Machine Learning-Based Counseling to Predict Psychological Readiness for Aspiring Entrepreneurs

Nesi Syafitri*[1], Syarifah Farradinna[2], Yudhi Arta[3], Icha Herawati[4], Wella Jayanti[5]
[1,3,] Teknik Informatika, Fakultas Teknik, Universitas Islam Riau
[2,4,5] Psikologi, Fakultas Psikologi, Universitas Islam Riau
e-mail: ***[1]nesisyafitri@eng.uir.ac.id**, [2] syarifah.farrdinna@psy.uir.ac.id , [3]yudhiarta@eng.uir.ac.id, [4]ichaherawati@psy.uir.ac.id, [5]Wellajayanti17@gmail.com

Abstract

Machine learning (ML) has shown significant potential in the field of psychological counseling, including its application in counseling for psychological readiness for entrepreneurship. This study introduces an intelligent application developed using machine learning to assess psychological readiness for entrepreneurship, based on the Entrepreneurship Psychological Readiness (EPR) instrument. To identify the most suitable machine learning model, two models—Naïve Bayes (NB) and k-Nearest Neighbor (k-NN)—were compared using a dataset of 1,095 training samples. The application provides predictions in four categories of readiness: not ready, requires training, recommended, and ready for entrepreneurship. The EPR instrument consists of 33 items designed to measure eight parameters used as inputs for the prediction process. The data was randomized, and the experiment was repeated five times to evaluate the consistency of the model performance. Eighty percent of the data was allocated for training, while the remaining 20% was used for testing. The results from the five trials showed that the Naïve Bayes model consistently outperformed k-NN, achieving an accuracy rate of 89.58% during testing. Based on these findings, the Naïve Bayes model is recommended for use in psychological counseling to assess and predict an individual's readiness for entrepreneurship effectively.

Keywords— Psychological Readiness, Entrepreneurship, Machine Learning, Naïve Bayes, K-NN

1. INTRODUCTION

Machine learning is part of Artificial Intelligence (AI). Machine learning (ML) refers to the ability of computers to learn without explicit programming. With the continuous increase in computational power and the availability of large datasets, ML approaches aim to surpass human performance in various tasks by extracting insights from vast amounts of data[1].

AI has found widespread application across numerous fields, including medical diagnosis, mathematical and physical modeling, image classification in biology and chemistry, and many others. In recent years, psychology has increasingly explored the use of AI to predict and classify various phenomena. For example, AI has been employed to assess pain levels from brain scans[2], analyze personality traits using machine learning techniques[3], predict harmful social media use[4],[5], and support the diagnosis and prognosis of mental illnesses and disorders[6]. It is also used to detect depression levels, evaluate the risk of suicidal and self-harming behaviors[6], and aid in suicide prevention efforts [7]. Additionally, facial expression recognition has been leveraged for psychological identification[8]. Researchers continue to investigate ways to enhance AI models, particularly within the realm of psychology, to improve their accuracy and reliability[9].

Entrepreneurship has become a cornerstone of modern economic growth, driving innovation, job creation, and technological advancements. However, the journey of entrepreneurship is fraught with challenges, requiring not only financial resources and strategic planning but also psychological readiness. The latter is a crucial determinant of entrepreneurial success, encompassing resilience, adaptability, risk tolerance, and motivation.

Entrepreneurial readiness is described as a set of individual characteristics that separate individuals by demonstrating curiosity and competence in observing and analyzing their creative and productive potential and running a business by directing all abilities for self-achievement[10]. Entrepreneurial readiness is a system that identifies individual interests, assesses a person's preparedness to start a business, and increases socioeconomic welfare. Individual traits influence entrepreneurial intentions[11], In addition to the contextual elements that affect individuals' intentions and behaviors as entrepreneurs [12].

Psychological readiness is defined as a set of descriptions of three elements: motivation, which includes entrepreneurial drive and the content of ideas to see opportunities; characteristics, which include personality traits that reflect psychological characteristics; and competence, which is a description that demonstrates creativity, productivity, and efficiency in business operations[13].

Traditional assessment methods, such as questionnaires and self-report interviews, are often subjective and susceptible to bias. Developing an objective and measurable approach to accurately evaluate these diverse psychological traits remains a significant challenge. Experts are continuously working on developing various tools to measure psychological readiness and are beginning to innovate using advanced technologies. One notable advancement is the application of artificial intelligence (AI), including machine learning models, to assess entrepreneurial readiness among college graduates[5]. AI technology not only aids in predicting social and psychological readiness to start a business, but it also delivers a level of assurance in the forecasts made.

The primary goal of this research is to develop an intelligent counseling tool, referred to as Neuris Technology, by incorporating a machine learning model to assist experts in evaluating an individual's psychological readiness for entrepreneurship. Neuris Technology leverages metrics from the Entrepreneurial Psychological Readiness (EPR) assessment, which comprises 33 categorically scored questions. The EPR instrument refers to the Codarus instrument, which was designed and tested using Confirmatory Factor Analysis (CFA), yielding 33 question items. The Entrepreneurship Psychological Readiness (EPR) can be used to assess an individual's readiness for entrepreneurship by examining knowledge, personal ability, action, willingness to learn, relationships with others, self-development, enthusiasm, and external assistance[14]. The EPR instrument is specifically designed to determine whether individuals intending to start a business are mentally prepared for the challenges of entrepreneurship.

2. RESEARCH METHODS

The research methodology employed in this study is depicted in the diagram below, outlining the sequential stages beginning with data collection. Psychological readiness for entrepreneurship data was gathered using the EPR questionnaire. After data collection, the next stage involved data preprocessing, ensuring the data is prepared for machine learning. Following preprocessing, machine learning models were developed by comparing the performance of two popular algorithms: K-Nearest Neighbors (KNN) and Naïve Bayes. The model demonstrating superior performance was selected and subsequently integrated into the Neuris Tech application. The final stage involved testing the application's performance to evaluate its effectiveness. The research stages can be described in an organized fashion, as illustrated in Figure 1.

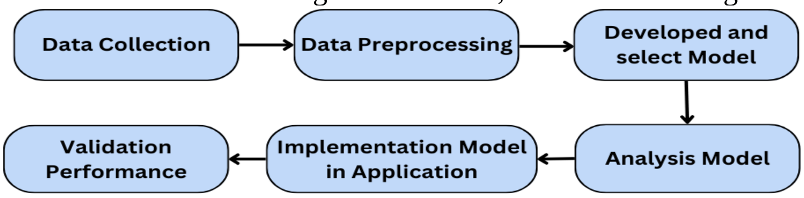


Figure 1. Research stages to develop EPR intelligent application

2. 1. Data Collection

The data used in this study consists of primary data collected through questionnaires distributed to students from various higher education institutions in Riau, using the EPR instrument. This instrument, referred to as the EPR-33, comprises eight indicators and 33 question items. The EPR-33 is designed to assess an individual's psychological and social readiness to determine their preparedness for entrepreneurship. It evaluates eight parameters, with responses to the 33 question items recorded on a five-point Likert scale: Strongly Disagree (1), Disagree (2), Undecided (3), Agree (4), and Strongly Agree (5). A detailed description of the eight parameters and their corresponding question items is provided in Table 1.

Table 1. Parameters of The EPR-33 Instrument

Parameters	**Number of question items**
Personal Knowledge	4
Personal Adversity	7
Committed Certain Action	4
Willingness to learn	2
Personal Relationship to Others	4
Personal growth	6
Passion Achieved	3
Related Person Support	3
Total	33

Table 1 presents the parameters measured by the Entrepreneurial Psychological Readiness (EPR-33) instrument. The EPR-33 assessment consists of 33 question items distributed across eight key parameters that evaluate different aspects of psychological readiness for entrepreneurship. These parameters include:

1) Personal Knowledge (4 items): Assesses the individual's understanding and preparedness for entrepreneurship
2) Personal Adversity (7 items): Evaluates resilience and the ability to cope with challenges
3) Committed Certain Action (4 items): Measures actions taken towards entrepreneurial goals.
4) Willingness to Learn (2 items): Gauges openness to acquiring new knowledge and skills.
5) Personal Relationship to Others (4 items): Assesses interpersonal relationships and their impact on entrepreneurial readiness.
6) Personal Growth (6 items): Evaluates self-development and progression in a personal capacity.
7) Passion Achieved (3 items): Measures the level of passion achieved for entrepreneurial endeavors.
8) Related Person Support (3 items): Assesses support from others, such as mentors or close networks.

During the data collection stage, questionnaires were delivered to students from several universities in the Riau region. Characteristics of respondents aged 17 to 25 years include both males and women. The total number of datasets obtained was 1095. 80% of the dataset, or 876 datasets, is utilized for the machine learning model construction process, while 20%, or 219 data, is used to test the performance of the resulting model.

2. 2. Data Preprocessing

Managing missing data is an essential step in the preprocessing stage when working with survey instruments like the EPR questionnaire. With 33 question items spread across eight parameters, any missing responses can compromise the validity and reliability of the results if not appropriately managed.

The first step in addressing missing data is to calculate the percentage of completed items for each respondent and identify cases where the incompleteness rate exceeds the acceptable threshold of 20%. Respondents with a high rate of incomplete answers are excluded to ensure the reliability of the data.

In the second step, the validated data will undergo a scoring process. The first step in the scoring process involves assigning weights to responses for each question item. The total score is divided into five levels: very high, high, medium, low, and very low, based on the weighted responses. Once the scores for individual parameters are calculated, the next step involves summing these scores to derive an overall score. This overall score represents a composite measure of the respondent's psychological readiness for entrepreneurship, taking into account all eight parameters. By consolidating the individual scores, the overall score provides a holistic view of the individual's preparedness. After obtaining the overall score, the final step involves classifying the respondent into one of four psychological readiness categories. These categories are determined using a set of classification norms[15] that relate the overall score to the population's mean (μ) and standard deviation (σ), as shown in Table 2.

Table 2. Categorization norms for classifying EPR scores

Categorization	Categorization Norms
Not Ready	$X \leq \mu - 1\sigma$
Requires training	$\mu - 1\sigma < X \leq \mu$
Recommended	$\mu < X \leq \mu + 1\sigma$
Ready for Enterpreneurship	$X > \mu + 1\sigma$

Where X is the parameter score, μ is the average value, and σ is the deviation distribution value.

2. 3. Development and Selection of Machine Learning Model

Machine learning models undergo two main phases: the training phase and the testing phase. The dataset is divided into two subsets: the training dataset and the testing dataset. Approximately 80% of the dataset is allocated to the training phase, while the remaining 20% is reserved for the testing phase.

The training phase involves developing the machine learning model by learning patterns and relationships within the training data. In contrast, the testing phase evaluates the performance of the trained model by using the testing dataset to assess its accuracy and effectiveness.

The machine learning model chosen for the application is based on a performance evaluation of two prominent categorization models: Naïve Bayes and k-nearest neighbor (k-NN).

2.3.1. Naive Bayes

Classification allows learning patterns from historical data, such as traits, variables, and features, on various characteristics of pre-labeled items to place new unidentified objects into their respective classes or groups.

The Naive Bayes classification algorithm is a standard Bayes classification technique with a straightforward algorithm structure and good processing efficiency. The Naïve Bayes classifier has the advantage of requiring minimal training data to estimate the essential parameters (variable mean and variance) [16].

Naïve Bayes is a supervised learning algorithm based on Bayes' theorem. It is commonly used for classification tasks through a probabilistic approach. It was chosen for this study because it requires processing a relatively smaller dataset. The Naïve Bayes algorithm is represented by the following equation (1).

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \tag{1}$$

Where X is data with an unknown class and H is a hypothesis that the data belongs to a specific class. P(H|X) is the posterior probability of hypothesis H conditional on X, P(X|H) is the probability of X given the condition of hypothesis H, P (H) is the prior probability of hypothesis H, and P(X) is the prior probability of variable X.

The Naïve Bayes classifier follows four steps:

a) Calculate the prior probability value P(Hi) for each hypothesis Hi
b) Calculate the conditional probability value P (X | Hi) for each hypothesis Hi on each attribute X
c) Calculate the posterior probability value (likelihood) with the equation (2):

$$likelihood(Hi) = P(H_i) \prod_{f=1}^{n} P(X_f | H_i) \tag{2}$$

d) Classify a new piece of data based on the highest posterior probability value of all existing hypotheses.

2.3.2. K-Nearest Neighbour

The K-Nearest Neighbor (KNN) algorithm is a supervised machine learning method used for classification tasks[17], [18]. It works by calculating the distance between a query sample and predefined classes[19]. KNN employs a case-based reasoning approach, relying on stored examples to generate predictions[20]. For a given query, it identifies the nearest data points and applies a majority voting rule to determine the most frequent class among the neighbors[17].

The following are the steps of the k-nearest neighbor algorithm used to determine the class of new data:

1) Determine parameter k.
2) Calculating the distance between evaluation data and training data.
3) Sort the distances formed.
4) Categorize the data class of k-nearest neighbors.
5) Identify the majority of data classes using k-nearest neighbors. Furthermore, the majority data class is used to represent the newly evaluated data.

The most commonly used distance calculation in the K-NN method is the Euclidean distance, represented by the following equation (3):

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{3}$$

Where *xi* is sample data, *yi* is data testing, *i* is variable data, *d(x,y)* is dissimilarity, and *n* is dimensions of data.

The process of developing and selecting machine learning models consists of two phases: the training phase and the testing phase. In the training phase, training data is used to build the machine learning models, while in the testing phase, testing data is used to evaluate and validate the performance of the developed models. Figure 2 describes the process of constructing and selecting machine learning models.

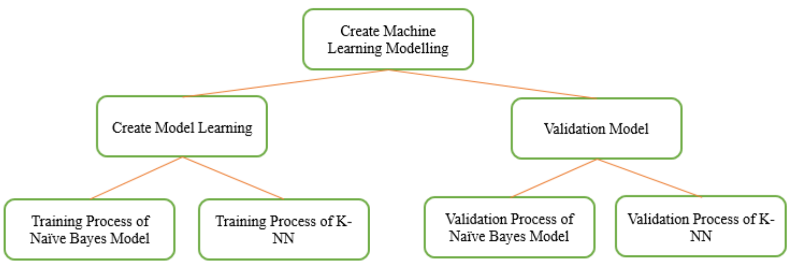


Figure 2. Formation and selection process of machine learning models

The resulting machine-learning model is then tested and validated using a confusion matrix. A confusion matrix is a table used to evaluate the performance of a classification model. It provides a summary of the model's predictions by comparing the predicted values with the actual values. It includes four key metrics: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The results are shown in Table 3.

Table 3. Confusion Matrix Table

Actual Class	**Predicted Class**	
	YES	**NO**
YES	True Positive (TP)	False Negative (FN)
NO	False Positive (FP)	True Negative (TN)

Using these values, various performance metrics can be calculated, including accuracy, precision, recall/sensitivity, and specificity as shown in equations (4), (5), (6), and (7).

$$Accuracy = \frac{(TP+TN)}{(TP+FN+FP+TN)} \quad (4)$$

Accuracy is a common metric used to evaluate the performance of a classification model. It measures the proportion of correctly predicted instances (both positive and negative) out of all instances in the dataset. Accuracy measures how often the model makes correct predictions.

$$Precision = \frac{(TP)}{(TP+FP)} \quad (5)$$

Precision measures the accuracy of the positive predictions made by the model. In other words, precision measures how accurately the model predicts positive instances. When precision is high, it means the model usually makes correct positive predictions. When precision is low, it means the model frequently makes incorrect positive predictions.

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

Recall, also known as sensitivity or true positive rate, measures the model's ability to identify all relevant positive instances. In other words, recall measures the proportion of true positive examples correctly identified by the model out of all actual positive instances.

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

Specificity, also known as the true negative rate, is a metric used to evaluate the performance of a classification model. In simpler terms, specificity measures the proportion of true negative examples correctly classified as negative by the model.

3. RESULT AND DISCUSSION

In this research, RapidMiner software was used to develop and select machine learning models. Two models, Naïve Bayes (NB) and K-Nearest Neighbor (K-NN) were compared to identify the most effective classification model. Table 4 provides a detailed overview of the dataset used for model development. The predicted outcomes for psychological readiness for entrepreneurship are classified into four categories: *Not Ready, Requires Training, Recommended, and Ready for Entrepreneurship*.

Table 4. Dataset Overview

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Categories
Very High	High	Very Low	Medium	Low	Low	Medium	Medium	Recommended
Very High	High	Very Low	Medium	Medium	Very High	Very Low	Very High	Ready for Entrepreneurship
Very High	High	Very Low	High	High	Very High	Very Low	Very Low	Recommended
Medium	Medium	Low	Low	Very Low	Low	Medium	Medium	Requires Training
Very High	Very High	Very High	Very High	High	High	Medium	Medium	Ready for Entrepreneurship
High	Medium	Medium	High	Medium	Medium	Medium	Very High	Recommended
Medium	Very High	Very Low	Very Low	Medium	Very High	High	Very Low	Requires Training
Low	Medium	Medium	Very High	Medium	Low	Medium	Low	Requires Training
Very High	Very High	Medium	High	Very High	Very High	Very High	Very High	Ready for Entrepreneurship
High	High	Low	Medium	High	High	Low	High	Recommended
Very High	Very High	Very High	Medium	High	High	High	High	Ready for Entrepreneurship
Very High	High	Low	Very High	High	High	Medium	Low	Recommended
High	Very High	Medium	Medium	High	Very High	Medium	Very High	Ready for Entrepreneurship
Medium	Very High	Very Low	Medium	Low	Very High	High	High	Recommended
Low	High	Low	Medium	High	Medium	Very Low	Very Low	Requires Training
Very High	Very High	Very High	Very High	High	Very High	Very High	Very High	Ready for Entrepreneurship
Medium	Medium	Low	Low	Low	Medium	Medium	High	Requires Training
Medium	Medium	Medium	Medium	Medium	Medium	Medium	High	Recommended
High	Very High	Very Low	Medium	Low	Medium	Very High	High	Recommended
Medium	Very High	Low	Medium	Very High	High	High	Very High	Ready for Entrepreneurship

Table 4 illustrates the 1095 datasets used in this study. 80% of the data was utilized as the training dataset, with the remaining 20% used as the testing dataset. Experiments were conducted on 2 (two) machine learning models for classification: Naïve Bayes (NB) and K-Nearest Neighbor. The data were randomized, and the experiment was repeated five (5) times to ensure that all models performed consistently. Accuracy is based on a confusion matrix, which is used as a performance indicator to determine which technique is most effective in modeling an individual's psychological preparation for entrepreneurship.

3.1. Naïve Bayes Model Training and *Testing Results*

The initial classification model in this study is developed using the Naïve Bayes (NB) algorithm. The process begins by training the Naïve Bayes model with a dataset consisting of 876 training samples. To evaluate the model's performance, it is tested five times using 219 distinct testing datasets. The results of these five trials are presented in Table 5.

Table 5. Testing Results of Naïve Bayes Model

Performance	Testing Result (%)					Average
	Exp1	Exp2	Exp3	Exp4	Exp5	
Accuracy	87.21	89.95	91.78	89.50	89.50	89.58
Precision	83.07	81.69	92.22	90.30	81.25	85.71
Recall / Sensitivity	87.83	82.55	77.24	75.46	81.89	80 .89

Performance	Testing Result (%)					Average
	Exp1	Exp2	Exp3	Exp4	Exp5	
Specificity	86.35	90.66	90.91	88.35	89.45	89.14

Table 5 presents the testing results for the Naïve Bayes model across five experiments, detailing performance metrics like Accuracy, Precision, Recall/Sensitivity, and Specificity. Each metric holds significance for evaluating the model's performance and its implications for decision-making.

Accuracy measures the overall correctness of the model's predictions, including both positive and negative classifications. A value of 89.58% indicates that nearly 90% of all classifications, regardless of category, are correct. High accuracy implies the model performs well in general classification tasks.

Precision reflects how accurately the model predicts positive cases (e.g., "Ready for Entrepreneurship"). A higher precision means fewer false positives, while lower values indicate the model misclassifies more negative cases as positive. An average precision of 85.71% means the model is relatively reliable in predicting readiness while minimizing incorrect classifications of unprepared individuals. This ensures better resource allocation for entrepreneurship training and development.

Recall measures how effectively the model identifies true positives out of all actual positive cases. With an average recall of 80.89%, the model balances identifying readiness while leaving room for improvement in reducing missed cases.

Specificity evaluates how well the model identifies true negatives out of all actual negative cases. High specificity means the model effectively avoids false positives, ensuring that individuals classified as "Not Ready" genuinely lack readiness. The average specificity of 89.14% demonstrates the model's strength in correctly identifying unprepared individuals,

3.2. K-Nearest Neighbour Model Training and *Testing Results*

The next classification model developed in this study employs the K-Nearest Neighbor (K-NN) algorithm. The K-Fold Cross-Validation technique is utilized to evaluate the accuracy of different K values during the model construction process. Testing is conducted five times using k-fold values of 3, 5, 10, 15, and 20. For each experiment, the model is tested using three different values for k (the number of nearest neighbors): 5, 10, and 15.

Table 6. Testing Results of K-NN Model

Result K-NN			Performance			
Exp	**k-fold**	**K**	**Precision**	**Recall**	**Accuracy**	**Specificity**
1	3	5	83.51	74.55	84.84	87.98
		10	87.34	69.26	85.75	87.58
		15	86.59	63.99	84.84	88.22
2	5	5	81.03	74.13	84.84	88.21
		10	87.74	71.68	86.03	88.23
		15	85.92	66.09	84.29	87.30
3	10	5	81.42	74.46	84.67	87.84
		10	85.14	73.64	86.12	88.63
		15	86.92	67.42	85.21	87.26
4	15	5	81.62	76.90	84.20	87.89
		10	87.92	73.20	85.84	88.10
		15	87.36	65.99	85.75	87.67
5	20	5	81.34	75.36	84.11	87.86
		10	85.53	73.58	86.31	88.76
		15	83.29	68.66	84.85	87.09
Average			**84.84**	**71.26**	**85.18**	**87.90**

Table 6 shows that the average precision (84.84%) reflects the model's strong capability in minimizing false positives. The average recall (71.26%), while Medium, indicates potential for improvement in identifying all true positive cases. The average accuracy (85.18%) demonstrates a balanced and reliable performance in predicting both positive and negative cases. Additionally, the average specificity (87.90%) highlights the model's effectiveness in accurately identifying negative cases, significantly reducing the occurrence of false positives.

Five tests were performed using the Naïve Bayes and K-Nearest Neighbor (K-NN) classification models on a dataset of 1,095 samples. The results highlight the accuracy of each model in making correct predictions, as determined by the confusion matrix presented for each experiment. Table 7 provides a comparison of the performance metrics for both learning models based on their testing outcomes.

Table 7. Table of Comparison Machine Learning Model Performance

Learning Model	**Performance**			
	Avg Accuracy	**Avg Precision**	**Avg Recall**	**Avg Specificity**
Naïve Bayes	89.58	85.71	80.89	89.14
K-NN	77.90	84.84	71.26	87.90

Table 7 shows that Naïve Bayes outperforms K-NN in terms of accuracy, showcasing its robustness and reliability for this classification task. With an accuracy of 89.58%, Naïve Bayes demonstrates strong performance across both positive and negative predictions. While both models exhibit strong precision, Naïve Bayes slightly surpasses K-NN in accurately predicting positive cases.

In terms of recall, Naïve Bayes significantly outperforms K-NN, making it more dependable for identifying positive cases. K-NN, on the other hand, has a lower recall, indicating that it occasionally misses positive cases and is less sensitive to detecting all positive outcomes.

Both models are effective in handling negative classifications, but Naïve Bayes has a slight advantage in avoiding false positives. Although K-NN shows good specificity at 87.90%, it is slightly lower than Naïve Bayes, which achieves a specificity of 89.14%.

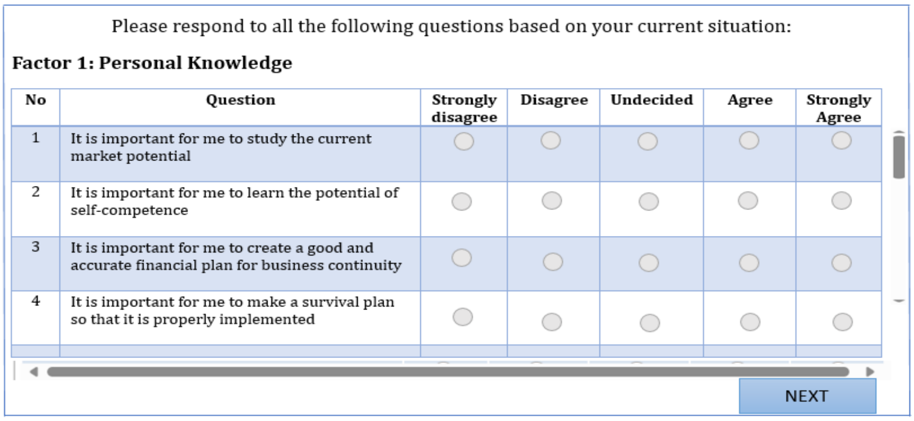
The confusion matrix's accuracy results are then expressed in the form of ROC (Receiver Operating Characteristic) analysis, and the categorization of the accuracy value produced by each learning model adheres to the following provisions [21]:

a. Accuracy score 0,90 – 1,00 = excellent classification
b. Accuracy score 0,80 – 0,90 = good classification
c. Accuracy score 0,70 – 0,80 = fair classification
d. Accuracy score 0.60 – 0,70 = poor classification
e. Accuracy score 0.50 – 0.60 = failure

Table 7 demonstrates that Naïve Bayes is better overall. It outperforms K-NN in all metrics, including precision, recall, accuracy, and specificity. This makes it a more reliable choice for this classification problem. The Naïve Bayes model achieves the greatest accuracy of 89.58% (0.8958), placing it in the "Good Classification" category.

3.3. Implementation of the model in counseling application

In this section, the chosen Naïve Bayes model is implemented in a web-based system. Figure 3 shows the interface used by the user to answer each question asked during the counseling session.



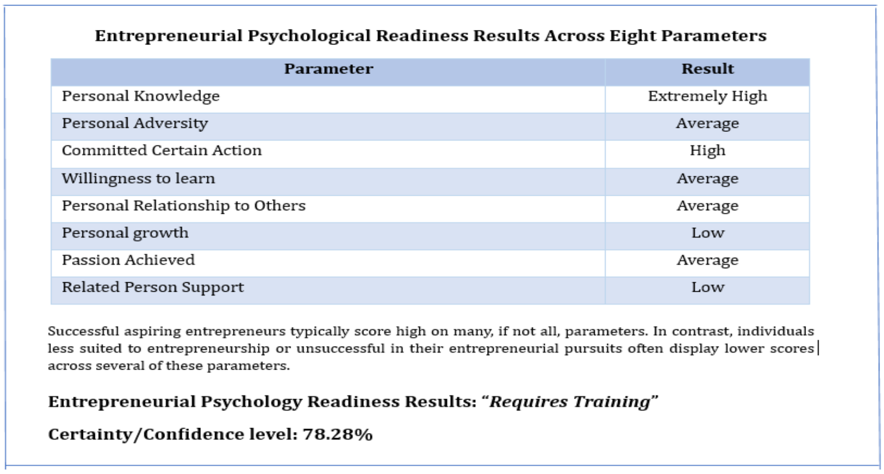
Please respond to all the following questions based on your current situation:

Factor 1: Personal Knowledge

No	Question	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
1	It is important for me to study the current market potential					
2	It is important for me to learn the potential of self-competence					
3	It is important for me to create a good and accurate financial plan for business continuity					
4	It is important for me to make a survival plan so that it is properly implemented					

Figure 3. Interface of Counseling Session

Figure 4 shows the prediction result interface for program users who have answered 33 questions measuring 8 (eight) indicators of psychological preparation for entrepreneurship.



Entrepreneurial Psychological Readiness Results Across Eight Parameters

Parameter	Result
Personal Knowledge	Extremely High
Personal Adversity	Average
Committed Certain Action	High
Willingness to learn	Average
Personal Relationship to Others	Average
Personal growth	Low
Passion Achieved	Average
Related Person Support	Low

Successful aspiring entrepreneurs typically score high on many, if not all, parameters. In contrast, individuals less suited to entrepreneurship or unsuccessful in their entrepreneurial pursuits often display lower scores across several of these parameters.

Entrepreneurial Psychology Readiness Results: "*Requires Training*"

Certainty/Confidence level: 78.28%

Figure 4. Counseling Prediction Result Interface

In addition to the entrepreneurial readiness categories given, the prediction results interface shows the percentage of one's trust in the prediction results. The percentage of confidence level displayed is an advantage of developing machine learning-based counseling applications.

4. CONCLUSION

The Naïve Bayes learning model outperformed the K-NN model in this study, achieving the highest accuracy of 89.58%. This indicates that the Naïve Bayes model is a reliable choice for predicting an individual's psychological readiness for entrepreneurship, as it demonstrates strong performance across key metrics like precision, recall, and specificity. Its ability to balance positive and negative predictions makes it particularly suitable for this classification task.

This study used a dataset with a higher proportion of data belonging to the "Ready for Entrepreneurship" and "Recommended" classes, while the other two classes, "Not Ready" and "Requires Training," had fewer instances. This class imbalance likely influenced the performance of the Naïve Bayes model, as it might have been better at predicting outcomes for the majority classes than for the minority classes. Such imbalances can create challenges for accurate classification across all categories.

One of the primary challenges in this study was dealing with an imbalanced dataset. The uneven distribution of data among classes may have skewed the model's performance, making it biased toward the majority classes. This highlights the need for strategies such as oversampling minority classes, undersampling majority classes, or employing algorithms designed to handle imbalanced data.

Future research should explore the incorporation of other machine learning models, such as Decision Trees, Random Forests, Support Vector Machines (SVM), and ensemble methods, to improve prediction performance and robustness. These models could provide alternative perspectives on the classification problem and address the limitations observed in this study.

The proposed counseling application, which utilizes the Naïve Bayes algorithm to predict psychological readiness for entrepreneurship, effectively helps users evaluate their preparedness to start a business.

5. ACKNOWLEDGMENTS

The authors would like to thank Universitas Islam Riau (UIR) for their financial support of this study.

REFERENCES

[1] S. Hornstein, V. Forman-Hoffman, A. Nazander, K. Ranta, and K. Hilbert, "Predicting therapy outcome in a digital mental health intervention for depression and anxiety: A machine learning approach," *Digital Health*, vol. 7, 2021. doi: 10.1177/20552076211060659.

[2] T. Goto, C. A. Camargo Jr, M. K. Faridi, R. J. Freishtat, and K. Hasegawa, "Machine Learning–Based Prediction of Clinical Outcomes for Children During Emergency Department Triage," *JAMA Netw. Open*, vol. 2, no. 1, pp. e186937–e186937, Jan. 2019, doi: 10.1001/jamanetworkopen.2018.6937.

[3] W. Bleidorn and C. J. Hopwood, "Using Machine Learning to Advance Personality Assessment and Theory," *Personal. Soc. Psychol. Rev.*, vol. 23, no. 2, pp. 190–203, May 2018, doi: 10.1177/1088868318772990.

[4] M. Savci, A. Tekin, and J. D. Elhai, "Prediction of problematic social media use (PSU) using machine learning approaches," *Curr. Psychol.*, vol. 41, no. 5, pp. 2755–2764, May 2022, doi: 10.1007/s12144-020-00794-1.

[5] N. Syafitri, S. Farradinna, W. Jayanti, and Y. Arta, "Machine learning to create decision tree model to predict outcome of entrepreneurship psychological readiness (EPR)," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 2, pp. 381-390, 2023.

[6] R. Dave, K. Sargeant, M. Vanamala, and N. Seliya, "Review on Psychology Research Based on Artificial Intelligence Methodologies," *J. Comput. Commun.*, vol. 10, no. 05, pp. 113–130, 2022, doi: 10.4236/jcc.2022.105008.

[7] F. N. R. Putri and J. Riyono, "Teknologi Artificial Intelligence dalam Upaya Pencegahan Bunuh Diri," *Metr. Ser. Hum. dan Sains*, vol. 3, no. 1, pp. 11–18, Apr. 2022.

[8] S. Bahri, R. Samsinar, and P. S. Denta, "Pengenalan Ekspresi Wajah untuk Identifikasi Psikologis Pengguna dengan Neural Network dan Transformasi Ten Crops," *Resist. (Elektronika Kendali Telekomun. Tenaga List. Komputer)*, vol. 5, no. 1, p. 15, May 2022, doi: 10.24853/resistor.5.1.15-20.

[9] R. Jacobucci, "Pairing machine learning and clinical psychology: How you evaluate predictive performance matters," *OSF Preprints*, Jan. 15, 2021. [Online]. Available:

https://osf.io/73pnk

[10] J. D. Elhai and C. Montag, "The compatibility of theoretical frameworks with machine learning analyses in psychological research," *Current Opinion in Psychology*, vol. 36, pp. 83-88, 2020.

[11] S. Farradinna, T. N. Fadhlia, and A. Azmansyah, "Entrepreneurial Personality in Predicting Self-Regulation on Small and Medium Business Entrepreneurs in Pekanbaru, Riau, Indonesia," *GATR J. Manag. Mark. Rev.*, vol. 3, no. 1, pp. 34–39, Feb. 2018, doi: 10.35609/jmmr.2018.3.1(5).

[12] B. Hermanto and S. E. Suryanto, "Entrepreneurship ecosystem policy in Indonesia," *Mediterranean Journal of Social Sciences*, vol. 8, no. 1, 2017.

[13] E. Kallas, "Environment-Readiness Entrepreneurship Intention Model: The Case of Estonians and the Russian-Speaking Minority in Estonia," *SAGE Open*, vol. 9, no. 1, p. 2158244018821759, Jan. 2019, doi: 10.1177/2158244018821759.

[14] H. M. da Silva Veiga, G. Demo, and E. R. Neiva, "The psychology of entrepreneurship," in *Organizational Psychology and Evidence-Based Management: What Science Says About Practice*, E. R. Neiva, C. V. Torres, and H. Mendonça, Eds., Springer International Publishing, 2017, pp. 135–156. doi: 10.1007/978-3-319-64304-5

[15] S. Azwar, "Penyusunan skala psikologi (II)," *Pustaka Belajar*, 2021.

[16] H. Chen, S. Hu, R. Hua, and X. Zhao, "Improved naive Bayes classification algorithm for traffic risk management," *EURASIP J. Adv. Signal Process.*, vol. 2021, no. 1, Dec. 2021, doi: 10.1186/s13634-021-00742-6.

[17] 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, p. 6256, Apr. 2022, doi: 10.1038/s41598-022-10358-x.

[18] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., The Morgan Kaufmann Series in Data Management Systems, 2011.

[19] A. Ali, M. A. T. Alrubei, L. F. M. Hassan, M. A. M. Al-Ja'afari, and S. H. Abdulwahed, "Diabetes diagnosis based on KNN," *IIUM Engineering Journal*, vol. 21, no. 1, pp. 175–181, 2020. doi: 10.31436/iiumej.v21i1.1206.

[20] A. A. Nababan, M. Khairi, and B. S. Harahap, "Implementation of K-Nearest Neighbors (KNN) Algorithm in Classification of Data Water Quality," *J. Mantik*, vol. 6, no. 1, pp. 30–35, May 2022.

[21] T. Rosandy, "Perbandingan Metode Naive Bayes Classifier Dengan Metode Decision Tree (C4.5) Untuk Menganalisa Kelancaran Pembiayaan (Study Kasus : Kspps / Bmt Al-fadhila," *J. Teknol. Inf. Magister Darmajaya*, vol. 2, no. 01, pp. 52–62, 2016.