substitute for experts for teenagers, especially high school students, in identifying their interests from an early stage. It also aims to provide knowledge and job recommendations that align with their interests. In case users desire more detailed insights into their interests and comprehensive suggestions, they are encouraged to consult directly with a psychologist.

Third Stage - Analysis Phase: (1). Processing Parameter Data: The processing of parameter data (personality traits for determining interests) is used to establish interests. This involves operations

such as inserting, updating, and deleting data across the entire dataset. This feature enhances the dynamism of the outcomes. (2). Calculating Personality Probability Values: Probability values of personality traits are computed to determine interests in alignment with the RIASEC model based on Holland's theory, developed by Holland. (3). Validation by Experts: The computed values are checked and validated by experts in the field.

Advancements in Technology and Artificial Intelligence with the rapid evolution of technology, particularly in Artificial Intelligence (AI), the use of machine learning in recommendation systems has become increasingly prominent. This technology enables the development of effective recommendation systems by employing AI techniques that allow computer systems to learn from data and make predictions or recommendations based on discovered patterns. Leveraging machine learning techniques offers an efficient solution for providing professional recommendations, particularly in the rapidly evolving domain of Information Technology and Communication (ITC).

Designing a Profession Recommendation System for the IT Field The design of a professional recommendation system is centered around professions within the IT domain. The implementation involves utilizing machine learning to generate professional recommendations based on behavioral patterns. This is done using the Python programming language on the Google Colab platform. The adopted algorithm model is Random Forest, which attains an accuracy level of 0.99 or 90%. This system holds the potential to significantly aid users in discovering relevant IT-related professional recommendations, including previously unconsidered options.

Accessibility and implementation to access the machine learning-based system via Google Colab, users require a stable internet connection and a web browser. As a result, the system is designed to be showcased as a website. This approach ensures that users can conveniently utilize the system's features and receive professional recommendations within the IT field.

Acknowledgments

The research team would like to thank all parties who have contributed to the implementation of this research.

References

Alpaydin, E. (2020). *Introduction to machine learning*. MIT Press.

Badan Nasional Sertifikasi Profesi. (2020). *Peraturan Badan Nasional Sertifikasi Profesi Nomor 2 Tahun 2020 tentang Standar Kompetensi Kerja Nasional Indonesia*.

Bertalanffy, L. V. (2015). General System Theory: Foundations, Development, Applications, revised ed.; George Braziller. *Inc.: New York, NY, USA*.

Burke, R. (2017). Hybrid web recommender systems. *The Adaptive Web: Methods and Strategies of Web Personalization*, pp. 377–408.

Cappelli, A., Mugnaini, M., & Cini, E. (2020). Improving roller milling technology using the break, sizing, and reduction systems for flour differentiation. *LWT*, *133*, 110067.

Davenport, T. H., & Prusak, L. (2017). *Working knowledge: How organizations manage what they know*. Harvard Business Press.

Hammouda, K., & Karray, F. (2010). A comparative study of data clustering techniques. *University of Waterloo, Ontario, Canada*, *1*.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2017). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

Hidayat, F. K., & Wahyuni, S. N. (2019). Pendeteksian Minat Dan Bakat Menggunakan Metode Riasec. *Indonesian Journal of Business Intelligence (IJUBI)*, *2*(1), 32–39.

Johnson, R. W. (2020). *Challenges and Opportunities in Designing Occupation-based Career Recommendation Systems in the IT Industry.* 112–119.

Liao, H. Y., Huang, Y. H., & Lin, Y. T. (2016). *The Effect of RIASEC on Career Decision Making Self-Efficacy and Outcome Expectations of Senior High School Students*. Journal of Career Development

Mar'i, F., Mahmudy, W. F., & Yusainy, C. (2019). Sistem Rekomendasi Profesi Berdasarkan Dimensi Big Five Personality Menggunakan Fuzzy Inference System Tsukamoto. *J. Teknol. Inf. Dan Ilmu Komput*, *6*(5), 457–466.

Parewe, A. M. A. K., Mahmudy, W. F., Ramdhani, F., & Anggodo, Y. P. (2018). Dental disease detection using hybrid fuzzy logic and evolution strategies. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *10*(1–8), 27–33.

Sampson Jr, J. P., Lenz, J. G., Dozier, V. C., Osborn, D. S., Peterson, G. W., Reardon, R. C., Lenz, J. G., Morrison, A. R., Galles, J. A., & Melvin, B. R. (2020). *A Bibliography of CIP Theory, Research, and Practice*. Florida State.

Sarawak, G. A., et al. (2014). *System Design Principles and Guidelines for Development Organizations. International Conference on Computer Science and Computational Intelligence*.

Smith, J., Brown, A., & Chen, C. (2021). SA Machine Learning Approach to Career Recommendation in the IT Sector Based on BNSP Occupational Framework. *International Journal of Artificial Intelligence*, pp. *31*, 127–143.

Swanson, J. L., & Fouad, N. A. (2014). *Career theory and practice: Learning through case studies*. Sage publications.

Witten, I. H., & Frank, E. (2016). Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record*, *31*(1), 76–77.