

Klasifikasi Rumah Tangga Penerima Subsidi Listrik di Provinsi Gorontalo Tahun 2019 dengan Metode *K-Nearest Neighbor* dan *Support Vector Machine*

Yohana Madame Hutahaean^{a1}, Arie Wahyu Wijayanto^{a2}

^aProgram Studi Statistika, Politeknik Statistika STIS

Jl. Otto Iskandardinata No. 64C, Jatinegara, Jakarta Timur, Daerah Khusus Ibukota 13330

¹211810664@stis.ac.id

³ariewahyu@stis.ac.id

Abstrak

Program subsidi listrik merupakan salah satu program pemerintah untuk penanganan kemiskinan, dimana keluarga tidak mampu mendapatkan bantuan subsidi listrik yang dibayarkan pemerintah ke PT Perusahaan Listrik Negara (PLN). Permasalahannya adalah masih terdapat rumah tangga yang mampu secara ekonomi namun tetap mendapatkan subsidi listrik. Penelitian ini bertujuan untuk melakukan klasifikasi rumah tangga penerima subsidi listrik menggunakan *data mining* serta melakukan perbandingan hasil klasifikasi dengan metode *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM). Alasan pemilihan metode ini dibandingkan metode lainnya dalam *data mining*, KNN merupakan metode yang dapat mewakili lazy learning dan SVM merupakan metode klasifikasi yang dapat memberikan generalisasi. Data yang digunakan adalah data Susenas Provinsi Gorontalo tahun 2019. Variabel yang digunakan adalah status penerimaan subsidi listrik sebagai kelas dan variabel penjelas (atribut) mencakup jumlah anggota rumah tangga, status kepemilikan bangunan, luas lantai rumah, bahan atap rumah terluas, bahan dinding terluas, bahan lantai rumah terluas, sumber air minum utama, bahan bakar utama untuk memasak, dan tempat pembuangan akhir tinja. Program yang digunakan dalam pengolahan data adalah R. Hasil penelitian menunjukkan bahwa metode KNN memiliki akurasi yang lebih baik dalam melakukan klasifikasi yaitu sebesar 98,07%. Secara keseluruhan, terdapat perbedaan yang signifikan dari klasifikasi KNN dan SVM, dimana kinerja KNN jauh lebih baik dari SVM dalam melakukan klasifikasi.

Kata kunci: Klasifikasi, *K-Nearest Neighbor*, Subsidi Listrik, *Support Vector Machine*

Electricity Subsidy Recipient Households Classification in Gorontalo Province in 2019 using K-Nearest Neighbor and Support Vector Machine

Abstract

The electricity program is a government program which aims to alleviate poverty, where underprivileged households are paid by the Government to PT. PLN. The problem that still occur is the nontargeted household receive electricity subsidies. This study aims to classify electricity subsidy recipient households using data mining and to compare methods which are KNN and SVM methods. The reason for choosing these two methods than other methods in data mining are KNN is a method that can represent lazy learning and SVM is a method that can make the generalizations. The data used is the Susenas data for Gorontalo Province in 2019. The variables used are the status of electricity subsidy recipient as a class and the explanatory variables (attributes) consisting the number of household members, building ownership status, floor area of the house, the widest roof material, the widest wall material, floor material, water source, cooking fuel, and final disposal of feces. The program used for data processing was R. The results showed that the KNN method has better accuracy in classification, which is 98,07%. Overall, there are significant differences between the classification with KNN and SVM, where the performance of KNN is much better than SVM in classifying.

Keywords: classification, K-Nearest Neighbor, Electricity Subsidy, Support Vector Machine

I. PENDAHULUAN

Kemiskinan adalah isu nasional dan global yang terus menjadi bahasan semua Negara[1]. Indonesia sebagai

Negara berkembang juga memberikan perhatian yang besar terhadap penanganan kemiskinan agar tercipta masyarakat yang adil dan makmur seperti termuat dalam alinea

keempat UUD 1945. Kemiskinan merupakan keadaan dimana seseorang atau rumah tangga memiliki pendapatan yang tidak cukup untuk memenuhi kebutuhan hidup minimum [2]. BPS menyebutkan bahwa 40% masyarakat Indonesia berada dalam ekonomi terendah. Salah satu langkah penanganan kemiskinan yang dilakukan pemerintah adalah pemberlakuan subsidi listrik. Program subsidi listrik diharapkan dapat menjadi salah satu solusi untuk membantu rumah tangga miskin agar tetap mendapatkan arus listrik dengan bayaran yang lebih murah [3].

Listrik menjadi salah satu industri terbesar di dunia. Dalam kehidupan modern, listrik merupakan salah satu kebutuhan vital yang sangat penting untuk keberlangsungan aktivitas manusia maupun kegiatan industri [4]. Kebutuhan akan listrik diatur oleh Negara melalui PT. PLN. Kemajuan teknologi dan informasi yang dinamis mendorong manusia menggunakan peralatan dengan tenaga listrik yang menyebabkan konsumsi energi meningkat. Sehingga, setiap rumah tangga mengharapkan sambungan ke jaringan listrik tanpa gangguan agar kehidupan dapat berjalan dengan baik [5]. Dalam hal ini, semakin banyak konsumsi tenaga listrik suatu rumah tangga, maka akan semakin banyak pula tarif listrik yang harus dibayarkan. Upaya yang dilakukan oleh pemerintah dalam mengontrol harga tenaga listrik adalah penyesuaian tarif listrik dan pemberlakuan subsidi tepat sasaran. Kebijakan Pemerintah melalui Peraturan Menteri ESDM Nomor 29 Tahun 2016 mengenai Mekanisme Pemberian Subsidi Tarif Tenaga Listrik untuk Rumah Tangga mengatur penerapan subsidi tenaga listrik agar tepat sasaran, sehingga rumah tangga mampu seharusnya tidak lagi mendapatkan subsidi. Bantuan subsidi listrik yang diterima juga berbeda tiap rumah tangga tergantung pemakaian energi listrik rumah tangga tersebut.

Pemerintah Indonesia selalu memberikan perhatian untuk distribusi kesejahteraan masyarakat yang tidak mampu dengan berbagai subsidi listrik, sehingga subsidi listrik mendapat perhatian khusus dalam penerapannya. Bantuan ini juga dipertimbangkan untuk mendorong industrialisasi dan kemajuan seluruh masyarakat sampai pelosok negeri. Namun, kebijakan yang sudah berlaku ini sering diperdebatkan karena dampak yang ditimbulkan bila kebijakan ini dihapuskan maupun terus diberlakukan. Pemerintah sendiri telah menyadari bahwa adanya subsidi yang terlalu besar untuk tenaga listrik dapat menggerus keuangan Negara. Sehingga, harus dilakukan perubahan arah kebijakan sebagai upaya mengendalikan beban subsidi listrik dengan pengurangan subsidi [6].

Konsumsi masyarakat akan listrik dipengaruhi oleh subsidi listrik pemerintah, dimana subsidi listrik meningkatkan konsumsi rumah tangga sebab jika rumah tangga tersebut mendapatkan subsidi listrik maka harga yang dibayarkan untuk tarif listrik menjadi lebih sedikit. Akibatnya, akan meningkatkan konsumsi penggunaan listrik sebab tidak ada rasa khawatir dengan biaya yang akan dikeluarkan.

Penduduk Provinsi Gorontalo 15,83 persen hidup di bawah Garis Kemiskinan pada tahun 2019. Sesuai dengan UU No. 30 Tahun 2019, pemerintah menyediakan dana

bagi kelompok masyarakat yang tidak mampu. Penyediaan dana bagi masyarakat kategori miskin diberikan dalam bentuk subsidi terhadap tarif tenaga listrik untuk rumah tangga dengan daya listrik 450 VA dan 900 VA. Meskipun penggunaan tenaga listrik sama, subsidi ini membuat rumah tangga penerima bantuan membayar dengan tarif yang lebih rendah dibandingkan dengan konsumen yang tidak menerima subsidi [7]. Adanya program subsidi tenaga listrik dari pemerintah diharapkan menjadi salah satu langkah pengentasan kemiskinan di Provinsi Gorontalo. Program ini dapat dikatakan berhasil apabila terdapat realisasi kebijakan dengan harapan yang dibuat [7], yaitu keluarga tidak mampu mendapatkan subsidi tenaga listrik dan keluarga mampu seharusnya tidak mendapatkan subsidi tenaga listrik.

Namun, program ini belum sepenuhnya berjalan efektif seperti yang telah dirancang sebelumnya. Masih terdapat rumah tangga sasaran penerima subsidi listrik yang tidak mendapatkan bantuan sosial ini [3]. Subsidi listrik masih banyak jatuh kepada orang-orang yang mampu secara ekonomi [8]. Ukuran keberhasilan program sesuai pemikiran David C. Korten adalah adanya kesesuaian antara tiga unsur meliputi kesesuaian antara program dan pemanfaat, kesesuaian antara program dan organisasi yang melakukan, dan kesesuaian antara pemanfaat dan organisasi yang melakukan. Bantuan subsidi listrik oleh pemerintah tentu membutuhkan anggaran yang tepat, sehingga bantuan ini diharapkan keluarga penerima manfaat berasal dari keluarga miskin dan tidak mampu. Namun, di berbagai daerah di Indonesia khususnya Gorontalo masih terjadi masalah tidak tepat sasaran bantuan subsidi listrik dan adanya ketidakadilan dalam pemberlakuan subsidi ini. Belum tepatnya rumah tangga sasaran penerima bantuan listrik memerlukan metode analisis yang lebih kompleks dan efisien. *Data mining* merupakan analisis statistik yang dapat mengelompokkan rumah tangga yang berhak sebagai keluarga penerima subsidi listrik secara lebih efisien.

Berdasarkan latar belakang di atas, peneliti tertarik meneliti lebih jauh klasifikasi rumah tangga penerima subsidi listrik menggunakan pendekatan *data mining*. Penelitian ini bertujuan untuk:

- Mengetahui klasifikasi/pengelompokan rumah tangga penerima bantuan subsidi listrik menggunakan *data mining*.
- Melakukan perbandingan metode antara KNN (*K-Nearest Neighbor*) dan SVM (*Support Vector Machine*)

II. METODOLOGI

A. *Data dan Sumber Data*

Penelitian ini menggunakan data mikro Susenas Provinsi Gorontalo tahun 2019. Data ini berjumlah 756 rumah tangga di Provinsi Gorontalo. Tabel 1 menunjukkan variabel-variabel yang digunakan pada penelitian ini. Program yang digunakan untuk pengolahan data adalah R. Metode yang digunakan adalah *k-nearest neighbour* (KNN) dan *support vector machine* (SVM).

TABEL I
DAFTAR VARIABEL DAN KODE PERTANYAAN

Deskripsi	Kode	Variabel	Tipe Data
Status Penerimaan Bantuan Subsidi Listrik	M1905	Dependen	Kategorik
Jumlah Anggota Rumah Tangga	M1401	Independen	Numerik
Status Kepemilikan Bangunan	M1402	Independen	Kategorik
Luas Lantai Rumah	M1503	Independen	Numerik
Bahan Atap Rumah Terluas	M1504	Independen	Kategorik
Bahan Dinding Terluas	M1505	Independen	Kategorik
Bahan Lantai Rumah Terluas	M1506	Independen	Kategorik
Sumber Air Minum Utama	M1508A	Independen	Kategorik
Bahan Bakar Utama untuk Memasak	M1601	Independen	Kategorik
Tempat Pembuangan Akhir Tinja	M1507C	Independen	Kategorik

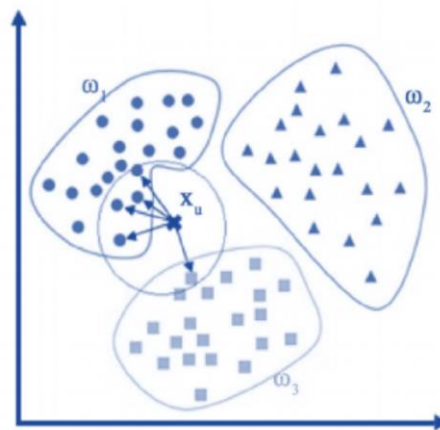
B. Klasifikasi

Klasifikasi merupakan salah satu metode dalam *data mining* yang bertujuan untuk mengidentifikasi pola dan kesamaan karakteristik dari basis data [9], sehingga model yang terbentuk diharapkan dapat memprediksi nilai kelas yang belum diketahui pada objek tertentu. Klasifikasi memerlukan *training set* untuk menentukan model dan *testing set* untuk mengetahui akurasi dari model yang terbentuk [10]. Ada berbagai jenis metode *data mining* untuk klasifikasi diantaranya *Support Vector Machine (SVM)*, *K-Nearest Neighbor (KNN)*, *Random Forest*, *AdaBoost*, *Naïve Bayes Classifier (NBC)*, *Classification and Regression Trees (CART)*, *ID3*, dan *C4.5*. pada penelitian ini, metode klasifikasi yang akan digunakan adalah *K-Nearest Neighbor (KNN)* dan *Support Vector Machine (SVM)* dalam melakukan klasifikasi rumah tangga penerima subsidi listrik.

C. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) adalah metode klasifikasi berbasis jarak dengan prinsip dasar dan sederhana yang dapat diimplementasikan secara luas [11]. Algoritma ini banyak digunakan untuk estimasi dan klasifikasi titik data terutama ketika dataset didistribusikan dalam beberapa kelas. Metode ini mengklasifikasikan sampel baru berdasarkan kemiripannya dengan semua sampel pada data pembelajaran [12]. KNN merupakan teknik *lazy learning* dimana sebagian besar perhitungan dilakukan pada *training data* untuk mendapatkan objek yang paling dekat pada kelompok *k* di *testing data* [13]. Langkah-langkah pada algoritma KNN:

1. Tetapkan jumlah tetangga terdekat atau *k* sebagai parameter
2. Hitung jarak sampel baru ke semua sampel dalam data latih
3. Pilih *k* sampel atau *k* tetangga dalam data latih yang memiliki jarak terdekat dengan sampel baru
4. Gunakan label kelas suara terbanyak dari *k* tetangga terdekat sebagai label sampel baru.



Gambar 1. Ilustrasi KNN [14]

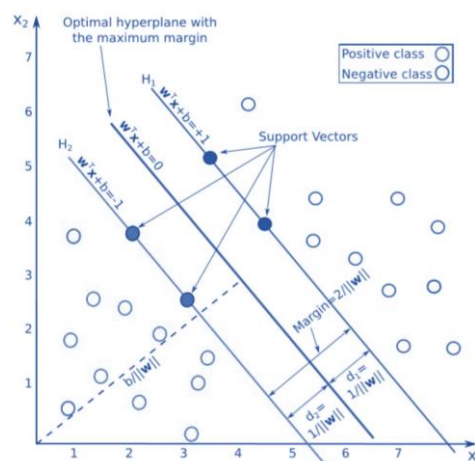
Pada gambar 1, ada 3 label kelas yang diwakili oleh berbagai warna yaitu: $\omega_1, \omega_2, \omega_3$. Selanjutnya terdapat sampel x dengan 5 tetangga terdekat ($k=5$), yang diklasifikasikan oleh KNN sebagai bagian dari ω_1 . Ini karena mayoritas dari tetangga terdekat x adalah ω_1 , sehingga ω_1 adalah label dari sampel x .

Label kelas sampel ditentukan berdasarkan kesamaannya. Selain itu, diperlukan jarak terukur untuk perhitungannya. Jika kesamaan yang dimiliki dengan sampel besar, maka jaraknya juga akan besar [15]. Oleh karena itu, penelitian ini mengimplementasikan *Euclidean Distance* sebagai metode pengukuran untuk menghitung jarak terdekat antar sampel. Jarak dapat diperoleh dengan rumus di bawah ini [16]:

$$D(x, y) = (\sum_{k=1}^n |x_k - y_k|^2)^{1/2} \tag{1}$$

D. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah metode *machine learning* yang bertujuan untuk memisahkan vektor input ke dalam ruang dimensi yang lebih tinggi serta membangun *hyperplane* yang optimal untuk memisahkan data menggunakan fungsi kernel [17]. Fungsi pemisah (*hyperplane*) terbaik didapatkan dengan mengukur jarak antara vektor terdekat dengan *hyperplane*.



Gambar 2 Ilustrasi SVM [19]

Diberikan $\{x_i, y_i\}_{i=1}^M$ sebagai kumpulan data, dimana x_i adalah *input vector* dan y_i adalah label kelas. Metode SVM akan membuat *hyperplane* terbaik yang memisahkan 2 kelas yang didefinisikan sebagai:

$$w^T \cdot x + b = 0 \tag{2}$$

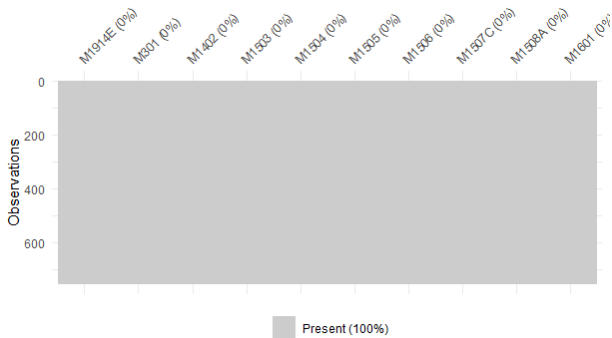
dimana W adalah vektor ortogonal, b adalah jarak antara titik origin dengan *hyperplane*, dan x adalah sampel [18]. Gambar 2 menunjukkan ilustrasi SVM dengan dua kelas, positif dan negatif.

Kelebihan dari metode SVM adalah mampu melakukan generalisasi yaitu klasifikasi data lain yang tidak masuk dalam data yang digunakan pada *data mining*. Selain itu, SVM dapat bekerja untuk mengklasifikasikan data yang linier maupun nonlinier.

E. Preprocessing

Pengolahan data dilakukan setelah memastikan data yang akan diolah “bersih” melalui rangkaian *preprocessing* sehingga model yang dihasilkan memiliki kualitas yang baik. *Preprocessing* terdiri dari pengecekan *missing value*, redundansi, standardisasi,

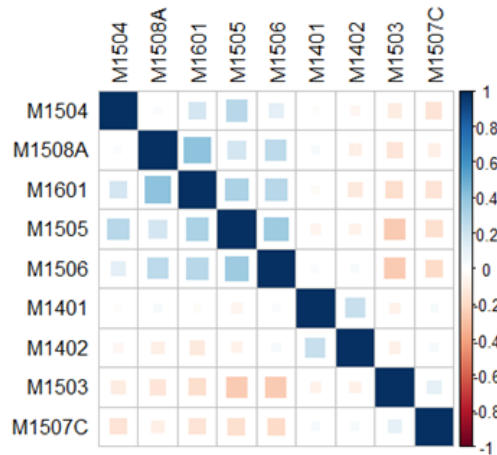
1) *Missing Value*: Adanya *missing value* pada data yang digunakan dapat menyebabkan terbatasnya kinerja dari model yang dibentuk [20], sehingga perlu penanganan *missing value* sebelum pengolahan data agar model prediksi yang dihasilkan maksimal. Beberapa teknik penanganan *missing value* adalah menghapus baris yang memiliki isian yang kosong dan imputasi dengan mean, median, dan sebagainya.



Gambar 3. Plot kelengkapan dataset

Berdasarkan pengecekan *missing value*, didapatkan hasil pada gambar 3 bahwa dataset yang digunakan pada penelitian ini memiliki isian yang lengkap. Sehingga, tidak dibutuhkan penanganan *missing value* untuk dataset yang digunakan.

2) *Redundansi*: Berdasarkan gambar 1 terlihat bahwa tidak terdapat warna yang pekat di dalam plot. Hal ini menunjukkan bahwa variabel penjelas tidak memiliki korelasi yang tinggi. Sehingga dapat dikatakan bahwa tidak terdapat redundansi pada seluruh variabel penjelas dalam data dan data ini dapat digunakan untuk klasifikasi.



Gambar 4. Plot korelasi variabel independen

3) *Standardisasi*: Setelah dipastikan tidak terdapat *missing value* dan redundansi, perlu dilakukan standardisasi pada data numerik. Dalam data ini, dilakukan standardisasi variabel Luas Lantai Rumah (M1503) dan jumlah ART (M1401). Metode standardisasi yang digunakan adalah normalisasi dengan *z-score*.

F. Pengujian Model

Pengujian model dilakukan dengan membagi data pada *training set* dan *testing set* dengan perbandingan 50 persen dan 50 persen. *training set* tersebut akan digunakan untuk melihat seberapa cocok metode *data mining* yang digunakan dalam memodelkan data. Kecocokan memodelkan data dilihat dari hasil kinerja dari metode yang digunakan.

G. Evaluasi dan Validasi Hasil

Evaluasi diperlukan untuk menilai kinerja model yang terbentuk dari data yang digunakan. Data yang digunakan untuk menguji adalah *testing set*. Selanjutnya dilakukan validasi hasil untuk menilai hasil prediksi yang didapatkan dari model. Evaluasi dan validasi hasil yang digunakan adalah dengan

1) *Confusion Matrix*: matriks yang digunakan untuk mengukur kinerja dari metode data mining yang digunakan. Selain itu, matriks ini juga menunjukkan perbandingan hasil klasifikasi dari model yang dihasilkan dengan hasil yang sebenarnya.

TABEL II
CONFUSION MATRIX PADA KLASIFIKASI 2 KELAS

Kelas Prediksi	Kelas Sebenarnya	
	Ya	Tidak
Ya	TP	FP
Tidak	FN	TN

Berdasarkan tabel 2, evaluasi model dilakukan dengan menghitung akurasi, sensitivitas, spesifisitas, PPV, dan NPV. Rumus perhitungannya ditunjukkan oleh persamaan di bawah ini:

$$Akurasi = \frac{TP+TN}{TP+FN+TN+FP} \tag{3}$$

$$Sensitivitas = \frac{TP}{TP+FN} \tag{4}$$

$$Spesifisitas = \frac{TN}{TTN+FP} \tag{5}$$

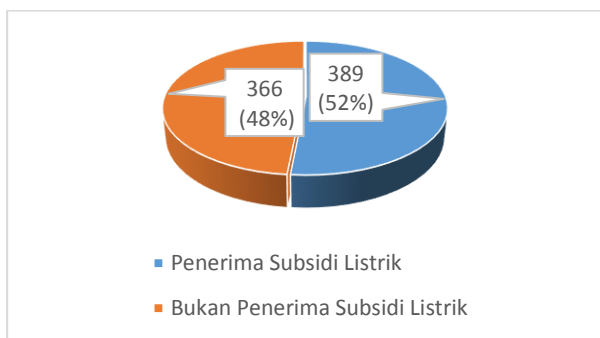
$$PPV = \frac{TP}{TP+FP} \tag{6}$$

$$NPV = \frac{TN}{FN+TN} \tag{7}$$

dimana TP (*True Positive*) mewakili jumlah rumah tangga penerima bantuan subsidi listrik yang diprediksi dengan benar. Sementara itu FN (*False Negative*) mewakili jumlah rumah tangga penerima bantuan subsidi listrik yang diprediksi salah. TN (*True Negative*) mewakili jumlah rumah tangga bukan penerima bantuan subsidi listrik yang diprediksi benar dan FP (*False Postive*) mewakili jumlah rumah tangga bukan penerima bantuan subsidi listrik yang diprediksi secara salah.

2) *ROC Curve*: metode yang digunakan untuk melihat performa dari hasil klasifikasi model yang dibuat berdasarkan nilai yang telah didapatkan pada perhitungan *confusion matrix*.

III. HASIL DAN PEMBAHASAN



Gambar 5. Proporsi penerima bantuan subsidi listrik

Gambar 5 menunjukkan bahwa proporsi penerima bantuan subsidi listrik adalah 389 rumah tangga atau sebesar 52%. Sedangkan proporsi rumah tangga bukan penerima bantuan subsidi listrik adalah sebesar 48%.

Klasifikasi memerlukan asumsi bahwa data yang digunakan seimbang. Kondisi ketidakseimbangan data adalah keadaan dimana satu kelas memiliki jumlah sampel besar dan kelas lainnya memiliki jumlah sampel yang kecil. Berdasarkan gambar 2, tidak terdapat indikasi *imbalanced* pada variabel dependen (*class*).

TABEL III
CONFUSION MATRIX PADA KLASIFIKASI METODE KNN DENGAN SPLIT VALIDATION

	Penerima Subsidi	Bukan Penerima Subsidi
Penerima Subsidi	188	0
Bukan Penerima Subsidi	5	192

Berdasarkan tabel 3, diketahui bahwa dengan metode KNN terdapat 188 rumah tangga penerima subsidi listrik dan 192 rumah tangga bukan penerima subsidi listrik yang diprediksi dengan benar.

TABEL IV
CONFUSION MATRIX PADA KLASIFIKASI METODE SVM DENGAN SPLIT VALIDATION

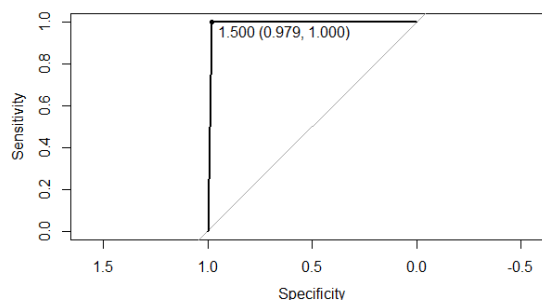
	Penerima Subsidi	Bukan Penerima Subsidi
Penerima Subsidi	142	95
Bukan Penerima Subsidi	52	88

Berdasarkan tabel 4, diketahui bahwa dengan metode SVM terdapat 142 rumah tangga penerima subsidi listrik dan 88 rumah tangga bukan penerima subsidi listrik yang diprediksi dengan benar.

TABEL V
PERBANDINGAN EVALUASI MODEL DENGAN SPLIT VALIDATION

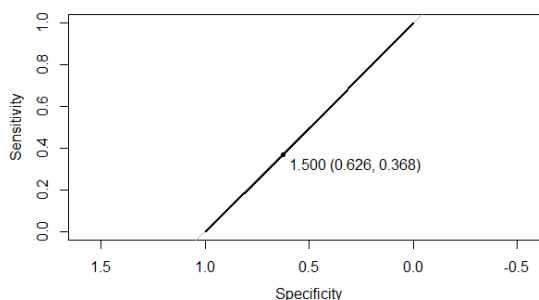
	KNN	SVM
Akurasi	0,9870	0,6101
Kappa Statistic	0,9740	0,2143
Sensitivity	0,9741	0,7320
Specificity	1,0000	0,4809
Positive Pred. Value	1,0000	0,5992
Negative Pred Value	0,9746	0,6286

Dari tabel 5 terlihat bahwa akurasi yang dihasilkan KNN jauh lebih baik yakni sebesar 98,70%. Sedangkan akurasi yang dihasilkan dengan metode SVM adalah 61,01%. Nilai dari akurasi, *Kappa statistic*, *sensitivity*, *specificity*, *positive predicted value*, *negative predicted value* dengan menggunakan metode KNN juga lebih baik dari SVM. Model yang dihasilkan dari kedua metode ini sangat berbeda signifikan. Oleh karena itu, berdasarkan *split validation* dapat dikatakan bahwa metode klasifikasi dengan KNN memiliki performa yang lebih baik dari SVM dalam melakukan klasifikasi.



Gambar 6. Kurva ROC metode KNN

Berdasarkan gambar 6 terlihat bahwa metode KNN memiliki nilai AUC sebesar 97,9%. Hal ini menunjukkan bahwa kinerja algoritma KNN untuk memprediksi seluruh dataset sudah sangat baik



Gambar 7. Kurva ROC metode SVM

Garis ROC pada metode KNN dan SVM sangat berbeda. Berdasarkan gambar 7 terlihat bahwa metode SVM memiliki nilai AUC sebesar 62,6%. Hal ini menunjukkan bahwa ketelitian hasil klasifikasi dengan metode SVM masih tergolong kurang baik.

Berdasarkan hasil evaluasi dengan *split validation* dan kurva ROC, metode KNN memiliki ketelitian yang lebih baik dari metode SVM dalam melakukan klasifikasi. Ini menunjukkan metode KNN sebaiknya digunakan dalam menyelesaikan kasus ini dibandingkan metode SVM.

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilakukan, dapat diambil kesimpulan bahwa akurasi yang dihasilkan dengan metode KNN lebih baik dari SVM. Selain itu, didapatkan pula bahwa model yang dibentuk untuk klasifikasi rumah tangga penerima bantuan subsidi listrik dengan KNN sudah baik Hasil evaluasi menunjukkan terdapat perbedaan performa yang signifikan dari hasil yang dibentuk dengan metode KNN dan SVM, dimana performa metode KNN jauh lebih baik. Sehingga, metode klasifikasi yang sebaiknya digunakan dalam melakukan klasifikasi rumah tangga penerima subsidi listrik di Provinsi Gorontalo tahun 2019 adalah metode KNN.

Saran yang dapat diberikan kepada pemerintah berdasarkan penelitian ini adalah dalam menentukan rumah tangga sasaran penerima bantuan subsidi listrik perlu melakukan kajian dan koordinasi dengan pakar yang ahli di *data mining*, sehingga setiap bantuan yang diberikan pemerintah dapat tersalurkan dengan tepat sasaran.

DAFTAR PUSTAKA

- [1] A. Levanon, E. Lavee, and R. Strier, "Explaining the Factors Shaping the Likelihood of Poverty Among Working Families by Using a Concurrent Mixed," *Soc. Indic. Res.*, vol. 157, no. 3, pp. 1089–1109, 2021, doi: 10.1007/s11205-021-02689-5.
- [2] R. Hanna and B. A. Olken, "Universal Basic Incomes versus Targeted Transfers: Anti-Poverty Programs in Developing Countries," *J. Econ. Perspect.*, vol. 32, no. 4, pp. 201–226, 2018.
- [3] B. H. Yanto and Y. Yunus, "Evaluasi Penentuan Kelayakan Pemberian Subsidi Listrik dengan Metode MFEP," *J. Inform. Ekon. Bisnis*, vol. 3, pp. 109–114, 2021, doi: 10.37034/infkeb.v3i3.91.
- [4] H. Foy and I. Bosman, "Electricity Situation," 2021.
- [5] Y. Abulkher, "Tripoli 's Electricity Crisis and its Politicisation Policy Brief," *Clingendael Netherlands Inst. Int. Relations*, no. April, 2020.
- [6] S. Sugiartiningsih, "Pengaruh Subsidi Listrik Terhadap Konsumsi Rumah Tangga Di Indonesia Periode 2004-2017," *J. Accounting, Financ. Taxation, Audit.*, vol. 1, no. 2, pp. 64–73, 2020, doi: 10.28932/jafta.v1i2.2455.
- [7] D. A. Ermawaty, "Model Kesesuaian Kebijakan Subsidi Listrik," *J. Ilm. Indones.*, vol. 5, no. 1, pp. 54–70, 2020.
- [8] A. Widarma and H. Kumala, "Sistem Pendukung Keputusan Dalam Menentukan Pengguna Listrik Subsidi Dan Nonsubsidi Menggunakan Metode Fuzzy Mamdani (Studi Kasus : PT. PLN Tanjung Balai)," *J. Teknol. Inf.*, vol. 2, no. 2, p. 165, 2019, doi: 10.36294/jurti.v2i2.432.
- [9] V. Plotnikova, M. Dumas, and F. Milani, "Adaptations of data mining methodologies: A systematic literature review," *PeerJ Comput. Sci.*, vol. 6, pp. 1–43, 2020, doi: 10.7717/PEERJ-CS.267.
- [10] A. ACHARYA and D. SINHA, "An Educational Data Mining Approach to Concept Map Construction for Web based Learning," *Inform. Econ.*, vol. 21, no. 4/2017, pp. 41–58, 2017, doi: 10.12948/issn14531305/21.4.2017.04.
- [11] J. Salvador-Meneses, Z. Ruiz-Chavez, and J. Garcia-Rodriguez, "Compressed kNN: K-nearest neighbors with data compression," *MDPI*, vol. 21, no. 3, pp. 1–21, 2019, doi: 10.3390/e21030234.
- [12] A. Kataria and M. D. Singh, "A Review of Data Classification Using K-Nearest Neighbour Algorithm," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 3, no. 6, pp. 354–360, 2013.
- [13] R. Ullah, A. H. Khan, and S. M. Emaduddin, "ck-NN: A Clustered k-Nearest Neighbours Approach for Large-Scale Classification," *ADCAIJ Adv. Distrib. Comput. Artif. Intell. J.*, vol. 8, no. 3, pp. 67–77, 2019, doi: 10.14201/adcaij2019836777.
- [14] W. Zhang, "Machine Learning Approaches to Predicting Company Bankruptcy," *J. Financ. Risk Manag.*, vol. 06, no. 04, pp. 364–374, 2017, doi: 10.4236/jfrm.2017.64026.
- [15] Z. Zhang *et al.*, "Modulation signal recognition based on information entropy and ensemble learning," *Entropy*, vol. 20, no. 3, pp. 1–18, 2018, doi: 10.3390/e20030198.
- [16] H. A. Abu Alfeilat *et al.*, "Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review," *Big Data*, vol. 7, no. 4, pp. 221–248, 2019.
- [17] Arfiani, Z. Rustam, J. Pandelaki, and A. Siahaan, "Kernel Spherical K-Means and Support Vector Machine for Acute Sinusitis Classification," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 546, no. 5, 2019, doi: 10.1088/1757-899X/546/5/052011.
- [18] J. Abukhait, A. M. Mansour, and M. Obeidat, "Classification based on Gaussian-kernel support vector machine with adaptive fuzzy inference system," *Prz. Elektrotechniczny*, vol. 94, no. 5, pp. 14–22, 2018, doi: 10.15199/48.2018.05.03.
- [19] A. Tharwat, "Parameter investigation of support vector machine classifier with kernel functions," *Knowl. Inf. Syst.*, vol. 61, no. 3, pp. 1269–1302, 2019, doi: 10.1007/s10115-019-01335-4.
- [20] S. Lim, S. J. Kim, Y. J. Park, and N. Kwon, "A deep learning-based time series model with missing value handling techniques to predict various types of liquid cargo traffic," *Expert Syst. Appl.*, vol. 184, p. 115532, 2021.