

PERSONALITY ANALYSIS THROUGH SOCIAL MEDIA BASED ON MACHINE LEARNING

1st M. Jaka Firdaus Teguh Yahya
Fakultas Teknik Elektro
Telkom University
Bandung, Indonesia

jakafirdaus@student.telkomuniversity.ac.id

2nd Kris Sujatmoko
Fakultas Teknik Elektro
Telkom University
Bandung, Indonesia

krissujatmoko@telkomuniversity.ac.id

3rd Thomhert Suprpto S
Fakultas Teknik Elektro
Telkom University
Bandung, Indonesia

thomhert@telkomuniversity.ac.id

Abstrak - Ketidakcocokan antara keterampilan dan pekerjaan seringkali menjadi masalah di dunia profesional. Ketidakcocokan ini dapat menyebabkan penurunan kinerja dari yang diharapkan. Penelitian ini bertujuan untuk memberi solusi atas masalah tersebut dengan sebuah website yaitu Psyche yang dapat membantu seseorang untuk meningkatkan kesadaran diri dalam mengenali kekuatan dan kelemahan pribadi. Psyche terintegrasi dengan model machine learning untuk melakukan prediksi kepribadian berdasarkan The Myers-Briggs Type Indicator (MBTI) melalui tweet dari media sosial twitter pengguna yang kemudian menghasilkan rekomendasi pekerjaan. Psyche melalui dua tahap pengujian: pengujian fungsional utama dan pengujian staging. Fokus pengujian pertama adalah pada fungsi utama dan memastikan semua fitur berfungsi dengan baik. Pengujian kedua melibatkan responden dan pengumpulan umpan balik melalui kuesioner. Berdasarkan temuan kuesioner, diperoleh kesimpulan bahwa responden cenderung setuju bahwa Psyche berfungsi dengan lancar dan diharapkan dapat menjadi solusi atas masalah yang diangkat.

Kata Kunci - Ketidakcocokan keterampilan dan pekerjaan, Psyche, Kepribadian, MBTI

Abstract - A mismatch between skills and occupations is often a problem in the professional world. This mismatch can lead to less performance than expected. This research aims to provide a solution to the problem with a website, Psyche, which can help a person increase self-awareness by recognizing personal strengths and weaknesses. Psyche is integrated with a machine learning model to predict personality based on the Myers-Briggs Type Indicator (MBTI) through tweets from the user's Twitter social media, which then produces occupation recommendations. Psyche went through two stages of testing: main functional testing and staging testing. The focus of the first test was on the main functionality and ensuring all features were functioning properly. The second test involved respondents and feedback collection through questionnaires. Based on the findings of the questionnaire, it was concluded that respondents tend to agree that Psyche functions smoothly and is expected to be a solution to the problems raised.

Keywords - Mismatch between skills and occupation, Psyche, Personality, MBTI

I. INTRODUCTION

One of the issues in Indonesia's employment sector is the mismatch or gap between job opportunities and industrial demands, which is created by a mismatch in education and skills [1]. The employment demand and supply imbalance forced many higher education graduates in Indonesia to seek jobs outside their field of study, regardless of the major they studied in college. This scenario is referred to as a horizontal mismatch. A mismatch between education and employment is an issue because when an employee works in an area that does not match their educational background, they must work harder to acquire the necessary skills or competencies for the job. It means that employees must learn new things and dive into different information than what they previously knew, which may make them feel uncomfortable with their employment and lead to poor levels of job satisfaction. This will also lead to diminished productivity and firm growth [2].

One of the causes of the mismatch is a lack of high-quality human resources, which is also caused by individuals' lack of self-knowledge. Self-development is one strategy for developing human resources [3].

This study focuses on the website that was created, Psyche. Users can utilize the Psyche website to do MBTI-based personality detection, which will also provide an explanation of the career path depending on the user's personality type. Psyche is the solution that is supposed to be the first step toward resolving the issues identified.

The programming languages used for developing the Psyche website are Python for the back-end and CSS, HTML, and JavaScript for the front-end. Open-source Python libraries are utilized in the back-end, such as ntscraper, which helps with the back-end website development. The flask framework is utilized to make route organization easier on the back-end. MongoDB will be used to store user data, personality prediction results, and tweets. Additionally, machine learning model will be integrated into the backend so that users can perform personality detection.

II. THEORY REVIEW

A. Myers-Briggs Type Indicator (MBTI)

The MBTI is a personality inventory. Career development specialists have adopted it to aid a client in picking the best career path as part of a full self-assessment [4]. The MBTI was developed by Katharine Cook Briggs and her daughter Isabel Briggs Myer, deriving ideas from Carl Gustav Jung's theory of personality. Using a model based on four major factors—each of which corresponds to a major dimension of the MBTI personality type theory—the profile of these actions is diversified. The aspects of Sensing (S) vs. Intuition (N), Thinking (T) vs. Feeling (F), Judging (J) vs. Perceiving (P), and Extrovert (E) vs. Introvert (I). Combining the four dimensions yields 16 different personality types, including ESTJ, ENTJ, ESFJ, ENFJ, ESTP, ENTP, ESFP, ENFP, INFP, ISFP, INTP, ISTP, INFJ, ISFJ, INTJ, and ISTJ.

The 16 MBTI Types			
ISTJ	ISFJ	INFJ	INTJ
ISTP	ISFP	INFP	INTP
ESTP	ESFP	ENFP	ENTP
ESTJ	ESFJ	ENFJ	ENTJ

Figure 2.1 The 16 MBTI Types

B. Front-end

The portion of the system that gives the user a view is called the front-end. It sets up the system's interface and is in charge of creating the visual components. HTML, CSS, and JavaScript are used in the front-end for the user interface and user experience of the Psyche website.

The front-end consists of two major components: the user interface (UI) and the user experience (UX). UI's primary responsibilities include handling color composition, processing visuals, and displaying animations. In other words, the user interface (UI) is in charge of improving the appearance of the system. When a system is being developed, the user interface is created based on their requests. UX encompasses how people feel when using a system, its ease of use, and its impact on achieving goals. UX is determined after the system is installed and used by users.

C. Back-end

The portion of the software that runs on the server side and is in charge of interacting with the database and program logic is called the back-end. An Application Programming Interface (API) is used to facilitate communication between the client and the back-end. Flask is the framework used in

the back-end; using this framework will make it much easier to organize API endpoints in the back-end.

III. RESEARCH METHOD

A. Website Overview

According to the outcomes of the team's requirements assessment, actors and the system interact continuously in order to construct information systems that meet the functional requirements of an existing system.

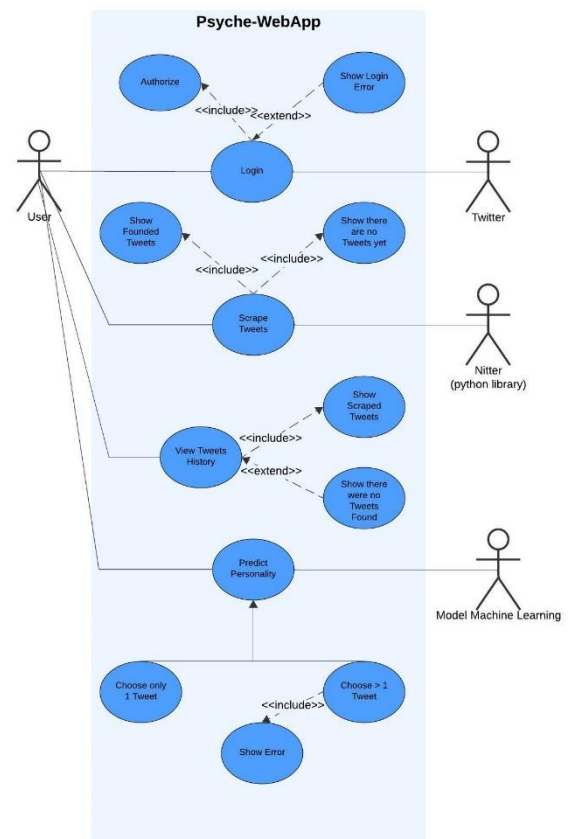


Figure 3.1 Use Case Diagram

Figure 3.1 is the work activities between the user and the system from the Psyche website. This use case diagram depicts the primary functionality of the Psyche website, which is aimed at predict user personality using tweet from Twitter. This diagram depicts various players and key interactions in the program, including the login procedure, tweet retrieval, tweet analysis, and personality prediction.

A flowchart is a diagram that shows the steps and decisions required to complete a program's workflow. Each stage is depicted in diagram form and connected by lines or arrows. Flowcharts are important for determining the flow of a process, enhancing comprehension, and connecting technical and non-technical requirements [5]. Flowchart is made into 2 parts to make it easier to visualize the flowchart of the Psyche web application. The first flowchart is the home

page section, can be seen in the image below and continued with the flowchart for the dashboard page in figure 3.3.

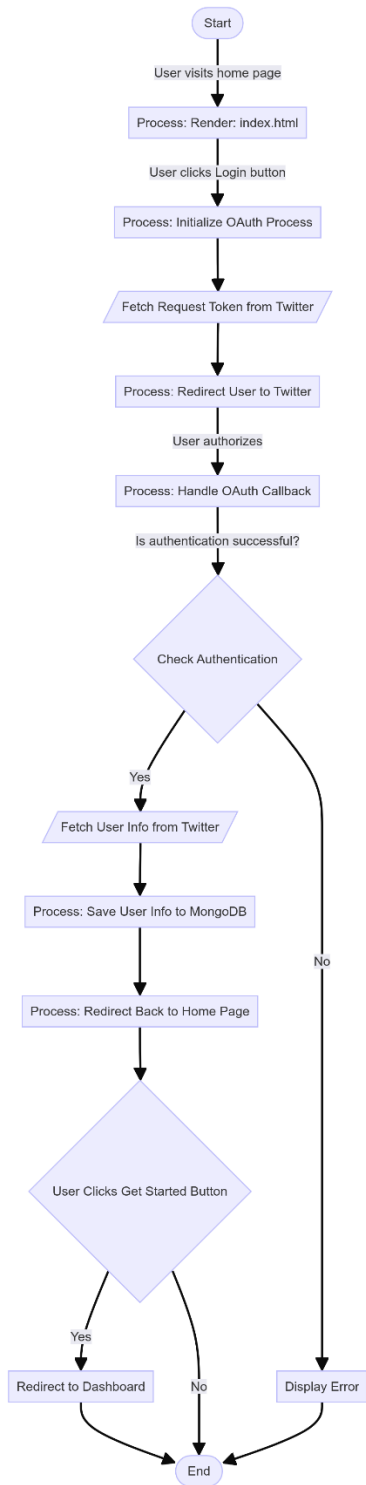


Figure 3.2 Home Page Flowchart

This flowchart describes the workflow that occurs when a user visits the home page of the web application. This flowchart includes the authentication process through OAuth using Twitter, callback handling, as well as the decision flow that directs the user to the dashboard page or displays an error

message. This flowchart provides a clear visualization of how the system handles the initial interaction with the user and ensures the logged-in user has legitimate access rights before proceeding to the dashboard. The authentication process is comprehensive, ensuring that only authorized users can proceed to the Psyche dashboard. The flow also handles potential errors well, providing a structured and secure user experience. This visualization helps in understanding the logic behind managing user access at the initial stage of interaction with the application.

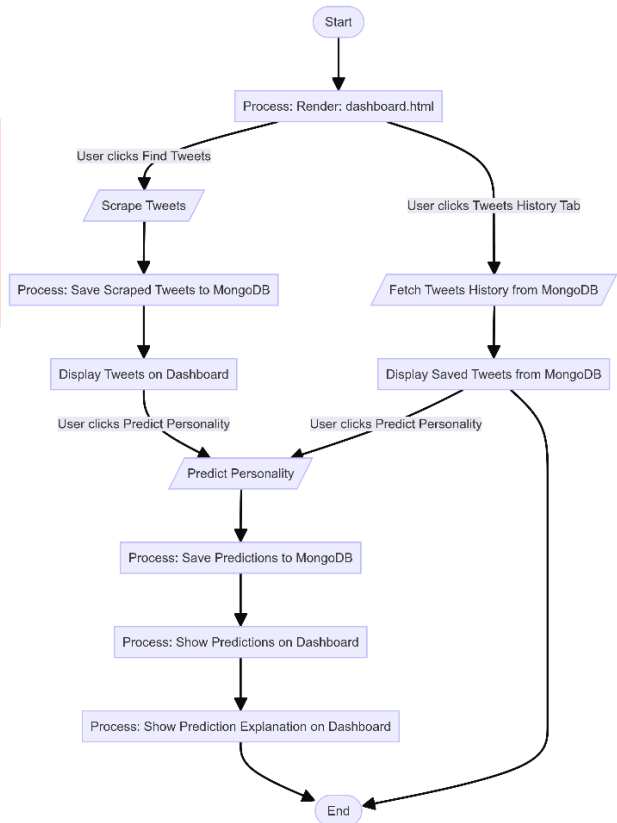


Figure 3.3 Dashboard Page Flowchart

The flowchart above illustrates the workflow of the dashboard page in the web application that was created. This dashboard page performs multiple duties, including collecting tweets, storing data in MongoDB, predicting personality, and showing prediction results. This flowchart shows how the application's dashboard page operates in a straightforward and ordered manner. Each step, from locating and storing tweets to identifying personality and showing prediction results and explanations, has been carefully planned to give the user with a pleasant experience.

B. Implementation

The Psyche web application's implementation begins with identifying a mechanism for scraping user tweets from Twitter, followed by deciding the programming language and framework to be utilized on the backend. The client-side interface and user interaction with Psyche are created in HTML, CSS, and JavaScript. A design is required for the user

interface (UI) and user interaction (UX). This stage's actions include looking for design references from websites such as Dribbble, which already supply UI and UX designs, in order to speed up the work of web apps on the client side.

Python was selected as the server-side programming language. Flask is a Python framework for web development and application construction. Flask was used to construct Psyche because it is lightweight to operate. This is because Flask's core is straightforward and modular.

C. Website Testing

Testing is divided into two steps. The first step involves testing the core functionality of the Psyche web application, which is also known as the development phase. At the development phase, testing is solely done on localhost and does not include web application responses; the goal is to guarantee that everything works properly. The second step is the staging phase involves respondents, and testing is carried out by asking respondents for input in the form of a questionnaire. The questionnaire responses are presented on a Likert scale

IV. RESULT AND DISCUSSION

A. Implementation

The initial step in the implementation process is to discover the user's tweet scraping technique. Some common tweet scraping methods are:

1. Create program code from scratch using Python and utilize the open-source library Selenium.
2. Utilize the open-source library snsrape available for Python.
3. The third method is the method used in PSYCHE, which is by using the open-source library ntscraper.
4. This method can be used if using node.js on the server side is to use the open-source library puppeteer available for node.js.
5. As opposed to the previous four techniques, this approach connects directly to Twitter/X. However, this solution needs a \$100 USD charge to gain access to Twitter/X's API endpoint, which allows you to retrieve user tweets directly.

Psyche will be using the open-source library ntscraper. Besides being free, ntscraper may still be utilized today, and there is no need to write program code from scratch, making server-side work easier. ntscraper requires a nitter to scrape user tweets. Nitter is a free, open-source alternative Twitter front-end. PSYCHE uses nitter, <https://nitter.privacydev.net>, which is publicly available. The problem with this strategy is that nitter has a time restriction for collecting tweets and must wait before being utilized again.

Psyche uses Twitter/X for login. The open-source library OAuth1Session is used to authenticate PSYCHE with the user's Twitter/X account, which requires the following

information: API Key, API Secret Key, and Callback URI, all of which may be obtained from the Twitter Developer. Then, it will seek tokens from PSYCHE users. Finally, users will be sent to the Uniform Resource Locator (URL) to allow PSYCHE to access their data. Authorization is the process of providing access permissions to third parties in order for them to have restricted access to the data possessed; this process may be completed if authentication is performed so that the information supplied matches the user who completed the login procedure. OAuth allows users to grant third-party sites access to information held by other service providers without disclosing access privileges or all of their data or providing a login and password.

The Psyche website, created using the Flask framework, features a main page and a dashboard.

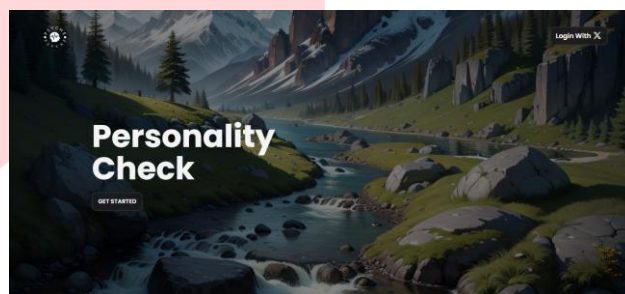


Figure 4.1 Home Page

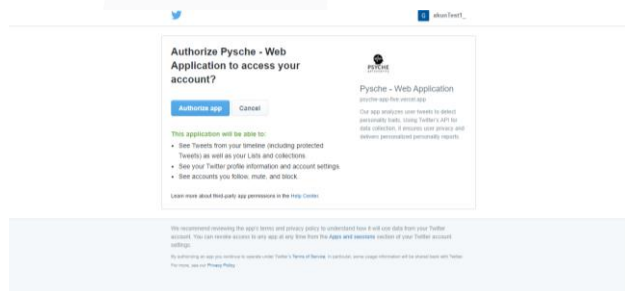


Figure 4.2 Authorization Page

First, on the main page, there is a login with an X button that allows users to authenticate with Twitter. Users will be directed to the Twitter/X authorization page, where they will grant Psyche permission to access the necessary data from their Twitter account.

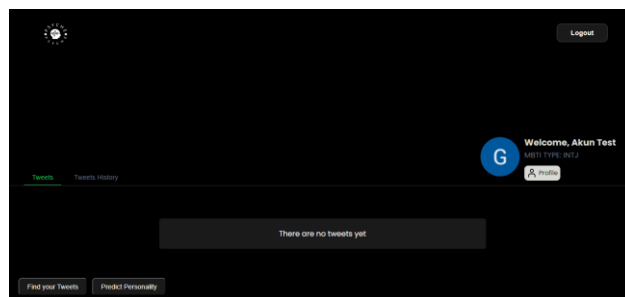


Figure 4.3 Dashboard Tweets

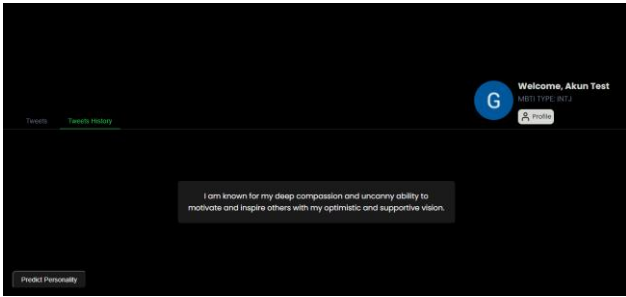


Figure 4.4 Dashboard Tweets History

On the Dashboard, there are two-page options: Tweets and Tweet History. If the user is using PSYCHE for the first time, they can start by scraping tweets from the Tweets page. Scraped tweets are saved in the database and may be seen in the tweet history. Psyche provides a Tweets History page to help users make predictions based on previously scraped tweets.

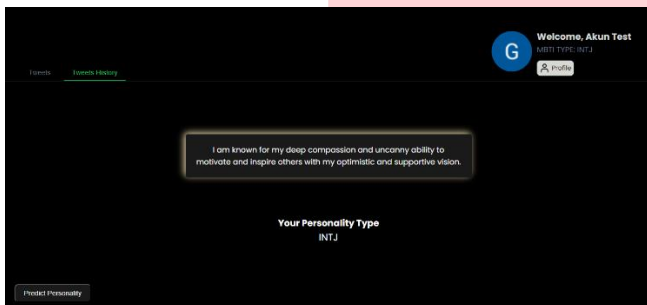


Figure 4.5 Predict Personality

The machine learning model that has been connected with the backend may receive the selected tweet text from the frontend, generate predictions, and then provide the predictions to the frontend. To view the prediction findings, users can hit the predict personality button.

Psyche helps users identify personality types using the Myers-Briggs Type Indicator (MBTI) and gives a brief description of each personality type's career path. The following is the procedure for using the Psyche on the web application:

1. The initial step is to access the Psyche online application.
2. On the Home Page, users will notice two buttons: login with X in the upper right corner and get started. To authenticate using their Twitter/X account, users can first press the login with X button.
3. Users will be forwarded to the permission page, which is handled by Twitter/X. If the user has not previously signed in with a Twitter/XX account on a web browser, he or she is required to do so first. Logging in can be done immediately from the authorization page or by opening the Twitter/X page from the browser. If the user has previously

signed in to the authorization page, they may provide authority to the Psyche web application.

4. If authentication is successful, the user will be routed to the home page. When the user presses the get started button, they will be routed to the dashboard.
5. On the dashboard, users have two-page options: Tweets and Tweets History (green). If the user is new to Psyche, please click the Find Your Tweets option on the Tweets page to search for user tweets and wait a few seconds for the tweets to appear on the dashboard.
6. Users may guess their personality by clicking the "predict your personality" button. Before proceeding, ensure that you have chosen the tweets you wish to utilize for personality type prediction. After clicking the predict your personality button, the results will be presented on the dashboard.
7. The Tweets History page saves previously searched user tweets. Users can also make predictions using the Tweets History page, much as they did in step 6.
8. Users can view a brief explanation of their predicted personality by clicking the profile button.

B. Website Testing

1. Development Phase

User Authentication Test with Twitter/X. Table 4.1 below explains the process of user authentication. The result of this procedure is that the user is able to navigate to the authorization page and then back to the home page of the Psyche web application, and data from the user's Twitter/X account, such as their Twitter ID, username, name, and profile photo, is stored in a database.

Table 4.1 Test A

No	Data Input	Results to be achieved	Observation Results
1	Connecting Psyche web app with Twitter/X	Psyche users can access authorization page	The user is successfully directed to the authorization page of Twitter/X
2	User provides authorization to Psyche web	The user is redirected to the home page of Psyche	The user is successfully redirected back to the home page and the data from the user is successfully stored in the database.

Personality Detection and Machine Learning Integration Tests. Table 4.2 describes the tests for personality detection and the integration of machine learning models. The exam went smoothly, and the results are as predicted. Integration with machine learning models is seamless, and prediction results may be saved in a database.

Table 4.2 Test B

No	Data Input	Results to be achieved	Observation Results
1	Create a separate file with the backend file, model_machineLearning.py, which contains the machine learning model.	Machine learning models can be used for personality type prediction	The machine learning model successfully connects with the backend and successfully provides predictions.
2	User uses the predict personality button on the dashboard	Display the prediction result from the backend to the frontend	The prediction results were successfully displayed on the frontend without any problems.

Tweets scraping Tests. The Psyche web application uses the ntscraper library to conduct user tweet scraping. The find your tweets button on the Psyche dashboard page allows users to search for tweets. The test is described in Table 4.3

Table 4.3 Test C

No	Data Input	Results to be achieved	Observation Results
1	Searching for public Nitter	The nitter obtained can still be used for tweet scraping	Nitter still works fine and can be used for scraping with ntscraper
2	Use the Ntscraper library on the backend and import the Nitter class from Ntscraper to connect with	Getting tweets from users	Successfully get user tweets from the Ntscraper library using public Nitter

	the public Nitter.		
--	--------------------	--	--

Database Integration Tests. Psyche website will use MongoDB for the database. Testing is performed to guarantee that the backend can store data received from the frontend. Before testing, a database named Psyche App was built in MongoDB, and four collections were constructed within the database to contain the data required by the Psyche web application. Collections developed include collection user to store user information, collection tweets to store user tweets, collection predictions to store prediction results, and collection career to store data for a brief description of each personality's professional path.

Table 4.4 Test D

No	Data Input	Results to be achieved	Observation Results
1	Using the pymongo library to initiate a connection with MongoDB from the backend	Connect with MongoDB	successfully connected with MongoDB
2	Trying to log in, search for tweets, and make predictions to see if the data can be stored on the backend	Data can be stored in the database	All data can be stored in the database without any problems
3	Trying to retrieve data from the collection career and display it on the frontend	Data can be displayed on the frontend	Data in the collection career can be displayed on the frontend properly

The results of all tests A, B, C and D are presented in table 4.5 below, all input data from all tests are given a successful status based on the observation results.

Table 4.5 Test Results

Results			
A	B	C	D
Succeed	Succeed	Succeed	Succeed

1. Staging Phase

The staging phase step requires responders to test the Psyche web application that has passed the previous testing stage. At this level, testing is done using a questionnaire. The questionnaire is issued regardless of the responder

Respondent’s responses are in the Likert scale form, which is a psychometric scale extensively used in surveys and the most widely used scale in survey research. Table 4.6 provides an explanation of the scale utilized.

Table 4.6 Likert Scale

Score	Description
1	Strongly Disagree
2	Disagree
3	Neutral
4	Agree
5	Strongly Agree

The staging stage testing was implemented online using Google Form from July 24, 2024, to July 25, 2024, with 32 responders. The Table 4.7 respondent's replies to 10 items on the questionnaire

Table 4.7 Respondent’s Replies

Questions	Points				
	1	2	3	4	5
I will use this App frequently	0%	9,4%	18,8%	40,6%	31,3%
I found this App too complicated	31,1%	40,6%	18,8%	3,1%	6,3%
I find this App easy to use	0%	0%	12,5%	31,3%	56,3%
I think I need technical support to use this App	15,6%	28,1%	34,4%	9,4%	12,5%
I found the various functions in this App well integrated.	0%	3,1%	18,8%	31,3%	46,9%
I think there are too many inconsistencies in this App	18,8%	40,6%	31,3%	0%	9,4%
Most people will be able to learn to use this App quickly	0%	0%	21,9%	21,9%	56,3%
I find this App very confusing to use	28,1%	43,8%	21,9%	0%	6,3%
I feel confident using this App	0%	3,1%	15,6%	37,5%	43,8%
I need to learn many things before I can use this App	15,6%	18,8%	21,9%	12,5%	31,3%

Based on the ten questions presented to responders, it is possible to infer that the test findings were effective.

Table 4.8 summarizes the calculation with a Likert scale.

Table 4.8 Likert Scale Recap

Questions	Results	Question Types	Remark
I will use this App frequently	73.12%	Positive	Agree
I find this app too complicated	18.4%	Negative	Strongly Disagree
I find this App easy to use	88.75%	Positive	Strongly Agree
I think I need technical support to use this App	32.5%	Negative	Disagree
I found the various functions in this App well integrated.	84.37%	Positive	Agree
I think there are too many inconsistencies in this App	24.06%	Negative	Disagree
Most people will be able to learn to use this App quickly	86.87%	Positive	Strongly Agree
I find the App very confusing to use	21.25%	Negative	Disagree
I feel confident using this App	71.87%	Positive	Agree
I need to learn many things before I can use this App well	27.5%	Negative	Disagree

Question types are divided into two categories: positive and negative, which is critical for Likert scale computations. Here are the various scores for the two categories:

Positive Questions:

- Rating Scale 1: Strongly Disagree
- Rating Scale 2: Disagree
- Rating Scale 3: Neutral
- Rating Scale 4: Agree
- Rating Scale 5: Strongly Agree

Negative Question:

- Rating Scale 1: Strongly Agree
- Rating Scale 2: Agree
- Rating Scale 3: Neutral

- Rating Scale 4: Disagree
- Rating Scale 5: Strongly Disagree

This difference can affect the results of the Likert scale value calculation. Table 4.9 below shows sections that can be used to determine the description for each question, ranging from strongly agree to strongly disagree

Table 4.9 Interval Points

Interval Points	Remarks
0 - 19.99%	Strongly Disagree
20 - 39.99%	Disagree
40 - 59.99%	Neutral
60 - 79.99%	Agree
80 - 100%	Strongly Agree

The question's description is determined by the interval value picked. Table 4.8 demonstrates that the answers for all negative questions range from 0 to 39.99%, meaning that the questions provided are not permitted by the majority of respondents, while the opposite occurs for positive questions.

V. CONCLUSION

The web application Psyche has two stages of development: the first is functional development, and the second is staging. During the first inspection, focus on the most important feature, determining whether all features can function properly. The test conditions for the first successful stage can be seen in Table 4.5.

The second testing phase required respondents to execute tests on the Psyche web application and submit comments on Psyche through a questionnaire. Based on the questionnaire's findings, it was determined that respondents were more likely to agree that Psyche could function smoothly.

Psyche finished two stages of testing and was declared successful in each. Psyche can currently determine personality traits from user tweets and offer a quick description of the individual's personality type. In the future, it is believed that the solutions proposed in this study would aid in the development of human resources (HR) in Indonesia and minimize the number of individuals who encounter horizontal mismatches.

REFERENCES

- [1] L. Mahdiah, "Ketidaksesuaian antara Pendidikan dan Kebutuhan Tenaga Kerja Masih Besar," databoks. Diakses: 22 Oktober 2023. [Daring]. Tersedia pada: <https://databoks.katadata.co.id/datapublish/2019/05/28/ketidaksesuaian-antarapendidikan-dan-kebutuhan-tenaga-kerja-masih-besar>
- [2] T. Hoturu, A. Dilly, G. Papuling, P. Studi, A. Bisnis, dan U. H. Namotemo, "Dampak Mismatch Pendidikan-Pekerjaan Terhadap Pengembangan Keahlian Karyawan Di Halmahera," *Jurnal Ilmu Manajemen dan Akutansi*, vol. 10, no. 2, hlm. 102–103, 2022
- [3] A. Saoqillah, "PENGEMBANGAN SDM MELALUI MANAJEMEN DIRI," *Jurnal Riset Manajemen dan Bisnis (JRMB) Fakultas Ekonomi UNIAT*, vol. 1, no. 2, hlm. 145–152, Okt 2016, doi: 10.36226/jrmb.v1i2.18.
- [4] S. Dee and C. Mallari, "Myers-Briggs Type Indicator (MBTI) Personality Profiling and General Weighted Average (GWA) of Nursing Students Myers-Briggs Type Indicator (MBTI) Personality Profiling and General Weighted Average (GWA) of Nursing Students Myers-Briggs Type Indicator (MBTI) Personality Profiling and General Weighted Average (GWA) of Nursing Students 2," 2017.
- [5] Flowchart Adalah: Fungsi, Jenis, Simbol, Dan Contohnya. Dicoding Blog. (2021, July 26). <https://www.dicoding.com/blog/flowchart-adalah/>

