

Klasifikasi Rumah Tangga Desa Rigi Berbasis Web Menggunakan Metode *Support Vector Machine* Dan *Grid Search*

Yan Sofyan Andhana Saputra ^{1*}, Prodensio Veto Meo²

¹ Dosen Program Studi Teknologi Informasi Fakultas Teknik, Universitas Darma Persada,

² Program Studi Teknologi Informasi Fakultas Teknik, Universitas Darma Persada, (Mahasiswa)

Jl. Taman Malaka Selatan No.22, Pondok Kelapa, Duren Sawit, DKI Jakarta, Indonesia 13450

*Koresponden : yansofyan@gmail.com

Abstrak

Sistem informasi desa (SID) digunakan oleh pemerintah Desa Rigi untuk membantu pekerjaan yang ada di desa dalam hal layanan surat, info desa, kependudukan, dan website desa yang berisi informasi kegiatan dan program yang berada di Desa Rigi. Namun sistem informasi desa yang digunakan belum mampu untuk mengklasifikasikan status kemiskinan dari suatu rumah tangga. Penelitian ini bertujuan untuk membuat aplikasi klasifikasi rumah tangga berbasis web. Aplikasi ini dibuat menggunakan metode *Support Vector Machine* dan *Grid Search*. *Support Vector Machine* digunakan untuk mengklasifikasi status kemiskinan dari suatu rumah tangga dan *Grid Search* digunakan untuk mencari parameter terbaik pada *Support Vector Machine* sehingga hasil klasifikasi menjadi semakin akurat. Setelah melakukan penelitian, terdapat peningkatan akurasi sebesar 0.0182 % dari model yang dilatih dengan menggunakan *Support Vector Machine* dan *Grid Search*. Penelitian pada Desa Rigi ini bisa menghasilkan solusi untuk mengklasifikasikan status rumah tangga pada sistem informasi yang ada di Desa Rigi.

Kata kunci: *Grid Search*; *Klasifikasi Rumah Tangga*; *Sistem Informasi Desa*; *Support Vector Machine*.

Abstract

The village information system (SID) is used by the Rigi Village government to assist with work in the village in terms of mail services, village information, population, and a village website which contains information on activities and programs in Rigi Village. However, the village information system used is not yet able to classify the poverty status of a household. This research aims to create a web-based household classification application. This application was created using the *Support Vector Machine* and *Grid Search* methods. *Support Vector Machine* is used to classify the poverty status of a household and *Grid Search* is used to find the best parameters on the *Support Vector Machine* so that the classification results become more accurate. After conducting research, there was an increase in accuracy of 0.0182% from the model trained using *Support Vector Machine* and *Grid Search*. It is hoped that this research can be a solution for classifying household status in the existing information system in Rigi Village.

Keywords: *Grid Search*; *Household Classification*; *Village Information System*; *Support Vector Machine*.

1. Pendahuluan

Rumah tangga adalah kelompok orang yang tinggal bersama dalam satu unit perumahan dan berbagi sumber daya, tanggung jawab, dan interaksi sehari-hari[1]. Di Indonesia, profil rumah tangga miskin dapat ditentukan berdasarkan ciri tempat kediaman atau rumah, ketenagakerjaan, sosial demografi, dan tingkat pendidikan[2].

Klasifikasi rumah tangga mengacu pada proses pengelompokan atau kategorisasi rumah tangga berdasarkan berbagai kriteria yang digunakan untuk tujuan analisis atau pengambilan keputusan. Berdasarkan status kemiskinan rumah tangga dibedakan menjadi rumah tangga miskin dan rumah tangga tidak miskin. Dalam menentukan suatu rumah tangga tergolong kedalam rumah tangga miskin dan rumah tangga tidak miskin, Badan Pusat Statistik menggunakan 14 variabel kemiskinan [4] yaitu:

1. Kepemilikan aset.
2. Ukuran bangunan dalam luas.
3. Jenis atau tipe lantai
4. Jenis atau tipe dinding
5. Sarana pembuangan air besar

6. Sumber penerangan.
7. Sumber air minum
8. Pendidikan terakhir kepala keluarga.
9. Akses ke puskesmas atau poliklinik
10. Akses ke lapangan pekerjaan
11. Jenis atau tipe bahan bakar untuk memasak
12. Frekuensi dalam sepekan pembelian ayam, daging dan susu
13. Frekuensi makan dalam sehari
14. Jumlah pakain baru (dalam stel) yang dibeli dalam setahun

Data Mining adalah salah satu bidang dari proses penemuan pengetahuan yang mampu memberikan jalur-jalur inovatif dalam menginterpretasi data, ekstraksi informasi yang sebelumnya tidak teridentifikasi, dan mendeteksi pola yang tersembunyi di antara sekelompok data, dalam set data yang besar untuk memprediksi hasil sesuai dengan kebutuhan yang ingin dicapai dari suatu proses.

Support Vector Machine dapat diimplementasikan untuk klasifikasi status kemiskinan dari suatu rumah tangga. sedangkan *Grid Search* digunakan untuk mencari parameter terbaik pada *Support Vector Machine*. Vapnik dan Alexey Ya menemukan metode *Support Vector Machine* disingkat SVM sebagai model pembelajaran mesin berbasis kernel untuk mengerjakan tugas klasifikasi dan regresi [8].

Keunggulan SVM terletak pada kenyataan bahwa SVM memperoleh subset vektor dukungan selama fase pembelajaran, yang seringkali hanya merupakan bagian kecil dari himpunan data asli. Set vector dukungan ini mewakili tugas klasifikasi tertentu dan terdiri dari himpunan data yang kecil.

Sedangkan *Grid search* adalah pencarian yang menyeluruh berdasarkan pada subset yang telah didefinisikan, menemukan nilai hyperparameter yang optimal. Nilai hyperparameter ditetapkan dengan sejumlah langkah, penentuan batas bawah atau nilai minimal dan penentuan batas atas atau nilai maksimal[9].

2. Metodologi

Metodologi penelitian adalah suatu kerangka kerja untuk mengembangkan penelitian atau serangkaian tahapan kerja dengan bantuan pendekatan tertentu dan alat serta asumsi yang digunakan. Adapun metode yang digunakan pada penelitian adalah penggunaan teknik atau prosedur yang spesifik untuk menganalisis data yang ada [13].

Pengumpulan dan pembuatan dataset yang diperlukan pada proses klasifikasi diperoleh melalui interviu, pengamatan dan studi pustaka. Setelah pengumpulan data, selanjutnya dianalisis sesuai standar proses pada CRISP-DM. Sedangkan metode *Waterfall* digunakan untuk pengembangan sistem. *Support Vector Machine* digunakan untuk mengklasifikasi status kemiskinan dari suatu rumah tangga dan *Grid Search* digunakan untuk mencari parameter terbaik pada *Support Vector Machine* sehingga hasil klasifikasi menjadi semakin akurat

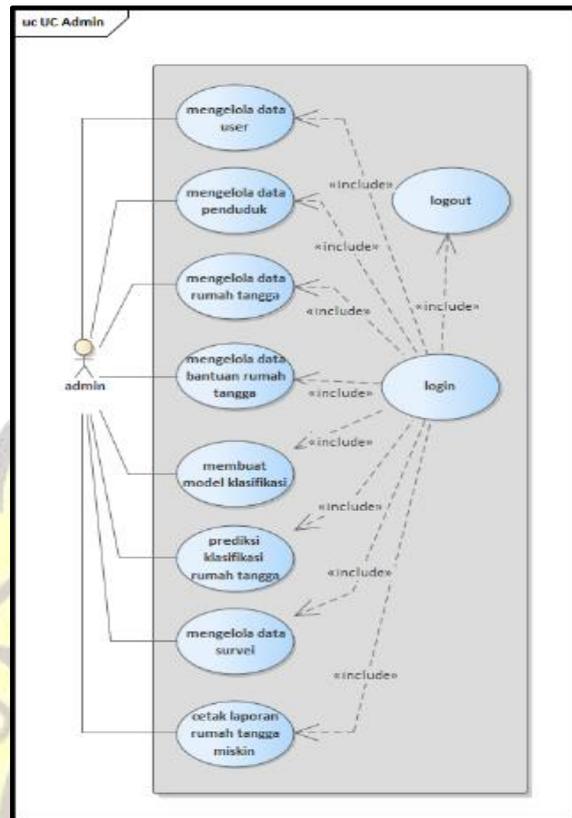
3. Hasil dan Pembahasan

3.1. Sistem Yang Sedang Berjalan

Sistem yang digunakan untuk mengelola data penduduk, data keluarga dan data rumah tangga yang ada di Desa Rigi adalah sistem Informasi Desa (SID). Pada sistem SID data rumah tangga yang disimpan adalah data rumah tangga yang tidak memiliki status rumah tangga miskin atau tidak miskin.

Proses penentuan status rumah tangga belum dapat dilakukan didalam sistem informasi desa yang digunakan. Hal ini dikarenakan pada sistem informasi desa belum memiliki fungsi untuk melakukan proses klasifikasi rumah tangga.

3.2. Sistem Yang Diusulkan



Gambar 1. Use Case Sistem Usulan

Sistem yang diusulkan adalah aplikasi klasifikasi rumah tangga berbasis *web* menggunakan metode *support vector machine* dan *Grid Search*.

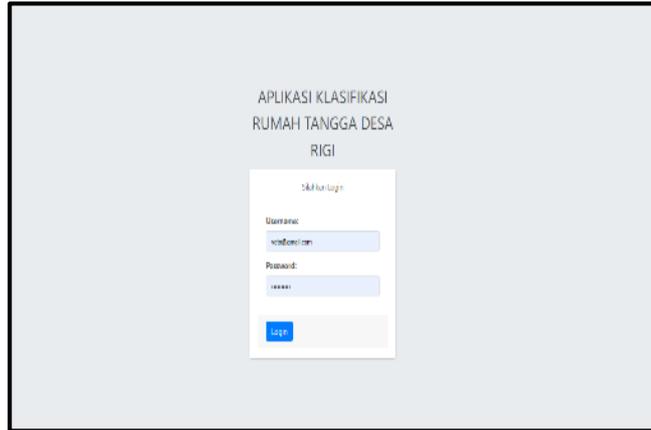
3.3. Penerapan Algoritma

Berikut ini adalah tahapan kegiatan yang dilaksanakan untuk menerapkan algoritma Support Vector Machine (SVM) dan Grid Search :

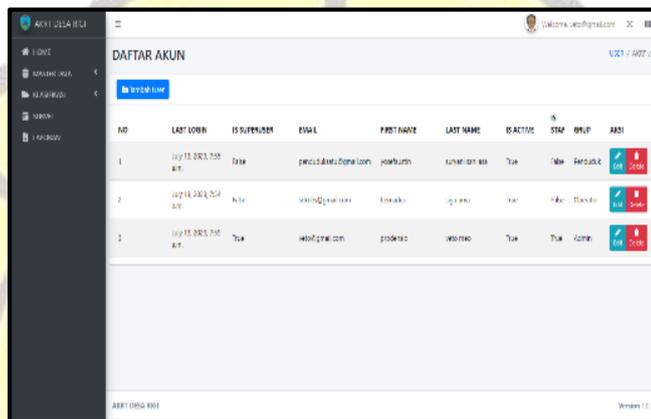
1. Mempersiapkan dataset
2. Inisialisasi model SVM
3. Tentukan Fungsi Objektif SVM
4. Terapkan Grid Search
5. Hitung nilai fungsi objektif SVM untuk setiap kombinasi parameter
6. Buat model SVM baru dengan parameter terbaik
7. Latih model SVM dengan parameter terbaik menggunakan set pelatihan
8. Evaluasi model SVM pada set pengujian
9. Melakukan klasifikasi pada data baru menggunakan model yang telah ditetapkan.

3.4. Pembuatan User Interface

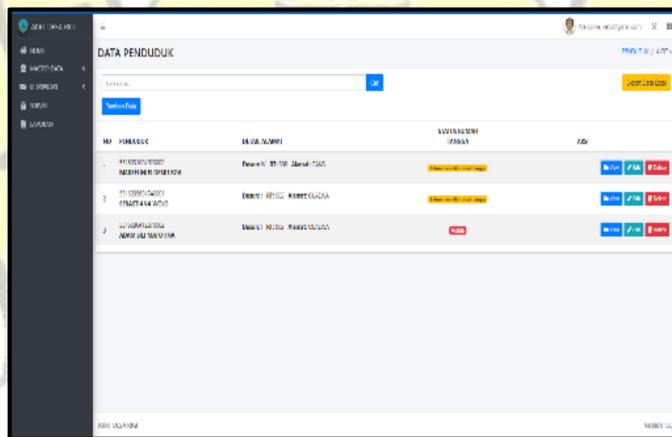
User interface dibuat untuk digunakan oleh *user* (admin, operator, penduduk)



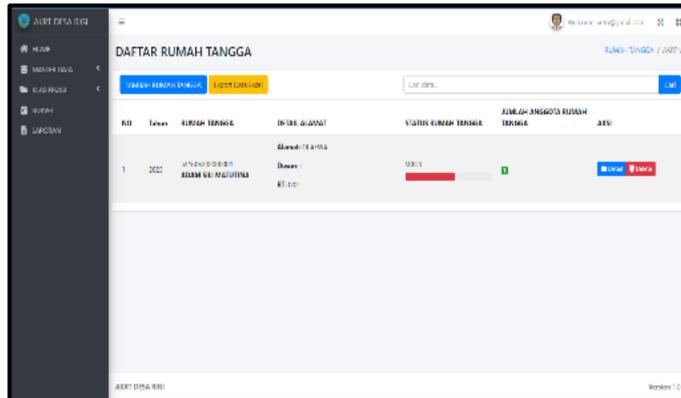
Gambar 2. Halaman Login
Halaman *login* digunakan untuk *user* masuk ke dalam sistem.



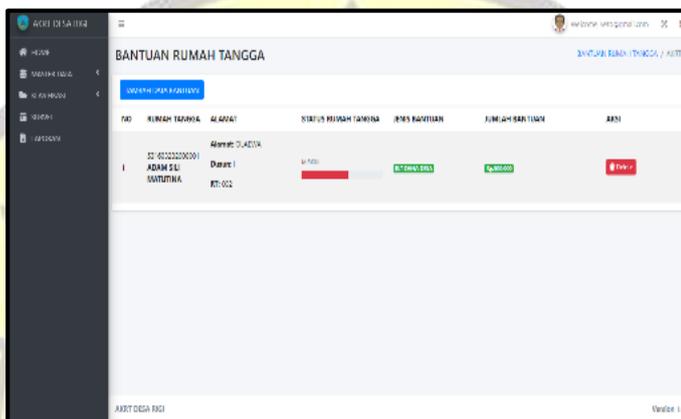
Gambar 3. Halaman Data User
Halaman data *user* digunakan untuk mengelola *user* yang ada didalam sistem.



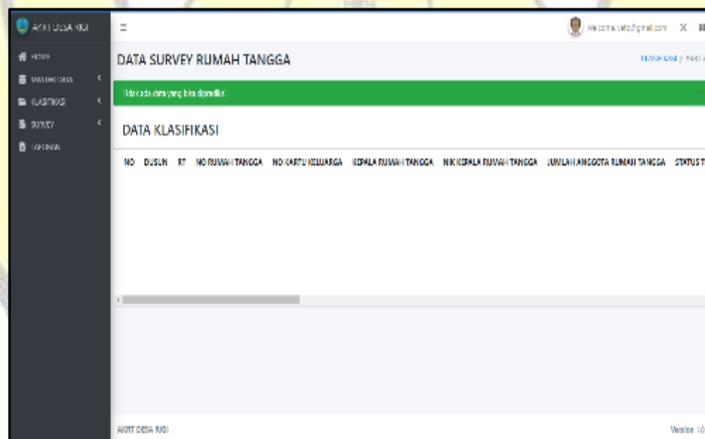
Gambar 4. Halaman Data Penduduk
Halaman data penduduk digunakan untuk mengelola data penduduk.



Gambar 5. Halaman Data Rumah Tangga
Halaman data rumah tangga digunakan untuk mengelola data rumah tangga



Gambar 6. Halaman Data Bantuan
Halaman data bantuan digunakan untuk mengelola data bantuan rumah tangga



Gambar 7. Halaman Prediksi
Halaman prediksi digunakan untuk melakukan klasifikasi status rumah tangga.

4	Sarana pembuangan air besar	Kepemilikan fasilitas buang air besar rumah tangga	Jamban bersama/tetangga, jamban umum, jamban sendiri
5	Sumber air minum	Sumber air minum untuk konsumsi rumah tangga	Mata air /perpipaan
6	Sumber penerangan rumah	Sumber penerangan rumah tangga	Lampu minyak/lilin, listrik PLN
7	Bahan bakar untuk memasak	Bahan bakar yang digunakan untuk memasak sehari-hari	Kayu bakar, LPG, minyak tanah
8	Jumlah dalam sepekan, konsumsi daging/ayam/susu	Frekuensi rumah tangga dalam mengkonsumsi daging/susu/telur dalam jangka waktu selama satu minggu	Satu kali, lebih dari satu kali
9	Jumlah makan dalam sehari	Frekuensi makan rumah tangga dalam satu hari	Lebih dari dua kali, kurang dari sama dengan dua kali
10	Jumlah dalam setahun, pembelian pakaian baru	Frekuensi rumah tangga dalam membeli pakaian baru dalam waktu satu tahun	Satu kali, lebih dari satu kali
11	Penghasilan kepala rumah tangga	Penghasilan kepala keluarga per bulan	
12	Pendidikan kepala rumah keluarga	Pendidikan terakhir kepala keluarga	Tidak sekolah, SD, SMP, SMA, diploma, sarjana
13	Mampu membayar biaya pengobatan	Kemampuan rumah tangga untuk membayar biaya pengobatan di fasilitas kesehatan (puskesmas atau poliklinik)	Mampu, tidak mampu
14	Memiliki simpanan aset	Memiliki simpanan atau barang yang gampang dijual senilai minimal lima ratus ribu rupiah (Rp. 500.000,-) seperti sepeda, mesin, kendaraan bermotor, emas ataupun aset lainnya	Ya, tidak
15	Status rumah tangga	Status kemiskinan rumah tangga	Miskin, tidak miskin

3.5.3. Data Preparation

Tabel 2. Atribut Data

No	Atribut	Tipe Data	Keterangan
1	Ukuran lantai dalam luas	numerik	fitur
2	Jenis atau tipe lantai	ordinal	fitur
3	Jenis atau tipe dinding	ordinal	fitur
4	Sarana pembuangan air besar	ordinal	fitur
5	Sumber air minum	ordinal	fitur
6	Sumber penerangan rumah	ordinal	fitur
7	Bahan bakar untuk memasak	ordinal	fitur
8	Jumlah konsumsi daging susu telur dalam seminggu	ordinal	fitur
9	Jumlah makan dalam sehari	ordinal	fitur
10	Jumlah membeli pakain baru dalam setahun	ordinal	fitur
11	Penghasilan kepala keluarga	numerik	fitur
12	Pendidikan kepala keluarga	ordinal	fitur
13	Mampu membayar biaya pengobatan	nominal	fitur
14	Memiliki simpanan aset	nominal	fitur
15	Status rumah tangga	nominal	label

Data rumah tangga yang digunakan berjumlah 182 *record* data. kemudian data dibagi menjadi data *training* 70% (145 record) dan data *testing* 30% (37 record).

Setelah data dibagi dua menjadi data latih atau *training* dan *data testing*, proses selanjutnya adalah melakukan transformasi data. untuk data numerik menggunakan *Standard Scaler*, untuk data ordinal menggunakan *Ordinal Encoder*, dan untuk data nominal menggunakan *One-Hot Encoding*

3.5.4. Modeling

```

def latih_kode(request):
    user_groups = request.user_groups.all()
    if request.method == "GET":
        data = request.session.get('data')
        df = pd.DataFrame(data)
        df.drop(['id'], axis=1, inplace=True)
        X = df.drop(columns='status rumah tangga')
        y = df['status rumah tangga']
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, stratify=y, random_state=2)
        numeric_features = ['luas lantai tempat tinggal', 'penghasilan kepala rumah tangga']
        numeric_transformer = StandardScaler()
        ordinal_features = ['pendidikan_kepala_rumah_tangga']
        ordinal_transformer = OrdinalEncoder()
        nominal_features = ['jenis lantai tempat tinggal', 'jenis dinding tempat tinggal', 'fasilitas buang air besar',
                           'sumber penerangan rumah', 'bahan bakar untuk memasak', 'jumlah konsumsi daging_susu_sayam_dalam_seminggu',
                           'jumlah kamar dalam rumah', 'jumlah kendaraan pribadi dalam rumah', 'sumber air minum',
                           'mampu membayar biaya pengobatan', 'memiliki simpanan awal']
        nominal_transformer = OneHotEncoder()
        preprocessor = ColumnTransformer(
            transformers=[
                ('numeric', numeric_transformer, numeric_features),
                ('ordinal', ordinal_transformer, ordinal_features),
                ('nominal', nominal_transformer, nominal_features)
            ]
        )
        svm = SVC()
        pipeline = Pipeline(steps=[('preprocessor', preprocessor), ('svm', svm)])
        pipeline.fit(X_train, y_train)
    
```

Gambar 10. Implementasi SVM

Pada tahap modeling data rumah tangga dilatih dengan *Support Vector Machine*

```

# mencari parameter terbaik untuk svm
param_grid = {
    'svm_C': [0.001, 0.01, 0.1, 1, 10, 100],
    'svm_kernel': ['linear', 'rbf', 'poly'],
    'svm_gamma': [0.1, 0.01, 0.001],
}

grid_search = GridSearchCV(estimator=pipeline, param_grid=param_grid, cv=10)
grid_search.fit(X_train, y_train)
best_model = grid_search.best_estimator_
best_params = grid_search.best_params_

# evaluasi model dengan melihat akurasi data training
    
```

Gambar 11. Implementasi Grid Search

Setelah dilakukan *modeling Support Vector Machine* dilatih kembali menggunakan *Grid Search* untuk mencari kombinasi parameter terbaik.

3.5.5. Evaluation

Tabel 3. Confusion Matrix

Aktual	Prediksi	
	Miskin	Tidak miskin
Miskin	TP = 22	FP = 0
Tidak miskin	FN = 1	TN = 32

Confusion Matrix digunakan untuk mengevaluasi model yang telah dilatih dalam memprediksi label status rumah tangga dengan data *training* 70% dan data *testing* 30 %. label status rumah tangga dibagi menjadi miskin dan tidak miskin. Berikut ini adalah hasil perhitungannya :

1. Akurasi = 0,9818
2. *Precision*
 - a. Miskin = 1
 - b. Tidak miskin = 0,9696

3. *Recall*
 - a. Miskin = 0,9565
 - b. Tidak miskin = 1
4. *F1-Score*
 - a. Miskin = 0,98
 - b. Tidak miskin = 0,98

3.5.6. Pengujian

Uji coba aplikasi ini menggunakan cara yaitu uji coba fungsional. Pada tahap ini, pengujian dilaksanakan untuk memastikan bahwa masing-masing bagian dapat berfungsi dengan baik dengan sistem saat ini.

Tabel 4. Uji Coba

No	Menu	Hasil Pengujian
1	Halaman Login	Sesuai
2	Halaman Home Admin	Sesuai
3	Halaman Home Operator	Sesuai
4	Halaman Home Penduduk	Sesuai
5	Halaman Data User	Sesuai
6	Halaman Tambah Data user	Sesuai
7	Halaman Data Penduduk	Sesuai
8	Halaman Tambah Data Penduduk	Sesuai
9	Halaman Data Rumah Tangga	Sesuai
10	Halaman Tambah Data Rumah Tangga	Sesuai
11	Halaman Bantuan	Sesuai
12	Halaman Tambah Data Bantuan	Sesuai
13	Halaman Data Klasifikasi	Sesuai
14	Halaman Hasil Klasifikasi	Sesuai
15	Halaman Modeling	Sesuai
16	Halaman Data Survei	Sesuai
17	Halaman Survei Rumah Tangga	Sesuai
18	Halaman Laporan	Sesuai
19	Halaman Cari Data	Sesuai
20	Halaman Profil	Sesuai
21	Halaman Edit Profil	Sesuai

4. Kesimpulan

Berikut adalah hasil kesimpulan penelitian ini, berdasarkan pelaksanaan dan hasil pembahasan sistem klasifikasi rumah tangga Desa Rigi berbasis web menggunakan metode Support Vector Machine dan Grid Search.

1. Cara membangun sistem klasifikasi rumah tangga menggunakan Support Vector Machine dan Grid Search adalah dengan berbasis web menggunakan bahasa pemrograman Python dan database MySQL lalu dengan library Scikit-Learn digunakan untuk membangun model prediksi yang optimal untuk melakukan prediksi status rumah tangga.
2. Adanya peningkatan akurasi 0.0182 % dari model yang dibangun menggunakan Support Vector Machine dan Grid Search jika dibandingkan dengan model yang dibangun hanya menggunakan Support Vector Machine.
3. Perbandingan hasil antara sistem yang dibangun dengan sistem manual yang ada di Desa Rigi menunjukkan bahwa hanya terdapat satu data yang diprediksi salah jika menggunakan data latih yang sudah ada. Namun jika diujikan dengan data baru yang belum pernah ada sistem belum dapat melakukan prediksi dengan tepat. Hal disebabkan karena jumlah data latih yang digunakan masih sangat sedikit dan terbatas

Daftar Pustaka

- [1] A. B. P. Samudro, *Pengeluaran Untuk Konsumsi Penduduk Indonesia Per Provinsi*. Jakarta: Badan Pusat Statistik RI/BPS-Statistics Indonesia, 2022.
- [2] N. D. Santi, T. Mumtaz, A. D. Fatmawati, and L. Retnosari, *Penghitungan dan Analisis Kemiskinan Makro Indonesia Tahun 2022*. Jakarta: BADAN PUSAT STATISTIK, 2022.
- [3] R. T. Handayanto and Herlawati, *Data Mining dan Machine Learning Menggunakan Matlab dan Python*. Bandung: Bandung : Informatika, 2020.
- [4] F. Effendy and P. Purbandini, "Klasifikasi Rumah Tangga Miskin Menggunakan Ordinal Class Classifier," *J. Nas. Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 30–36, 2018, doi: 10.25077/teknosi.v4i1.2018.30-36.
- [5] S. S. Cembranel, F. Lezama, J. Soares, S. Ramos, A. Gomes, and Z. Vale, "A Short Review on Data Mining Techniques for Electricity Customers Characterization," *2019 IEEE PES GTD Gd. Int. Conf. Expo. Asia, GTD Asia 2019*, pp. 194–199, 2019, doi: 10.1109/GTDAsia.2019.8715891.
- [6] H. A. Issad, R. Aoudjit, and J. J. P. C. Rodrigues, "A comprehensive review of Data Mining techniques in smart agriculture," *Eng. Agric. Environ. Food*, vol. 12, no. 4, pp. 511–525, 2019, doi: 10.1016/j.eaef.2019.11.003.
- [7] M. K. Gupta and P. Chandra, "A comprehensive survey of data mining," *Int. J. Inf. Technol.*, vol. 12, no. 4, pp. 1243–1257, 2020, doi: 10.1007/s41870-020-00427-7.
- [8] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, no. xxxx, 2020, doi: 10.1016/j.neucom.2019.10.118.
- [9] Sulistiana and M. A. Muslim, "Support Vector Machine (SVM) Optimization Using Grid Search and Unigram to Improve E-Commerce Review Accuracy," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 8–15, 2020.
- [10] T. Z. Jasman, E. Hasmin, Sunardi, C. Susanto, and W. Musu, "Perbandingan Logistic Regression, Random Forest, dan Perceptron pada Klasifikasi Pasien Gagal Jantung," *CRSID J.*, vol. 14, no. 3, pp. 271–286, 2022, [Online]. Available: <https://www.doi.org/10.22303/csrid.14.3.2022.271-286>.
- [11] M. K. Dahouda and I. Joe, "A Deep-Learned Embedding Technique for Categorical Features Encoding," *IEEE Access*, vol. 9, pp. 114381–114391, 2021, doi: 10.1109/ACCESS.2021.3104357.
- [12] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognit.*, vol. 91, pp. 216–231, 2019, doi: 10.1016/j.patcog.2019.02.023.
- [13] Z. A. Hasibuan, "Metodologi Penelitian Pada Bidang Ilmu Komputer Dan Teknologi Informasi," *Konsep, Tek. Dan Apl.*, no. Universitas Indonesia, p. 194, 2007.