Identifying Fear of Missing Out (FOMO) in Adolescents Using K-Nearest Neighbors: An Experimental Study of k-Values and Distance Metrics

Ricco Wahyu Pamungkas [1]*, Mochammad Anshori [2]*, Wahyu Teja Kusuma [3]*
* Informatics study program, Institut Teknologi, Sains, dan Kesehatan RS. DR. Soepraoen Kesdam V/BRW
riccopamungkas@gmail.com [1], moanshori@itsk-soepraoen.ac.id [2], wtkusuma@itsk-soepraoen.ac.id [3]

Article Info

Article history:

Received 2026-04-28
Revised 2026-05-19
Accepted 2026-05-25

Keyword:

Classification,
Fear of Missing Out,
K-Nearest Neighbor,
KNN,
Psychoinformatics.

ABSTRACT

Fear of Missing Out (FOMO) is a psychological phenomenon commonly experienced by teenagers due to the high intensity of social media use, and has the potential to cause emotional and social impacts if not identified early. The main problem in identifying FOMO is its internal nature and the difficulty in measuring it objectively using conventional methods. This research proposes a data mining-based classification approach using K-Nearest Neighbor (KNN) to identify the level of FOMO in adolescents. The dataset was obtained from 136 respondents through a questionnaire that included demographic data and the ON-FoMO scale. The research stages include data preprocessing (encoding and Min-Max normalization), data splitting using stratified holdout (80:20), and experiments varying K (3–19) and distance metrics (Euclidean, Manhattan, Chebyshev). The experimental results show that the combination of Euclidean distance with K=11 yields the best performance with an accuracy of 85.71%, ROC AUC of 0.786, Precision–Recall AUC of 0.826, and sensitivity of 100%. The experimental results indicate that the selection of the K parameter and the distance method significantly affect classification performance. Overall, this study concludes that the KNN algorithm with the optimal configuration is effective as an initial screening method for the level of FOMO in adolescents in an systematic and data-based manner.

I. Introduction

The rapid development of information and communication technology has brought teenagers into the digital era, where social media has become an important part of daily activities [1]. One of the psychological phenomena that arises due to the high use of social media is Fear of Missing Out (FOMO), which is a feeling of anxiety when someone feels they are missing out on important information, activities, or experiences in their social environment [2]. Teenagers are the group most vulnerable to experiencing FOMO because of the high need to be accepted and recognized by social groups [3]. The impact of FOMO has been proven through various studies, including a significant relationship with impulsive buying behavior, post-purchase regret, and tendencies towards digital consumer behavior [4][5][6]. The latest data also shows the urgency of this problem, such as findings from the University of Indonesia Maju (UIMA) and the Institute of Social and Political Sciences (IISIP) Jakarta (2025), which revealed that FOMO contributes to emotional mental disorders and even the risk of suicidal ideation in junior high school students [7]. Additionally, a recent study in East Java found that 110 teenagers in Jombang identified themselves as having FOMO in 2023 [8], showing that this phenomenon is not only a global issue, but is also a real problem in Indonesia.

Although FOMO has a serious psychological impact, identifying it remains challenging because it is an internal condition that cannot be observed directly. As a result, traditional methods such as interviews and observations are often not systematic and rely heavily on respondents' perceptions [9]. On the other hand, the high use of social media among teenagers in Indonesia increases the risk of FOMO, so early preventive measures are needed [10]. For this reason, the ability to predict the level of FOMO is very important because the characteristics of FOMO are complex, involving emotional, cognitive, and social aspects, making it difficult to measure accurately using conventional methods [11]. The development of data-based predictive models is necessary so that the FOMO identification process can be carried out more objectively, quickly, and accurately, so that at-risk adolescents can be treated earlier.

The current development of machine learning methods provides opportunities for psychological research to carry out classification processes more objectively and efficiently. Machine learning approaches have been used in various studies to detect psychological conditions, including the classification of depression and other mental disorders [12]. One of the widely used algorithms is K-Nearest Neighbor (KNN), which is a classification method that determines the label of new data based on its proximity to previous data [13][14]. Several studies have shown that KNN is effective in predicting mental health conditions with competitive levels of accuracy [15][16]. In addition, KNN has the advantage of simplicity, flexibility in various types of data, and does not require certain distribution assumptions, so it remains effective for use on non-linear data patterns [17].

In developing a KNN model, selecting the right parameters is an important factor in achieving optimal classification performance. The two main parameters that determine the quality of the model are the choice of distance measure and the value of k, as the number of nearest neighbors. Distance measure is used to calculate the proximity between data, which can be calculated using various methods such as Euclidean, Manhattan, or Chebyshev. Each distance method has different characteristics and can produce varying model performance depending on the data structure. In addition, the value of k also plays an important role in determining classification decisions, where a k value that is too small risks making the model sensitive to noise, while a k value that is too large can reduce prediction accuracy. Research shows that the selection of appropriate KNN parameters, both in terms of distance methods and the value of k, can significantly improve the accuracy and stability of the model [18][19][20].

Based on the aforementioned problem, this research aims to develop a classification system for the level of FOMO in adolescents using the KNN algorithm. This research examines several distance measures and variations of the value of k to determine the parameter combination that yields the best performance. The main result of this research is the discovery of the most optimal KNN parameter configuration in predicting FOMO levels, so the resulting model can be used as a more data-driven and systematic initial screening tool to assist in the handling process related to FOMO.

The novelty of this study lies in the comparative analysis of multiple distance metrics and K-value configurations in KNN for adolescent FOMO classification using the validated Indonesian ON-FoMO scale. In addition, this study evaluates model performance using multiple evaluation metrics, including ROC Curve and Precision–Recall Curve, to obtain a more comprehensive understanding of classification performance.

II. METHOD

A. Research Design

This research uses a workflow that consists of model evaluation. Each stage is designed so that the FOMO level classification process can be carried out systematically, structurally, and methodologically valid. In general, the research flow is illustrated in Figure 1.

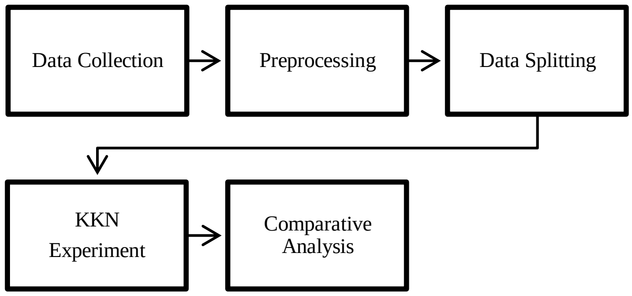

Figure 1. Research Flow of the KNN Method for FOMO Level Classification

Figure 1 shows the research process flow used in this study. The research began with data collection using questionnaires, followed by preprocessing to ensure the data is in a clean condition. Next, the data is divided into training data and test data.

After that, the classification process using the K-Nearest Neighbor (KNN) algorithm is performed on the training data. Various K values are tested to determine the best K value based on the resulting accuracy. The final stage is model evaluation using accuracy, confusion matrix, precision, recall, and F1-score to assess the model's performance in predicting the level of FOMO.

B. Data Collection

Data collection in this study was conducted primarily through the distribution of questionnaires to respondents. Respondents were selected based on their alignment with the research criteria, namely, teenagers who actively use social media and are willing to fill out the questionnaire. The research instrument consists of two main parts, namely demographic data and the FOMO scale.

The demographic data section contains questions that do not use a Likert scale. The information collected includes age, gender, the most frequently used social media, and the duration of social media use per day. Additionally, there is a direct question regarding the respondents' self-perception related to FOMO experiences, namely "Do you feel like you are experiencing FOMO?" with the answer options yes or no. This demographic data is used to describe the basic profile of respondents and support the analysis of social media user characteristics.

The FOMO scale section in this study uses the Indonesian version of the Online Fear of Missing Out (ON-FoMO) Scale, which has been validated by Kurniawan and Utami (2022). The ON-FoMO Scale is a psychological measurement tool specifically designed to measure the level of Fear of Missing Out in the context of online activities and social media use, making it relevant to the current conditions of teenagers. This scale consists of 20 statements answered using a 1–5 Likert scale, with categories 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, and 5 = Strongly Agree.

Table 1 maps 20 items of the ON-FoMO scale statements into four main dimensions, namely Anxiety (Anx), Need to Belong (Bel), Addiction (Add), and Need for Popularity (Pop), in accordance with the measurement tool structure developed in previous research [2].

The Anxiety Dimension (Anx) reflects feelings of anxiety, restlessness, and negative affect when individuals cannot access the internet or social media. The Need to Belong (Bel) dimension describes the individual's need to remain connected, accepted, and be part of a social group on social media. The Addiction (Add) dimension measures the tendency to use social media excessively and the difficulty in controlling online behavior. Meanwhile, the Need for Popularity (Pop) dimension relates to the need to obtain attention, likes, comments, and social validation from other social media users.

The results of the validation study indicate that the ON-FoMO scale has good construct validity and reliability, with a Cronbach's alpha value of 0.89, making it suitable for measuring the level of FOMO in the online context among the adolescent population in Indonesia.

The ON-FoMO scale was selected in this study because it has been culturally adapted and validated for Indonesian adolescents. In addition, the scale demonstrates high internal consistency, indicated by a Cronbach's alpha value of 0.89, and has been used in previous studies related to online social behavior and Fear of Missing Out in Indonesia.

The total ON-FoMO score was calculated by averaging the responses to all Likert-scale questionnaire items. The respondents were then classified into two categories, namely low FOMO and high FOMO, based on the median value of the total score distribution. Respondents with scores above the median were categorized as high FOMO, while respondents with scores equal to or below the median were categorized as low FOMO.

TABLE 1
MAPPING FOMO SCALE ITEMS BASED ON FOUR DIMENSIONS

Items	Anx	Bell	Add	Pop
I feel anxious when I can't access social media.	✓			
If I don't have access to social networks, I will think of ways to connect.	✓			
I keep thinking about social media when I can't access it.	✓			
I usually feel annoyed if I stay disconnected from social media for too long.	✓			
I feel anxious when my phone doesn't have an internet signal.	✓			
I feel distant from people when I see them looking happy in their posts.		✓		
Often, I feel sad seeing people on social media looking happier than I am.		✓		
I felt upset (disappointed) when my friend didn't tag me in their post.		✓		
I feel sad knowing from a post that my friend attended an event and I wasn't invited.		✓		
I feel unhappy knowing on social media that my friend is at a place I also want to visit.		✓		
When I start looking at social media updates, I find it hard to stop.			✓	
My family and friends complain because I spend a lot of time on social media.			✓	
In social situations, I pay more attention to my cellphone than to my friends.			✓	
I arrived late to an appointment because I was too engrossed in social media.			✓	
When I'm on social media, I forget all my problems.			✓	
I want people to like and comment on my post				✓
I want to get more "likes" and comments on my posts.				✓
I only post photos and videos that I think my friends will like.				✓
I feel annoyed when my posts don't get many likes and comments.				✓
I am not interested in the reactions my friend gives to my posts.				✓

C. Preprocessing Data

The preprocessing stage is carried out to ensure that the data used in the classification process is clean, consistent, and ready to be processed by the algorithm. At this stage, data cleaning is carried out, which includes the removal of duplicate data, checking the consistency of answers, and handling missing values or empty data to maintain the quality of the dataset. After the data was cleaned, all features were normalized using the Min–Max method with the aim of standardizing the scale between variables so that they fall within the range [0–1]. Normalization is necessary because the KNN algorithm is very sensitive to differences in feature scales, so larger values can dominate the distance calculation process. The normalization formula is shown in Equation (1):

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

In that formula, x is the original value of a feature, x' is the normalized value, whereas $x_{\min}$ and $x_{\max}$ each shows the minimum and maximum values of that feature in the entire dataset. Through this transformation, the smallest value will be mapped close to 0 and the largest value close to 1, so that all features are on a uniform scale and ready to be used in the classification process.

D. Data Distribution

At this stage, the data that has undergone preprocessing is divided into two parts, namely 80% as training data and 20% as test data. Data splitting is performed using the Stratified Holdout Splitting technique, which is a separation method that maintains the proportion of each FOMO category (low and high) in the training and test data to ensure that its distribution remains consistent with the original data. This technique ensures that class representation remains balanced, so the KNN model's learning process is not biased and its performance evaluation becomes more accurate [21].

E. K-Nearest Neighbor (KNN) Experiment

At the KNN experiment stage, the model is built by testing various combinations of K values and distance methods to determine the best configuration for classifying FOMO levels. The K-Nearest Neighbor (KNN) algorithm is a classification method that determines the label of a new data point by comparing it to the training data based on a specific distance metric. The classification process is carried out by identifying a number of k nearest neighbors and assigning a class based on majority voting.

In this study, odd values of K ranging from 3 to 19 were selected to avoid tie voting in binary classification and to evaluate model performance across small to moderate neighborhood sizes. Smaller K values tend to be more sensitive to noise, while larger K values may reduce sensitivity to local data patterns. Therefore, multiple K values were tested to identify the most optimal parameter configuration for FOMO classification.

In this study, several distance metrics are used to measure the proximity between data points because the choice of distance method can affect classification performance. Previous research shows that KNN is generally evaluated using three main distance metrics, namely Euclidean, Manhattan, and Chebyshev, each of which has different calculation characteristics [22].

1) *Euclidean Distance:* is the most commonly used distance measurement method. Distance is calculated based on the square root of the sum of the squared differences between features. This method is suitable for continuous scale numerical data.

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \quad (2)$$

In Equation (2), the symbol *d(p,q)* represents the distance between two data points, namely the points *p* and *q*. Each point consists of several features, so p_i dan q_i showing feature value ke-*i* from each of those points. Euclidean Distance provides a measure of the "straight" distance between two points in multidimensional space, making this method widely used for continuously scaled numerical data.

2) *Manhattan Distance:* calculating distance based on the sum of absolute differences between features. This distance illustrates a path similar to movement in a grid-shaped city (forward-backward, left-right).
This method is more tolerant of outliers compared to Euclidean.

$$d(p,q) = \sum_{i=1}^{n} | p_i - q_i | \quad (3)$$

In Equation (3), the Manhattan Distance is calculated by summing all the absolute differences between the feature values at two data points, namely point p and point q.The symbol $d(p,q)$ represents the distance between the two points. Each feature is represented by p_i dan q_i, which are the values of the ke-*i* pada titik p dan q. The absolute difference $| p_i - q_i |$ indicates the extent to which the value of a feature differs between the two points, regardless of the direction of the difference (greater or smaller). The entire absolute difference from the first feature to the n-th feature is summed to produce the total Manhattan distance. This method is relevant for data with linear feature differences and is more tolerant of outliers compared to Euclidean Distance.

3) *Chebyshev Distance:* calculating distance based on the largest difference value between features. This method is suitable for conditions where the largest difference in one feature has a significant impact on classification.

$$d(p,q) = max_i \ | p_i - q_i | \quad (4)$$

In Equation (5), Chebyshev Distance defines the distance between two points p and q as the largest difference value among all the compared features. The symbol d(p,q) indicates the distance between the two points, while p_i and q_i represent the value of the i-th feature at each point. Thus, Chebyshev Distance considers that the distance between points is determined by the difference in the most dominant features. This method is suitable for use in conditions where one feature with the greatest difference is considered to have a significant impact on classification.

4) *Determination of K Value:* In the K-Nearest Neighbor (KNN) algorithm, the K value is an important parameter that determines the number of nearest neighbors used to classify a data point. Choosing the right K value is necessary for the model to optimally classify the data. In this study, the value of K was tested with several variations, namely K = 3, 5, 7, 9, 11, 13, 15, 17, and 19 [23].

F. Comparative Analysis

A comparative analysis was conducted to compare the performance of each KNN parameter configuration, both based on variations in K values and distance methods. The comparison was made using accuracy metrics, confusion matrix, sensitivity, and specificity to determine the parameter combination that yields the best performance [24].

1) *Confusion Matrix:* used to describe the model's ability to classify data into low and high FOMO classes. This matrix consists of four components, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), which indicate the number of correct and incorrect predictions. Through the confusion matrix, the classification error patterns made by the model can be analyzed in more detail [25].

TABLE 2
STRUCTURE CONFUSION MATRIX

	Positive Prediction	Negative Prediction
Actual Positive	TP	FN
Actual Negative	FP	TN

2) *Accuracy:* used to measure the overall level of the model in the classifying test data. The accuracy value is obtained from the comparison of the number of correct predictions against all the tested data. This metric provides an overview of the model's performance, but it is not sufficient to explain the model's performance on each class. overall [26]. The accuracy formula is shown in Equation (5):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

3) *Sensitivity:* used to measure the model's ability to detect data that truly belongs to the positive class. The sensitivity value shows the proportion of positive data that is correctly predicted compared to all the data that is actually positive. This metric is important to determine the extent to which the model can recognize all positive cases and minimize false negative predictions. In this study, sensitivity plays a role in measuring the model's ability to detect respondents with a high FOMO category. The sensitivity formula is shown in Equation (6):

$$Sensitivity = \frac{TP}{TP + FN} \quad (6)$$

4) *Specificity:* used to measure the model's ability to identify data that truly belongs to the negative class. The specificity value shows the proportion of negative data that is correctly predicted compared to all the data that is actually negative. This metric is important to determine the extent to which the model can avoid false positive predictions. In this study, specificity is used to measure the model's ability to identify respondents with low FOMO category [27]. The specificity formula is shown in Equation (7):

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

III. RESULTS AND DISCUSSION

In this section, the results of descriptive analysis and preprocessing of respondent data are presented, including demographic characteristics, social media usage patterns, and FOMO questionnaire scores. Statistics such as minimum, maximum, average, and standard deviation values are used to describe the distribution of each feature before the modeling process is carried out.

The dataset consisted of 68 respondents categorized as low FOMO and 68 respondents categorized as high FOMO, resulting in a balanced class distribution with a proportion of 50% for each class. This balanced distribution helps reduce classification bias and supports a more reliable evaluation of the machine learning model.

Table 3 shows the research dataset consists of 136 respondents with various features, including demographic characteristics, social media usage patterns, and FOMO questionnaire scores. All features have undergone an encoding process to become numerical. Although the number of samples in each variable is the same, the distribution of values in each feature shows varying scales and levels of dispersion. Some features have a wider range of values compared to other features, which could potentially affect the distance calculation process in the classification algorithm.

TABLE 3
DISTRIBUTION OF THE FOMO DATASET

	Count	**Mean**	**Min**	**25 %**	**50 %**	**75 %**	**Max**	**Std**
Age	136	21.83	15	20	21	23	38	3.53
Gender	136	0.419	0	0	0	1	1	0.49
Q1	136	3.448	1	3	4	4	5	1.14
Q2	136	3.507	1	3	4	4	5	1.08
Q3	136	3.139	1	2	3	4	5	1.05
Q4	136	3.25	1	2	3	4	5	1.16
Q5	136	3.485	1	3	4	4	5	1.16
Q6	136	2.139	1	1	2	3	5	1.08
Q7	136	2.117	1	1	2	3	5	1.12
Q8	136	2.051	1	1	2	3	5	1.09
Q9	136	2.110	1	1	2	3	5	1.07
Q10	136	2.066	1	1	2	3	5	1.09
Q11	136	2.919	1	2	3	4	5	1.12

Q12	136	2.632	1	2	3	3	5	1.12
Q13	136	2.286	1	2	2	3	5	1.00
Q14	136	2.029	1	1	2	3	5	1.08
Q15	136	2.595	1	2	3	3	5	1.10
Q16	136	2.654	1	2	3	3	5	1.26
Q17	136	2.845	1	2	3	4	5	1.32
Q18	136	2.441	1	1	2	3	5	1.20
Q19	136	2.191	1	1	2	3	5	1.10
Q20	136	2.705	1	2	3	3	5	1.04
Social Media	136	2.963	1	2	3	4	6	1.46
Social Media Duration	136	2.911	1	2	3	4	4	0.93
Etiqueta Fomo?	136	0.625	0	0	1	1	1	0.48

The distribution of FOMO classes in the dataset is shown in Figure 2. The graph illustrates the comparison of the number of respondents in each FOMO category, which is used to assess data balance before the modeling process is carried out. This information is important because class distribution can affect the performance of the classification model.

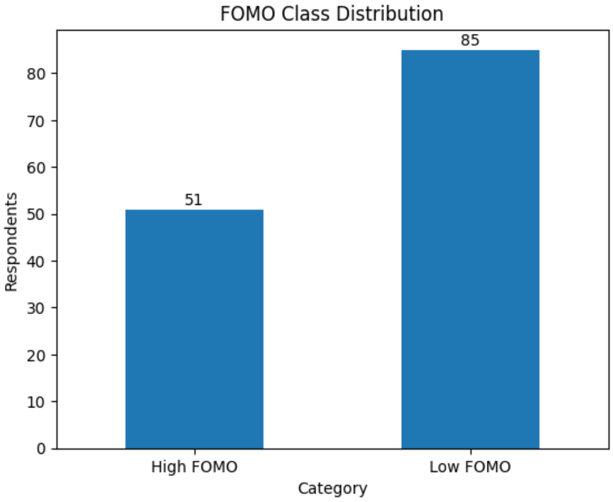

Figure 2. Fomo Class Distribution Graph

Considering the characteristics of feature distribution and class distribution, a normalization process is carried out to standardize the scale of all features into the range of 0–1. This stage aims to ensure that each variable has a balanced contribution and to prevent the dominance of certain features, given that the K-Nearest Neighbor (KNN) algorithm uses a distance metric that is sensitive to differences in data scale. After the normalization process, the dataset is then divided into training and testing data using the stratified holdout method with an 80:20 ratio. This division aims to train the model on the training data and objectively evaluate the model's performance on the test data that has never been seen before. The use of stratification ensures that the distribution of the FOMO class remains balanced, making the evaluation results more representative of the model's generalization capability. The next stage involves conducting classification experiments using the KNN algorithm with variations in K values and distance metrics. In this study, the K values used are K = 3, 5, 7, 9, 11, 13, 15, 17, and 19, while the distance metrics tested include Euclidean, Manhattan, and Chebyshev. These parameter variations are applied to analyze the influence of the number of nearest neighbors and the distance measurement method on classification performance. The purpose of testing various parameter combinations is to obtain the most optimal KNN configuration in classifying the level of FOMO in adolescents based on accuracy values and other evaluation metrics.

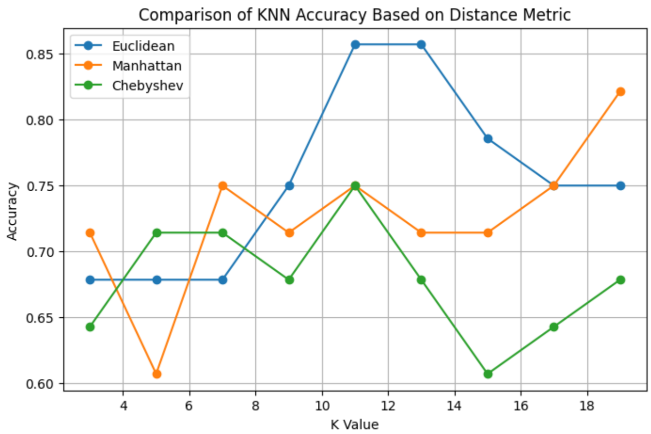

Figure 3. Accuracy Comparison Graph

Based on the accuracy comparison graph in Figure 3, it is evident that the performance of the K-Nearest Neighbor (KNN) algorithm varies with each combination of K value and distance metric. Changes in the value of K have been proven to affect the accuracy level of the model for each distance method. In the Euclidean method, accuracy gradually increases until it reaches its highest value at K=11, then decreases as the value of K becomes larger. This indicates that there is a balance point between the model's ability to capture data patterns and the stability of classification results. Meanwhile, the Manhattan and Chebyshev methods show an accuracy pattern that tends to fluctuate but cannot surpass the performance of the Euclidean method. Overall, these results affirm that the selection of the K value and the distance method are important factors that determine the quality of the classification model's performance.

Based on the accuracy calculation results presented in Table 4, each combination of K values and distance metrics yields different performance. The Euclidean method shows the highest accuracy of 85.71% at K=11 and K=13. However, considering the stability of performance and the principle of selecting a K value that is not too large, K=11 is chosen as the optimal configuration. In the Manhattan method, the highest accuracy was achieved at K=19 with a value of 82.14%, while the Chebyshev method produced a maximum accuracy of 75.00% at several variations of K values. Overall, the results in the table show that the Euclidean method provides superior classification performance compared to Manhattan and Chebyshev on this research dataset.

TABLE 4
ACCURACY OF ALL K + METRIX COMBINATIONS

K Value	Euclidean	Manhattan	Chebyshev
3	0.6786	0.7143	0.6429
5	0.6786	0.6071	0.7143
7	0.6786	0.7500	0.7143
9	0.7500	0.7143	0.6786
11	**0.8571**	0.7500	0.7500
13	0.8571	0.7143	0.6786
15	0.7857	0.7143	0.6071
17	0.7500	0.7500	0.6429
19	0.7500	0.8214	0.6786

These findings indicate that the characteristics of the normalized data are more suitable for analysis using Euclidean distance, so the combination of Euclidean with K=11 is established as the best model in this study.

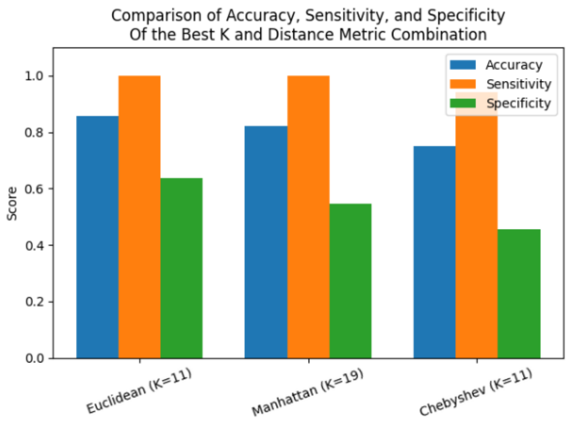

Figure 4. Comparison Bar Chart

Figure 4 shows the comparison of KNN model accuracy based on variations in K values and three distance methods, namely Euclidean, Manhattan, and Chebyshev, making it easier to see the most optimal method for data classification. Based on the graph, the Euclidean method yields the highest accuracy at almost all K values, peaking at K = 11, indicating that this method is most suitable for normalized data. The Manhattan method has fairly stable performance, but generally still falls below Euclidean and achieves the best results at K = 19. Meanwhile, the Chebyshev method shows lower accuracy and tends to be unstable, making it less suitable for the dataset used. Overall, this graph confirms that the best combination is Euclidean with K = 11, and it shows that the choice of distance method significantly affects the classification results.

The evaluation of the best KNN model's performance was then conducted using the confusion matrix displayed in Figure 5. Based on these results, the True Positive (TP) value was 17, the True Negative (TN) value was 7, the False Positive (FP) value was 4, and the False Negative (FN) value was 0. These results indicate that the majority of the test data were correctly classified by the model.

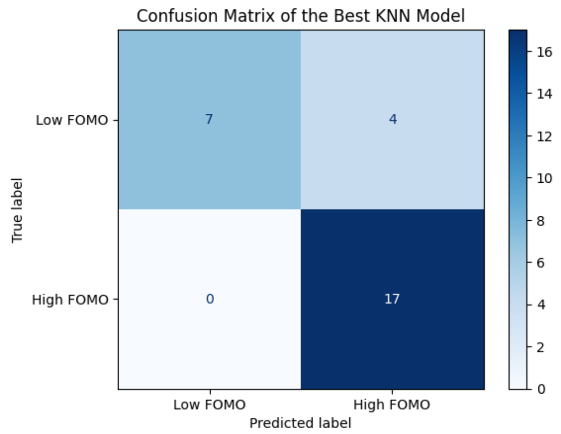

Figure 5. Euclidean Confusion Matrix

Quantitatively, the model achieved an accuracy of 85.71%, indicating a good overall classification ability. A sensitivity value of 100% indicates that all respondents with a high FOMO category were detected without any classification errors in that class. This is important because it shows that the model is very effective in recognizing individuals with a high level of FOMO. Meanwhile, a specificity value of 63.64% indicates that most respondents with a low FOMO category were also correctly identified, although there were still some prediction errors in this class.

Based on the confusion matrix results, the model produced several false positive predictions, where respondents categorized as low FOMO were classified as high FOMO. This may occur because some respondents exhibited questionnaire response patterns similar to individuals with high FOMO characteristics. However, the absence of false negative predictions indicates that the model was highly effective in detecting respondents with high FOMO levels.

Although the model achieved 100% sensitivity, this result should be interpreted cautiously due to the relatively small dataset size, which may increase the risk of overfitting. Therefore, further validation using larger datasets and cross-validation techniques is recommended to ensure the generalizability and robustness of the model.

Overall, the results of this evaluation indicate that the combination of Euclidean parameters with K=11 is capable of providing good and balanced classification performance, making it suitable for use as a final model in diagnosing the level of FOMO in adolescents.

The Euclidean distance metric achieved better classification performance because the dataset had been normalized using Min-Max normalization, making the numerical feature space more suitable for straight-line distance calculations. Euclidean distance is effective in capturing overall similarity patterns among respondents' questionnaire scores, while Manhattan and Chebyshev distances may be less sensitive to subtle variations in the data distribution.

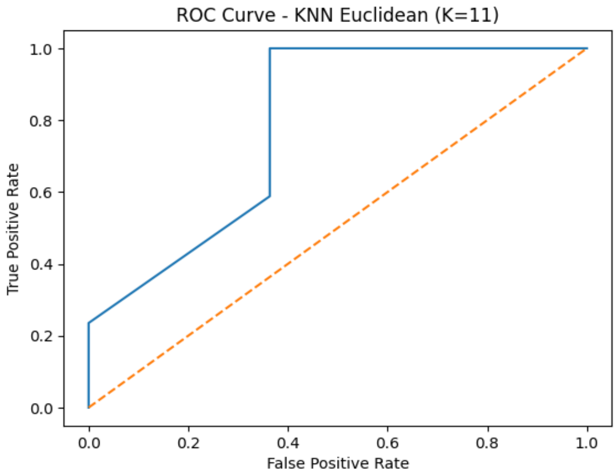

Figure 6. ROC Curve

The evaluation of the best KNN model's performance is then conducted using the Receiver Operating Characteristic (ROC) Curve shown in Figure 6. The ROC Curve is used to measure the model's ability to distinguish between high FOMO and low FOMO classes through the relationship between True Positive Rate (sensitivity) and False Positive Rate at various classification threshold values.

Based on the calculation results, an Area Under Curve (AUC) value of 0,786 was obtained. The value indicates that the model has good discriminative ability in classifying the two FOMO categories. Interpretatively, there is a 78.6% probability that the model will assign a higher probability score to respondents who truly belong to the high FOMO category compared to randomly selected respondents from the low FOMO category. This result indicates that the KNN model with a combination of Euclidean distance and K=11 has a fairly stable performance in distinguishing between the two classes.

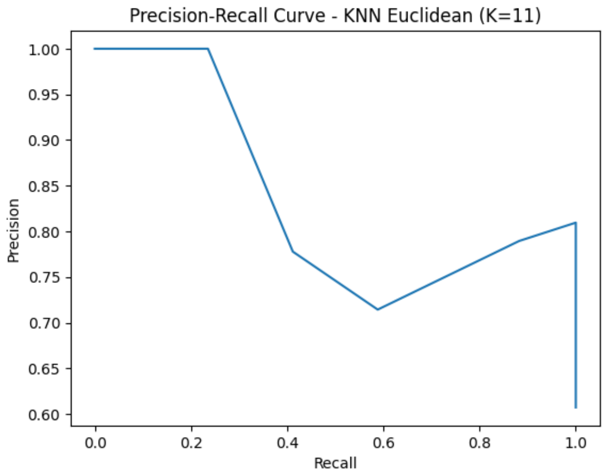

Figure 7. Precision-Recall Curve

The next performance evaluation of the model was conducted using the Precision–Recall Curve shown in Figure 7. The Precision–Recall Curve is used to analyze the balance between precision and recall in detecting the positive class, which is respondents with a high FOMO category. This curve is highly relevant when the focus of the research is on the model's ability to accurately identify positive cases.

Based on the calculation results, an Area Under Curve (AUC) value of 82.6% was obtained. This value indicates that the model has a very good ability to maintain a balance between precision and recall across various classification thresholds. Interpreting this result indicates that the model is not only capable of detecting all high FOMO cases (high recall) but also consistently maintains a high level of positive prediction accuracy (precision).

The higher Precision–Recall AUC value compared to the ROC AUC also indicates that the model has a stronger performance in specifically identifying the high FOMO class. Thus, these results further reinforce that the combination of Euclidean distance and K=11 is the optimal configuration in this study. Overall, the experimental results indicate that the K-Nearest Neighbor (KNN) algorithm with the combination of Euclidean distance and K=11 is capable of providing optimal classification performance in diagnosing the level of FOMO in adolescents. Evaluation using a confusion matrix, ROC Curve, and Precision–Recall Curve shows that the model has good and stable discriminative ability in detecting high FOMO categories. Therefore, the KNN-based approach in this study can be an alternative method that is simple, systematic, and efficient to support the initial screening process of FOMO levels in adolescents based on data mining.

IV. CONCLUSION

This study aims to classify the level of Fear of Missing Out (FOMO) among adolescents using the K-Nearest Neighbor (KNN) algorithm based on questionnaire data thru preprocessing stages (encoding, Min-Max normalization, and stratified holdout 80:20) as well as experiments with variations in K values (3–19) and distance metrics (Euclidean, Manhattan, Chebyshev). The evaluation results show that the combination of Euclidean with K = 11 provides the best performance with an accuracy of 85.71%, ROC AUC of 76.6%, Precision–Recall AUC of 82.6%, and sensitivity of 100%, making it optimal as it balances local pattern capture and classification stability and is effective as an initial screening tool for FOMO levels among adolescents.

Ethical considerations are important in psychological classification studies involving adolescents. Data privacy, informed consent, and the potential risk of misclassification should be carefully considered before implementing machine learning models in real-world screening applications.

The proposed model is intended only as an initial screening tool and not as a replacement for professional psychological assessment. Further validation involving psychologists or mental health professionals is recommended to ensure the reliability and appropriateness of the system in practical applications.

For future research, it is recommended to use a larger sample size and apply cross-validation techniques to improve model generalization. Additionally, the research can be developed by comparing the KNN algorithm with other classification methods, such as Support Vector Machine or Random Forest, resulting in a more comprehensive and robust predictive model.

REFERENCES

[1] M. Benvenuti, M. Wright, J. Naslund, and A. C. Miers, "How technology use is changing adolescents ' behaviors and their social , physical , and cognitive development," pp. 16466–16469, 2023, doi: doi.org/10.1007/s12144-023-04254-4.

[2] R. Kurniawan and R. H. Utami, "Validation of Online Fear of Missing Out (ON-FoMO) Scale in Indonesian Version," *J. Neo Konseling*, vol. 4, no. 3, p. 1, 2022, doi: 10.24036/00651kons2022.

[3] A. Alabri, "Fear of Missing Out (FOMO): The Effects of the Need to Belong , Perceived Centrality , and Fear of Social Exclusion," vol. 2022, 2022, doi: 10.1155/2022/4824256.

[4] E. S. Handayani and R. Haryadi, "Korelasi Kepribadian Fear Missing Out Terhadap Adiksi Impulsive Online Buying Di Masa Pandemi Pada Anak dan Remaja," *Bull. Couns. Psychother.*, vol. 4, no. 3, pp. 761–768, 2023, doi: 10.51214/bocp.v4i3.439.

[5] P. Dissonance, "SHOPPING BEHAVIOR : The Impact On (With a Gender Perspective)," vol. 23, no. 1, pp. 134–157, 2025, doi: doi.org/10.21776/ub.jam.2025.023.1.08.

[6] D. N. Nguyen and D. Van Nguyen, "FOMO dan Perilaku Pembelian Implusif Anak Muda," *Bisnis dan Manaj.*, vol. 10, no. 3, pp. 41–47, 2025, doi: doi.org/10.24018/ejbmr.2025.10.3.2638.

[7] N. W. Koesmawardhani, "Riset: FOMO Picu Gangguan Mental & Risiko Bunuh Diri pada Pelajar SMP," detikedu. [Online]. Available: https://www.detik.com/edu/sekolah/d-8223797/riset-fomo-picu-gangguan-mental-risiko-bunuh-diri-pada-pelajar-smp

[8] W. Mufidah, E. Nursanti, U. Darul, and U. Jombang, "Fear Of Missing Out(FoMO) Pada Remaja Pengguna Instagram," *J. Clin. Ind. Soc. Educ. Psychol.*, vol. 000, pp. 46–57, 2023, doi: https://doi.org/ 10.32492/idea.v7i1.7105.

[9] R. Servidio, B. Koronczai, M. D. Griffiths, and Z. Demetrovics, "Problematic Smartphone Use and Problematic Social Media Use : The Predictive Role of Self-Construal and the Mediating Effect of Fear Missing Out," vol. 10, no. February, pp. 1–10, 2022, doi: 10.3389/fpubh.2022.814468.

[10] A. Pratiwi, A. Fazriani, D. STIKes YATSI Tangerang, and Stik. Tangerang, "Hubungan Antara Fear of Missing Out (Fomo) Dengan Kecanduan Media Sosial Pada Remaja Pengguna Media Sosial," *Artik. Penelit. J. Kesehat.*, vol. 9, no. 1, 2020, doi: 10.37048/kesehatan.v9i1.123.

[11] M. Gupta and A. Sharma, "Fear of missing out: A brief overview of origin, theoretical underpinnings and relationship with mental health," *World J. Clin. Cases*, vol. 9, no. 19, pp. 4881–4889, 2021, doi: 10.12998/wjcc.v9.i19.4881.

[12] S. Abimayu, N. Bahtiar, and E. A. Sarwoko, "Implementasi Metode Support Vector Machine (SVM) dan t-Distributed Stochastic Neighbor Embedding (t-SNE) untuk Klasifikasi Depresi," *J. Masy. Inform.*, vol. 14, no. 2, pp. 146–158, 2023, doi: https://doi.org/10.14710/jmasif.14.2.59513.

[13] S. Munawaroh, U. A. Rosyidah, and R. Yanuarti, "Klasifikasi Tingkat Kecemasan Atlet Sebelum Bertanding Menggunakan Algoritma K–Nearest Neighbor (KNN) Berbasis Website," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 5, no. 2, pp. 87–94, 2024, doi: 10.37148/bios.v5i2.120.

[14] W. T. K. Widyastuti Andriyani, Mochammad Anshori, Dwi Normawati, Risqy Siwi Pradini, Mohamad Zaenudin, Muhammad Iqbal Harisuddin, M. Syauqi Haris, Astuty, Anna Angela Sitinjak, *Matematika Pada Kecerdasan Buatan*. 2024. [Online]. Available: https://books.google.co.id/books?hl=en&lr=&id=cEYhEQAAQBAJ&oi=fnd&pg=PP1&ots=q4w9L__NSd&sig=p6pu2rV06VpfrZfkh6oJometj6k&redir_esc=y#v=onepage&q&f=false

[15] U. Madububambachu, A. Ukpebor, and U. Ihezue, "Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review," *Clin. Pract. Epidemiol. Ment. Heal.*, vol. 20, no. 1, pp. 1–16, 2024, doi: 10.2174/0117450179315688240607052117.

[16] A. Pamungkas, R. Rizal Isnanto, D. Mutiara, and K. Nugraheni, "Implementation of K-Nearest Neighbor in Case-Based Reasoning for Mental Health Diagnosis Systems," *Sci. J. Informatics*, vol. 11, no. 4, pp. 1109–1120, 2024, doi: 10.15294/sji.v11i4.19912.

[17] B. Sun and H. Chen, "A Survey of k Nearest Neighbor Algorithms for Solving the Class Imbalanced Problem," *Wirel. Commun. Mob. Comput.*, vol. 2021, 2021, doi: 10.1155/2021/5520990.

[18] É. O. Rodrigues, "Combining Minkowski and Chebyshev : New distance proposal and survey of distance metrics using k-nearest neighbours classifier," *Pattern Recognit. Lett.*, vol. 110, pp. 66–71, 2018, doi: 10.1016/j.patrec.2018.03.021.

[19] N. A. Saputra, K. Aeni, and N. M. Saraswati, "Indonesian Hate Speech Text Classification Using Improved K-Nearest Neighbor with TF-IDF- ICSρF," vol. 11, no. 1, pp. 21–30, 2024, doi: 10.15294/sji.v11i1.48085.

[20] N. Rikatsih, M. Anshori, and R. S. Pradini, "K-Nearest Neighbor Method for Early Detection of Diabetes Patients Based on Symptoms and Clinical Data," *J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 9, no. 2, pp. 187–192, 2024, doi: https://doi.org/10.25139/inform.v9i2.8582.

[21] D. Z. Abidin, M. Rosario, and A. Sadikin, "Improving Term Deposit Customer Prediction Using Support Vector Machine with SMOTE and Hyperparameter Tuning in Bank Marketing Campaigns," vol. 6, no. 3, pp. 1267–1278, 2025, doi: https://doi.org/10.52436/1.jutif.2025.6.3.4585 Improving.

[22] M. C. Saletia, M. Anshori, and M. S. Haris, "Comparing Different KNN Parameters Based on Woman Risk Factors to Predict the Cervical Cancer," vol. 9, no. 5, 2025, doi: doi.org/10.30871/jaic.v9i5.10746.

[23] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Sci. Rep.*, vol. 12, no. 1, pp. 1–11, 2022, doi: 10.1038/s41598-022-10358-x.

[24] N. Aulia, S. N. Sari, and N. Wakhidah, "Jurnal Informatika : Jurnal pengembangan IT Analisis Sentimen Aplikasi Get Contact di App Store Menggunakan Metode SVM (Support Vector Machine)," vol. 10, no. 1, pp. 139–148, 2025, doi: 10.30591/jpit.v10i1.8057.

[25] M. M. Mijwil and M. M. Mijwil, "A Comparative Analysis of Machine Learning Algorithms for Classification of Diabetes Utilizing Confusion Matrix Analysis A Comparative Analysis of Machine Learning Algorithms for Classification of Diabetes Utilizing Confusion Matrix Analysis," vol. 21, no. 5, 2024, doi: doi.org/10.21123/bsj.2023.9010.

[26] C. Chou and D. Hsu, "Predicting the Onset of Diabetes with Machine Learning Methods," 2023, doi: doi.org/10.3390/jpm13030406.

[27] F. S. Gomiasti, E. Kartikadarma, J. Gondohanindijo, and D. R. I. Moses, "Enhancing Lung Cancer Classification Effectiveness Through Hyperparameter-Tuned Support Vector Machine," *J. Comput. Theor. Appl.*, 2024, doi: https://doi.org/10.62411/jcta.10106.