

# Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Keluarga Sejahtera

1<sup>st</sup> Dwi Tian Tonara  
 Universitas Buana Perjuangan Karawang  
 Karawang, Indonesia  
[if18.dwitonara@mhs.ubpkarawang.ac.id](mailto:if18.dwitonara@mhs.ubpkarawang.ac.id)

2<sup>nd</sup> Ahmad Fauzi  
 Universitas Buana Perjuangan Karawang  
 Karawang, Indonesia  
[afauzi@ubpkarawang.ac.id](mailto:afauzi@ubpkarawang.ac.id)

3<sup>rd</sup> Hilda Yulia Novita  
 Universitas Buana Perjuangan Karawang  
 Karawang, Indonesia  
[hilda.yulia@ubpkarawang.ac.id](mailto:hilda.yulia@ubpkarawang.ac.id)

**Abstract**—Setiap keluarga menginginkan kehidupan yang sejahtera dan memenuhi setiap kebutuhan baik primer maupun sekunder. Kartu keluarga sejahtera merupakan salah satu program pemerintah untuk menanggulangi kemiskinan. Penentuan bantuan pemerintah agar tepat sasaran dapat dilakukan melalui klasifikasi keluarga yang layak dan tidak layak. Data didapat dari pemerintahan Desa Mekarmaya sebanyak 432 data, dengan data layak sebanyak 25 data dan tidak layak sebanyak 407 data, yang diterapkan dengan menggunakan algoritma *Support Vector Machine* (SVM). Diuji menggunakan *orange* dan *python* dengan menghasilkan data layak sebanyak 7 data dan data tidak layak sebanyak 80 data, dengan hasil accuracy sebesar 94%.

**Kata Kunci:** Klasifikasi, Data Mining, Keluarga Sejahtera, *Support Vector Machine* (SVM)

## I. PENDAHULUAN

Keluarga Sejahtera dibentuk oleh perkawinan yang sah, yang dilandasi dengan bertaqwah kepada Tuhan yang Maha Esa, mampu memenuhi kebutuhan hidup baik lahir, spiritual maupun sosial, memiliki tujuan yang seimbang antar anggota serta antar keluarga dengan masyarakat di lingkungannya. Setiap pemerintahan memiliki penduduk yang tingkat kesejahteraan berbeda, hal ini dapat dilihat berdasarkan upaya pemerintah untuk mencapai tujuannya dalam meningkatkan kesejahteraan penduduk. Oleh karena itu, pemerintah melaksanakan berbagai program demi meningkatkan kesejahteraan penduduknya. Namun, pemerintah terkadang masih terdapat kesalahan dalam penerimaan bantuan, salah satunya yaitu seringkali salah atau tidak tepat sasaran dalam menentukan antara penduduk yang layak untuk mendapatkan bantuan dan tidak layak mendapatkan bantuan[1].

Upaya pemerintah dalam menanggulangi kemiskinan yaitu dengan mengeluarkan Kartu Keluarga Sejahtera (KKS). Kartu keluarga Sejahtera merupakan salah satu program pemerintah untuk percepatan dalam menanggulangi kemiskinan. Hal ini tercantum dalam Peraturan Presiden (Perpres) Nomor 166 Tahun 2014. Program Kartu Keluarga Sejahtera ini memiliki berbagai macam kriteria, jumlah penduduk yang sangat banyak menyulitkan tim penyeleksi dalam menyeleksi masyarakat yang sudah tepat sasaran atau belum. Desa Mekarmaya merupakan salah satu daerah pedesaan yang berada di Kecamatan Cilamaya Wetan, dengan tingkat kemiskinan yang sangat tinggi. Tetapi dalam melakukan program Kartu Keluarga Sejahtera (KKS) masih terdapat kesalahan dalam memberikan bantuan pada masyarakat yang tidak tepat sasaran. Seperti masyarakat yang masuk dalam kategori mampu, masih mendapatkan program bantuan Kartu Keluarga Sejahtera (KKS), sedangkan yang tidak mampu masih banyak yang belum mendapatkan bantuan kartu keluarga sejahtera[2], dalam permasalahan ini penulis akan menyelesaikan masalah dengan membandingkan nilai akurasi antara layak dan tidak layak menggunakan algoritma *Support Vector Machine* (SVM).

Penelitian terkait telah dilakukan oleh Setiawan Menggunakan Fuzzy Tsukamoto di Desa Ngerawan Kecamatan Berbek Kabupaten Nganjuk, dalam penelitiannya mampu diperoleh nilai pengujian sebesar 88%[3]. Selanjutnya, Suhendra,, M. A., Ispriyanti, D dan Sudarsono,S Penelitian nilai accuracy dalam menggunakan Metode CHAID menghasilkan nilai sebesar 90.2% sedangkan dengan menggunakan regresi logistik biner menghasilkan nilai sebesar 88%.

Penelitian terkait lainnya telah dilakukan oleh Rahmayanti, Azhar, dan Pramudita, dalam penelitiannya memiliki 15 atribut. Dengan nilai accuracy sebesar 91.9%. Pengujian klasifikasi menghasilkan nilai *accuracy* dan 69.8% dengan menggunakan hasil seleksi fitur algoritma C5[4]. Kemudian, Permana, amril mutoi, dan anis fitri nur masruriyah dalam penelitiannya menggunakan algoritma C5.0 dengan 120 data. Penelitian ini memiliki 9 atribut. Hasil dari penelitian ini adalah algoritma C5.0 nilai *accuracy*nya yaitu 86.67% sedangkan untuk algoritma KNN mendapatkan nilai accuracy sebesar 83.33%.[5]

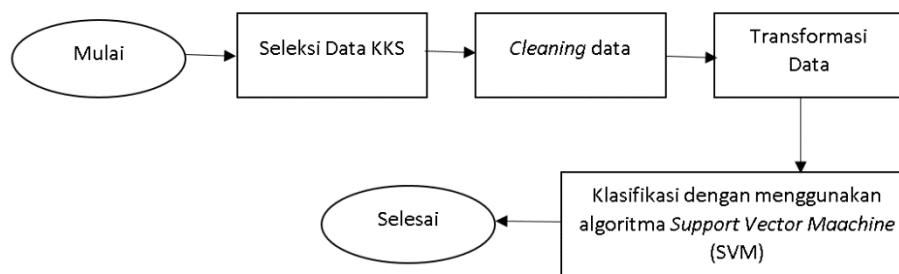
## II. METODOLOGI PENELITIAN

### A. Objek Penelitian

Klasifikasi Kartu Keluarga Sejahtera (KKS) dengan menerapkan metode *Support Vector Machine* (SVM) untuk mengklasifikasi Dataset yang didapat dari instansi pemerintahan Desa Mekarmaya. Jumlah data sebanyak 432 data klasifikasi yang dilakukan terdiri atas layak dan tidak layak

**B. Prosedur Penelitian**

Pada bagian prosedur penelitian dijelaskan proses penelitian sebagai berikut :

**C. Seleksi Data KKS**

Merupakan seleksi data yang tidak mengandung *noise*, dihapus atau dihilangkan Sebagian. Untuk mengaccuracykan kesalahan atau error dalam algoritma *Support Vector Machine* (SVM) terhadap klasifikasi kartu keluarga sejahtera data tersebut di *Cleaning* terlebih dahulu. Data diperoleh dari Intansi Pemerintahan Desa Mekarmaya yang berjumlah 432 data dengan 7 atribut, antara lain terdiri dari jumlah anggota keluarga, penghasilan kepala keluarga, pekerjaan kepala keluarga, Pendidikan kepala keluarga, status tempat tinggal, jenis lantai tempat tinggal, kondisi rumah, dan nilai dari layak atau tidak layaknya penerima kartu. Data yang cantumkan diambil 2 dusun dari 6 dusun yang terdapat di desa mekarmaya yang kemudian data tersebut dibagi menjadi dua bagian yaitu menjadi data *training* 80% dan data *testing* 20%. Sehingga menjadi 346 data *training* dan 87 data *testing* dan 4 atribut diantaranya yaitu jumlah anggota keluarga, penghasilan kepala keluarga, pekerjaan kepala keluarga, Pendidikan kepala keluarga

**D. Cleaning Data**

*Cleaning* pada tingkatan ini dilakukan *cleaning* data yang merupakan titik utama dalam data mining. *Cleaning* data rancang untuk mengganti data yang tidak lengkap atau terdapat *missing value*. Proses cleaning data dilaksanakan secara manual untuk menetapkan bahwa data yang terseleksi layak untuk diinput pada proses pemodelan.

**E. Transformasi Data**

Transformasi atau merombak data sehingga membuat data yang sesuai dengan ketentuan. Fitur-fitur dalam transformasi sangat bermanfaat untuk menerangkan data yang bergantung pada goals yang ingin dituju dan proses transformasi pada data yang terpilih, sehingga data sesuai dengan proses *data mining*. Data yang telah diseleksi kemudian di transformasi atau dibentuk sesuai dengan variabel yang sudah ditentukan. Seleksi data dilakukan secara manual dengan menggunakan Microsoft excel. Atribut penghasilan kepala keluarga, pekerjaan kepala keluarga, Pendidikan kepala keluarga diganti menjadi variabel *numeric*.

**F. Klasifikasi Data Menggunakan Algoritma Support Vector Machine (SVM)**

Berikut merupakan tahapan klasifikasi data menggunakan algoritma *Support Vector Machine* (SVM) :

- A. Klasifikasi Data Dengan Hitung Manual
  1. Menerapkan perhitungan eliminasi
  2. Melakukan eliminasi  $w_4$  (dengan perkalian silang indeks  $w_4$ )
  3. Melakukan eliminasi  $w_3$  (dengan perkalian silang indeks  $w_3$ )
  4. Mencari nilai  $w_2$  pada persamaan 13 dan 14
  5. Mencari nilai  $w_2$  dengan menggunakan persamaan 14 dan  $w_1$
  6. Mencari nilai  $w_3$  dengan menggunakan persamaan 12,  $w_1$  dan  $w_2$
  7. Mencari nilai  $w_4$  dengan menggunakan persamaan 12,  $w_1$ ,  $w_2$ , dan  $w_3$
  8. Mencari bias ( $b$ ) dengan menggunakan persamaan 5 dengan mensubstitusikan  $w_1$ ,  $w_2$ ,  $w_3$ , dan  $w_4$
- B. Klasifikasi dengan menggunakan *tools orange*
  1. *Import Dataset*
  2. Pengukuran ROC (*Receiver Operating Characteristic*) dan *Confusion Matrix*
  3. *Prediction* Algoritma SVM
  - Visualisasi ROC
- C. Klasifikasi Data menggunakan *python*.
  1. *Import Library* dan *dataset*
  2. Menampilkan data yang terdapat pada dataset
  3. *Drop* atau menghilangkan kolom atau variabel yang tidak dibutuhkan
  4. Pilih variabel yang akan dihilangkan, pada penelitian ini akan menghilangkan variabel ‘nama’ dan ‘keterangan’
  5. Membagi dataset kedalam data *training* dan *testing* i set sebesar 20%
  6. Membuat prediksi terhadap data i
  7. Mengevaluasi algoritma untuk menghitung *precision*, *recall*, dan *accuracy*

8. Melatih algoritma SVM dengan kernel linear
9. Menampilkan koordinat *support vector*
10. Membuat grafik histogram

### III. HASIL DAN PEMBAHASAN

#### A. Seleksi Data KKS

Data didapat dari Intansi Pemerintahan Desa Mekarmaya yang berjumlah 432 data dengan 7 atribut, antara lain terdiri dari jumlah anggota keluarga, penghasilan kepala keluarga, pekerjaan kepala keluarga, Pendidikan kepala keluarga, status tempat tinggal, jenis lantai tempat tinggal, kondisi rumah, dan nilai dari layak atau tidak layaknya penerima kartu. Berikut adalah data masyarakat di Desa Mekarmaya :

No. KK	nama	istri	jumlah	penghasilan k	Nilai
3215152506070042	AGUS SALIM	YAYAH	3	3000000	TIDAK LAYAK
3215151601160009	HJ.ROSIDAH	HJ.ROSIDAH	3	1000000	TIDAK LAYAK
3215152105120004	HODIJAH	HODIJAH	3	3500000	TIDAK LAYAK
3215150907070087	HAMSYAH	CUCUNNAYAH	4	4500000	TIDAK LAYAK
3215151007070005	H.DADANG KOMARU	H.JAAN NURAN	3	4000000	TIDAK LAYAK
3215150907070084	DITA	TUTI KURNIATI	3	2500000	TIDAK LAYAK
3215150309130002	AJO	EMAYANA	4	3500000	TIDAK LAYAK
3215151310100002	DADAN SOBARNA	IMAS MASRYAH	5	3000000	TIDAK LAYAK
3215150809140005	SEHUDIN, S E	SUSILAWATI,SH	4	4500000	TIDAK LAYAK
3215153001090001	UJANG ARIPIN	DAENAH	6	1500000	TIDAK LAYAK
3215151612110016	KARTIM	YAYAH	4	2900000	TIDAK LAYAK
3215152003090007	KASMAN	DESIH	5	3000000	TIDAK LAYAK
3215150507100025	DEDI MULYADI	SITI KHODIJAH	4	3000000	TIDAK LAYAK
3215151609070026	EDI JUNAEDI	RUMSARI	4	4500000	TIDAK LAYAK
3215152012110016	WAWAN HERMAWAN	INTA ERLINA	4	3500000	TIDAK LAYAK
3215151007070049	WAHAB	TOYAH	4	2400000	LAYAK

Gambar 2 Hasil Pengumpulan Data

#### B. Cleaning Data

Data yang sebelumnya memiliki 7 variabel setelah di proses cleaning data menjadi 4 variabel. Seperti data berikut :

jumlah ang	penghasilan	pekerjaan	pendidikan kepala keluarga	nama
6	4200000	WIRASWASTA	SMA/SMA	ELIS ELI SUSANTI
2	2000000	BURUH TANI/PERKEBUNAN	SD/SD	YAKUB
4	4200000	WIRASWASTA	SMA/SMA	SAEPUL ROHMAN
1	3500000	WIRASWASTA	SMA	LISNAWATI
4	3250000	WIRASWASTA	SMP/SMA	RIYA SUPRIYATIN
4	3000000	BURUH HARIAN LEPAS	SD/SD	RUMINAH
4	2800000	PEDAGANG	SD	RAONAH
3	4500000	KARYAWAN SWASTA	SMA/SMA	IWAN HERMAWAN
5	3750000	WIRASWASTA	SMA/SMA	ABDUL RAHMAT

Gambar 3 Data setelah di Cleaning

#### C. Hasil Transformasi Data

Data yang telah diseleksi kemudian di transformasi atau dibentuk sesuai dengan variabel yang sudah ditentukan. Seleksi data dilakukan secara manual dengan menggunakan Microsoft excel. Atribut penghasilan kepala keluarga, pekerjaan kepala keluarga, Pendidikan kepala keluarga diganti menjadi variabel numeric. Hasil dari transformasi data adalah sebagai berikut :

nama	jumlah anggota keluarga	penghasilan kepala keluarga	pekerjaan	pendidikan kepala keluarga	
AGUS SALIM	3	3	3	3	3
HJ.ROSIDAH	3	7	9		3
HODIJAH	3	1	7		3
HAMSYAH	4	1	5		3
HJ.AAN NURANILAH	3	1	3		1
TUTI KURNIATI	3	3	7		7
EMAYANA	4	1	3		7
IMAS MASRYAH	5	3	3		3
SUSILAWATI,SH	4	1	7		1
UJANG ARIPIN	6	5	9		7
KARTIM	4	3	3		7
KASMAN	5	3	3		7
SITI KHODIJAH	4	3	3		3
RUMSARI	4	3	7		7
INTA ERLINA	4	1	3		7
WAHAB	4	3	7		7
WIHARJA	3	5	5		5

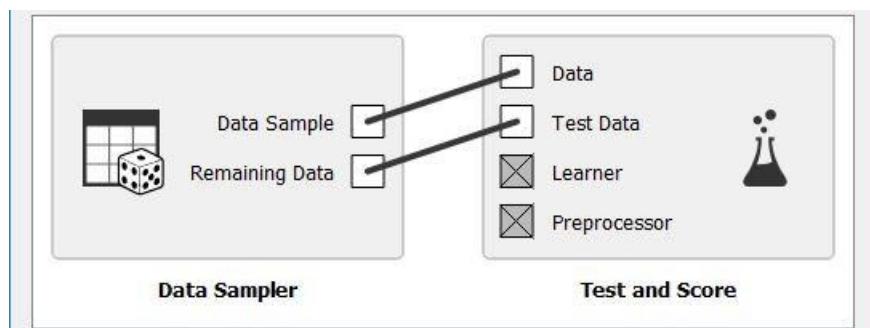
**D. Klasifikasi Data Data Menggunakan Algoritma Support Vector Machine (SVM)**a. Klasifikasi Menggunakan *tools orange*

## 1. Import Dataset



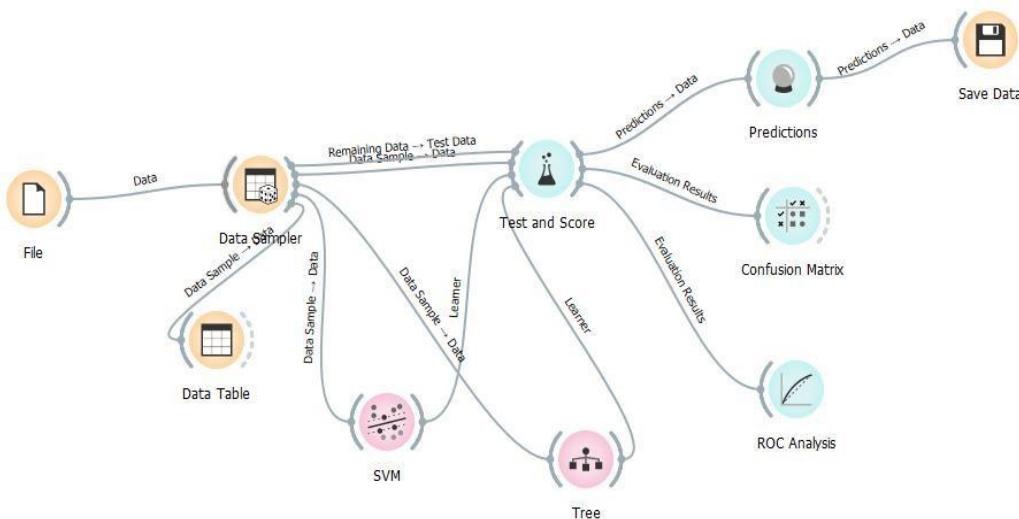
## 2. Pengukuran ROC dan Confusion Matrix

- Buka aplikasi orange kemudian klik file lalu memilih data yang akan diuji
- Setelah itu *drag and drop* button file lalu sambungkan dengan data *sampler*. Lalu disambungkan dengan testing and score. Setelah tersambung kemudian double klik untuk *settings* data *training* agar tergolong sebagai data sampel dan Sebagian data lainnya sebagai data testing seperti gambar *tools sampler* dan *testing* berikut ini :



- Hubungkan data sampler dengan algoritma SVM dan Tree kemudian dihubungkan ke dalam testing and score.
- Tambahkan button *Predictions*, *ROC Analysis*, dan *Confusion Matrix*.
- Button Testing and Score* disambukan dari arah kanan ke arah kiri *button predictions*, *ROC Analysis*, dan *confusion matrix*

Ilustrasi dari tahapan-tahapan predictions, ROC Analysis, dan confusion matrix diatas, ada pada gambar dibawah ini :



b. Klasifikasi Menggunakan *python*

- Import Library dan dataset
- Menampilkan data yang terdapat pada dataset
- Drop atau menghilangkan kolom atau variabel yang tidak dibutuhkan
- Pilih variabel yang akan dihilangkan, pada penelitian ini akan menghilangkan variabel ‘nama’ dan ‘keterangan’
- Membagi dataset kedalam data training dan testing i set sebesar 20%
- Membuat prediksi terhadap data i
- Mengevaluasi algoritma untuk menghitung precision, recall, dan accuracy
- Melatih algoritma SVM dengan kernel linear
- Menampilkan koordinat support vector
- Membuat grafik histogram

E. Pengujian dengan menggunakan *orange* dan *python*

1. Pengujian dengan *orange*

Berdasarkan data *training* sebanyak 80% dan data *testing* sebanyak 20%. *Confusion matrix* pada tools *orange* digunakan untuk mendeteksi data yang layak dan tidak layak. Hasil nilai dari *confusion matrix* seperti pada gambar *confusion matrix* dibawah ini :

		Predicted		$\Sigma$
		LAYAK	TIDAK LAYAK	
Actual	LAYAK	6	1	7
	TIDAK LAYAK	0	80	80
		$\Sigma$	81	87

*Confusion matrix* dihitung menggunakan perhitungan secara manual. Berdasarkan data yang sudah diolah oleh *orange* menghasilkan dua nilai yaitu layak dan tidak layak dari masing-masing nilai *accuracy*, *precision*, *recall*, sebagai berikut :

F. Hasil Pengujian

1. Hasil Pengujian dengan *orange*

Implementasi algoritma SVM pada <i>orange</i>	Accuracy	Recall	Precision
layak	98,9%	100%	98,7%
Tidak layak	98,9%	85%	100%

Perhitungan hasil *accuracy confusion matrix* menggunakan *tools orange* dengan mengimplementasikan algoritma *Support Vector Machine* (SVM) dapat dilihat bahwa *accuracy* pada *Support Vector Machine* memiliki *accuracy* sebesar 98,9% baik layak maupun tidak layak. Hasil *precision* pada *Support Vector Machine* layak menghasilkan angka 98,7% dan tidak layak 100%. Hasil *recall* pada *Support Vector Machine* nilai layak 100% dan nilai tidak layak 85%.

## 2. Hasil Pengujian dengan *python*

Implementasi <i>python</i> SVM	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>
layak	94%	0%	0%
Tidak layak	94%	100%	94%

Perhitungan hasil *accuracy confusion matrix* menggunakan *python* dengan mengimplementasikan algoritma *Support Vector Machine*. Dapat dilihat bahwa *accuracy* pada *Support Vector Machine* mendapatkan nilai sebesar 94%. Hasil *precision* pada *Support Vector Machine* layak menghasilkan angka 0% dan tidak layak 94%, kemudian hasil *recall* pada *Support Vector Machine* nilai layak 0% dan nilai tidak layak 100%.

## IV. KESIMPULAN DAN SARAN

### A. Kesimpulan

Berdasarkan hasil dan pembahasan diperoleh beberapa kesimpulan sebagai berikut :

1. Dengan menerapkan algoritma *Support Vector Machine* untuk klasifikasi suatu data, dapat disimpulkan bahwa algoritma *Support Vector Machine* (SVM) dapat membantu dalam menentukan masyarakat yang mampu dan tidak mampu, atau layak dan tidak layak dalam penerimaan bantuan kartu keluarga sejahtera (KKS).
2. Dengan menggunakan nilai *accuracy* yang diterapkan dengan menggunakan *python*, menghasilkan nilai *accuracy* 94%. Sedangkan nilai *accuracy* yang diterapkan menggunakan *orange* menghasilkan *accuracy* sebesar 98,9% .

### B. Saran

Perlu adanya implementasi menggunakan tools lain selain *orange* dan *python google colab* untuk mengetahui nilai dari algoritma *Support Vector Machine* (SVM). Serta dapat menggunakan algoritma dengan metode yang lain untuk membandingkan nilai *accuracy* pada klasifikasi.

## PENGAKUAN

Naskah ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Dwi Tian Tonara, dengan judul Penerapan Algoritma *Support Vector Machine* (SVM) untuk Klasifikasi Keluarga Sejahtera (studi kasus: Desa Mekarmaya Karawang) yang dibimbing oleh Ahmad Fauzi dan Hilda Yulia Novita.

## DAFTAR PUSTAKA

- [1] Widyadara, M. A. D., & Irawan, R. H. (2019). Implementasi Metode Naïve Bayes dalam Penentuan Tingkat Kesejahteraan Keluarga. RESEARCH : Computer, Information System & Technology Management, 2(1), 19.
- [2] Suhendra, M. A., Ispriyanti, D., & Sudarno, S. (2020). Ketepatan Klasifikasi Pemberian Kartu Keluarga Sejahtera Di Kota Semarang Menggunakan Metode Regresi Logistik Biner Dan Metode Chaid. Jurnal Gaussian, 9(1), 64–74.
- [3] Setiawan, A. B., & Kom, M. (2018). Sistem Pendukung Keputusan Menentukan Penerimaan Kartu Keluarga Sejahtera (KKS) Menggunakan Fuzzy Tsukamoto Didesa Ngerawan Kec. Berbek Kab. Nganjuk Oleh : Syeh Lendi Artana Dibimbing Oleh : Program Studi Teknik Informatika Fakultas Teknik (FT) Univ.
- [4] Rahmayanti, V., Azhar, Y., & Pramudita, A. E. (2020). Penerapan algoritma C5.0 pada analisis faktor-faktor pengaruh kelulusan tepat waktu mahasiswa Teknik Informatika UMM. Jurnal Repositor, 1(2), 131.
- [5] Permana, T., Siregar, A. M., Masruriyah, A. F. N., & Juwita, A. R. (2020). Perbandingan Hasil Prediksi Kredit Macet Pada Koperasi. Conference on Innovation and Application of Science and Technology, 3(1), 737–746.