JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 6, No. 1. December 2023 DOI: https://doi.org/10.34288/jri.v6i1.260
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

APPLICATION OF XGB CLASSIFIER FOR OBESITY RATE PREDICTION

Duwi Cahya Putri Buani[1], Nia Nuraeni[2*]

Informatika
Universitas Nusa Mandiri
duwi.dcp@nusamandiri.ac.id[-1] , nia.nee@nusamandiri.ac.id[-2*]
(*) Corresponding author

ABSTRACT

According to the Ministry of Health, the percentage of the population in Indonesia who are overweight is 13.5% for adults aged 18 years and over, while 28.7% are obese with BMI>=25 and obese with BMI>=27 as much as 15.4%. Meanwhile, at the age of children 5-12 years, 18.8% were overweight and 10.8% were obese. From these data, early detection of obesity levels is needed. From these data, prevention is needed so that the percentage of the population who experience obsediness can decrease, one of the efforts that can be done is to do early detection of obesity, to do early detection of obesity can be done using Machine Learning. In this study, it was discussed about the prediction of obestias levels using 7 (seven) models, namely Naive Bayes (NB), Random Forest (RF), K-NN, Decision Tree Classifier (DTC), SVM, XGB Classifier (XGB), Logistic Regression (LR) from the seven models used to predict the obesity level of XGB Classifier (XGB) which has the highest accuracy, namely Accurasy 0.96, with an f1-score of 0.96, Precission and recall 0.96.

Keywords: Obesity; XGB Classifier; CRIPS-DM

Abstract

Menurut Kementerian Kesehatan Prosentase jumlah penduduk di Indonesia yang mengalami kelebihan berat badan sebesar 13.5% untuk orang dewasa berusia 18 tahun ke atas, sementara 28.7% mengalami obesitas dengan IMT>=25 dan mengalami obesitas dengan IMT>=27 sebanyak 15.4%. Sementara itu pada usia anak 5-12 tahun 18.8% mengalami kelebihan berat badan dan 10.8% mengalam obesitas. Dari data tersebut maka sangat diperlukan deteksi dini tingkat obesitas. Dari data tersebut maka perlu adaya pencegahan agar prosentase penduduk yang mengalami obseitas dapat menurun, salah satu upaya yang dapat dilakukan adalah dengan melakukan deteksi dini obesitas, untuk melakukan deteksi dini obesitas dapat dilakukan dengan menggunakan Machine Learning. Dalam penelitian ini dibahas mengenai prediksi tingkat obestias dengan menggunakan 7 (tujuh) model yaitu Naive Bayes (NB), Random Forest (RF), K-NN, Decision Tree Classifier (DTC), SVM, XGB Classifier (XGB), Logistic Regression (LR) dari ketujuh model yang digunakan untuk prediksi tingkat obesitas XGB Classifier (XGB) yang memiliki akurasi paling tinggi yaitu Accurasy 0,96, dengan f1-score sebesar 0,96, Precission dan recall 0,96.

Kata kunci: obesitas; xgb classifier; CRIPS-DM

INTRODUCTION

Overweight or obesity is an abnormal or excessive accumulation of fat that threatens health. According to the WHO standard Body Mass Index (BMI) if it is greater than 25 then it is considered overweight, and if the Body Mass Index (BMI) is greater than 30 then it is considered obese (Fuadi & Meidian, 2017; James et al., 2020; Jiang et al., 2016; Setiyo Nugroho et al., 2018) . Obesity is already considered an epidemic, with more than 4 million deaths per year from being overweight or obese in 2017, in line with the global burden of disease (WHO, 2023)

Overweight and obesity are on the rise in adults and children. From 1975 to 2016, the prevalence of overweight or obese children and adolescents aged 5 to 19 worldwide more than quadrupled, from 4% to 18%.(Sherly & Sitanggang , 2022)

Obesity is one aspect of the malnutrition problem, with more people obese than underweight in all regions except Africa and Asia Once considered only a problem in high-income countries, overweight and obesity are now rising dramatically in low- and middle-income countries, especially in urban areas

Most overweight or obese children live in developing countries, where growth rates are 30%

higher than in developed countries. From the description above, it is necessary to have early detection of obesity so that the population in Indonesia and even in the world can prevent endemic obesity or obesity,

The degree of obesity can be determined based on eating habits and physical condition consists of several parameters that can be used to measure the accuracy of predictions of obesity rates (Setiyani Et Al. , 2023).

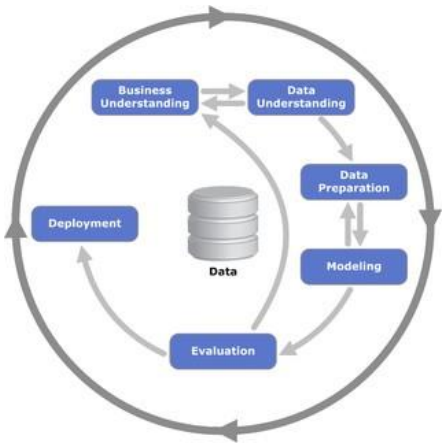
One way to do early detection is to predict obesity levels using Machine Learning Algorithms. Machine Learning is a method that is often used in research, Machine learning is used to improve the results of prediction or automatic detection (Hasanah & Nurhopipah, 2023; Roihan et al., 2019; Supriyadi et al., 2020; Yusa & Sindu, 2017)

In this study the author uses the XGB Classifier (XGB) algorithm, this study is expected to prevent the increase in the number of people who are obese by measuring and improving lifestyle. The limitation in this study is only up to finding the best model for early detection of obesity.

Previous research with Obesity Rate Prediction has been conducted using several algorithms including research using the KNN classification algorithm with an accuracy result of 79.96%, (Sibi & Widiarti , 2005) using a naïve bayer algorithm the accuracy results obtained 80%, using the SVM algorithm the accuracy results obtained 71.80%. From previous research, it can be seen that the accuracy results are still not optimal so it is necessary to use other models that can produce more accurate accuracy.(Muslehatin & Ibnu, 2017; Sitanggang & Sherly, 2022)

RESEARCH METHODS

In this study the author used the CRISP-DM method, the CRISP-DM method is still very capable of being used to conduct research(Kurniawan Et Al. , 2023). The stages of the CRISP-DM method include *Business* Understanding, *Data Understanding, Data Preparation, Modelling, Evaluation and Deployment*(Fadillah, 2015). The Flow Chart of the CRIP-DM Method can be seen in figure 1.



Source: (Ministry of Finance, 2022)
Figure 1. CRISP-DM Flow Chart

Here is an explanation from Figure 1:

a. *Business Understanding*
Business Understanding is the interpretation of business goals and requirements, which is then translated into knowledge to define key problems that can be addressed through *data mining* (Yudha, 2021).
b. *Data Understanding*
This stage is done by collecting data, evaluating data and checking data quality.(Fadillah, 2015)
c. *Data Preparation*
It is a stage to prepare the data from cleaning the data, removing the same data, deleting empty data and selecting attributes so that the data is ready for use. (Wurijanto Et al., 2022)
d. *Modeling*
Modeling is a stage where the application of mainning data models that have been determined according to data characteristics (Hananto, 2017)
e. *Evaluation*
In this stage what is done is to evaluate the models to be used and make sure there are no obstacles and run well.(Yudiana Et Al. , 2023)
f. *Deployment*
During the implementation phase, reports are generated regarding the results of all processed data, and the data is further developed and visualized.(Yudiana Et Al. , 2023)

RESULTS AND DISCUSSION

1. *Business Understanding*

According to the Ministry of Health, the percentage of the population in Indonesia who are overweight is 13.5% for adults aged 18 years and

over, while 28.7% are obese with BMI>=25 and obese with BMI>=27 as much as 15.4%. Meanwhile, at the age of children 5-12 years, 18.8% were overweight and 10.8% were obese. From these data, early detection of obesity levels is needed.

2. *Data Understanding*

The data used is public data which is an estimate of obesity rates in people from Mexico, Peru and Colombia, with ages between 14 and 61 years and diverse eating habits and physical conditions. Then the information is processed so that 17 attributes are obtained, the data amounts to 2111 records. The following is an explanation of the attributes in Table 1.

Table 1. Attributes from obesity data

Attributes	Information
Gender	Respondent's Gender
Age	Age of Respondents
Height	Respondent's Height
Weight	Respondent's Weight
family_history_with_overweight	Family History of Being Overweight
FAVC	Frequent consumption of high caloric food
FCVC	Frequency of consumption of vegetables
NCP	Number of main meals
CAEC	Consumption of food between meals
SMOKE	Smoke
CH2O	Consumption of water daily
SCC	Calories consumption monitoring
FAF	Physical activity frequency
TUE	Time using technology devices
CALC	Consumption of alcohol
MTRANS	Transportation used
NObeyesdad	(Obesity Rate) – with data classification using insufficient weight, normal weight, overweight level I, overweight level II, Type I obesity, Type II obesity, type II obesity, and type III obesity.

Source: (Palechor & Manotas, 2019)

Of the 17 attributes, there is 1 label, namely NObeyesdad and there are 7 classifications of which can be seen in figure 2.

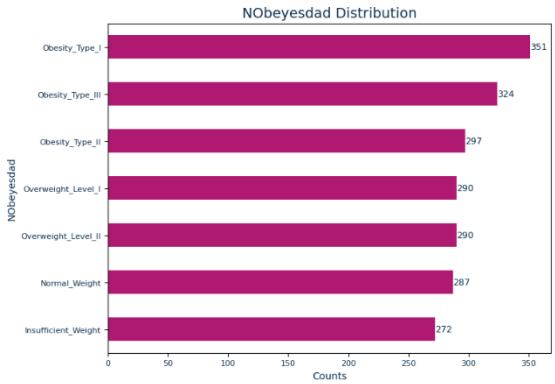


Figure 2. Number of NObeyesdad attribute data

In figure 2 of the NObeyesdad attribute there are 7 classes including Obesity Type 1 there are 351 data, ObesityType 3 there are 324, ObesityType 3 is 297, Overweight Level I is 290, Overweight Level 2 is 290 data, Normal Weight is 287 data and Insufficient Weight is 272 data.

3. *Data Preparation*

From the data obtained, analysis is carried out by checking the same data, blank data, and transforming the data and dividing the data into testing and trending data by dividing 70% of testing data and 30% of traning data so that the data used is trending 1477 and data for testing 634.

4. *Modeling*

The modeling process is carried out by making several models to get the best model, the algorithm used in this study is:

a. *Naive Bayes (NB)*

This method is believed to be useful in providing decision-making recommendations. (Silvana Et Al. , 2020)

b. Random Forest (RF)

Random Forest (RF) is a method that can improve the accuracy of results when generating attributes randomly for each node.(Nusrang , 2022)

c. K-NN

The KNN algorithm is a method of classifying data based on the shortest distance to the data object.(Cholil Et Al. , 2021)

d. Decision Tree Classifier (DTC)

C45 or also called Decision Tree Algorithm is an algorithm that can be used to form decision trees, decision trees are one method that is very easy to interpret by humans.(Nasrullah, 2021)

e. SVM

*Support Vector Machine (*SVM) is a technique that can separate two data sets from two different classes by maximizing the limits of the separator function (*hyperplane).*(Dasmasela Et al., 2022).

f. XGB Classifier (XGB)

XGBoost or *Extreme Gradient Boosting* is a decision tree learning-based classification method that applies ensemble enhancement techniques. The classification has several methods, one of which is the eXtreme Gradient Boosting (XGBoost) method. In the use of several examples of research using classification techniques, these calculations can produce excellent precision .(Rizky Mubarok Et Al. , 2022)(Yulianti Et al., 2022)

g. Logistic Regression (LR)

Logistic regression is a method of statistical analysis used to describe the relationship between a dependent variable with two or more categories and one or more independent variables on a categorical scale.(Featured Et Al. , 2017), 7 (Seven) The model used to conduct research is a classification algorithm that has the ability to make predictions, the author wants to test the algorithm or model that has the best accuracy. In this study Testing was carried out using a 10-fold cross-validation technique for NB, RF, K-NN, DT, SVM and XGB classification algorithms resulting in matrix confusion, matrix confusion is a method that can be used to measure the performance of classification algorithms. The following is the comparison result of the 7 algorithms used can be seen in figure 3.

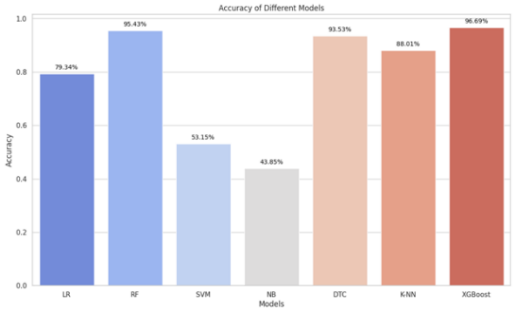


Figure 3. Model Accuracy Results

Figure 3 is a graph of the accuracy results obtained after the algorithm is applied to obesity data the results are NB has an accuracy of 43.85%, RF has an accuracy of 95.43%, K-NN has an accuracy of 88.01%, DTC has an accuracy of 93.53%, SVM has an accuracy of 53.15%, XGBoost has an accuracy of 96.69% and LR has an accuracy of 79.34% of these results, the model that has the highest accuracy is XGBoost.

5. Evaluation

The fifth stage of the CRISP-DM method is Evaluation, evaluation is carried out by looking at the Accuracy, Precission, Recall, and F-1 Score values. Here's how the Accuracy, Precission, Recall, and F-1 Score are calculated:

$$\text{Accuracy} = \frac{TP}{Jumlah\ Data} \quad \text{.....(1)}$$

$$\text{Precission} = \frac{TP}{TP+FP} \quad \text{.....(2)}$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{.....(3)}$$

$$\text{F} - 1\ \text{Score} = \frac{2 * (Recall * Precission)}{(Recall + Precission)} \quad \text{.....(4)}$$

Because the data used is multiclass data, there are only True Positive (TP), Positive Falsh (FP) and Negative Falsh (FN), which can be used to calculate Accuracy, Precission, Recall and F1-Score. The following are the evaluation results of the 7 (Seven) Models used:

Table 2. Model Evaluation Results

Algo	TP	FN	FP	Accry	F1-S	Prec	Rcall
NB	278	356	356	0,44	0,44	0,22	0,44
RF	600	34	34	0,95	0,95	0,95	0,95
K-NN	558	76	76	0,88	0,88	0,88	0,88
DTC	596	38	38	0,94	0,94	0,94	0,94
SVM	337	297	297	0,53	0,53	0,53	0,53
LR	503	131	131	0,79	0,79	0,79	0,79
XGB	**606**	**28**	**28**	**0,96**	**0,96**	**0,96**	**0,96**

Table 2 shows that the Model Using the XGBoost Algorithm has the highest accuracy value where True Positive (TP) with 606, Falsh Negative (FN) of 28, Positive Falsh of 28 Accurassy 0.96, f1-score of 0.96, Precission and recall 0.96. Thus it can be concluded that the Model using XGBoost has better results.

CONCLUSIONS

This study uses the Method CRISP-DM to determine the best model in predicting obesity levels, the CRISP-DM Model has a very clear flow or step so that the results of the study are valid. In this study, the author used 7 (seven) models to predict the level of obesity from 7 (seven) models obtained as follows: Naive bayes with True Positive (TP) with 278, Falsh Negative (FN) with 356, Falsh Positive with 356 Accurasy 0.44, f1-score 0.44, Precission and recall 0.44, Random Forest with True Positive (TP) with 600, Falsh Negative (FN) with 34, Positive Falsh 34 Accurasy 0.95, f1-score of 0.95, Precission and recall 0.95, K-NN with True Positive (TP) with 558, Falsh Negative (FN) with 76, Falsh Positive with 76 Accurasy 0.88, f1-score 0.88, Precission and recall 0.88, Decision Tree Classifier with True Positive (TP) with 596, Falsh Negative (FN) with 38, Positive Falsh 38 Accurasy 0.94, f1-score of 0.94, Precission and recall 0.94, SVM with True Positive

(TP) with 337, Falsh Negative (FN) of 297, Falsh Positive of 297 Accurasy 0.53, f1-score of 0.53, Precission and recall 0.53, Logistic Regression with True Positive (TP) with 503, Falsh Negative (FN) of 131, Falsh Positive of 131 Accurasy 0.79, f1-score of 0.79, Precission and recall 0.79, and XGBoost with the highest accuracy value where True Positive (TP) with 606, Falsh Negative (FN) of 28, Positive Falsh of 28 Accurasy 0.96, f1-score of 0.96, Precission and recall 0.96. Thus the best model of the study is XGBoost or XGB Classifier.

REFERENCES

Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, *6*(2), 118–127. https://ejournal.bsi.ac.id/ejurnal/index.php/ijcit/article/view/10438

Dasmasela, R., Tomasouw, B. P., & Leleury, Z. A. (2022). *Penerapan Metode Support Vector Machine (SVM) Untuk Mendeteksi Penyalahgunaan Narkoba*. Jurnal Matematika, Statistik Dan Terapannya. https://ojs3.unpatti.ac.id/index.php/parameter/article/view/3630/2847

Fadillah, A. P. (2015). Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ). *Jurnal Teknik Informatika Dan Sistem Informasi*, *1*(3), 2443–2229. https://media.neliti.com/media/publications/133327-ID-penerapan-metode-crisp-dm-untuk-prediksi.pdf

Fuadi, D. F., & Meidian, A. C. (2017). Pengaruh Obesitas Terhadap Kebugaran Anak Sekolah Dasar Usia 10-12 Tahun. *Fisioterapi : Jurnal Ilmiah Fisioterapi*, *17*(2), 78–83. https://doi.org/10.47007/FISIO.V17I2.2216

Hananto, V. R. (2017). *Analisis Penentuan Metode Data Mining Untuk Prediksi Kelulusan Mahasiswa Sebagai Penunjang Angka Efisiensi Edukasi.*

Hasanah, U., & Nurhopipah, A. (2023). Prediksi Persentase Body Fat Menggunakan Algoritma CART dan M5'. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, *4*(4), 351–363. https://doi.org/10.35746/JTIM.V4I4.316

James, S. L., Castle, C. D., Dingels, Z. V., Fox, J. T., Hamilton, E. B., Liu, Z., Roberts, N. L. S., Sylte, D. O., Bertolacci, G. J., Cunningham, M., Henry, N. J., Legrand, K. E., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abedi, A., Abegaz, K. H., Abosetugn, A. E., Abushouk, A. I., … Vos, T. (2020). Estimating global injuries morbidity and mortality: methods and data used in the Global Burden of Disease 2017 study. *Injury Prevention : Journal of the International Society for Child and Adolescent Injury Prevention*, *26*(Supp 1), I125–I153. https://doi.org/10.1136/INJURYPREV-2019-043531

Jiang, S. Z., Lu, W., Zong, X. F., Ruan, H. Y., & Liu, Y. (2016). Obesity and hypertension. *Experimental and Therapeutic Medicine*, *12*(4), 2395. https://doi.org/10.3892/ETM.2016.3667

Kemenkeu. (2022). *CRISP DM Sebagai Salah Satu Standard untuk Menghasilkan Data Driven Decision Making yang Berkualitas*. https://www.djkn.kemenkeu.go.id/artikel/baca/15134/CRISP-DM-Sebagai-Salah-Satu-Standard-untuk-Menghasilkan-Data-Driven-Decision-Making-yang-Berkualitas.html

Kurniawan, I., Buani, D. C. P., Abdussomad, A., Apriliah, W., & Saputra, R. A. (2023). Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *10*(2), 421–428. https://doi.org/10.25126/JTIIK.20231026225

Muslehatin, W., & Ibnu, M. (2017). *Penerapan Naïve Bayes Classification untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau*. 2579–5406. https://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/3276

Nasrullah, A. H. (2021). *Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris*. *7*(2). http://ejournal.fikom-unasman.ac.id

Nusrang, M. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, *4*(3), 121–127. https://doi.org/10.35580/VARIANSIUNM31

Palechor, F. M., & Manotas, A. de la H. (2019). Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. *Data in Brief*, *25*, 104344. https://doi.org/10.1016/J.DIB.2019.104344

Rizky Mubarok, M., Herteno, R., Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lambung Mangkurat Jalan Ahmad Yani Km, I., & Selatan, K. (2022). Hyper-

Parameter Tuning Pada Xgboost Untuk Prediksi Keberlangsungan Hidup Pasien Gagal Jantung. *KLIK - KUMPULAN JURNAL ILMU KOMPUTER*, *9*(2), 391–401. http://klik.ulm.ac.id/index.php/klik/article/view/484

Roihan, A., Abas Sunarya, P., & Rafika, A. S. (2019). IJCIT (Indonesian Journal on Computer and Information Technology) Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. In *IJCIT (Indonesian Journal on Computer and Information Technology)* (Vol. 5, Issue 1). http://ejournal.bsi.ac.id/ejurnal/index.php/ijcit

Setiyani, L., Indahsari, A. N., Roestam, R., 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.I1.2022.139-146

Setiyo Nugroho, P., Catur Wijayanti, A., Studi Kesehatan Masyarakat, P., & Kapuas Raya, Stik. (2018). Body Mass Index Dan Kaitannya Dengan Diabetes Melitus Pada Umur > 15 Tahun Di Indonesia, Studi Data Indonesian Family Life Survey V. *An-Nadaa: Jurnal Kesehatan Masyarakat (e-Journal)*, *5*(1), 7–10. https://doi.org/10.31602/ANN.V5I1.1643

Sherly, & Sitanggang, D. (2022). *Model Prediksi Obesitas dengan Menggunakan Support Vector Machine*. http://jurnal.unprimdn.ac.id/index.php/JUSIKOM/article/view/2443/1497

Sibi, S. Y., & Widiarti, A. R. (2005). *Klasifikasi Tingkat Obesitas Mempergunakan Algoritma KNN*.

Silvana, M., Akbar, R., Syahnum, A., & Informasi, J. S. (2020). Pemanfaatan Metode Naïve Bayes dalam Implementasi Sistem Pakar Untuk Menganalisis Gangguan Perkembangan Anak. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *6*(2), 74–81. https://doi.org/10.25077/TEKNOSI.v6i2.2020.74-81

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

Supriyadi, R., Gata, W., Maulidah, N., Fauzi, A., Komputer, I., & Nusa Mandiri Jalan Margonda Raya No, S. (2020). *Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah*. *13*(2), 67–75. http://journal.stekom.ac.id/index.php/E-Bisnis■page67

Tampil, Y., Hanny, K., & Langi, Y. (2017). *Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado*. https://ejournal.unsrat.ac.id/index.php/decartesian/article/view/17023/16560

WHO. (2023). *Obesity*. https://www.who.int/health-topics/obesity#tab=tab_1

Wurijanto, T., Setiawan, H. B., & Tjandrarini, A. B. (2022). Penerapan Model CRISP-DM pada Prediksi Nasabah Kredit yang Berisiko Menggunakan Algoritma Support Vector Machine. *Jurnal Ilmiah Scroll (Jendela Teknologi Informasi)*, *10*(1), 1–6. https://doi.org/10.30640/EJOURNALSCROLL.V10I1.291

Yudha, M. A. (2021). *CRISP-DM, Pendekatan Proses dalam Data Mining | by Muhammad Andi Yudha | Medium*. Medium.Com. https://andiyudha.medium.com/crisp-dm-pendekatan-proses-dalam-data-mining-68bf8c2dc908

Yudiana, Y., Yulia Agustina, A., & Nur Khofifah, dan. (2023). *Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan*. *8*(1), 1–20. http://e-journal.lp2m.uinjambi.ac.id/ojp/index.php/ijoieb

Yulianti, E. H., Soesanto, O., & Sukmawaty, Y. (2022). Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit. *JOMTA Journal of Mathematics: Theory and Applications*, *4*(1), 21–26.

Yusa, M., & Sindu, W. (2017). Evaluasi Model Decision Tree C4.5 Guna Prediksi Posibilitas Resiko Obesitas. *Seminar Nasional Informatika (SNIf)*, *1*(1), 147–152. https://e-journal.potensi-utama.ac.id/ojs/index.php/SNIf/article/view/257