NUTRITIONALSTATUSINDICATOROFTODDLERSUSING AMACHINELEARNINGAPPROACH

Nurjanni Hidayati Tanjung[1], Diva Nabillah[2], Intan Thiyas Intarita[3], Febria Sri Handayani[4]

[1,2,3,4] Institut Teknologi dan Bisnis PalComTech

Jalan Basuki Rahmat No.05, Palembang 30127

nurjanni.tan12@gmail.com[1*], divanabillah495@gmail.com[2], intanthiyas@gmail.com[3], febria_sri@palcomtech.ac.id[4]

Abstract

Using a machine learning-based methodology that incorporates visualization, classification, and clustering techniques, this study attempts to identify and categorize the nutritional status of toddlers. The dataset, which was obtained from Kaggle, has characteristics like height (in centimeters), gender, and age (in months). The RapidMiner software was used to do the analysis. The K-Means approach was utilized for clustering, while the K-Nearest Neighbor (K-NN) algorithm was utilized for classification. K-NN classification performed well in identifying all nutritional status groups, with an accuracy of 99.93%. From the visualization results using scatter plots and bar charts, it is known that toddlers with the nutritional status of "severely stunted" and "stunted" are mostly experienced by toddlers aged 24 months and above and toddlers with a height of less than 90 cm at the age of over 12 months are most often included in the "stunted" and "severely stunted" categories. K-Means effectively divided the data into four clusters, each of which represented a distinct growth trend according to height and age. The study does not account for other possible affecting factors like weight or socioeconomic position, and is restricted to a dataset with basic attributes (age, gender, and height). By providing a machine learning model for early stunting detection that can be used in healthcare systems and mobile health applications, this study advances the field of health informatics.

Keywords: *nutritional status, clustering, toddlers, machine learning, classification*

1. INTRODUCTION

One of the most significant public health indicators is the nutritional status of toddlers. Stunting is a significant issue in Indonesia because toddlers are particularly vulnerable to chronic malnutrition. Stunting can hinder both long-term physical growth and cognitive development (Ramadhani & Ramadhanu, 2024; Julianti & Elni, 2020). Early detection is therefore crucial. Traditional methods sometimes rely on manual evaluations, which are time-consuming and prone to errors. Malnutrition problems are also greatly influenced by sociodemographic and medical factors, including parental education, family income, and access to healthcare facilities (Kusumajaya et al., 2023; Kassie & Workie, 2020).

According to Ningsih et al. (2022), multifactorial factors such as feeding habits and hygienic circumstances frequently impact nutritional issues in toddlers between the ages of 12 and 59 months, underscoring the need for more comprehensive diagnostic tools. Stunting among toddlers aged 6 to 23

months is still common in areas like East Nusa Tenggara, necessitating more efficient data-driven treatments (Wulandary & Sudiarti, 2024). Although impacted by distinct sets of socioeconomic and access-related factors, a nationwide study also found that stunting affects both Indonesian urban and rural populations (Siramaneerat et al., 2024).

Tools for accurately and automatically classifying nutritional status are provided by machine learning. Machine learning models may use simple factors like height, gender, and age to identify early indicators of malnutrition. This study suggests a thorough method for analyzing toddlers' nutritional data that includes visualization, classification, and clustering algorithms.

One of the recurring obstacles in addressing stunting is the lack of timely and structured intervention systems at the community level. The implementation of integrated stunting programs is hampered by the time, workload, and coordination constraints that community health center health workers frequently experience (Ginting et al., 2023). Inadequate community awareness, unused growth tracking tools, and fragmented stakeholder communication all contribute to these obstacles. This demonstrates the pressing need, particularly in under-resourced locations, to implement supportive technology that can help with data analysis, case prioritizing, and real-time decision-making.

The promise of machine learning (ML) for the early diagnosis of toddlerhood malnutrition is being highlighted by international research. A recent meta-analysis of machine learning applications on Demographic and Health Surveys (DHS) data showed stable diagnostic performance across low- and middle-income countries, with an average accuracy of 68.9% for stunting prediction and 84.4% for wasting prediction (Rao et al., 2025). According to these findings, ML could improve community nutrition programs by detecting toddlers who are at risk before physical symptoms appear, particularly in settings with low resources like rural Indonesia.

2. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Prior research has demonstrated the efficacy of machine learning in health-related forecasting. Gustriansyah et al. (2024), for instance, showed how machine learning may be used to predict nutritional status. K-NN and other classification algorithms were compared in comparable situations by Dambe et al. (2023). It has been demonstrated that the K-NN algorithm performs more accurately than other models, such as Naive Bayes and Decision Trees (Putri et al., 2024). The K-NN algorithm combined with other algorithms also has the most optimal accuracy value, reaching 99% and the possibility of a very low error value (Insany et al., 2023). This shows that the K-NN algorithm is very effective in previous studies. Nevertheless, a lot of research solely looks at categorization, leaving out clustering and visualization. However, some studies carry out clustering and categorization simultaneously, but not with visualization. Such as study conducted by Fahik et al. (2018), showed that the K-Means clustering method was able to group village data based on TBGM with a value of 98.83% and the KNN method reached a value of 93.10 %.

The use of machine learning in nutrition monitoring is still developing in addition to these studies. In order to predict toddlers nutritional status, Insany et al. (2023) used both K-NN and Artificial Neural Networks (ANN). They found that K-NN was more effective and easier to implement, whereas ANN provided better scalability but required more processing resources. This emphasizes how crucial it is to select models that strike a compromise between practicality and accuracy, especially in public health settings with inadequate digital infrastructure.

In Malaka Regency, Fahik et al. (2018) showed that it is feasible to classify nutritional data at the village level by combining K-NN and K-Means. Although their investigation demonstrated excellent performance, it lacked visualization and instead concentrated on spatial grouping. Our study goes beyond this by presenting a three-pronged strategy that integrates clustering, classification, and visualization into a unified model pipeline.

This research makes the hypothesis that combining data visualization with classification and clustering methods can increase the accuracy of early stunting diagnosis based on the results of these studies. Compared to single-method models, this combination approach can assist identify nutritional hazards more clearly, especially for field practitioners and health officers.

3. RESEARCH METHODOLOGY

RapidMiner is used in this study's quantitative experimental methodology. Data collecting, preprocessing, machine learning algorithm modeling, and model evaluation are the steps.

3.1 Research Procedures

1. Data Collecting

   The dataset, which included characteristics like age (months), gender, height (cm), and nutritional condition, was obtained from Kaggle.

   (https://www.kaggle.com/datasets/rendiputra/stunting-balita-detection-121k-rows).

2. Data Preprocessing

   When necessary, normalization was performed, gender was encoded to numeric format, and missing values were handled as part of the data cleaning process. The goal variable in the classification process was the nutritional status attribute.

3. Modeling

   The model applied for the classification algorithm is K-Nearest Neighbor (K-NN) where the data is divided into 70:30 for training and testing, and the clustering algorithm uses KMeans without the label and the model built using the RapidMiner application.

4. Model Evaluation

   Evaluation is done using the following metrics: accuracy, precision, recall, and F1-score. Data is divided using the split validation method to measure model performance fairly using the principles of supervised learning (Witten et al., 2016). As for clustering, evaluation is done by visualizing the results and analyzing the suitability of the cluster distribution to the natural category.

3.2 Data Visualization

Finding early trends in nutritional issues requires the use of visualization. Bivariate correlations, namely the association between height and age across dietary categories, are displayed using scatter plots. In order to identify irregularities or imbalances, histograms provide a summary of the height values' distribution throughout the complete dataset. Comparative comparisons of average height within each nutritional status group are made possible using bar charts.

RapidMiner's built-in charting module was used to build each visualization. Different nutritional labels were represented by color encoding, enabling simple visual mapping. Preprocessing procedures also made sure that there was no visual bias brought on by missing data or outliers.

3.3 K-Nearest Neighbor Classification

Classification is done using the K-Nearest Neighbor (K-NN) algorithm. The K-NN algorithm is an approach to grouping based on the distance between data (Yuliansyah et al., 2022). This algorithm was chosen because of its simplicity in calculating the proximity between data from the elimination of the Euclidean framework. This algorithm also has advantages in the process of understanding and implementing it (Putri et al., 2024). In this study, the functions used as input are age (months), height (cm), and gender (converted to numeric format). The target label is the nutritional status of toddlers. The data is divided into training and testing data using the split validation method from a ratio of 70:30. Performance assessment is carried out by measuring accuracy, recall, and F1 score to identify the quality of the model when categorizing nutritional status categories.

After testing a number of choices (3, 5, 7, 9) to balance model stability and sensitivity, the value of k used in the K-NN model was set to 5. Since Euclidean distance works well with normalized numerical data, it was chosen as the similarity metric. RapidMiner's Split Validation operator was used to divide the dataset into 70% training and 30% testing sets.

3.4 K-Means Clustering

Clustering was performed using the K-Means algorithm to group toddler data based on attribute similarities without using nutritional status labels. In the K-Means method, data is grouped into several groups where each group has similar or the same characteristics as the others but with other groups having different characteristics (Apriyani, 2023). The features used in the dataset are age, height, and gender (numeric). The determination of the number of clusters was set at four (k = 4), adjusting to the number of nutritional status categories in the original data in the dataset used, severely stunted, stunted, normal, and high. Before the clustering process was carried out, the nutritional status attribute was removed from the data so that the approach was truly unsupervised. The results of the clustering will later be visualized in the form of a scatter plot to evaluate the cluster formation pattern and its suitability to the natural distribution in the data.

4. RESULTS AND DISCUSSIONS

4.1 Research Procedures

1. Data Collecting

Row No.	Status Gizi	Umur (bulan)	Jenis Kela...	Tinggi Bada...
1	stunted	0	laki-laki	44.592
2	tinggi	0	laki-laki	56.705
3	normal	0	laki-laki	46.863
4	normal	0	laki-laki	47.508
5	severely stun...	0	laki-laki	42.743
6	stunted	0	laki-laki	44.258
7	tinggi	0	laki-laki	59.573
8	severely stun...	0	laki-laki	42.702
9	stunted	0	laki-laki	45.252
10	tinggi	0	laki-laki	57.202
11	normal	0	laki-laki	51.354
12	normal	0	laki-laki	53.050
13	severely stun...	0	laki-laki	43.545
14	normal	0	laki-laki	46.253
15	severely stun...	0	laki-laki	43.676
16	normal	0	laki-laki	52.640

ExampleSet (120,999 examples,1 special attribute,3 regular attributes)

Figure 1: Toddler stunting dataset table

From the figure above, it can be seen that the dataset consists of four attributes, namely nutritional status, age (months), gender, and height (cm).

2. Data Preprocessing

The obtained dataset will be cleaned by checking for empty values, encoding categorical variables (gender), and normalization required for the K-NN algorithm. The nutritional status attribute is used as a label.

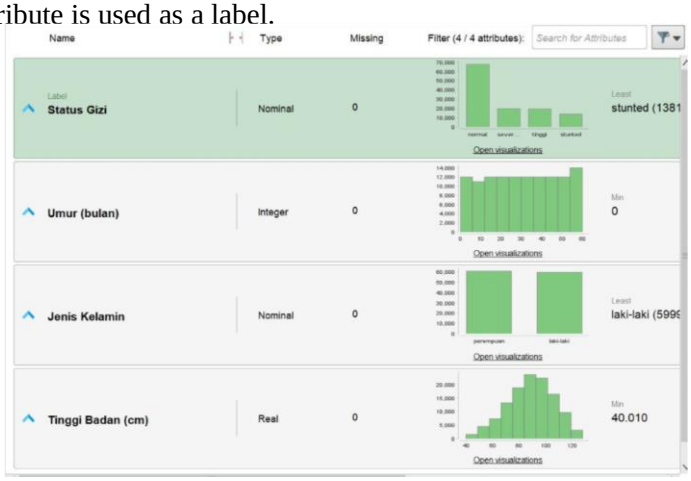

Figure 2: Missing value table

From the figure 2 above, it can be seen that the dataset used is in a clean condition, meaning there are no missing values, inappropriate values, or anomalous data.

3. Modelling

Modeling was performed in RapidMiner using the K-NN algorithm for classification and K-Means for clustering, with a structured process flow from preprocessing to automatic evaluation of results.

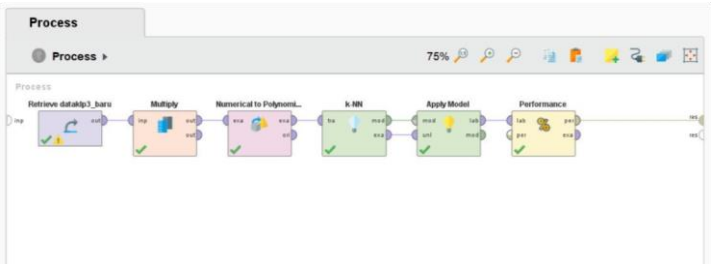

Figure 3: K-NN Algorithm Modelling

The K-NN algorithm uses the retrieve, multiply, numerical to polynomial, K-NN, apply model, and performance operators.

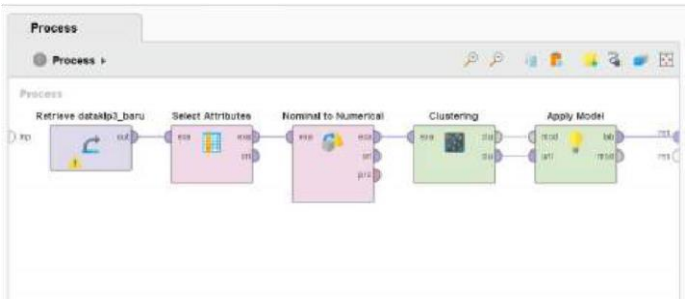

Figure 4: K-Means Algorithm Modelling

The K-Means algorithm uses the retrieve operator, select attributes, nominal to numerical, K-Means clustering, and apply model.

4. Model Evaluation

After processing the dataset with the operators that have been added to the previous modeling, it will produce a table of K-NN classification results as follows.

accuracy: 99.93%

	true stunted	true tinggi	true normal	true severely stunted	class precision
pred. stunted	13802	0	19	18	99.73%
pred. tinggi	0	19551	22	0	99.89%
pred. normal	6	9	67714	0	99.98%
pred. severely stun...	7	0	0	19851	99.96%
class recall	99.91%	99.95%	99.94%	99.91%	

Figure 5: K-NN classification result table

After processing the dataset using the operators that have been added to the previous modeling, the following table of K-Means clustering results will be produced.

Open in | Turbo Prep | Auto Model | Interactive Analysis | Filter (120.999 / 120.999 examples): all

Row No.	id	Status Gizi	cluster	Jenis Kela...	Jenis Kela...	Umur (bulan)	Tinggi Bada...
1	1	stunted	cluster_1	1	0	0	44.592
2	2	tinggi	cluster_1	1	0	0	56.705
3	3	normal	cluster_1	1	0	0	46.863
4	4	normal	cluster_1	1	0	0	47.508
5	5	severely stun...	cluster_1	1	0	0	42.743
6	6	stunted	cluster_1	1	0	0	44.258
7	7	tinggi	cluster_1	1	0	0	59.573
8	8	severely stun...	cluster_1	1	0	0	42.702
9	9	stunted	cluster_1	1	0	0	45.252
10	10	tinggi	cluster_1	1	0	0	57.202
11	11	normal	cluster_1	1	0	0	51.354
12	12	normal	cluster_1	1	0	0	53.050
13	13	severely stun...	cluster_1	1	0	0	43.545
14	14	normal	cluster_1	1	0	0	46.253
15	15	severely stun...	cluster_1	1	0	0	43.676
16	16	normal	cluster_1	1	0	0	52.640

ExampleSet (120.999 examples,3 special attributes,4 regular attributes)

Figure 6: K-Means clustering result table

4.2 Visualization

Visualization is the first step in the data exploration process to gain an initial understanding of the characteristics of the dataset. In this study, three visualizations were used to explore the relationship between the attributes of age, height, and nutritional status.

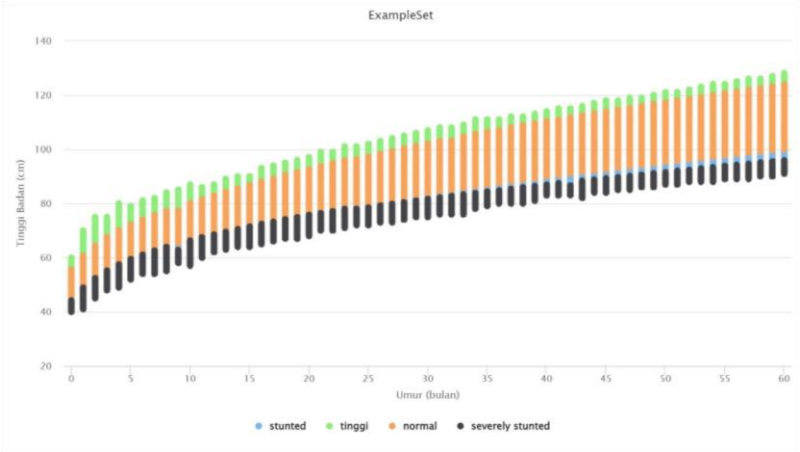

Figure 7: Visualization of toddler data (height by age) with scatter plot chart

The scatterplot figure shows the distribution of age and height with color based on nutritional status. From this graph, it can be seen that toddlers with "severely stunted" status generally have lower height and are at an early age. In contrast, toddlers with "high" status are in areas with heights above 110 cm. A height of less than 90 cm after 12 months may be a crucial cutoff point for identifying toddlers at risk of stunting, according to the scatter plot, which also clearly shows the division between nutritional groups. This knowledge can direct healthcare professionals' early screening standards.

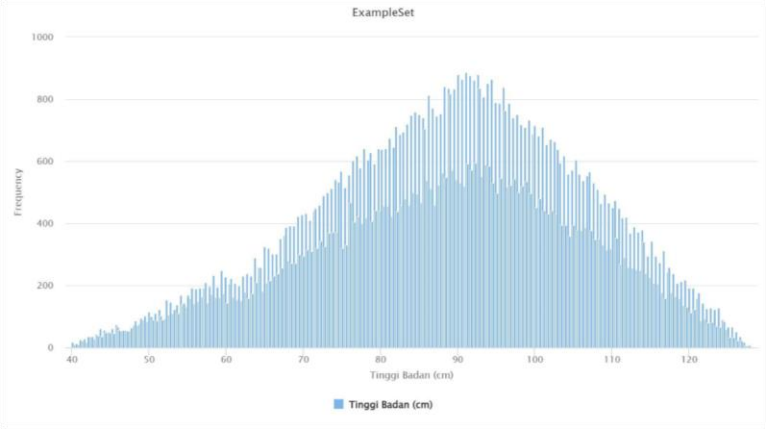

Figure 8: Visualization of toddler data (height) with histogram chart

The histogram of height shows a normal distribution with most toddlers being between 90–100 cm tall. The bar chart showing the average height per nutritional status category shows an upward trend from “severely stunted” to “high”, indicating the relevance of height as an indicator of nutritional status.

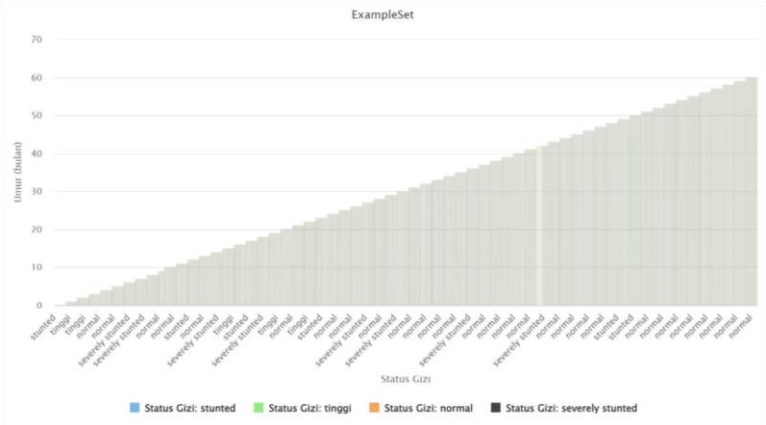

Figure 9: Visualization of toddler data (nutrition categories based on age) with bar chart

The bar chart figure above shows the distribution of nutritional status of toddlers based on age (months), with each bar representing one toddler, where the height of the bar indicates age and the color of the bar indicates nutritional status (high, normal, stunted, severely stunted). It can be seen that the nutritional status of "normal" is the most dominant, but there are still many toddlers who experience stunted and severely stunted, especially at the age of two years and above. Meanwhile, the status of "high" is rarely found. The prevalence of stunting and severe stunting rises sharply after 24 months, as this graphic illustrates, indicating that this age range is a crucial opportunity for intervention. It highlights the necessity of ongoing observation and dietary assistance, especially in the second and third years of life.

4.3 Classification (K-NN)

Classification using the K-NN algorithm achieved an accuracy of 99.93%. Other models such as Random Forest (71.81%), Naive Bayes (55.42%), and Decision Tree (64.97%) gave lower

results. The confusion matrix shows that K-NN can distinguish all categories of nutritional status accurately.

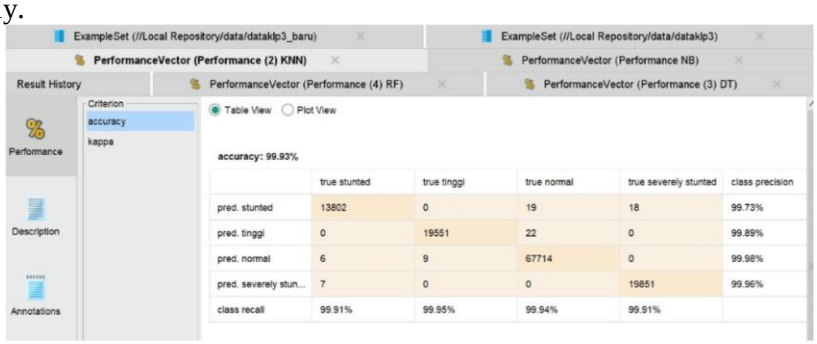

Figure 10: Results of analysis using the K-NN algorithm

The figure above shows the results of the K-NN (K-Nearest Neighbor) model evaluation in classifying nutritional status in toddlers. The K-Nearest Neighbors (KNN) classification model performed remarkably well, attaining a 99.93% accuracy rate. The fact that this number is almost flawless means that practically every entry in the dataset was appropriately categorized. Nonetheless, a closer look at the confusion matrix is necessary to make sure that this exceptional performance holds true for all classes. Based on the confusion matrix displayed, the model can classify any category of nutritional status (stunted, high, normal, severely stunted). With just 19 records from the "normal" class and 18 from the "severely stunted" class being incorrectly placed into this group, 13.802 records were accurately predicted for the "stunted" category.

This yields a 99.73% class precision and a 99.91% recall, showing that predictions for this category are both very accurate and thorough. Additionally, the "high" (tinggi) category performed exceptionally well. With the exception of just 22 records from the "normal" class that were mistakenly predicted to be "high," all 19.551 actual records from this class were correctly classified. As a result, the model's accuracy in recognizing this category was demonstrated by its extremely high precision of 99.89% and recall of 99.95%. Just six entries from the "stunted" and nine from the "high" categories were incorrectly categorized into the "normal" class out of 67.714 genuine records. With a precision of 99.98% and a recall of 99.94%, this demonstrated an almost flawless harmony between sensitivity and accuracy for this class. Finally, the "severely stunted" category performed exceptionally well as well. Only seven records from the "stunted" category were mistakenly projected to be "severely stunted," out of 19.851 actual records, which were all correctly classified. The accuracy and consistency of the model in determining this nutritional status were once again confirmed by the 99.96% precision and 99.91% recall that were obtained.

Overall, with precision and recall exceeding 99% for each class, the KNN model showed balanced and exceptional performance across all four nutritional status areas. This implies that the training and testing dataset was quite optimal, either in terms of feature space structure that complements KNN's distance-based methodology, appropriate feature selection, or balanced class distribution. Furthermore, it is essential to make sure that the right preprocessing and normalization have been used because KNN is sensitive to irrelevant characteristics and data scale. These results allow us to conclude that the K-NN algorithm is used very effectively to automatically identify nutritional status in toddlers based on age, gender and height.

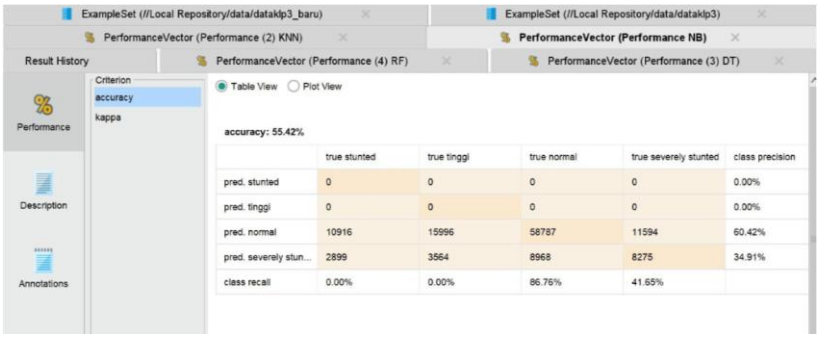

Figure 11: Results of analysis using the Naive Bayes algorithm

The figure above shows the results of nutritional status classification in toddlers using the Naive Bayes algorithm. This classification model's accuracy was 55.42%, which indicates that it accurately predicted only roughly half of the total data. Even though this number might seem reasonable at first glance, a deeper look at the confusion matrix shows that the model's performance is unbalanced and skewed toward some classes, which raises questions about how effective it is, especially for data from minority classes. The confusion matrix shows that the "normal" and "severely stunted" groups account for the bulk of predictions, whereas the "stunted" and "high" categories received no predictions at all. The completely 0 rows for "pred. stunted" and "pred. high," which shows that the model never allocated any data to these two categories, demonstrate this. The complete absence of predictions for these two important classes is a clear indication of classification bias.

The model's performance for the "normal" class was comparatively good. Out of the real "normal" class, 58.787 records were accurately classified. In addition, the model misclassified 11.594 data from "severely stunted," 15.996 from "high", and 10.916 from "stunted" into the "normal" group. Consequently, this class's precision was just 60.42%, which indicates that only 6 out of 10 "normal" predictions were correct. At 86.76%, the recall was comparatively high, suggesting that the model was able to identify the majority of "normal" records. On the other hand, the model predicted 2.899 records from "stunted", 3.564 from "high", and 8.968 from "normal" as "severely stunted" for the "severely stunted" class. Out of the actual "severely stunted" class, only 8.275 records were appropriately classified. As a result, the precision for this class was only 34.91%, suggesting that other categories accounted for the majority of the model's predictions for this class. Additionally, the recall was comparatively low at 41.65%, meaning that the model missed over half of the records that were indeed "severely stunted".

More importantly, the "high" and "stunted" classes showed 0.00% precision and recall, which means that neither a single record from either of these two categories was identified by the model, nor were any predictions made for them. In a medical setting, where misclassifying illnesses like stunting or aberrant growth can lead to missed diagnosis and a lack of care, this is extremely problematic. This result suggests that Naive Bayes may not be a suitable choice for nutritional status classification because it cannot recognize important categories such as stunting.

Figure 12: Results of analysis using Decision Tree algorithm

Based on the classification results shown in the figure 12, the model only achieved an accuracy of 56.00%. This means, of all the data tested, only 56% was correctly predicted. However, this accuracy figure does not accurately reflect the model's performance, as a closer look at the confusion matrix reveals that the model completely failed to recognize three of the four categories, namely stunted, high, and severely stunted. From the confusion matrix table analysis, it can be seen that all data, without exception, are classified as the "normal" category. This is seen in the "pred. normal" row which includes all data from the four original categories, namely 13.815 data from the "stunted" category, 19.560 data from the "high", 67.755 data from the "normal", and 19.869 data from the "severely stunted". Meanwhile, the other three rows such as "pred. stunted", "pred. high", and "pred. severely stunted", all have a value of zero, indicating that no data is expected to fall into these three categories. This situation results in a classification situation that is highly biased towards the majority category, namely normal. Although the model has a 100 % recall value for the "normal" category, indicating that all normal data were successfully identified, the model is completely unable to recognize other categories, with 0% recall values for stunted, high, and severely stunted. Similarly, the model's precision value for all categories, except normal, is 0%, indicating no accurate predictions for those categories. The precision for the "normal" category itself is 56%, indicating that of all "normal" predictions, only 56% actually come from that category. The dataset's class imbalance is most likely the root of this issue. The model may learn that classifying all data as "normal" will produce high accuracy even if the prediction quality is actually quite poor if the "normal" class is vastly overrepresented. When it comes to toddlers nutritional status, this is particularly risky because incorrectly classifying conditions as "stunted" or "severely stunted" can have major treatment repercussions.

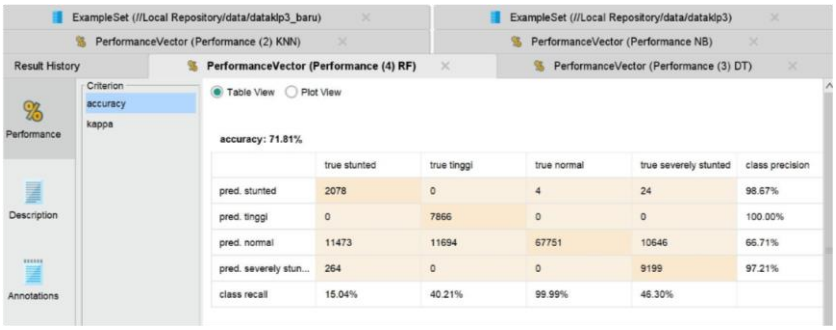

Figure 13: Results of analysis using Random Forest algorithm

The figure above shows the results of nutritional status classification in toddlers using the Random Forest algorithm, comparing the overall classification accuracy to the other algorithms examined, it is 71.81%, indicating a moderate level of predictive performance. The confusion matrix shows how accurately each category was categorized in relation to its actual nutritional status.

The model's recall of 99.99% and class precision of 66.71% demonstrate its remarkable performance in identifying the "normal" category. This suggests that practically every toddler who is actually categorized as "normal" was identified accurately. However, a comparatively poor precision counterbalances this high recall, thus many predictions that are classified as "normal" also contain data from other categories, particularly "high" and "stunted." Additionally, the "high" (tinggi) category performs well, with a recall of 40.21% and perfect precision of 100 %. This suggests that this class may be under-identified because, despite the model's high accuracy in

predicting a toodler to be "high," it only properly identifies roughly 40% of all tall toddlers. In contrast, the model's recall and precision for the "severely stunted" group were 46.30% and 97.21%, respectively. This indicates that while the model is highly likely to be accurate when it predicts a toddler to be severely stunted, it misses over half of the toddlers who are actually badly stunted. The model performs the worst in the "stunted" category, with a recall of only 15.04 %, suggesting that it has trouble accurately identifying toddlers who are stunted. Despite having a 98.67% accuracy rate, only a small percentage of actual stunted cases are recorded. This implies that the model may miss a large number of real cases of stunting since it is overly conservative in predicting this category.

In conclusion, the Random Forest algorithm exhibits high precision in the majority of classes, but recall problems in minority categories like "stunted" and "severely stunted." This emphasizes that in order to increase recall in underrepresented labels, either improved class balancing or ensemble tuning are required. Even though Random Forest is generally balanced, it could not be the best option in situations where ignoring real stunted instances would have a significant cost, including in public health settings where early intervention is crucial.

The K-NN algorithm shows the advantage that the closest approach to numerical data such as height and age is very effective. This result is true in the study of Ramadhani & Ramadhanu (2024) which shows that K-NN excels in classifying toddler health status data due to its simplicity and accuracy.

Qualitatively, the high classification results provide potential for practical use in decision support systems for early detection of toddlers. Due to its almost perfect accuracy, this K-NN model can be used in Puskesmas or toodler health services as an automatic identification tool for falls, allowing interventions to be targeted more quickly.

4.4 Clustering (K-Means)

After the classification process, this study continued the analysis using an unsupervised learning approach through the K-Means algorithm to find natural groups in toddler data.

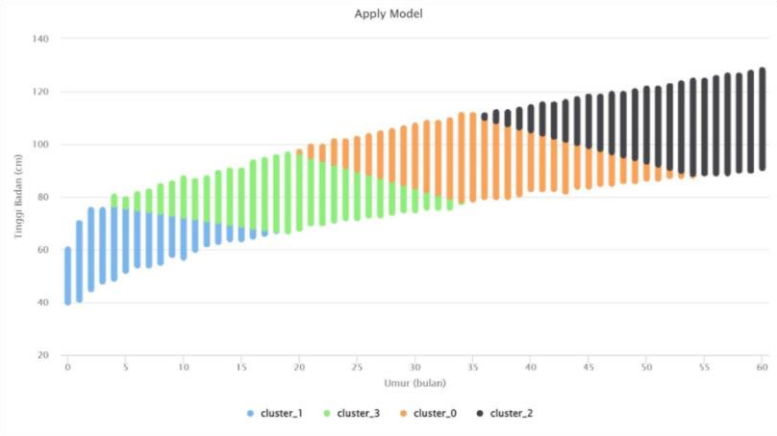

Figure 14: Results of clustering analysis using K-Means

The clustering results show the formation of four clusters based on similarities in age, height, and gender. The scatter plot of the clustering results shows a fairly clear separation pattern. Clusters with low height and early age are thought to be related to the "severely stunted" category, while clusters with high height and older age tend to reflect "high" status.

These findings indicate that K-Means is able to group toddler data into certain growth patterns even without label information. These results reinforce the importance of utilizing numerical data (age and height) in a more in-depth analysis of nutritional status.

5. CONCLUSION

This study successfully categorized the nutritional status of toddlers using data on age, gender, and height through machine learning methods. Through data visualization, a clear pattern can be seen between physical growth and nutritional status categories. From the visualization results using scatter plots and bar charts, it is known that toddlers with nutritional status "severely stunted" and "stunted" are mostly experienced by toddlers aged 24 months and above and toddlers with a height below 90 cm at the age of over 12 months are most often included in the "stunted" and "severely stunted" categories. From the classification analysis using the K-Nearest Neighbor (K-NN) algorithm, information was obtained that the algorithm showed the most precise results with an accuracy level of 99.93%. This indicates that the model is the most effective for automatically detecting the nutritional status of toddlers. Then through clustering analysis using the K-Means algorithm, the results showed that this algorithm can group existing data without any previously provided label information. These results indicate that the combination of visualization, classification, and clustering can be an integrated solution to detect the risk of stunting early by grouping and analyzing toddler nutritional status data.

In this way, the objectives set in the introductory chapter at the beginning can be realized, namely implementing and evaluating machine learning algorithms to classify nutritional status while determining the best model based on evaluation criteria.

The development prospects of the results of this study include several aspects, including:

1. As a tool to detect stunting problems early for health workers. By inputting toddler information such as age, height, and gender, this system is able to provide classification results quickly and objectively, and can be applied on a wider scale.
2. Supporting the decision-making process at the institutional level, for example in the health office, to determine age groups or areas at high risk of stunting.
3. Facilitating monitoring of toddler's nutritional status by parents through the integration of models in mobile applications, so that parents can monitor their toddler's nutritional development routinely and take preventive measures earlier.

LIMITATION AND STUDY FORWARD

These results open up opportunities for further research, such as adding data features (for example, weight or economic status) or applying the model to a wider population to improve the accuracy and generalizability of the system.

ACKNOWLEDGEMENT

The creators express appreciation to PalComTech Institute of Technology and Business for the bolster and offices given amid this investigate.

REFERENCES

Apriyani, P., Dikananda, A. R., & Ali, I. (2023). Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi. *Hello World Jurnal Ilmu Komputer, 2(1).* https://doi.org/10.56211/helloworld.v2i1.230

Dambe, M. L., Padang, S. Y., & Adha, M. S., (2023). Evaluasi K-Nearest Neighbors Untuk Klasifikasi Status Gizi Balita. *infinity, 3(1),* 34-41.

Fahik, B. Y. L., Djahi, B. S., & Rumlaklak, N. D. (2018). Data Mining untuk Klasifikasi Status Gizi Desa di Kabupaten Malaka Menggunakan Metode K-Nearest Neighbor. *J-ICON, 6(1),* 1-7.

Ginting, R., Girsang, E., Sinaga, M., & Manalu, P. (2023). Barriers to Stunting Intervention at a Community Health Center: A Qualitative Study. *Jurnal Penelitian Pendidikan IPA*, *9*(10), 8185–8191. https://doi.org/10.29303/jppipa.v9i10.4656

Gustriansyah, R., Suhandi, N., Puspasari, S., & Sanmorino, A. (2024). Machine Learning Method to Predict the Toddlers Nutritional Status. *INFOTEL, 16(1),* 32-43. https://doi.org/10.20895/infotel.v15i4.988

Insany, G. P., Yulistiana, I., & Rahmawati, S. (2023). Penerapan KNN dan ANN pada klasifikasi status gizi balita berdasarkan indeks antropometri. *CoSciTech, 4(2),* 385-393.

Julianti, E. & Elni, E. (2020). Determinants of Stunting in Children Aged 12-59 Months. *Nurse Media Journal of Nursing, 10(1),* 36-45. https://doi.org/10.14710/nmjn.v10i1.25770

Kassie, G. W. & Workie, D. L. (2020). Determinants of under-nutrition among children under five years of age in Ethiopia", *BMC Public Health, 20(399)*.

Kusumajaya, A. A. N., Mubasyiroh, R., Sudikno, S., Nainggolan, O., Nursanyoto, H., Sutiari, N. K., Adhi, K. T., Suarjana, I. M., & Januraga, P. P. (2023). Sociodemographic and Healthcare Factors Associated with Stunting in Children Aged 6-59 Months in the Urban Area of Bali Province, Indonesia 2018. *Nutrients*, *15*(2), 389.

Ningsih, T. M., Hardjito, K., & Yanuarini, T. A. (2022). Factors Associated with Nutrition of 12-59 Months Toddlers, *JNK, 9(1),* 084-091.

Putri, I. P., Terttiaavini, T., & Arminarahmah, N. (2024). Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak: Comparative Analysis of Machine Learning Algorithms for Predicting Child Stunting. *MALCOM, 4(1),* 257-265. https://doi.org/10.57152/malcom.v4i1.1078

Ramadhani & Ramadhanu. (2024). Metode Machine Learning untuk Klasifikasi Data Gizi Balita dengan Algoritma Naïve Bayes, K-NN dan Decision Tree. *Jurnal SIMETRIS, 15(1),* 57-67.

Rao, B., Rashid, M., Hasan, M. G., & Thunga, G. (2025). Machine Learning in Predicting Child Malnutrition: A Meta-Analysis of Demographic and Health Surveys Data. *International Journal of Environmental Research and Public Health*, *22*(3), 449. https://doi.org/10.3390/ijerph22030449

Siramaneerat, I., Astutik, E., Agushybana, F., Bhumkittipich, P., & Lamprom, W. (2024). Examining determinants of stunting in Urban and Rural Indonesian: a multilevel analysis using the population-based Indonesian family life survey (IFLS). *BMC Public Health, 24(1371)*.

Witten, I. H., Frank, E., Hall, M. A., Pal, C. J., & Foulds, J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (5th ed.). Morgan Kaufmann.

Wulandary, W. & Sudiarti, T. (2024). Stunting on Children Aged 6 - 23 Months in East Nusa Tenggara Province, *KEMAS, 19(4),* 585-595.

Yuliansyah, M. R., B. M., Franz. A. (2022). Perbandingan Metode K-Nearest Neighbors dan Naïve Bayes Classifier Pada Klasifikasi Status Gizi Balita di Puskesmas Muara Jawa Kota Samarinda. *ATASI, 1(1),* 08-20.