

Implementasi Algoritma Klasifikasi dengan Teknik *Discretization* dan *Bagging* untuk Meningkatkan Akurasi Prediksi Penyakit Stroke

Annisa Maulana Majid¹, Ismasari Nawangsih²

¹Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia, 17532
e-mail: ¹annisa.maulanamajid@pelitabangsa.ac.id

e-mail: ²ismasari.n@pelitabangsa.ac.id

Submitted Date: 2023-06-07
Revised Date: 2023-06-26

Reviewed Date: 2023-06-08
Accepted Date: 2023-06-29

Abstract

Stroke is a disease that causes death with a prevalence rate of 10.9% based on the 2018 RI Basic Health Research Results. Early diagnosis is needed to accurately predict stroke symptoms as an effort to prevent stroke. Research related to the prediction of stroke has been carried out using several methods but produces a level of accuracy that is less than optimal in the classification algorithm, so efforts to increase accuracy are needed to produce accurate information in predicting stroke. The purpose of this study is to implement a classification algorithm by applying discretization and bagging techniques to improve stroke prediction. The results obtained from this study are that the Naïve Bayes classification algorithm by combining discretization and bagging techniques has a higher level of accuracy with an accuracy of 95.21%, an increase of 7.36% compared to using only a single algorithm..

Keywords: Stroke; Classification Algorithm; Discretization; Bagging

Abstrakk

Penyakit stroke merupakan salah satu dari penyakit penyebab kematian dengan tingkat pravelensi 10,9% berdasarkan Hasil Riset Kesehatan Dasar RI tahun 2018. Perlu adanya diagnosa sejak dini untuk memprediksi gejala penyakit stroke yang akurat sebagai upaya pencegahan penyakit stoke . Penelitian yang terkait prediksi penyakit stroke telah dilakukan dengan beberapa metode namun menghasilkan tingkat akurasi yang kurang maksimal pada algoritma klasifikasi, sehingga diperlukan adanya upaya peningkatan akurasi untuk menghasilkan informasi yang akurat dalam memprediksi penyakit stroke. Tujuan dari penelitian ini yaitu melakukan implementasi algoritma klasifikasi dengan menerapkan teknik *discretization* dan *Bagging* dalam meningkatkan prediksi penyakit stroke. Hasil yang diperoleh dari penelitian ini adalah algoritma klasifikasi *Naïve Bayes* dengan menggabungkan teknik *discretization* dan *Bagging* memiliki tingkat akurasi lebih tinggi dengan akurasi sebesar 95.21%, meningkat sebesar 7.36% dari pada hanya menggunakan algoritma tunggal saja.

Keywords: Penyakit Stroke; Algoritma Klasifikasi; Discretization; Bagging

1. Pendahuluan

Teknik *data mining* telah diterapkan di berbagai macam bidang, salah satunya pada bidang kesehatan. Saat ini teknik *data mining* dapat digunakan untuk memprediksi penyakit yang mematikan untuk upaya mencegah tingginya tingkat kematian karena adanya prediksi dini terhadap penyakit tersebut. Salah satu penyakit mematikan yaitu penyakit stroke. Berdasarkan dari Hasil Riset Kesehatan Dasar yang dilaporkan pada

tahun 2018 bahwa Penyakit Stroke memiliki tingkat prevalensi berdasarkan diagnosis dokter sebesar 10.9% (Kemenkes RI, 2018). Dalam hal ini maka diperlukan adanya prediksi dini untuk upaya pencegahan penyakit stroke. Pentingnya prediksi dini terhadap gejala penyakit stoke akan menjadi pencegahan awal penyakit stroke. Prediksi penyakit stroke akan mempermudah tenaga medis juga dalam mendiagnosa pasien secara dini dan memberikan upaya pencegahan penyakir stroke.



Penelitian yang membahas tentang diagnosa atau prediksi penyakit stroke telah dilakukan dengan menggunakan berbagai macam algoritma klasifikasi, di antaranya algoritma *Support Vector Machine* (SVM), *Decision Tree*, *Naïve Bayes*, *Random Forest*, *K-nearest Neighbor* (KNN) dan lain-lain. Penelitian Nitish Biswas dkk, menghasilkan akurasi sebesar 96.90 % pada algoritma *Decision Tree*, 74.77% pada algoritma *Naïve Bayes*, 98.82% pada algoritma KNN, (Biswas et al., 2022). Penelitian oleh Soumyabrata Dev dkk, menghasilkan akurasi sebesar 74% pada algoritma *Decision Tree*, 77% pada algoritma *Neural Network*, 68% algoritma SVM (Dev et al., 2022). Penelitian oleh Hager Saleh dkk, menghasilkan akurasi sebesar 79% pada algoritma *Decision Tree*, 77% akurasi pada algoritma SVM dan *Logistic Regression* (Ahmed et al., 2019).

Namun hasil pada penelitian sebelumnya menunjukkan bahwa penerapan algoritma klasifikasi perlu adanya peningkatan pada akurasi dan penanganan atribut kontinyu pada dataset untuk menghasilkan data akurasi yang paling terbaik.

Bagging (bootstrap agregating) merupakan salah satu dari metode *ensemble* yang bertujuan untuk meningkatkan sebuah akurasi pada klasifikasi yang tidak stabil, *Bagging* dinilai peka terhadap perubahan kecil dalam data pelatihan sehingga dapat meningkatkan akurasi algoritma klasifikasi (Wu et al., 2020). Teknik *discretization* merupakan salah satu jenis teknik data *preprocessing* untuk mengubah atribut numerik kontinyu menjadi interval (Peker & Kubat, 2021).

Penelitian ini menggunakan tiga algoritma klasifikasi yaitu *Decision Tree*, KNN, dan *Naïve Bayes* dengan menerapkan teknik *Bagging* untuk meningkatkan akurasi dari algoritma klasifikasi dan teknik *discretization* untuk penanganan atribut kontinyu dalam memprediksi penyakit stroke.

2. Tinjauan Pustaka

Tinjauan Pustaka merupakan pembahasan materi yang terkait pada penelitian ini. Adapun materi-materi sebagai berikut:

a. Data Mining

Data Mining merupakan suatu proses pencarian pola dari kumpulan data yang besar untuk menemukan suatu informasi sehingga dapat digunakan dalam mengambil sebuah keputusan, (Osman, 2019). *Data mining* yaitu proses penambangan data pada suatu database besar untuk mengekstraksi informasi yang berguna, (Sudarsono

et al., 2021). *Data mining* adalah teknik yang dapat mengekstraksi pola bersembunyi pada *big data*, (Derisma, 2020).

b. Decision Tree

Decision Tree yaitu salah satu jenis algoritma klasifikasi pengambilan keputusan per bentuk pohon untuk mengembangkan strategi dalam mencapai suatu tujuan, (Moshkov, 2022). *Decision Tree* adalah metode klasifikasi yang terdiri dari node dan simpul yang diimplementasikan untuk mengambil sebuah keputusan dengan pemilihan atribut, (Panhalkar & Doye, 2022). *Decision tree* bagian dari algoritma *supervised learning* yang digunakan membangun model pelatihan untuk memprediksi suatu kelas atau nilai melalui data pelatihan, (Charbuty & Abdulazeez, 2021).

c. K-Nearest Neighbor (KNN)

Algoritma KNN adalah algoritma yang menggeneralisasi sesuai dengan aturan tetangga terdekat yang dapat diartikan sebagai k-sample yang diuji dengan kemiripan terdekat, (Xing & Bei, 2020). KNN merupakan algoritma sederhana dengan mengklasifikasikan gambar menggunakan data pelatihan terdekat, titik data ditentukan pada k terdekat yang diberikan tugas untuk klasifikasi ke titik kueri yang belum berlabel, (Srinivas & Sasibhushana Rao, 2019). KNN menghasilkan keputusan yang sangat fleksibel bersifat nonparametrik dengan tidak memberikan asumsi secara distribusi, klasifikasi objek yang berdasarkan pada data pelatihan yang telah memiliki jarak paling terdekat terhadap objek tersebut, (Yulianto Anggi Priliani & Darwis, 2021)

d. Naïve Bayes

Naïve bayes merupakan algoritma klasifikasi yang sederhana termasuk dalam kelompok pengklasifikasi probabilitas yang didasarkan teorema bayes, (Bhatia & Malhotra, 2021). Algoritma *Naïve Bayes* terkenal sebagai metode dengan pengklasifikasian lebih baik dari algoritma lain karena independensi sangat kuat, model yang dibuat sederhana, dan dapat diimplementasikan dengan jumlah dataset besar, (Salmi & Rustam, 2019). *Naïve bayes* merupakan algoritma klasifikasi yang memiliki tujuan utama yaitu menemukan pemetaan terbaik antara sebagian data baru dan satu set data dengan masalah tertentu, (Yang, 2018).

e. Teknik *Discretization*

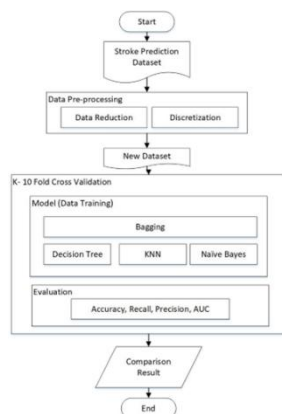
Teknik *discretization* merupakan proses partisi suatu data dengan atribut kontinu menjadi atribut diskrit dengan mengasosiasikannya ke nilai-nilai dengan kelompok interval, (García-Gil et al., 2018). *Discretization* merupakan suatu proses sebelum pemodelan algoritma dilakukan dengan mengubah nilai atribut kontinu menjadi interval dengan tipr nilai diskrit, (Saleh & Nasari, 2018). *Discretization* merupakan metode preprocessing yang digunakan untuk mengurangi jumlah nilai atribut kontinu dengan mengubah menjadi interval yang bernilai nominal, (Hacibeyoglu & Ibrahim, 2018).

f. *Bootstrap Aggregating (Bagging)*

Bootstrap Aggregating yang biasa disebut *Bagging* merupakan salah satu teknik *ensemble* yang menggunakan setiap proses model dalam vote yang sama besar. Cara kerja *Bagging* dengan menggabungkan semua hasil prediksi dengan voting dan tiap model memiliki bobot yang sama rata, (Kabari & Onwuka, 2019). *Bagging* merupakan algoritma ansemble yang efektif meningkatkan kinerja model dan dapat mengatasi masalah *overfitting*, (Lee et al., 2018). *Bagging* menghasilkan prediksi agregat yang diperoleh dengan aturan suara terbanyak dari setiap subruang yang dibangun, (Tuysuzoglu & Birant, 2020).

3. Metode Penelitian

Penelitian menggunakan algoritma klasifikasi yang dikombinasikan dengan teknik *discretization* dan metode *Bagging* untuk meningkatkan akurasi prediksi penyakit stroke. Algoritma klasifikasi dalam penelitian menggunakan *Decision Tree*, KNN, dan *Naïve Bayes*. Berikut merupakan tahapan-tahapan dalam penelitian:



Gambar 1. Tahapan penelitian Algoritma Klasifikasi dengan Teknik *Discretization* dan

Bagging untuk Meningkatkan Akurasi Prediksi Penyakit Stroke

a. *Dataset*

Dataset yang akan digunakan pada penelitian ini yaitu *stroke prediction dataset* bersumber dari kaggle.com dengan jumlah data sebanyak 5.110 data. *Dataset* stroke prediction terdiri dari 12 atribut, antara lain yaitu: *id*, *Gender*, *Age*, *Hypertension*, *Heart disease*, *Ever married*, *Work type*, *Residence type*, *Avg glucose level*, *body mass index (BMI)*, *Smoking status* dan *Stroke* sebagai Label.

b. *Data Pre-processing*

Penelitian ini melalui tahapan *data pre-processing* dalam mengolah *dataset* yang bertujuan untuk memaksimalkan tingkat akurasi pada prediksi penyakit stroke. *Pre-processing* data yang dilakukan yaitu *data reduction* dan teknik *discretization*. *Data reduction* adalah teknik *pre-processing* dengan cara mengurangi jumlah data tanpa mengubah integritas data, dapat dilakukan dengan pengurangan jumlah atribut namun tetap memilih atribut-atribut yang penting dalam proses *data mining*, (Alasadi & Bhaya, 2017). *Data Reduction* bertujuan untuk penyederhanaan data tanpa mengubah integritas data dan kualitas informasi yang didapatkan, (Virantika et al., 2022). Menurut (Dev et al., 2022) ada 4 faktor yang paling penting untuk mendeteksi penyakit stoke antara lain, usia, penyakit jantung, kadar glukosa rata-rata, dan hipertensi. Pada penelitian ini menggunakan 4 atribut penting untuk menghitung tingkat akurasi penyakit stroke yaitu *age*, *heart disease*, *average glucose level* dan *hypertension*. Teknik *discretization* juga dilakukan dalam penelitian ini untuk mengubah atribur kontinu kedalam nilai diskrit yang berbentuk interval

c. Validasi dan Evaluasi Data

Proses validasi yang dilakukan pada penelitian ini menggunakan *k-fold cross validation* dengan nilai k yaitu 10. *K-fold cross validation* merupakan teknik pengujian dilakukan dengan membagi data menjadi 2 bagian yaitu *data training* dan *data testing* yang dilakukan secara acak. *k-fold cross validation* akan mengulang proses sebanyak k-kali yang digunakan untuk membagi data ke sebuah himpunan seara acak, (Riyanto, 2019). *Dataset* akan dibagi menjadi 10 bagian dengan ketentuan 9/10 sebagai jenis *data training* dan 1/10 sebagai *data testing* kemudian iterasi dilakukan

berulang hingga 10 kali dengan kombinasi sampai semua *record* data mendapatkan bagian untuk menjadi *data testing*, (Buntoro, 2017). Alur kerja *k-fold cross validation* dimulai dari keseluruhan total *dataset* dibagi menjadi 10 bagian, bagian pertama menjadi iterasi/*fold* ke 1 sebagai *data testing*, 9 bagian lainnya menjadi *data training* kemudian menghitung akurasi, begitupun dengan *fold* ke 2, bagian kedua menjadi *data testing* dan bagian lainnya menjadi *data training*, proses tersebut akan terus menerus berulang sampai mencapai *fold* ke 10. Berikut adalah ilustrasi dari proses *k-fold cross validation*.

Pengujian	Dataset									
fold-1	Test	Train	Train	Train	Train	Train	Train	Train	Train	Train
fold-2	Train	Test	Train	Train	Train	Train	Train	Train	Train	Train
fold-3	Train	Train	Test	Train	Train	Train	Train	Train	Train	Train
fold-4	Train	Train	Train	Test	Train	Train	Train	Train	Train	Train
fold-5	Train	Train	Train	Train	Test	Train	Train	Train	Train	Train
fold-6	Train	Train	Train	Train	Train	Test	Train	Train	Train	Train
fold-7	Train	Train	Train	Train	Train	Train	Test	Train	Train	Train
fold-8	Train	Train	Train	Train	Train	Train	Train	Test	Train	Train
fold-9	Train	Train	Train	Train	Train	Train	Train	Train	Test	Train
fold-10	Train	Train	Train	Train	Train	Train	Train	Train	Train	Test

Gambar 2. Simulasi validasi dataset menggunakan 10-Fold Cross Falidation

Pada proses validasi terdapat model yaitu penerapan metode *Bagging* dan algoritma klasifikasi *Decision tree*, KNN, dan *Naïve Bayes*. Selanjutnya proses evaluasi menggunakan confusion table dan *ROC curve*. *Confusion table* digunakan untuk menganalisa tingkat seberapa baiknya klasifikasi terhadap *tuple* dari berbeda kelas, (Riyanto, 2019). Berikut adalah tabel *confusion*:

Tabel 1. Confusion Table

Confusion Table		Prediction	
		Yes	No
Actual	yes	TP	FN
	no	FP	TN

Keterangan tabel:

TP = True Positive TN = True Negative
 FP = False Positive FN = False Negative

Hasil dari *confusion table* berupa *accuracy*, *recall*, dan *precision*. *Accuracy* adalah perbandingan dari total prediksi benar dari total semua data. Rumus *Accuracy* adalah sebagai berikut:

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

Recall adalah persentase nilai benar (*true positive*) terhadap total keseluruhan nilai yang benar positif. Rumus *recall* sebagai berikut:

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

Precision merupakan persentase antara true positive dengan total data yang positif. Rumus *precision* sebagai berikut:

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

Receiver Operating Characteristics atau singkatan dari ROC akan menunjukkan akurasi secara visual berbentuk grafik. Kurva ROC akan menghasilkan gambaran nilai *Area Under Curve* (AUC). Kelompok nilai AUC dibagi menjadi 5, (Widaningsih, 2019), antara lain sebagai berikut:

- 1) 0.90-1.00 memiliki arti sangat baik
- 2) 0.80-0.90 memiliki arti baik
- 3) 0.70-0.80 memiliki arti cukup
- 4) 0.60-0.70 memiliki arti buruk
- 5) 0.50-0.60 memiliki arti gagal

4. Hasil dan Pembahasan

Penelitian ini dilakukan dengan pengujian pertama yaitu pengujian mandiri pada algoritma klasifikasi *Decision Tree*, KNN, dan *Naïve Bayes*. Pengujian kedua yaitu algoritma klasifikasi dengan menerapkan teknik *discretization* dan *Bagging*. Pengujian menghasilkan nilai *accuracy*, *recall*, *precision*, dan juga nilai AUC. Berikut table hasil pengujian yang dilakukan dalam penelitian.

Tabel 2. Hasil Penelitian

	Acc	Recall	Prec.	AUC
<i>Decision Tree</i>	95.01%	99.84%	95.16%	0.708
<i>Decision Tree + Discretize+ Bagging</i>	95.13%	100%	95.13%	0.571
KNN	94.52%	99.26%	95.19%	0.656
KNN+ Discretize+ Bagging	76.69%	77.08%	97.99%	0.895
<i>Naïve Bayes</i>	87.85%	90.39%	96.62%	0.826
<i>Naïve Bayes+</i>	95.21%	97.51%	97.46%	0.971

Discretize+				
Bagging				

Berdasarkan pada pengujian yang telah dilakukan dalam penelitian ini mendapatkan hasil bahwa adanya peningkatan dari nilai *accuracy*, nilai *recall*, nilai *precision* dan AUC. Algoritma klasifikasi *Decision tree* dan *Naïve Bayes* meningkat pada tingkat *accuracy*, *recall*, dan *precision* berbeda halnya dengan algoritma klasifikasi KNN meningkat pada hasil nilai AUC. Peningkatan terjadi ketika algoritma klasifikasi digabungkan dengan teknik *discretization* dan juga metode *Bagging* dibandingkan menggunakan metode secara mandiri. Pengujian yang menghasilkan hasil terbaik yaitu algoritma *Naïve Bayes* yang digabungkan dengan teknik *discretization* dan *Bagging* menghasilkan *accuracy* sebesar 95.21%.

5. Kesimpulan dan Saran

Pengujian yang menghasilkan metode paling terbaik yaitu menerapkan algoritma klasifikasi *Naïve Bayes*, dengan teknik *discretization* dan metode *Bagging* dalam memprediksi penyakit stroke. Hasil akurasi yang didapatkan yaitu sebesar 95.21% meningkat sebesar 7.36% daripada menggunakan algoritma klasifikasi *Naïve Bayes* tunggal yang mendapatkan tingkat nilai akurasi yaitu sebesar 87.85%.

Pengujian yang akan datang dapat mengimplementasikan algoritma klasifikasi *Naïve Bayes* dengan teknik *discretization* dan metode *Bagging* menggunakan *dataset* dengan objek lain untuk menguji tingkat akurasi. Hasil pengujian selanjutnya juga dapat diimplementasikan menjadi bentuk aplikasi untuk mendiagnosis penyakit stroke sejak dini. Pengembangan lain juga dapat dilakukan dengan menggunakan metode-metode lain untuk menghasilkan akurasi terbaik.

References

Ahmed, H., Abd-el, S. F., Youn, E. M. G., & Omran, N. F. (2019). *Stroke Prediction using Distributed Machine Learning Based on Apache Spark Stroke Prediction using Distributed Machine Learning Based on Apache Spark*. January. <https://doi.org/10.13140/RG.2.2.13478.68162>

Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*, 12(16), 4102–4107. <https://doi.org/10.3923/jeasci.2017.4102.4107>

Bhatia, S., & Malhotra, J. (2021). *Naïve bayes classifier*

for predicting the novel coronavirus. *Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021, Icicv*, 880–883. <https://doi.org/10.1109/ICICV50876.2021.9388410>

Biswas, N., Uddin, K. M. M., Rikta, S. T., & Dey, S. K. (2022). A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach. *Healthcare Analytics*, 2(August), 100116. <https://doi.org/10.1016/j.health.2022.100116>

Buntoro, G. A. (2017). *Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter*. 2(1), 32–41.

Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01), 20–28. <https://doi.org/10.38094/jastt20165>

Derisma, D. (2020). Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining. *Journal of Applied Informatics and Computing*, 4(1), 84–88. <https://doi.org/10.30871/jaic.v4i1.2152>

Dev, S., Wang, H., Shamrock, C., Jain, N., Veeravalli, B., & John, D. (2022). A predictive analytics approach for stroke prediction using machine learning and neural networks. *Healthcare Analytics*, 2(December 2021), 100032. <https://doi.org/10.1016/j.health.2022.100032>

García-Gil, D., Ramírez-Gallego, S., García, S., & Herrera, F. (2018). Principal Components Analysis Random Discretization Ensemble for Big Data. *Knowledge-Based Systems*, 150, 166–174. <https://doi.org/10.1016/j.knosys.2018.03.012>

Hacibeyoglu, M., & Ibrahim, M. H. (2018). EF_Unique: An Improved Version of Unsupervised Equal Frequency Discretization Method. *Arabian Journal for Science and Engineering*, 43(12), 7695–7704. <https://doi.org/10.1007/s13369-018-3144-z>

Kabari, L. G., & Onwuka, U. C. (2019). Comparison of Bagging and Voting Ensemble Machine Learning Algorithm as a Classifier. *International Journals of Advanced Research in Computer Science and Software Engineering*, 9(3), 19–23. www.ijarcsse.com

Kemenkes RI. (2018). Hasil Riset Kesehatan Dasar Tahun 2018. *Kementrian Kesehatan RI*, 53(9), 1689–1699.

Lee, S. J., Xu, Z., Li, T., & Yang, Y. (2018). A novel bagging C4.5 algorithm based on wrapper feature selection for supporting wise clinical decision making. *Journal of Biomedical Informatics*, 78(October 2017), 144–155. <https://doi.org/10.1016/j.jbi.2017.11.005>

Moshkov, M. (2022). *Decision trees for regular factorial*

- languages. *Array*, 15(January), 100203. <https://doi.org/10.1016/j.array.2022.100203>
- Osman, A. S. (2019). Data mining techniques: Review. *International Journal of Data Science Research*, 2(1), 1–4.
- Panhalkar, A. R., & Doye, D. D. (2022). Optimization of decision trees using modified African buffalo algorithm. *Journal of King Saud University - Computer and Information Sciences*, 34(8), 4763–4772. <https://doi.org/10.1016/j.jksuci.2021.01.011>
- Peker, N., & Kubat, C. (2021). Application of Chi-square discretization algorithms to ensemble classification methods. *Expert Systems with Applications*, 185(June 2020), 115540. <https://doi.org/10.1016/j.eswa.2021.115540>
- Riyanto, U. (2019). Analisis Perbandingan Algoritma Naive Bayes Dan Support Vector Machine Dalam Mengklasifikasikan Jumlah Pembaca Artikel Online. *JIKA (Jurnal Informatika)*, 2(2), 62–72. <https://doi.org/10.31000/v2i2.1521>
- Saleh, A., & Nasari, F. (2018). Implementation Equal-Width Interval Discretization in Naive Bayes Method for Increasing Accuracy of Students' Majors Prediction. *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, September, 104. <https://doi.org/10.24843/lkjiti.2018.v09.i02.p05>
- Salmi, N., & Rustam, Z. (2019). Naïve Bayes Classifier Models for Predicting the Colon Cancer. *IOP Conference Series: Materials Science and Engineering*, 546(5). <https://doi.org/10.1088/1757-899X/546/5/052068>
- Srinivas, B., & Sasibhushana Rao, G. (2019). A hybrid CNN-KNN model for MRI brain tumor classification. *International Journal of Recent Technology and Engineering*, 8(2), 5230–5235. <https://doi.org/10.35940/ijrte.B1051.078219>
- Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE - Journal of Business and Audit Information Systems*, 4(1), 13–21. <https://doi.org/10.30813/jbase.v4i1.2729>
- Tuysuzoglu, G., & Birant, D. (2020). Enhanced bagging (eBagging): A novel approach for ensemble learning. *International Arab Journal of Information Technology*, 17(4), 515–528. <https://doi.org/10.34028/iajit/17/4/10>
- Virantika, E., Kusnawi, K., & Ipmawati, J. (2022). Evaluasi Hasil Pengujian Tingkat Clusterisasi Penerapan Metode K-Means Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia. *Jurnal Media Informatika Budidarma*, 6(3), 1657. <https://doi.org/10.30865/mib.v6i3.4325>
- Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, 13(1), 16–25. <https://doi.org/10.36787/jti.v13i1.78>
- Wu, Y., Ke, Y., Chen, Z., Liang, S., Zhao, H., & Hong, H. (2020). Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. *Catena*, 187(February 2018), 104396. <https://doi.org/10.1016/j.catena.2019.104396>
- Xing, W., & Bei, Y. (2020). Medical Health Big Data Classification Based on KNN Classification Algorithm. *IEEE Access*, 8, 28808–28819. <https://doi.org/10.1109/ACCESS.2019.2955754>
- Yang, F. J. (2018). An implementation of naive bayes classifier. *Proceedings - 2018 International Conference on Computational Science and Computational Intelligence, CSCI 2018*, 301–306. <https://doi.org/10.1109/CSCI46756.2018.00065>
- Yulianto Anggi Priliani, & Darwis, S. (2021). Penerapan Metode K-Nearest Neighbors (kNN) pada Bearing. *Jurnal Riset Statistika*, 1(1), 10–18. <https://doi.org/10.29313/jrs.v1i1.16>