Paradigma, Vol. 26, No. 1, Maret 2024, P-ISSN 1410-5063, E-ISSN: 2579-3500, Page. 51-57
Published: LPPM Universitas Bina Sarana Informatika
PARADIGMA
website: http://jurnal.bsi.ac.id/index.php/paradigma/

Phyton-Based Machine Learning Algorithm to Predict Obesity Risk Factors in Adult Populations

Mari Rahmawati[1], Ade Fitria Lestari[2], Sri Hardani[3]

[1,2,3] Faculty of Technic and Informatics, Universitas Bina Sarana Informatika Jakarta, Indonesia

ARTICLE INFORMATION

Artikel History:
Received: February 23, 2024
Revised: March 18, 2024
Accepted: March 29, 2024

Keyword:

Obesity
Classification
K-Nearest Neighbor
Decision Tree
Naïve Bayes

ABSTRACT

Obesity is a serious problem in the world of health because it can cause various diseases. Adults are prone to obesity. There are many factors that are assumed to be risk factors for obesity in adults. There is a need for identification to determine the risk factors for obesity in adults so that prevention can be carried out. This research uses the Decision Tree, Naïve Bayes, and K-Nearest Neighbor algorithms to identify risk factors for obesity in adults. 2111 is processed using the Python programming language and applies these three algorithms. The Decision Tree algorithm identifies Weight, Age, Gender, MTRANS (primary mode of transportation), and SCC (monitors their caloric intake) as the main risk factors for obesity in adults. Naïve Bayes and K-Nearest Neighbor provide similar results when identifying risk factors for obesity in adults. Naïve Bayes and K-Nearest Neighbor identify Weight, Age, Gender, MTRANS (primary mode of transportation), and Family History with Overweight as risk factors for obesity in adults. This research also shows that the Decision Tree algorithm produces higher accuracy than the Naïve Bayes and K-Nearest Neighbor algorithms in adult diabetes classification models. The Decision Tree algorithm produces 94% accuracy, K-Nearest Neighbor 80%, and Naïve Bayes 60%.

Corresponding Author:
Mari Rahmawati
Faculty of Technic and Informatics,
Universitas Bina Sarana Informatika,
Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450,
Email: mari.mrw@bsi.ac.id

INTRODUCTION

Obesity has become a serious problem facing the whole world including Indonesia is now an emergency because it can cause a wide range of diseases such as heart disease, diabetes, hypertension, cancer, stroke and some other disorders. Obesity is caused by overweight than normal weight, generally due to excessive accumulation of triacyl glycerol in fatty tissue (Toar et al., 2023). According to the WHO and research Jiang et al that overweight or obesity is a condition of abnormality or excess fat in individuals that plays a role as one of the disease factors that threaten a person's health(Sitanggang & Sherly, 2022). According to Renew Bariatrics, data from the Centers for Disease Control and Prevention (CDC) indicate that by 2022 there are 22 states with a prevalence of obesity among adults over 35%, while in Indonesia according to the Central Statistical Agency (BPS) which processes data from Health Ministry's Basic Health Research (Kemenkes Riskdas) indicates that obesities in 2018 the prevalency of the population over the age of 18 including the adult category has increased by 21.80%, of which the data is likely to be a serious adult problem that needs to be addressed immediately and to be tried to reduce the rate of Obesity for the sake of public health.Many factors that cause the Indonesian population to suffer from obesity pose a high risk of becoming the trigger of all diseases such as body mass index (BMI), age, gender, genetic factors, biology, diet, physical activity, certain medical problems and so on. Previous research explains that there are several factors that cause overweight (obesity) including biological factors, development, behavioral environment and heredity (Alpiansah & Ramdhani, 2023). To suppress increased obesity in adults,

DOI: https://doi.org/10.31294/p.v26i1.3242

classification algorithms are needed to identify risk factors for obesities using machine learning. Machine learning is part of artificial intelligence that works to make computers have the ability to learn about new data without having to be programmed (Asri et al., 2024). Machine learning is the study of computational methods to identify complex in millions of data to build predictive models. The main focus of machine learning is to build a computer application that can study data, then create a model that is ready to use.

Different machine learning algorithms like Gradient Boosting (GB), Bagging meta estimator (BME), XB Boost (XGB), Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) are used in predicting risk obesity is based on a person's physical description and eating habits (Kaur et al., 2022). Prediction of obesity risk using machine learning algorithms by collecting 367 data from various age groups with obesity or non-obese classification based on blood test results using the BayesNet, Naïve Bayes, SMO, Simple Logistic, IBk, Kstar, J48, Random Forest and Random Tree Algorithms algorithms (Nur Cuhadar et al., 2023). K-Nearest Neighbor (KNN) Classification Method to determine the level of obesity based on eating habits and fairly good physical condition (Dewi & Dwidasmara, 2020). Predict obesity levels by comparing machine learning algorithms (Random Forest, Decision Tree, Logistic Regression, Naïve Bayes, KNN) and deep learning using datasets of obesity level data in individuals from Mexico, Peru and Colombia based on food habits and physical condition(Setiyani et al., 2023) . The ability of machine learning methods, namely Logistic Regression, Classification and Regression Trees (CART) and Naïve Bayes to identify obesity status in adults based on risk factors available in health datasets (Thamrin et al., 2021a).

RESEARCH METHOD

This research uses experimental methods, according to Winarni, which are systematic, thorough and logical to control a condition(Akbar et al., 2023). This research involves manipulating independent variables to test their influence on dependent variables by controlling other variable factors.

Overview of the stages of the research carried out as follows:

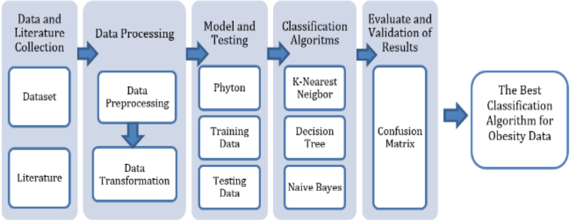

Figure 1. Research Stages

1. Data and Literature Collection

The study used a data set of obesity rates based on a person's dietary habits and physical condition. This data set is obtained from https://www.kaggle.com where this dataset is a data set from several countries such as Mexico, Peru and Colombia. The data set consists of 17 attributes and 2111 data as shown in Figure 2.

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
0	Female	21.000000	1.620000	64.000000	yes	no	2.0	3.0	Sometimes	no	2.000000	no	0.000000	1.000000	no	Public_Transportation	Normal_Weight
1	Female	21.000000	1.520000	56.000000	yes	no	3.0	3.0	Sometimes	yes	3.000000	yes	3.000000	0.000000	Sometimes	Public_Transportation	Normal_Weight
2	Male	23.000000	1.800000	77.000000	yes	no	2.0	3.0	Sometimes	no	2.000000	no	2.000000	1.000000	Frequently	Public_Transportation	Normal_Weight
3	Male	27.000000	1.800000	87.000000	no	no	3.0	3.0	Sometimes	no	2.000000	no	2.000000	0.000000	Frequently	Walking	Overweight_Level_I
4	Male	22.000000	1.780000	89.800000	no	no	2.0	1.0	Sometimes	no	2.000000	no	0.000000	0.000000	Sometimes	Public_Transportation	Overweight_Level_II
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2106	Female	20.976842	1.710730	131.408528	yes	yes	3.0	3.0	Sometimes	no	1.728139	no	1.676269	0.906247	Sometimes	Public_Transportation	Obesity_Type_III
2107	Female	21.982942	1.748584	133.742943	yes	yes	3.0	3.0	Sometimes	no	2.005130	no	1.341390	0.599270	Sometimes	Public_Transportation	Obesity_Type_III
2108	Female	22.524036	1.752206	133.689352	yes	yes	3.0	3.0	Sometimes	no	2.054193	no	1.414209	0.646288	Sometimes	Public_Transportation	Obesity_Type_III
2109	Female	24.361936	1.739450	133.346641	yes	yes	3.0	3.0	Sometimes	no	2.852339	no	1.139107	0.586035	Sometimes	Public_Transportation	Obesity_Type_III
2110	Female	23.664709	1.738836	133.472641	yes	yes	3.0	3.0	Sometimes	no	2.863513	no	1.026452	0.714137	Sometimes	Public_Transportation	Obesity_Type_III

2111 rows × 17 columns

Figure 2. Sample Obesity Data Set

Of the 17 attributes available, one attribute will be used as a label or target and 16 attributs are used as parameters. The attribute to be used as a label is NObeyesdad. This attribute contains a classification of the body time index, which consists of Insufficient Weight, Normal Weights, Overweight Level I, Overheight Level II, Obesity Type I, obesity type II, and obesities type III. The data distribution of each body time index class of this data set is described in figure 3 and figure 4 below.

Obesity Classification	count
Obesity_Type_I	351
Obesity_Type_III	324
Obesity_Type_II	297
Overweight_Level_I	290
Overweight_Level_II	290
Normal_Weight	287
Insufficient_Weight	272

Figure 3. Amount of data on each class

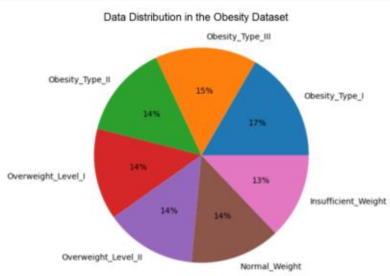

Figure 4. Data Distribution on Obesity Dataset

In addition to datasets, literature in the form of journals, books or internet references is required to support and complement research themes.

2. Data Processing

The processing of datasets is divided into two: data preprocessing and data transformation. Data preprocesing by means of missing value verification, data inconsistency verification and data mismatch verification.

3. Data preprocessing

Data preprocessing is carried out with the aim of improving the quality of the data and in accordance with the classification methods used (Ramadhan & Mandala, 2023). Generally, data from https://www.kaggle.com can be processed without preprocessing. However, in order to ensure quality, the study performed several checks including missing value checks, inconsistency data, and mismatched data.

4. Checking Missing Value

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeye
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
2109	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
2110	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	

Figure 5. Missing Value Verification Result

From Figure 5 above, you can see that there is no missing value.

5. Checking Incosistency Data

```
data.Gender.unique()

array(['Female', 'Male'], dtype=object)

data.Age.unique()

array([21.      , 23.      , 27.      , ..., 22.524036, 24.361936,
       23.664709])

data.Height.unique()

array([1.62    , 1.52    , 1.8     , ..., 1.752206, 1.73945 , 1.738836])
```

Figure 6. Data Incosistency Verification Results

Results of data inconsistency checks in Figure 6 show that each column contains the corresponding data.

6. Checking Data Mismatched

```
check_gender = data.applymap(lambda x: isinstance(x, (object)))['Gender']
check_gender

0       True
1       True
2       True
3       True
4       True
        ...
2106    True
2107    True
2108    True
2109    True
2110    True
Name: Gender, Length: 2111, dtype: bool
```

Figure 7. Data Verification Results Mismatched

Figure 7 shows the results of the verification data mismatched on the gender and age columns where in both the column, contains the data that corresponds to the data type in that column.

7. Data Transformation

Data transformation is a stage in which data is transformed and consolidated so that the data format is in accordance with data mining needs.

```
# Ubah nilai string menjadi nilai numerik
# update gender
g = {'Female': 0, 'Male': 1}
data['Gender'] = data['Gender'].map(g)

# update family_history_with_overweight
f = {'no': 0, 'yes': 1}
data['family_history_with_overweight'] = data['family_history_with_overweight'].map(f)

# update FAVC
data['FAVC'] = data['FAVC'].map(f)

# update CAEC
c = {'no': 0, 'Sometimes': 1, 'Frequently': 2, 'Always': 3}
data['CAEC'] = data['CAEC'].map(c)

# update SMOKE
data['SMOKE'] = data['SMOKE'].map(f)

# update SCC
data['SCC'] = data['SCC'].map(f)

# update CALC
data['CALC'] = data['CALC'].map(c)

# update MTRANS
t = {'Public_Transportation': 0, 'Walking': 1, 'Automobile': 2, 'Motorbike': 3, 'Bike': 4}
data['MTRANS'] = data['MTRANS'].map(t)

# update NObeyesdad
o = {'Insufficient_Weight': 0, 'Normal_Weight': 1, 'Overweight_Level_I': 2, 'Overweight_Level_II': 3, 'Obesity_Type_I': 4, 'Obes:
data['NObeyesdad'] = data['NObeyesdad'].map(o)
```

Figure 8. Data Transformation Process

8. Modelling and Testing

Formation of classification models using Phyton programming in data processing. Phyton is an interactive, object-oriented programming language. It supports a variety of programming standards beyond object programming (Kumar et al., 2023). Data testing with data training and data testing. Data training is a process that forms a model based on a data set (Kurniawan, 2020).

9. Split Dataset in Features and Target Variable

```
#split dataset in features and target variable
feature_cols = ['Gender','Age','Height','Weight','family_history_with_overweight','FAVC','FCVC','NCP','CAEC','SMOKE','CH2O','SCC'
X = data[feature_cols] # Features
y = data.NObeyesdad # Target variable
```

Figure 9. Process Separating Datasets into Features and Variables

Figure 9 above is a process of separating datasets into features and variables, where the column NObeyesdad is used as the target variable and the other column as the features.

10. Split Dataset into Training Set and Test Set

```
# Split dataset into training set and test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1) # 70% training and 30% test
```

Figure 10. Process separating datasets into training and testing data

In Figure 10, the data to be used is divided into two: 70% for data training and 30% for data testing, so that 1,478 data will be used for model coding and 633 data for testing and validation.

11. Classification Algorithms

There are three classification algorithms to predict adult obesity factors: K-Nearest Neighbor, Decision Tree, and Naïve Bayes. The K-nearest neighbor (KNN)-algorithm forms a model not based on mining data but using all the data and processing it directly (Fajri et al., 2020). The KNN algorithm is excellent in handling noise, simple and easy to use for data processing on a large scale (Okfalisa et al., 2017).

```
from sklearn.metrics import mean_squared_error
from math import sqrt
train_preds = knn_model.predict(X_train)
mse = mean_squared_error(y_train, train_preds)
rmse = sqrt(mse)
rmse
```

Figure 11. Model Formation with K-Nearest Neighbor Method

Decision Tree is one of the most powerful and best methods of data excavation in prediction and classification, besides the decision tree embracing a little pre-processing so that it easily controls categorical features(Ferdowsy et al., 2021) (Ridwan, 2022).

```
# Create Decision Tree classifer object
clf = DecisionTreeClassifier()

# Train Decision Tree Classifer
clf = clf.fit(X_train,y_train)

#Predict the response for test dataset
y_pred = clf.predict(X_test)
```

Figure 12. Modelling with the Decision Tree Method

Naïve Bayes is a method for predicting past events and working with a characteristic assumption used to estimate values not dependent on the values being estimated, this algorithm for classifying a set of statistical data to predict all possibilities of each member of a class. (Amien et al., 2023) (Dirik, 2023).

```
from sklearn.naive_bayes import GaussianNB

# Build a Gaussian Classifier
model = GaussianNB()

# Model training
model.fit(X_train, y_train)

# Predict Output
# predicted = model.predict([X_test[6]])

# print("Actual Value:", y_test[6])
# print("Predicted Value:", predicted[0])
# Make predictions
y_pred = model.predict(X_test)
```

Figure 13. Model Formation by Naïve Bayes Method

12. Evaluate and Validation of Results

This classification model is validated using cross validation, while the accuracy of the three methods, both K-Nearest Neighbor, Decision Tree and Naïve Bayes, is measured using the Confusion Matrix. The Confusion matrix supports in the quality evaluation of the classification model and is represented by a matrix that allows in visualization the performance of each class of the predictive model, in addition it is useful to measure accuracy and test the algorithms performance level against the used datasets (Rodríguez et al., 2021)(Thamrin et al., 2021b)

RESULT AND DISCUSSION

From modeling using the algorithms of Decision Tree, Naïve Bayes, and KNN, the results are as follows:

1. Decision Tree

No. Feature	Feature	Score
3	Weight	9.909,08
1	Age	390,41
0	Gender	231,91
15	MTRANS	145,29
11	SCC	88,00
4	Family History With Overweight	77,24
12	FAF	56,05
8	CAEC	52,53
14	CALC	49,20
6	FCVC	43,92
9	SMOKE	25,18
7	NCP	21,54
5	FAVC	20,86
13	TUE	16,22
10	CH2O	13,69
2	Height	0,70

Figure 14. Obesity Risk Factor Detection with Decision Tree

Figure 14 shows that the 5 main risk factors causing obesity in adults based on the decision tree algorithm are Weight, Age, Gender, MTRANS, and SCC.

```
Decision Tree Accuracy: 0.9369085173501577

Confusion Matrix
[[ 90   2   0   0   0   0   0]
 [  3  66   8   0   0   0   0]
 [  0   6  79   2   2   0   0]
 [  0   0   1  78   6   0   0]
 [  0   0   0   3 108   3   0]
 [  0   0   0   0   3  81   1]
 [  0   0   0   0   0   0  92]]
              precision    recall  f1-score   support

           0       0.97      0.98      0.97        92
           1       0.89      0.86      0.87        77
           2       0.90      0.89      0.89        89
           3       0.94      0.92      0.93        85
           4       0.91      0.95      0.93       114
           5       0.96      0.95      0.96        85
           6       0.99      1.00      0.99        92

    accuracy                           0.94       634
   macro avg       0.94      0.93      0.94       634
weighted avg       0.94      0.94      0.94       634
```

Figure 15. Confusion Matrix Decision Tree

The application of the Decision Tree algorithm to this study obtained a 94% accuracy as seen in Figure 15.

2. Naïve Bayes

No. Feature	Feature	Score
3	Weight	102.620,31
1	Age	415,53
0	Gender	228,13
15	MTRANS	154,98
4	family_history_with_overweight	83,46
11	SCC	78,64
8	CAEC	50,63
14	CALC	46,74
6	FCVC	43,98
12	FAF	41,35
7	NCP	25,50
13	TUE	24,21
5	FAVC	22,40
9	SMOKE	18,08
10	CH2O	15,64
2	Height	0,80

Figure 16. Obesity Risk Factor Detection with Naïve Bayes

The results of detecting risk factors for obesity in adults using the Naïve Bayes algorithm show that Weight, Age, Gender, MTRANS, and Family History with Overweight are the 5 main risk factors for obesity in adults as seen in Figure 16.

```
Accuracy Naïve Bayes :  0.6009463722397477

Confusion Matrix
[[57  6  5  1  4  0  0]
 [25 37 10  5 13  0  1]
 [ 5  9 27  7 41  0  0]
 [ 0  6  5 12 64  5  0]
 [ 0  0  2  4 79 20  0]
 [ 0  0  0  1 13 75  0]
 [ 0  0  0  1  0  0 94]]
              precision    recall  f1-score   support

           0       0.66      0.78      0.71        73
           1       0.64      0.41      0.50        91
           2       0.55      0.30      0.39        89
           3       0.39      0.13      0.20        92
           4       0.37      0.75      0.50       105
           5       0.75      0.84      0.79        89
           6       0.99      0.99      0.99        95

    accuracy                           0.60       634
   macro avg       0.62      0.60      0.58       634
weighted avg       0.62      0.60      0.58       634
```

Figure 17. Confusion Matrix Naeïve Bayes

Figure 17 shows that the Naïve Bayes algorithm used in this study produces an accuracy of 60%.

3. K-Nearest Neighbor

No. Feature	Feature	Score
3	Weight	102.620,31
1	Age	415,53
0	Gender	228,13
15	MTRANS	154,98
4	family_history_with_overweight	83,46
11	SCC	78,64
8	CAEC	50,63
14	CALC	46,74
6	FCVC	43,98
12	FAF	41,35
7	NCP	25,50
13	TUE	24,21
5	FAVC	22,40
9	SMOKE	18,08
10	CH2O	15,64
2	Height	0,80

Figure 18. Obesity Risk Factor Detection With K-Nearest Neighbor

Like Naive Bayes, KNN also detects that Weight, Age, Gender, MTRANS, and Family History with Overweight are the main factors causing obesity in adults. This is shown in Figure 18.

```
K-Nearest Neighbors Accuracy: 0.7965299684542587

Confusion Matrix
[[68  4  1  0  0  0  0]
 [21 38 11 12  7  1  1]
 [ 2 10 65  4  7  1  0]
 [ 4  3 10 62 11  1  1]
 [ 1  0  4  3 90  6  1]
 [ 0  0  0  0  1 88  0]
 [ 0  1  0  0  0  0 94]]
              precision    recall  f1-score   support

           0       0.71      0.93      0.80        73
           1       0.68      0.42      0.52        91
           2       0.71      0.73      0.72        89
           3       0.77      0.67      0.72        92
           4       0.78      0.86      0.81       105
           5       0.91      0.99      0.95        89
           6       0.97      0.99      0.98        95

    accuracy                           0.80       634
   macro avg       0.79      0.80      0.79       634
weighted avg       0.79      0.80      0.79       634
```

Figure 19. Confusion Matrix K-Nearest Neighbor

The accuracy result of using the K-Nearest Neighbor algorithm on obesity datasets is shown in Figure 19, where this algorytm produces an accurate 80%.

This research uses accuracy, precision, recall, and F-1 to evaluate the results of the three methods. The results of research using the K-Nearest Neighbor algorithm, Decision Tree, Naïve Bayes, show that the Decision Tree algorithm is the most accurate algorithm, with an accuracy of 93.60%, precision 99.00%, recall 100.00%, and F-1 99.00%. In second place is the K-Nearest Neighbor algorithm with an accuracy of 79.60%, precision 97.00%, recall 99.00%, and F-1 98.00%, while the Naïve Bayes algorithm is in third place with an accuracy of 60.00%. precision of 99.00%, recall of 99.00%, and F-1 of 99.00%. The comparison results between the K-Nearest Neighbor, Decision Tree, Naïve Bayes methods are in Table 1 and visualized the graph in Figure 20.

Table 1. Comparison of Classification Results

Algorithm	Accuracy (%)	Precision (%)	Recalls (%)	F-1 (%)
K-Nearest Neighbor	79,6	97	99	98
Decision Tree	93,6	99	100	99
Naïve Bayes	60	99	99	99

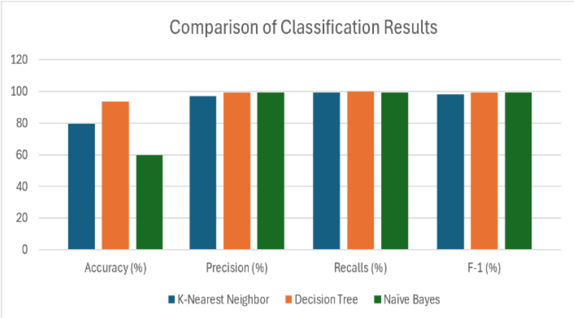

Figure 20. Comparison Graph of Classification Results.

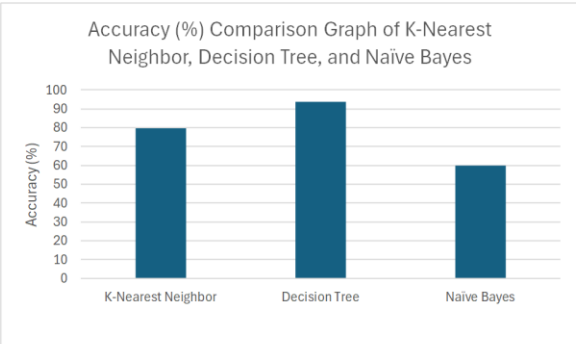

Figure 21. Accuracy (%) Comparison Graph of K-Nearest Neighbor, Decision Tree, and Naïve Bayes

Figure 21 shows a quite significant difference in accuracy, so it is clear that the accuracy of the Decision Tree is far superior to the other two algorithms.

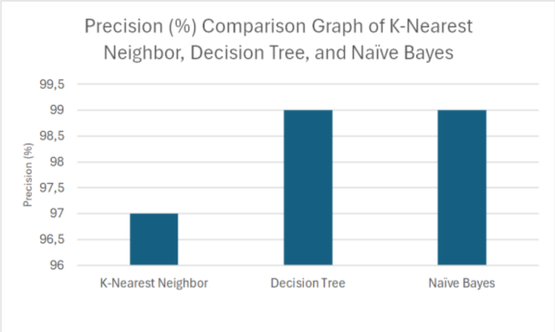

Figure 22. Precision (%) Comparison Graph of K-Nearest Neighbor, Decision Tree, and Naïve Bayes

In Figure 22 it can be seen that the Decision Tree and Naïve Bayes algorithms have the same high precision, 99%, its means that both algorithms succeeded in predicting 86% of the data correctly. This is different from the K-Nearest Neighbor algorithm which has a precision of 97%.

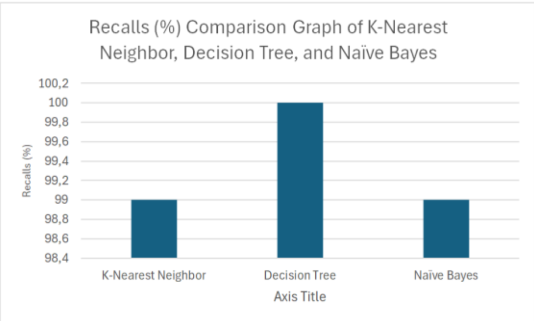

Figure 23. Recall (%) Comparison Graph of K-Nearest Neighbor, Decision Tree, and Naïve Bayes

Figure 23 displays the recall outcomes forthe three algorithms utilized in this study, with Decision Tree algorithm has the greatest recall value (100.00%), followed by Nearest Neighbor, Decision Tree and Naïve Bayes which have recall 99.00%. Recall shows how much the algorithm succeeds in predicting data that is labeled obesity.

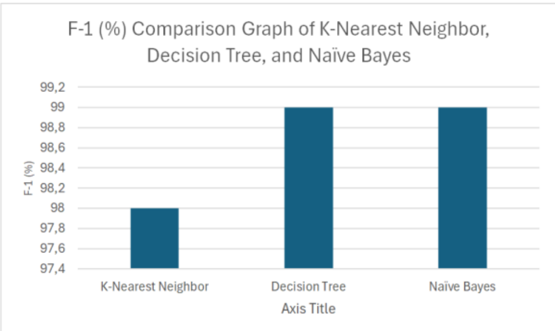

Figure 24. F-1 (%) Comparison Graph of K-Nearest Neighbor, Decision Tree, and Naïve Bayes

Figure 24, both Decision Tree, and Naïve Bayes both have an F1-Score of 99.00%, while the F1-Score of K-Nearest Neighbor is 98.00%.

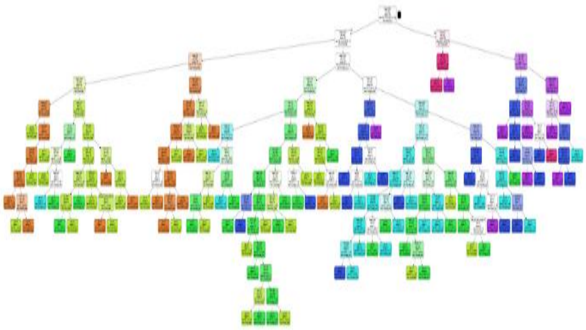

Figure 25. Model Shape Generated by Decision Tree Algorithm

CONCLUSION

This study compared the three algorithms K-Nearest Neighbor, Decision Tree, Naïve Bayes, which are widely used in the process of classification of data. The comparison results of the classification process obtained the highest accuracy of the three of these algorítms in a row: the decision tree algorytm with the greatest accuratity of 93.6%, K-nearest neighbor with the accuration of 79.6% and Naïva Bayes with the precision of 60% so that it could be concluded in the prediction of the risk of obesity in adult populations from the comparison of the third of the algorithms that decision tree has high interpretability and is able to work optimally so that gives accurate results. Based on a model made with the Decision Tree algorithm, it can be seen that the five (5) biggest risk factors for obesity in adults are weight, gender, age, family with a history of obese, and frequent consumption of high-calorie foods.

ACKNOWLEDGMENTS

The author thanks the Faculty of Engineering & Informatics, Bina Sarana Informatika University for funding the 2023 Foundation Lecturer Grant research through the Research & Community Service Information Base (BIPEMAS) with a research scheme in accordance with the assignment letter Number (156/6.02/UBSI/X/2023) dated October 18, 2023. So that this research can be useful for the community well.

REFERENCES

Akbar, R., Siroj, R. A., Win Afgani, M., & Islam Negeri Raden Fatah Palembang Abstract, U. (2023). Experimental Research Dalam Metodologi Pendidikan. *Jurnal Ilmiah Wahana Pendidikan*, *9*(2), 465–474. https://doi.org/10.5281/ZENODO.7579001

Alpiansah, A. B., & Ramdhani, Y. (2023). Optimasi Fitur dengan Forward Selection pada Estimasi Tingkat Obesitas menggunakan Random Forest. *SISTEMASI*, *12*(3), 860–873.

http://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/3125

Amien, I. L. F., Astuti, W., & Lhaksamana, K. M. (2023). Perbandingan Metode Naïve Bayes dan KNN (K-Nearest Neighbor) dalam Klasifikasi Penyakit Diabetes. *E-Proceeding of Engineering*, *10*(2), 1911–1920.

Asri, Y., Kuswardani, D., Horhoruw, L. F. M., & Ramadhana, S. A. (2024). *MACHINE LEARNING & DEEP LEARNING: Analisis Sentimen Menggunakan Ulasan Pengguna Aplikasi*. Uwais Inspirasi Indonesia. https://books.google.co.id/books?id=Yu7uEAAAQBAJ&newbks=0&printsec=frontcover&pg=PA20&dq=MACHINE+LEARNING&hl=id&source=newbks_fb&redir_esc=y#v=onepage&q=MACHINE LEARNING&f=false.

Dewi, A. M. S. I., & Dwidasmara, I. B. G. (2020). Implementation Of The K-Nearest Neighbor (KNN) Algorithm For Classification Of Obesity Levels. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, *9*(2), 277. https://doi.org/10.24843/jlk.2020.v09.i02.p15

Dirik, M. (2023). Application of machine learning techniques for obesity prediction: a comparative study. *Journal of Complexity in Health Sciences*. https://doi.org/10.21595/CHS.2023.23193

Fajri, M. S., Septian, N., & Sanjaya, E. (2020). Evaluasi Implementasi Algoritma Machine Learning K-Nearest Neighbors (kNN) pada Data Spektroskopi Gamma Resolusi Rendah. *Al-Fiziya: Journal of Materials Science, Geophysics, Instrumentation and Theoretical Physics*, *3*(1), 9–14. https://doi.org/10.15408/FIZIYA.V3I1.16180

Ferdowsy, F., Samsul Alam Rahi, K., Ismail Jabiullah, M., & Tarek Habib, M. (2021). A machine learning approach for obesity risk prediction. *Current Research in Behavioral Sciences*, *2*, 100053. https://doi.org/10.1016/j.crbeha.2021.100053

Kaur, R., Kumar, R., & Gupta, M. (2022). Predicting risk of obesity and meal planning to reduce the obese in adulthood using artificial intelligence. *Endocrine*, *78*(3), 458–469. https://doi.org/10.1007/S12020-022-03215-4

Kumar, M. S., Radhika, G., & Reddy, A. B. (2023). *Phyton Programming*. GCS Publishers.

Kurniawan, D. (2020). *Pengenalan Machine Learning Dengan Phyton Solusi Untuk Permasalahan Big Data*. PT. Elex Media Komputindo.

Nur Cuhadar, S., Karaduman, G., Uyanık, A., Durmaz, H., & Author, C. (2023). Performance Analysis of Machine Learning-Based Models for Early Diagnosis of Obesity Using Blood Test Parameters. *International Journal Of Engineering Science And Application S. N. Cuhadar et Al*, *7*(4).

Okfalisa, Gazalba, I., Mustakim, & Reza, N. G. I. (2017). Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification. *Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017*, *2018-January*, 294–298. https://doi.org/10.1109/ICITISEE.2017.8285514.

Ramadhan, Y., & Mandala, S. (2023). Analysis of Electrocardiogram Dynamic Features for Arrhythmia Classification. *Jurnal Online Informatika*, *8*(2), 204–212. https://doi.org/10.15575/JOIN.V8I2.1106

Ridwan, A. (2022). Penerapan Algoritma C4.5 Untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal Bisnis Digital Dan Sistem Informasi*, 41–48.

Rodríguez, E., Rodríguez, E., Nascimento, L., Da Silva, A., & Marins, F. (2021). Machine learning techniques to predict overweight or obesity. *4th International Conference on Informatics & Data Driven Medicine*.

Setiyani, L., Indahsari, A. N., & Roestam, R. (2023). Analisis Prediksi Level Obesitas Menggunakan Perbandingan Algoritma Machine Learning dan Deep Learning. *JTERA (Jurnal Teknologi Rekayasa)*, *8*(1), 139–146. https://doi.org/10.31544/jtera.v8.i1.2023.139-146

Sitanggang, D., & Sherly, S. (2022). Model Prediksi Obesitas dengan Menggunakan Support Vector Machine. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, *5*(2), 172–175. https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2443

Thamrin, S. A., Arsyad, D. S., Kuswanto, H., Lawi, A., & Nasir, S. (2021a). Predicting Obesity in Adults Using Machine Learning Techniques: An Analysis of Indonesian Basic Health Research 2018. *Frontiers in Nutrition*, *8*, 669155. https://doi.org/10.3389/FNUT.2021.669155/FULL

Thamrin, S. A., Arsyad, D. S., Kuswanto, H., Lawi, A., & Nasir, S. (2021b). Predicting Obesity in Adults Using Machine Learning Techniques: An Analysis of Indonesian Basic Health Research 2018. *Frontiers in Nutrition*, *8*, 669155. https://doi.org/10.3389/FNUT.2021.669155/BIBTEX

Toar, J., Telew, A., & Lumenta, G. (2023). Perbedaan tiga kategori aktivitas fisik pada status obesitas dan non obesitas. *Higeia Journal Of Public Health Research and Development*, *7*(3), 458–467.