

Early Detection System for Psychological Disorders Based on Supervised Learning with Case Based Reasoning Modeling in Specific Populations

I Wayan Supriana ^{1*}, Luh Arida Ayu Rahning Putri ^{2*}, I Gede Surya Rahayuda ^{3*}

^{*} Informatika, Universitas Udayana

wayan.supriana@unud.ac.id ¹, rahningputri@unud.ac.id ², igedesuryarahayuda@unud.ac.id ³

Article Info

Article history:

Received 2026-08-02

Revised 2026-09-11

Accepted 2026-10-09

Keyword:

*Early Detection,
Psychological Disorders,
Case-Based Reasoning,
Supervised Learning,
Data-Based Systems.*

ABSTRACT

Psychological disorders are conditions that can cause significant changes in an individual's life. However, the symptoms they experience are often more difficult to express because sufferers are unaware that they are experiencing psychological disorders. One of the biggest challenges in treating psychological disorders is detecting the risk of disorders at an early stage, especially in certain populations that may be more vulnerable to these conditions. These vulnerable populations can include workers with long working hours, students with high academic pressure, individuals who have experienced disasters or violence, and groups with a family history of mental illness. One approach to detecting the risk of psychological disorders is a data-driven detection system. This study used the Case-Based Reasoning (CBR) method, an approach that solves problems by leveraging previous case experiences. The aim of this study was to develop an early detection system for indications of psychological disorders by integrating CBR modeling with supervised learning-based data. Based on the test results, the developed CBR model achieved an average accuracy of 76.19% on 14 test cases representing various categories of indications and treatments. These results indicate that the CBR method has quite good performance and has the potential to be used as an aid in the early detection of psychological disorders.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

Psychology is a human mental process influenced by the environment, experiences, and biological conditions. Efforts to maintain a person's psychological well-being focus on aspects of human behavior and mental function [1]. Psychological disorders are conditions that affect a person's thought patterns, emotions, behavior, or mental functioning, thereby disrupting daily activities, social relationships, and quality of life. These disorders can be temporary or chronic, depending on the type, severity, and treatment. Psychological disorders are not simply weaknesses or character flaws, but rather health issues that require special understanding and treatment [2], [3]. Psychological disorders, such as depression, anxiety, and mental health disorders, are a global health issue that is increasingly gaining attention. The World Health Organization (WHO) reports that 1 in 8 people worldwide may experience a psychological disorder, making it a priority health issue in society. The 2024 Asia Care Survey

revealed that stress and burnout are the most common psychological and mental disorders experienced by people in Indonesia, with a prevalence of 56%, followed by sleep disorders at 42.6% and anxiety at 28.2%.

Despite extensive treatment and prevention efforts, one of the greatest challenges in addressing psychological disorders is detecting the risk of developing a disorder at an early stage. Early detection is crucial to prevent the disorder from progressing to a more severe stage, which often requires more complex and expensive interventions. Early detection of psychological disorders remains a challenge, particularly for certain populations that are more vulnerable to these conditions. These vulnerable populations may include workers with long work hours, students with high academic pressure, individuals who have experienced disasters or violence, and groups with a family history of mental illness.

One approach to detecting the risk of psychological disorders is a data-driven detection system. However, a major challenge often encountered in developing psychological

disorder risk prediction systems is the limited quality and availability of data [4]. Many prediction models rely on general population data, which does not reflect risk variability and is not representative of specific population groups. Therefore, the need for a more specific and adaptive psychological disorder detection model, which is addressed in this study, utilizes Case Based Reasoning (CBR) modeling, [5]. CBR can predict the risk of psychological disorders in certain populations based on factors such as age, gender, family history and environmental stressors.

Previous research has applied Case-Based Reasoning (CBR) to support indication categorization across various domains, including psychological disorders [6]. However, these approaches generally focus on case matching based on a case base without explicitly incorporating user population characteristics into the case representation. Conversely, supervised learning approaches offer classification capabilities based on labeled data, yet their performance relies heavily on the quality, representation, and alignment of the training data with the characteristics of the target population. This situation highlights a research gap in the development of screening mechanisms that integrate symptom-based case representations with specific population characteristics.

The objective of this research is to develop an early detection system for psychological disorders using Case-Based Reasoning (CBR) modeling on supervised learning data, serving as an initial identification process to categorize indications based on symptom similarity, techniques in CBR modeling using k-means clustering as case indexing, cosine similarity as calculation of case similarity. CBR is a method that studies and solves a problem based on past experience with a case-based approach that allows the system to handle risk variability more adaptively, according to the characteristics of a particular population [7]. Medical technology needs to receive broad support through the field of information technology so that the public can be more aware, especially of psychological disorders that require special understanding and treatment for certain populations at high risk. The fundamental problem in this research is how to develop an application for early detection of psychological disorders in certain populations by modeling problem solving of past cases to be able to solve new problems with Case Based Reasoning using reference data on psychological disorders on www.alodokter.com.

II. METHOD

This research uses the Cross Industry Standard Process for Data Mining (CRISP-DM) framework to develop a supervised learning-based early detection system for psychological disorders with a CBR approach. The stages involved include business understanding, data understanding, data preparation, modeling, evaluation, and deployment, which systematically support data processing, model development, and system implementation [8].

A. Business Understanding

The business understanding phase focused on formulating system requirements and specifying the problems to be modeled in the early detection system for psychological disorders. The main problem identified was the limitations of predictive models in capturing variations in characteristics across specific populations due to the use of generic data. Therefore, at this stage, the scope of the data to be used was determined, namely attributes relevant to the psychological condition of a specific population. Next, the modeling objective was defined as a classification problem to identify indications of psychological disorders based on historical data. The developed model is capable of accepting individual data inputs, calculating the degree of similarity to previous cases, and generating predictions regarding indication categories and recommendations for initial management.

B. Data Understanding

At this stage, the process carried out sequentially is to dig and explore the data, the steps taken are as follows.:

1. *Data collection:* In this study, the data used to model psychological disorders consisted of 72 disorder categories, with a single case representation for each category; each case was defined by 10 symptoms and corresponding treatment information, sourced from www.alodokter.com. The model testing phase involved two groups: university students and office workers. Attributes for students included credit hours (SKS), number of weekly assignments, organizational involvement, average sleep duration, deadline-related stress levels, leisure/rest time, gender, field of study, and academic semester. Attributes for office workers included age, gender, job position, length of employment, and average daily working hours. This study utilized symptom and treatment information available from public sources and did not use personal identities or patient medical records.

2. *Data analysis:* This process is conducted exploratively on text data to understand its structure and quality. This process includes identifying variations in symptom terms, text length, and word frequency. Techniques include analyzing term frequency, document length distribution, and identifying synonyms or terms with similar meanings.

3. *Validate data:* Data validation includes evaluating the completeness of symptom descriptions in each individual population, the consistency of terminology used, and the distribution of data across classes to identify potential data imbalances. The results of this stage form the basis for determining text preprocessing techniques, such as normalization, tokenization, and feature transformation, that will be used in the modeling stage.

C. Data Prepration

Data preparation aims to transform data to improve its quality. The main steps in data preparation are::

1. Selecting and sorting data: data will be selected based on the alignment between data quality and the predetermined analytical tasks.
2. Text data preprocessing: converting text data into a numerical format that can be processed by a model. This stage includes text normalization (converting characters to lowercase), tokenization, stopwords removal, and stemming.
3. Data cleaning: performed to improve data quality. This involves handling missing values, removing duplicate data, and addressing formatting inconsistencies.
4. Data reconstruction: This process modifies the data structure to better suit the CBR model; the steps involved are normalization, data transformation, and feature engineering.
5. Case representation: The case representation technique uses TF-IDF to calculate the importance weight of each word in each case. The resulting weighting is used to build a case base, facilitating the similarity calculation process with new cases.

$$W_{dt} = tf_{dt} \times IDF_t \quad (1)$$

$$IDF(t) = \log \frac{N}{df(t)} \quad (2)$$

$$TF - IDF(d, t) = TF(d, t) \times IDF(t) \quad (3)$$

where: d : the d -th sentence, t : the t -th word of the keyword, w : the weight of the d -th sentence with respect to the t -th word, tf : number of words searched in a sentence, IDF : inversed document frequency. After determining the weight of the case, the next stage is indexing which aims to index the case.

D. Modelling

The CBR model development technique is carried out sequentially as follows:

1. *Case indexing*: facilitates the arrangement of data structures in organizing records in storage to optimize several types of case retrieval operations [9], [10]. In this study, all attribute types will be indexed to facilitate searching. The method used for indexing is the k-means clustering algorithm. Each case is placed in a specific cluster that represents the symptom pattern within a specific population context, such as workers or students. The optimal number of k clusters formed in the k -means clusters is determined using the Elbow method to calculate inertia based on the optimal number of clusters. The k -means cluster flow is shown in Figure 1.

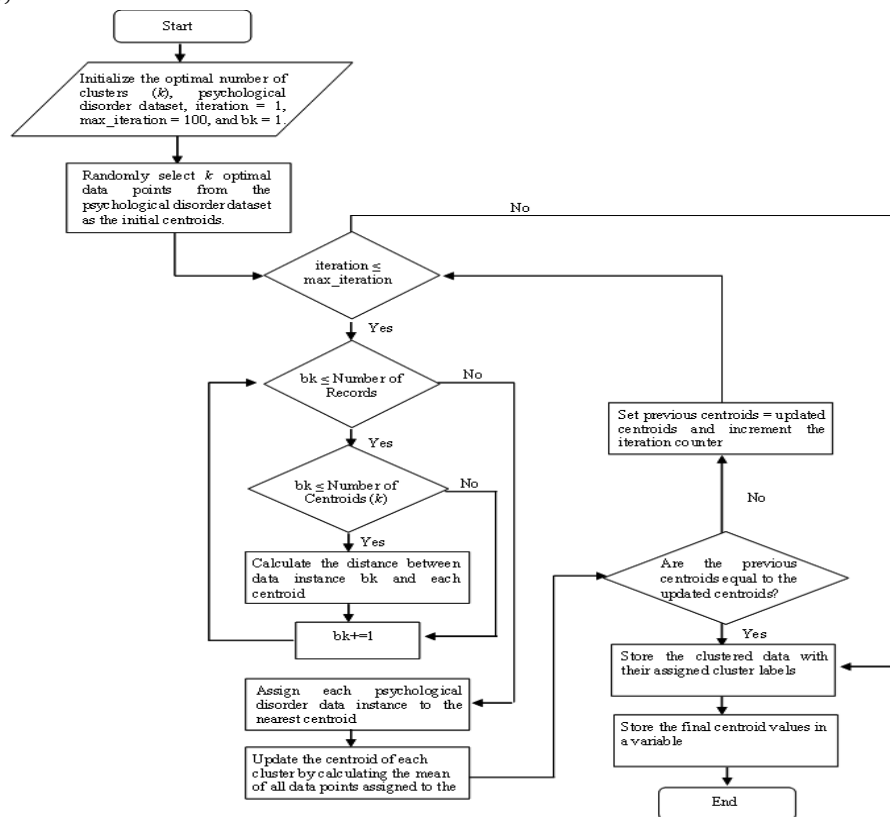


Figure 1. Indexing with K-Means

2. *Retrieve*: The retrieval stage aims to find previous cases in the case base that are most similar to the new case. To increase the relevance of the results, the search process is limited to clusters or groups of data that represent a specific population resulting from the previous clustering process

[11], [12]. Each case is represented as a TF-IDF-based feature vector, allowing for numerical similarity calculations. The method used is cosine similarity, by comparing the new case vector to all existing cases. The calculation is carried out through local similarities on each feature and accumulated

into a global similarity to obtain the overall similarity level. Local similarity is calculated on each feature between the new case and the old case. Because the data is represented as a TF-IDF vector, the local similarity is represented by the weight of each term in both cases.

$$sim_{lokal}(k) = w_{i,k} \times w_{j,k} \quad (4)$$

dimana:

- $w_{i,k}$ = TF-IDF weight of the k-th term in the new case
- $w_{j,k}$ = TF-IDF weight of the k-th term in the previous case

Global similarity is obtained by aggregating all local similarities using the cosine similarity method, resulting in an overall level of similarity between cases.

$$Sim_{global}(d_i, d_j) = \frac{\sum_{k=1}^n w_{i,k} \times w_{j,k}}{\sqrt{\sum_{k=1}^n (w_{i,k})^2} \times \sqrt{\sum_{k=1}^n (w_{j,k})^2}} \quad (5)$$

di mana:

- d_i = new cases
- d_j = old case
- n = number of features/terms

3. *Reuse*: new cases are compared with all cases in the case base to obtain a global similarity value [13], [14]. The old case with the highest similarity score will be selected as a candidate solution. The solution is determined based on a predetermined threshold, where only cases with a similarity score ≥ 0.8 are considered relevant. A similarity threshold of 0.8 is used as an initial criterion to determine whether a case bears sufficient resemblance to a case in the case base. Conversely, if no case meets the threshold, the new case will be processed further in the revision stage.

4. *Revision*: this stage is performed if the retrieval process fails to find a case with a global similarity value that meets the established threshold [15], [16]. In this situation, new cases cannot directly adopt the solution from the base case, requiring an adjustment process or search for alternative solutions. The revision process is carried out using the nearest neighbor algorithm, considering a number of nearby cases based on the highest similarity value. The solution is determined based on the proximity between cases, which can then be aggregated. The result of this stage is a solution that has been adjusted to the characteristics of the new case. In general, the set of nearest neighbors is defined as:

$$NN_k(C_{new}) = \{C_1, C_2, \dots, C_k\} \quad (6)$$

dengan:

- $NN_k(C_{new})$ = the set of k cases closest to a new case
- C_i = old cases with the highest similarity value

The second stage involves evaluating the nearest neighbor results, where the system assigns an "uncertain" or "low similarity" status and recommends further evaluation by an expert.

5. *Retain*: This stage is the process of storing new cases and their obtained solutions into the case base to enrich the knowledge base. New cases are represented in the form of TF-IDF-based feature vectors and are equipped with solution labels resulting from the retrieve or revise stage [17], [13]. Before being saved, validation is performed to ensure that new cases are not duplicates or have a very high degree of similarity to existing cases. If the case is declared valid, it is indexed and added to the case base. This process allows the system to continuously learn incrementally, thus improving performance and accuracy in handling new cases in subsequent iterations. New cases to be saved are defined as:

$$C_{new} = (X_{new}, S_{new}) \quad (7)$$

Where:

- X_{new} = feature vector (TF-IDF result)
- S_{new} = solution (result of retrieve or revise)

The process of adding cases to the case base is formulated as:

$$CB_{t+1} = CB_t \cup \{C_{new}\} \quad (8)$$

Where:

- CB_t = case base before addition
- CB_{t+1} = case base after addition

To avoid duplication, a condition is added:

$$\max(Sim(C_{new}, C_i)) < \delta \quad (9)$$

Where:

- δ = similarity threshold (e.g. 0.9)
- If it exceeds $\rightarrow$ it is not saved (considered a duplicate)

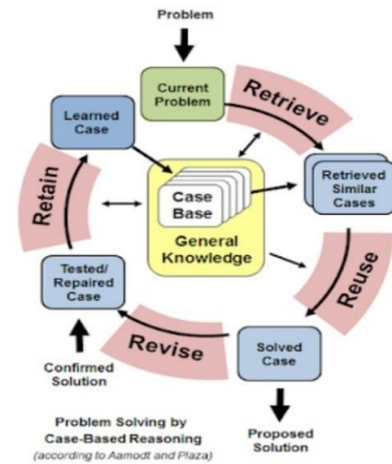


Figure 2. Case Based Reasoning Stages

E. System Design

The system is designed as a screening tool and is not a substitute for psychologists, psychiatrists, or healthcare professionals. If screening results indicate a risk or insufficient similarity, users are directed to seek further evaluation from a professional. As shown in Figure 3, the system workflow begins with data input comprising

information on psychological disorders sourced from alodokter.com and data on new cases from a specific population. In the initial stage, symptom text data undergoes preprocessing to generate a structured representation, followed by feature weighting using the TF-IDF method to convert the data into numerical vectors..

Next, to accommodate differences in characteristics between populations, a data clustering process is performed using the K-Means algorithm and the Elbow method is used to determine the optimal number of clusters. This process aims to group cases based on similar symptom patterns

relevant to a specific population, thus focusing the solution search process. In the retrieve stage, new cases are compared with cases in the appropriate cluster using cosine similarity. If the similarity value meets a threshold, the solution from the old case is reused. Otherwise, a revision stage is performed using the nearest neighbor approach to generate a solution that better matches the characteristics of the new case. The final stage is retain, which involves storing new cases and their solutions in the case base to enrich the representation of cases in a specific population, thereby improving the accuracy and relevance of predictions for subsequent use.

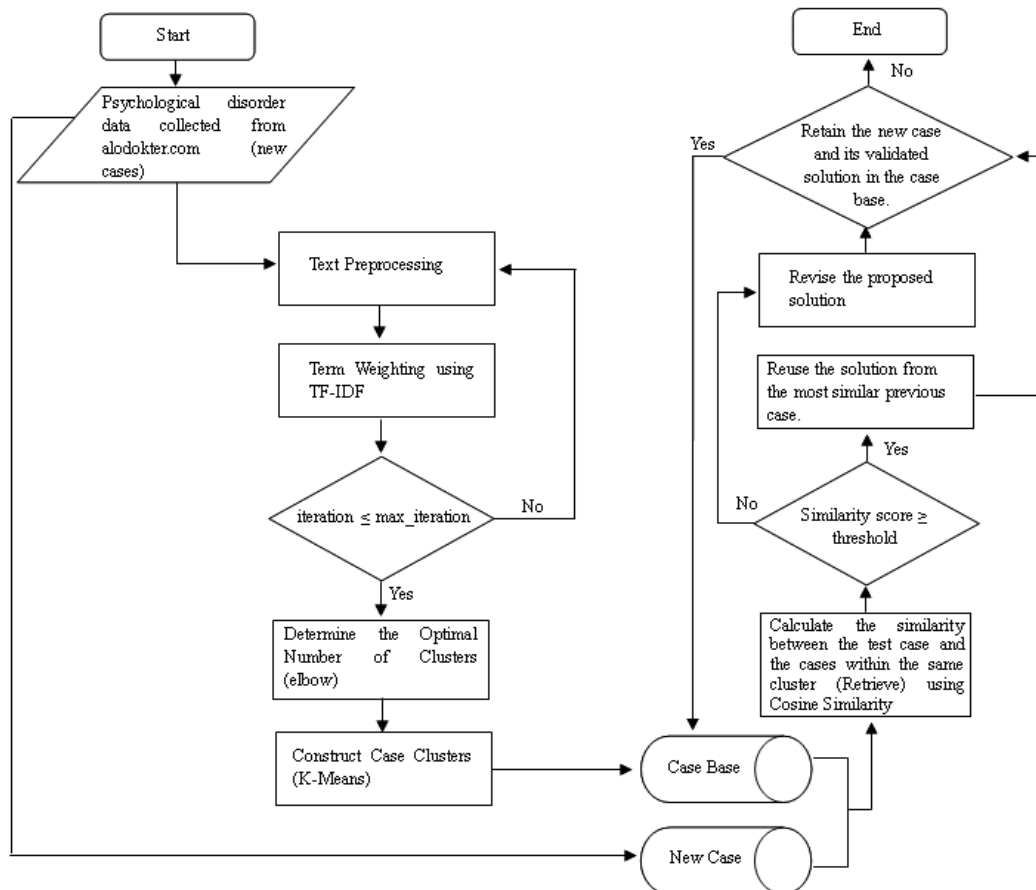


Figure 3. Research Flow

III. RESULTS AND DISCUSSION

A. Research Data

The data source for this study is www.alodokter.com, which comprises 72 types of psychological disorders with 5 to 10 symptoms. The data includes disease names, symptoms, and treatments. Each disorder is represented by a set of symptoms and is accompanied by treatment information as a reference for case-based solutions. Symptom data is the primary feature used in the modeling process, while disease name and treatment information are used as solution labels in the Case-Based Reasoning approach.

TABLE 1
PSYCHOLOGICAL DISORDER DATASET WWW.ALODOKTER.COM

Disease	Symptom	Handling
Schizophrenia	- Changes in thought patterns and behavior - Inability to socialize - Withdrawal from social environments - Difficulty sleeping - Decreased performance	- Antipsychotic drugs to treat symptoms of hallucinations and delusions, such as aripiprazole, clozapine, chlorpromazine - Psychotherapy to control

	• Loss of interest and motivation • Changes in facial expressions • Easily irritated and depressed	symptoms, which is generally followed by medication
Depression	• Excessive anxiety and worry • Emotional instability • Feelings of hopelessness or frustration • Constantly feeling tired and powerless • Unexplained dizziness and pain • Decreased appetite	• Undergoing psychotherapy or psychological therapy • Prescribing antidepressant medication • Hospitalization for severe depression
etc	etc	etc

TABLE 2
CERTAIN POPULATION CRITERIA

Student	
No	Criteria
1	Credits in one semester
2	Number of active assignments/courses per week
3	Involvement in campus organizations
4	Average nightly sleep time in the last week
5	Stress levels due to numerous deadlines
6	Free time to rest
7	Gender
8	Study program
9	Semester
Office workers	
1	Age
2	Gender
3	Title/position
4	Length of working
5	Average working hours per day

To accommodate specific population characteristics, this study uses additional attributes that represent individual conditions based on their groups, namely students and office workers, as presented in Table 2. These attributes are used as supporting features in the modeling process to improve the system's ability to capture contextual variations in the risk of psychological disorders. All data on symptoms and disease treatment will go through the process of segmentation, lowercase and remove punctuation, tokenization, filtering, and stemming to form a weight matrix with TF-IDF. Table 3 shows the results of an example of schizophrenia.

TABLE 3
TEXT PREPROCESSING RESULTS SEGMENTATION, LOWERCASE AND REMOVE PUNCTUATION, TOKENIZATION, FILTERING, AND STEMMING

Disease	Symptom	Handling
Schizophrenia	change, pattern, thought, behavior, inability, socialize, withdrawal, social, difficulty, sleep, decline, achievement,	antipsychotic medication, hallucination, delusion, Aripiprazole,

	loss, interest, change, expression, irritability, depression	psychotherapy, symptom control
--	--	--------------------------------

TABLE 4
RESULTS OF NORMALIZATION OF CERTAIN POPULATION ATTRIBUTES OF STUDENTS WITH SCHIZOPHRENIA DISORDERS

Attribute	Value	Preprocessing Results
credits	24	0.80
number of assignments/week	6	0.75
organizational involvement	active	1
average sleep time	5 hours	0.42
deadline pressure level	high	0.90

TABLE 5
RESULTS OF NORMALIZATION OF CERTAIN POPULATION ATTRIBUTES OF OFFICE WORKERS IN SCHIZOPHRENIA DISORDERS

Attribute	Value	Preprocessing Results
age	30	0.60
gender	male	1
Position	staff	0.50
length of working	5 years	0.50
working hours/day	9 hours	0.75

Both text data and population attributes have been transformed into structured and numeric forms. This representation is then used to generate feature vectors for the symptom data, combined with population attributes as additional features, as shown in Tables 4 and 5.

B. Case Representation

The representation technique uses TF-IDF to calculate the importance weight of each word in each case of psychological disorders. The resulting word weighting is used to build a case base, facilitating the process of calculating similarity with new cases, as presented in Tables 6 and 7.

TABLE 6
TF-IDF WEIGHT CALCULATION ON SCHIZOPHRENIA SYMPTOMS

Disease	Symptoms/terms	TF	TF-IDF
Schizophrenia	change	0.25	0.347
	pattern	0.25	0.520
	thought	0.25	0.520
	behavior	0.25	0.520
	inability	0.333	0.693
	patient	0.333	0.693
	socialize	0.333	0.693
	withdrawal	0.25	0.520
	self	0.25	0.520
	environment	0.25	0.520
	social	0.25	0.520
	difficulty	0.5	1.040
	sleep	0.5	1.040
	experience	0.333	0.693
	decline	0.333	0.693
	achievement	0.333	0.693
	loss	0.333	0.693

	interest	0.333	0.693
	motivation	0.333	0.693
	change	0.333	0.462
	expression	0.333	0.693
	face	0.333	0.693
	irritability	0.333	0.693
	anger	0.333	0.693
	depression	0.333	0.693

TABEL 7
TF-IDF WEIGHT CALCULATION FOR SCHIZOPHRENIA HANDLING

Penyakit	Treatment/Term	TF	TF-IDF
Schizophrenia	Medication	0.111	0.000
	Antipsychotic	0.111	0.077
	Manage	0.111	0.077
	Symptoms	0.111	0.000
	Hallucinations	0.111	0.077
	Delusions	0.111	0.077
	Aripiprazole	0.111	0.077
	Clozapine	0.111	0.077
	Chlorpromazine	0.111	0.077
	Psychotherapy	0.167	0.000
	Control	0.167	0.116
	Symptoms	0.167	0.116
	Generally	0.167	0.000
	Administration	0.167	0.116
	Medication	0.167	0.116

After determining the weights for psychological disorder cases, the next step is k-means indexing, which aims to quickly and easily index psychological disorder cases in the database. Model testing is performed eight times. The highest average SSE difference from the eight tests will be the optimal k value. This test will use a k value range of 1 to 10. The results of the optimal number of clusters test are shown in Table 8.

TABEL 8
SSE VALUE OF EACH K FOR 8 TESTING TIMES

K	Nilai SSE Pengujian ke-										Rata-rata	Std. Deviasi
	1	2	3	4	5	6	7	8	9	10		
1	4.647,980	4.647,980	4.647,980	4.647,980	4.647,980	4.647,980	4.636,794	4.636,794	4.636,794	4.636,794	4643.51	5,752
2	2.011,329	2.011,329	2.011,329	2.011,329	2.011,329	2.011,329	2.013,608	2.013,608	2.013,608	2.013,608	2012.24	1,174
3	1.759,238	1.653,078	1.931,329	1.759,238	1.930,641	1.989,094	1.651,398	1.777,134	1.651,398	1.651,398	1773.53	129,409
4	1.472,538	1.490,389	1.490,389	1.651,009	1.527,636	1.490,564	1.650,660	1.607,106	1.635,040	1.496,953	1546.84	69,741
5	1.670,828	1.395,941	1.477,500	1.387,603	1.465,172	1.387,603	1.465,748	1.532,896	1.430,223	1.395,695	1403.28	88,703
6	1.354,725	1.382,523	1.329,616	1.384,132	1.364,997	1.328,549	1.331,272	1.393,376	1.349,059	1.423,591	1362.73	30,805
7	1.317,248	1.355,218	1.289,972	1.305,668	1.321,354	1.331,739	1.247,412	1.246,311	1.347,719	1.410,097	1318.26	48,894
8	1.288,371	1.410,872	1.342,793	1.283,130	1.227,134	1.328,640	1.296,535	1.274,371	1.283,787	1.231,355	1302.22	48,887

Table 8 shows a continuous decrease in inertia without a clear "elbow point." However, there is a significant difference between 2 and 3 clusters. Therefore, the number of clusters used is 2, taking into account the total data set of 72 psychological disorders. Once the optimal k value is obtained, all old cases in the case base will be indexed according to their cluster. The test is performed eight times. The highest average SSE difference from the eight tests will be the optimal k value.

C. Similarity Analysis

Similarity calculations are performed at two levels: local similarity and global similarity. Local similarity measures the similarity of each feature, while global similarity is the aggregation of all features that represent a single case as a whole. [18], [19], [20].

1. Local similarity is calculated on each dimension of text features and population attributes. For text features, local similarity is represented by the contribution of each term's TF-IDF weight to the overall similarity. Terms that appear in both cases with high weights provide the greatest similarity value. Meanwhile, for population attributes, local similarity is calculated based on the proximity of values between attributes, such as the normalized difference in values. Cases with similarity values above 0.8 have a strong match on most local features, especially on dominant terms such as the main symptom.
2. Global similarity is the aggregation of all local similarities calculated using cosine similarity. This value reflects the overall level of similarity between two cases. The majority of cases have global similarity values above 0.8, indicating that the feature representation and clustering process successfully grouped cases with similar characteristics..

D. Model Evaluation

The evaluation phase involved testing several inputs to evaluate the model. The test data consisted of 14 new cases based on student and worker populations. The first evaluation used the student population with 7 cases, and the second used the worker population with 7 cases. The labels on the test cases do not represent clinical diagnoses; instead, they serve as case-based reference labels for evaluating the model's retrieval and classification capabilities regarding the categories available in the case base.

TABEL 9
STUDENT POPULATION EVALUATION VALUE

No	Case Code	Simil ar Case	Simi larit y	CBR Prediction	Label	Accu rate
1	PK_01	Case_020	0.93	Burnout	Burnout	Yes
2	PK_02	Case_011	0.88	Anxiety Disorder	Anxiety Disorder	Yes
3	PK_03	Case_020	0.76	Depressio n	Depressio n	Yes
4	PK_04	Case_020	0.81	Burnout	Stres	No
5	PK_05	Case_001	0.79	No Disorder	Anxiety Disorder	No
6	PK_06	Case_020	0.90	Depressio n	Depressio n	Yes
7	PK_07	Case_003	0.86	Stress	Stress	Yes

TABLE 10
OFFICE WORKER POPULATION EVALUATION VALUE

No	Case Code	Similar Case	Similarity	CBR Prediction	Label	Accuracy
1	Mhs_01	Case_003	0.96	Stress	Stress	Yes
2	Mhs_02	Case_010	0.91	Anxiety Disorder	Anxiety Disorder	Yes
3	Mhs_03	Case_002	0.75	Depression	Burnout	No
4	Mhs_04	Case_003	0.88	Stress	Stress	Yes
5	Mhs_05	Case_020	0.82	Burnout	Burnout	Yes
6	Mhs_06	Case_015	0.79	Anxiety Disorder	Anxiety Disorder	Yes
7	Mhs_07	Case_017	0.84	No Disturbance	No Disturbance	Yes

This evaluation was conducted to compare the performance of the CBR system. Accuracy values at the retrieval stage were compared between specific population types (students and office workers). The results of testing on the student population are shown in Table 9, and those on the office worker population are shown in Table 10. Based on Table 9, for the student population, the CBR model provided an accuracy of 71.42%. Table 4.10 shows that for the office worker population, the CBR model provided an accuracy of 85.71%. In the CBR model, the search for relevant cases only occurs on the basis of cases that have the same cluster as the new case. Therefore, computation time at the retrieval stage is faster. The accuracy value produced by the CBR model does not provide 100% optimal results because the system cannot handle two or more cases that have the same similarity value as the new case. If this occurs, the CBR model will select the case with the first serial number.

The overall performance of the CBR model achieved an accuracy of 76.19% across 14 test cases, representing 14 different indication categories from the case base. Most cases with similarity values above the threshold were correctly classified through the reuse mechanism. This indicates that the case base used was able to represent the symptom pattern quite well. Conversely, prediction errors generally occurred in cases with similarity values below the threshold, where the revision process using nearest neighbors was unable to produce a fully appropriate solution. This could be due to the high variation in symptoms between disorders, as well as significant differences in population attributes. The integration of population-specific attributes contributes to improving prediction accuracy, especially in cases with similar characteristics within the same population group. Cases originating from the same population tend to have a more accurate prediction rate than cases across populations.

IV. CONCLUSION

This study develops a CBR-based screening mechanism for indications of psychological disorders, taking into account the characteristics of student and office worker populations. Preliminary testing on 14 cases yielded an accuracy of 76.19%, with 71.42% accuracy for the student group and 85.71% for the office worker group. These results demonstrate that the CBR approach can serve as a proof of concept for case-similarity-based screening.

ACKNOWLEDGMENTS

The author would like to thank the Institute for Research and Community Service (LPPM) of Udayana University for the research grant funds provided and www.alodokter.com for information on psychological disorders data.

REFERENCES

- [1] M. Knapp and G. Wong, "Economics and mental health: the current scenario," *World Psychiatry*, vol. 19, no. 1, pp. 3–14, 2020, doi: 10.1002/wps.20692.
- [2] F. Porru, M. Schuring, U. Bültmann, I. Portoghese, A. Burdorf, and S. J. W. Robroek, "Associations of university student life challenges with mental health and self-rated health: A longitudinal study with 6 months follow-up," *J. Affect. Disord.*, vol. 296, no. September 2021, pp. 250–257, 2022, doi: 10.1016/j.jad.2021.09.057.
- [3] R. Rahim, W. Purba, M. Khairani, and R. Rosmawati, "Online Expert System for Diagnosis Psychological Disorders Using Case-Based Reasoning Method," *J. Phys. Conf. Ser.*, vol. 1381, no. 1, 2019, doi: 10.1088/1742-6596/1381/1/012044.
- [4] M. R. K. S. N. K. Weinberg, "Application of case-based reasoning in a fault detection system on production of drippers," *Appl. Soft Comput.*, vol. 75, pp. 227–232, 2019, [Online]. Available: <https://doi.org/10.1016/j.asoc.2018.11.017>
- [5] Z. Tocimáková, L. Puztová, J. Paralič, and D. Pella, "Case-based reasoning for support of the diagnostics of cardiovascular diseases," *Stud. Health Technol. Inform.*, vol. 270, pp. 537–541, 2020, doi: 10.3233/SHTI200218.
- [6] A. Mathew, L. Ma, and V. L. Narasimhan, "Case-based reasoning for data warehouse schema design," *36th Int. Conf. Comput. Ind. Eng. ICC IE 2006*, vol. 2006, pp. 3799–3810, 2006.
- [7] I. M. S. Bimantara and I. W. Supriana, "Case Based Reasoning (CBR) for Obesity Level Estimation Using K-Means Indexing Method," *J. Ilm. Kursor*, vol. 11, no. 4, pp. 145–156, 2023, doi: 10.21107/kursor.v11i4.268.
- [8] A. Aamodt and E. Plaza, "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches," vol. 7, pp. 39–59, 2018.
- [9] R. Houssin, J. Renaud, A. Coulibaly, D. Cavallucci, and F. Rousselot, "TRIZ theory and case based reasoning: synergies and oppositions," *Int. J. Interact. Des. Manuf.*, vol. 9, no. 3, pp. 177–183, 2015, doi: 10.1007/s12008-014-0252-1.
- [10] W. Y. C. Z.-M. D. C. P. Collet, "An ontology-based approach for inventive problem solving," *Eng. Appl. Artif. Intell.*, vol. 27, pp. 175–190, 2014, doi: <https://doi.org/10.1016/j.engappai.2013.07.005>.
- [11] V. E. Ekong, U. G. Inyang, and E. A. Onibere, "Intelligent Decision Support System for Depression Diagnosis Based on Neuro-fuzzy-CBR Hybrid," *Mod. Appl. Sci.*, vol. 6, no. 7, pp. 79–88, 2012, doi: 10.5539/mas.v6n7p79.
- [12] A. De Renzis, M. Garriga, A. Flores, A. Cechich, and A. Zunino, "Case-based Reasoning for Web Service Discovery and Selection," *Electron. Notes Theor. Comput. Sci.*, vol. 321, pp. 89–

- 112, 2016, doi: 10.1016/j.entcs.2016.02.006.
- [13] A. Smiti and Z. Elouedi, "Dynamic maintenance case base using knowledge discovery techniques for case based reasoning systems," *Theor. Comput. Sci.*, vol. 817, pp. 24–32, 2020, doi: 10.1016/j.tcs.2019.06.026.
- [14] W. Bannour, A. Maalel, and H. H. Ben Ghezala, "Case-based Reasoning for crisis response: Case representation and case retrieval," *Procedia Comput. Sci.*, vol. 176, pp. 1063–1072, 2020, doi: 10.1016/j.procs.2020.09.102.
- [15] A. Zidi, A. Bouhana, M. Abed, and A. Fekih, "An ontology-based personalized retrieval model using case base reasoning," *Procedia Comput. Sci.*, vol. 35, no. C, pp. 213–222, 2014, doi: 10.1016/j.procs.2014.08.101.
- [16] P. Zhang, A. Essaid, C. Zanni-Merk, and D. Cavallucci, "Case-based Reasoning for Knowledge Capitalization in Inventive Design Using Latent Semantic Analysis," *Procedia Comput. Sci.*, vol. 112, pp. 323–332, 2017, doi: 10.1016/j.procs.2017.08.004.
- [17] K. P. Tripathi, "A Review on Knowledge-based Expert System : Concept and Architecture," *Artif. Intell. Tech. - Nov. Approaches Pract. Appl.*, vol. 4, no. 4, pp. 19–23, 2011.
- [18] Z. P. Fan, Y. H. Li, X. Wang, and Y. Liu, "Hybrid similarity measure for case retrieval in CBR and its application to emergency response towards gas explosion," *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2526–2534, 2014, doi: 10.1016/j.eswa.2013.09.051.
- [19] K. A. J. Lu, "Ontology-supported case-based reasoning approach for intelligent m-Government emergency response services," *Decis. Support Syst.*, vol. 55, no. 1, pp. 79–97, 2013, [Online]. Available: <https://doi.org/10.1016/j.dss.2012.12.034>
- [20] S. Leanti, L. Rosa, D. B. Diep, and I. F. Nes, "Construction and Application of a," vol. 2, no. 17, pp. 1958–1963, 2012, [Online]. Available: http://dx.doi.org/10.2991/978-94-6463-172-2_217