

Komparasi Algoritma SVM Dan *Naive Bayes* Untuk Klasifikasi Kestabilan Jaringan Listrik

Sri Diantika¹⁾, Windu Gata²⁾, Hiya Nalatissifa³⁾, Mareanus Lase³⁾

^{1,2,3} Program Studi Magister Ilmu Komputer, STMIK Nusa Mandiri

Jalan Margonda Raya No. 545, Pondok Cina, Depok, Jawa Barat 16424, (021) 31908575

⁴ Program Studi Teknik Informatika, STMIK Nusa Mandiri

Jalan Margonda Raya No. 545, Pondok Cina, Depok, Jawa Barat 16424, (021) 31908575

ARTICLE INFO

Article history:

Received

Received in revised form

Accepted

Available online

ABSTRACT

A balance between supply and demand for electricity is needed to obtain a stable electricity network, so that the data pattern on the stability of the electricity network is needed for grouping or classifying data by utilizing data mining techniques to process information. To find a data mining method that can produce the best value in classifying the stability of the electricity network, in this study the application of the SVM and Naive Bayes classification is applied to the Electrical Grid Stability Simulated dataset taken from UCI Machine Learning. From the results of the classification test for the stability of the electricity network that has been carried out using the WEKA 3.8.2 application. The Support Vector Machine (SVM) method shows a better accuracy rate of 98.9% when compared to the Naive Bayes method which produces a value of 97.64% The accuracy of this result will show different results depending on the type of data, the number of instances, the class label and the percentage split data used.

Keywords: *Data mining, Naive Bayes, SVM, Classification, Electrical Grid Stability*

Abstrak

Keseimbangan antara pasokan dan permintaan listrik sangat diperlukan untuk mendapatkan jaringan listrik yang stabil, agar dapat diketahui pola data kestabilan jaringan listrik ini maka diperlukan pengelompokkan atau pengklasifikasian terhadap data dengan memanfaatkan teknik data mining guna mengolah informasi. Untuk mencari metode data mining yang bisa menghasilkan akurasi terbaik dalam mengklasifikasikan data Kestabilan jaringan listrik, maka pada penelitian ini dilakukan perbandingan penerapan algoritma klasifikasi SVM dan *Naive Bayes* terhadap dataset *Electrical Grid Stability Simulated* yang yang diambil dari *UCI Machine Learning*. Dari hasil pengujian klasifikasi kestabilan jaringan listrik yang telah dilakukan menggunakan aplikasi WEKA 3.8.2. Metode *Support Vector Machine* (SVM) menunjukkan tingkat accuracy yang lebih baik yaitu sebesar 98.9% jika dibandingkan dengan metode *Naive Bayes* yang menghasilkan nilai akurasi sebesar 97.64% Hasil akurasi ini akan menunjukkan hasil yang berbeda tergantung dengan jenis data, jumlah instance, *label class* dan *Percentage split data* yang digunakan.

Kata Kunci: *Data mining, Naive Bayes, SVM, Klasifikasi, Electrical Grid Stability*

Received September 30, 2020; Revised Oktober 2, 2020; Accepted Oktober 22, 2020

1. PENDAHULUAN

Kebutuhan akan energi listrik semakin meningkat seiring dengan kemajuan teknologi dan jumlah populasi manusia. Saat ini listrik bukan lagi suatu kebutuhan sekunder, melainkan telah beralih menjadi sebuah kebutuhan *primer*, terutama pada wilayah metropolitan, untuk itulah dibutuhkan jaringan listrik yang stabil. Untuk mendapatkan jaringan listrik yang stabil diperlukan keseimbangan antara pasokan dan permintaan listrik [1]

Untuk mengetahui faktor-faktor yang secara signifikan mempengaruhi kestabilan jaringan listrik, maka dilakukan penerapan teknik data *mining* untuk mengklasifikasikan kestabilan jaringan listrik. Data *mining* sendiri dapat diartikan sebagai ilmu yang digunakan untuk mengolah suatu informasi dan sekumpulan data yang memanfaatkan kecerdasan dalam membangun pola-pola untuk mengenali karakteristik dari data [2]. Atau dapat juga dikatakan bahwa *data mining* adalah proses mengekstraksi dan mengidentifikasi pengetahuan yang didapatkan dari sekumpulan data yang cukup besar [3]

Salah satu tugas yang dapat dilakukan dengan data *mining* adalah pengklasifikasian. Klasifikasi adalah tipe analisis data yang dapat dimanfaatkan untuk membantu orang menentukan kelas label dari sampel yang ingin di klasifikasi. Tujuan klasifikasi sendiri yaitu untuk meningkatkan kehandalan hasil yang diperoleh dari data [4]. Dalam menyelesaikan permasalahan klasifikasi, penggunaan metode atau teknik bertujuan untuk mempermudah proses klasifikasi. Beberapa teknik yang digunakan dalam kasus klasifikasi yaitu Pohon keputusan, Pengklasifikasi bayes atau naive bayes Jaringan saraf tiruan, Analisis statistik, Algoritma genetik, *Rough sets*, Pengklasifikasik-*nearest neighbour*, Metode berbasis aturan, *Memory based reasoning*, *Support vector machine*

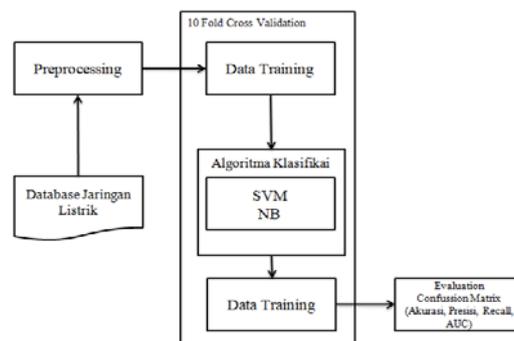
Peneliti melakukan *literature review* dan ditemukan fakta bahwa Penerapan teknik *data mining* untuk klasifikasi kestabilan jaringan listrik belum banyak dilakukan. Berdasarkan penelitian terkait yaitu pada eksperimen [1] menggunakan algoritma *Decision Tree*, menghasilkan akurasi sekitar 80%. Berdasarkan hasil *literature review* yang telah dilakukan maka, peneliti hendak menyelesaikan permasalahan mengenai pengelompokkan atau klasifikasi data kestabilan jaringan listrik ini dengan melakukan komparasi atau perbandingan antara algoritma klasifikasi *Support Vector Machine* (SVM) dan algoritma *Naive Bayes*. Perbandingan teknik klasifikasi ini bertujuan untuk menentukan algoritma klasifikasi *data mining* yang memberikan akurasi terbaik atau paling tinggi pada *dataset Electrical Grid Stability Simulated Data* dalam mengklasifikasikan kestabilan jaringan listrik. Dalam penelitian ini, peneliti menggunakan aplikasi WEKA untuk melakukan pengklasifikasian kestabilan jaringan listrik. Aplikasi WEKA sendiri merupakan aplikasi *data mining open source* (GPL) berbasis java [5] WEKA yang digunakan yaitu WEKA 3.8.2

2. METODOLOGI PENELITIAN

Metode penelitian yang digunakan mengacu pada *Cross-Standard Industry for Data Mining* (CRISP-DM) sebagai standarisasi pendekatan untuk *data mining* [6]. Dalam penelitian ini, tahapan CRISP-DM yang dilakukan hanya menggunakan lima tahapan, yaitu *business Understanding*, *data Understanding*, *data preparation*, *modeling dan evaluation*. tahapan CRISP-DM ini diterapka pada alur atau tahapan Penelitian yang dilakukan.

2.1. Alur Penelitian

Untuk memberikan arah atau acuan dalam melakukan penelitian ini, maka diperlukan adanya sebuah Alur penelitian yang jelas. Alur penelitian yang digunakan pada penelitian ini diilustrasikan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

Berikut penjabaran dari alur penelitian yang ditunjukkan pada Gambar 1

2.1.1. Dataset

Data merupakan objek yang akan diolah pada data mining, sebutannya adalah dataset. *Dataset* merupakan kumpulan data, dimana satu dataset mempresentasikan satu tabel pada database, atau bisa juga suatu matriks data dimana tiap kolom mewakili variabel, tiap baris mempresentasikan banyaknya data [7].

Dataset Kestabilan jaringan listik yang digunakan dalam penelitian ini termasuk kedalam jenis *dataset* yang adalah dataset terbuka atau *public dataset*, yang diambil dari UCI Machine Learning dengan url <https://archive.ics.uci.edu/ml/datasets/Electrical+Grid+Stability+Simulated+Data>. Nama dataset yang digunakan adalah *Electrical Grid Stability Simulated Data*, dataset ini terdiri dari *data training* dan *data test*, dimana *data training* merupakan data yang digunakan untuk perhitungan probabilitas dari data berdasarkan data pembelajaran yang dilakukan sedangkan *data test* adalah data yang digunakan untuk mengukur sejauh mana klasifikasi berhasil melakukan klasifikasi dengan benar [8]. *Dataset* ini memiliki 10000 *number of instances*, 14 *number of attributes* dan 2 *attribute class*.

2.1.2. Preprocessing

Sebelum memulai tahapan klasifikasi data dalam data mining, tahap *Preprocessing* menjadi salah satu tahapan yang penting, terutama jika data yang kita miliki terdapat data yang *redudance*, *Noisy* dan permasalahan data lainnya, maka tahapan *Preprocessing* ini sangat diperlukan sebelum memulai mining. *Preprocessing* sendiri merupakan tahapan dimana data akan dilakukan pengisian data yang kosong, menghilangkan duplikasi data, memeriksa inkonsistensi data, pembersihan data sertamemperbaiki kesalahan pada data. Proses pembersihan meliputi pengisian data yang kosong, menghilangkan duplikasi data, memeriksa inkonsistensi data, dan memperbaiki kesalahan pada data. Biasanya data yang kosong disebabkan oleh adanya data baru yang belum ada informasinya [9].

Pada penelitian ini *dataset* yang digunakan tidak memiliki *missing value* atau data yang kosong. Sehingga penelitian bisa langsung melanjutkan ketahapan klasifikasi berikutnya.

2.1.3. Algoritma Klasifikasi

Untuk megklasifikasikan data kestabilan jaringan listik yang memiliki dua kelas yaitu kelas stabil dan tidak stabil, peneliti menggunakan dua algoritma untuk klasifikasi yaitu:

a. Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan teknik klasifikasi. Konsep dari SVM adalah memaksimalkan margin, yang merupakan jarak pemisah antara kelas data dengan mencari *hyperlane* terbaik [10]. *Support Vector Machine* (SVM) adalah metode untuk memisahkan (mendiskriminasi) kumpulan data ke dalam beberapa kelompok. Pemisahan kelompok dilakukan dengan membangun garis pemisah yang tegas yang memisahkan suatu kelompok dengan kelompok yang lain. Pada umumnya, garis pemisah tersebut adalah linier [11]. Dari konsep dasar tersebut, algoritma SVM sering diaplikasikan dalam proses klasifikasi, dalam penelitian ini untuk memisahkan jaringan listik yang stabil dan tidak stabil.

b. Naïve Bayes

Metode *Naive Bayes* merupakan metode yang populer dalam klasifikasi, metode ini terkenal dengan prediksi probabilitas dari data yang disajikan. *Naive Bayes* memiliki nama lain seperti *idiot's Bayes*, *simple Bayes* dan *independence Bayes* karena kemudahannya serta tidak membutuhkan proses iterasi yang kompleks [12]. *Naive Bayes classifier* menggunakan dasar *teorema bayes* untuk memperkirakan kejadian bersyarat munculnya suatu kelas dengan asumsi bahwa variable-variabel yang mempengaruhinya bersifat independen [2]. Berikut bentuk persamaan umum dari *Naive Bayes* [13]:

$$P(C_i | X) = \frac{P(X | C_i)P(C_i)}{P(X)}$$

Keterangan:

- X : Kriteria suatu kasus berdasarkan maskan
- C_i : Kelas Solusi pola ke-i, diman i adalah jumlah label kelas
- P(C_i | X) : Probablilitas label kelas C_i dengan kriteria masukan X
- P(X |C_i) : Probablilitas kriteria masukan X dengan label kelas C_i
- P(C_i) : Probablilitas label kelas (C_i)

2.1.4. Evaluasi Kinerja

Evaluasi kinerja algoritma klasifikasi dapat dilihat dari berbagai aspek seperti dari akurasi, *Cross Validation*, *Confusion Matrix* dan *ROC curve*

Cross Validation merupakan salah satu teknik untuk menilai/memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. Data yang digunakan dalam proses pembangunan model disebut data latih/*training*, sedangkan data yang akan digunakan untuk memvalidasi model disebut sebagai data *test* [4].

Confusion Matrix merupakan suatu metode yang biasanya digunakan untuk melakukan perhitungan tingkat akurasi pada *Data mining* pada pengukuran kinerja dan sebagai representasi hasil proses klasifikasi [14].

ROC curve merupakan salah satu cara melakukan analisa terhadap model *classifier* yang telah dibuat. Penggunaan *ROC curve* ini untuk menentukan parameter model yang diinginkan sesuai dengan karakteristik dari model *classifier* yang digunakan [15]

3. HASIL DAN PEMBAHASAN

Setelah dilakukan *preprocessing* dan pengklasifikasian terhadap dataset *Grid Stability Simulated* menggunakan algoritma klasifikasi *Naïve Bayes* dan SVM maka didapatkan hasil pengujian berupa nilai *accuracy*, *precision* dan *recall*. Berikut penjabaran hasil pengujian dari kedua model algoritma klasifikasi sebagai berikut:

a. Hasil Klasifikasi Menggunakan Algoritma *Naïve Bayes*

Klasifikasi kestabilan jaringan listrik dengan Algoritma *Naïve Bayes* dan *test option* menggunakan beberapa varian *Percentage split* data *training* dan data *testing*, maka hasil pengujian ditampilkan pada tabel 1.

Tabel 1. Hasil pengujian dengan algoritma *Naïve Bayes*

Classifier Output:					
Data Training	Data Testing	Hasil Akurasi	Precision	Recall	
(%)	(%)	(%)	(%)	(%)	
50	50	97,64	97,6	97,6	
60	40	97,45	97,5	97,5	
70	30	97,47	97,5	97,5	
80	20	97,2	97,2	97,2	
90	10	97,5	97,5	97,5	

Dari tabel 1. Dapat diketahui bahwa hasil pengklasifikasian dataset dengan algoritma *Naïve Bayes*, *Percentage split* ikut andil dalam mempengaruhi hasil pengujian. Hasil pengujian dengan akurasi tertinggi didapat dari pengujian dengan nilai *Percentage split* yaitu data *training* sebesar 50% dan data *testing* sebesar 50%, jadi pengklasifikasian dengan algoritma *Naïve Bayes* menghasilkan akurasi sebesar 97,64%.

b. Hasil Klasifikasi Menggunakan Algoritma *Support Vector Machine* (SVM)

Klasifikasi kestabilan jaringan listrik dengan Algoritma *Support Vector Machine* (SVM) dan *test option* menggunakan beberapa varian *Percentage split*, maka hasil pengujian dengan menggunakan algoritma *Support Vector Machine* (SVM) ditampilkan pada tabel 2.

Tabel 2. Hasil pengujian dengan algoritma SVM

Classifier Output:					
Data Training	Data Testing	Hasil Akurasi	Precision	Recall	
(%)	(%)	(%)	(%)	(%)	
50	50	98,9	98,9	98,9	
60	40	98,93	98,9	98,9	
70	30	99,1	99,1	99,1	
80	20	99,35	99,4	99,3	
90	10	99,4	99,4	99,4	

Dari tabel 2. Dapat diketahui bahwa hasil pengklasifikasian dataset dengan algoritma *Support Vector Machine* (SVM), *Percentage split* ikut andil dalam mempengaruhi hasil pengujian. Hasil pengujian dengan akurasi tertinggi didapat dari pengujian dengan nilai *Percentage split* yaitu data *training* sebesar 90% dan

data *testing* sebesar 10%, jadi pengklasifikasian dengan algoritma *Naïve Bayes* menghasilkan akurasi sebesar 98,9%.

Setelah melakukan pengujian klasifikasi menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM), maka langkah selanjutnya adalah membandingkan algoritma mana yang menghasilkan tingkat akurasi tertinggi dalam mengklasifikasikan kestabilan jaringan listrik. Perbandingan hasil pengujian dengan menggunakan algoritma *Naïve Bayes* dan algoritma *Support Vector Machine* (SVM) ditampilkan pada tabel 3

Tabel 3. Hasil pengujian dengan algoritma SVM

Data Training (%)	Data Testing (%)	Akurasi %		Precision %		Recall %		ROC Area %	
		Naïve Bayes	SVM	Naïve Bayes	SVM	Naïve Bayes	SVM	Naïve Bayes	SVM
50	50	97,64	98,9	97,6	98,9	97,6	98,9	99,8	98,6
60	40	97,45	98,93	97,5	98,9	97,5	98,9	99,8	98,7
70	30	97,47	99,1	97,5	99,1	97,5	99,1	99,8	98,8
80	20	97,2	99,35	97,2	99,4	97,2	99,3	99,8	99,2
90	10	97,5	99,4	97,5	99,4	97,5	99,4	99,8	99,2

Berdasarkan dari tabel 3. Dapat diketahui bahwa dari hasil pengklasifikasian dataset dengan algoritma *Naïve Bayes*, *Percentage split* ikut andil dalam mempengaruhi hasil pengujian. Hasil pengujian algoritma *Naïve Bayes* akurasi tertinggi didapat dari pengujian dengan nilai *Percentage split* yaitu data *training* sebesar 50% dan data *testing* sebesar 50% akurasi sebesar 97,64%, sedangkan Hasil pengujian dengan algoritma *Support Vector Machine* (SVM) akurasi tertinggi didapat dari pengujian dengan nilai *Percentage split* yaitu data *training* sebesar 90% dan data *testing* sebesar 10% menghasilkan akurasi sebesar 98,9%.

4. KESIMPULAN DAN SARAN

Berdasarkan uraian hasil penelitian, maka didapatkan kesimpulan bahwa Algoritma *Support Vector Machine* (SVM) dapat mengklasifikasi kestabilan jaringan listrik lebih tinggi dibuktikan dengan nilai akurasi yang dimiliki yaitu sebesar 98,9% jika dibandingkan dengan algoritma *Naïve Bayes* yang menghasilkan nilai akurasi sebesar 97,64%. Dari hasil tersebut peneliti merekomendasikan penggunaan algoritma SVM pada klasifikasi kestabilan jaringan listrik. Hasil *accuracy* akan menunjukkan hasil yang berbeda tergantung dengan jenis data, nilai *k-folds* dan jumlah *instance*, label *class*, *Percentage split* yang digunakan

DAFTAR PUSTAKA

- [1] Arzamasov, V., Böhm, K., & Jochem, P, "Towards Concise Models Of Grid Stability. *Towards Concise Models Of Grid Stability*," 2018.
- [2] Sigit Adinugroho, Y. A. S, "Implementasi Data Mining Menggunakan WEKA," Universitas Brawijaya Press, Malang, 2018.
- [3] Kurniadi, A., & Novianto, Y, "Penerapan Metode Regresi Linier Untuk Memprediksi Kebiasaan Pelanggan Studi Kasus : PT. Mensa Binasukses," 2(2), 2020
- [4] Hendrian, S., "Algoritma Klasifikasi Data Mining Untuk Memprediksi," 11(3), 266–274, 2018.
- [5] Sulianta, F., & Juju, D., "Data Mining. Jakarta: PT Elex Media Komputering," 2018 .
- [6] Nurajijah, & Riana, D., "Algoritma Naïve Bayes , Decision Tree , Dan SVM Untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah Naïve Bayes , Decision Tree , And SVM Algorithm For Classification Of Sharia", 77–82, 2019
- [7] Nengsih, W., "Implementasi Data Mining Dalam Kasus Terapa," Politeknik Caltex, Riau, 2019
- [8] Anggraini, R. A., Widagdo, G., Budi, A. S., & Qomaruddin, M., "Penerapan Data Mining Classification Untuk Data Blogger Menggunakan Metode Naïve Bayes," 7(1), 47–51, 2019.
- [9] Pristyanto, Y., "Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset," 13(1), 11–16, 2019
- [10] Werdiningsih, I., Nuqoba, B., & Muhammadun, *Data Mining Menggunakan Android, Weka Dan SPSS*, UNAIR, Jawa Timur, 2020.
- [11] Kusuma, & Purba, D., *Machine Learning Teori, Program Dan Studi Kasus*, Deepublish, Yogyakarta, 2020.

- [12] Pujianto, U., & Ristanti, P. Y., “*Perbandingan Kinerja Metode C4.5 Dan Naive Bayes Dalam Klasifikasi Artikel Jurnal PGSD Berdasarkan Mata Pelajaran,*” 29(1), 50–67, 2019.
- [13] Andono, P. N., T.Sutojo, & Muljono, *Pengolahan Citra Digital*, ANDI, Yogyakarta, 2018
- [14] Mardhiyah, P. A., Siregar, R. R. A., & Palupiningsih, P., “*Jurnal Teknologi Klasifikasi Untuk Memprediksi Pembayaran Kartu Kredit Macet Jurnal Teknologi,*” 3(1), 91–101, 2020.
- [15] Dewi, S., “*Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan,*” XIII(1), 60–66, 2016