

PENERAPAN DATA MINING UNTUK MEMPREDIKSI TRANSAKSI NASABAH DENGAN ALGORITMA C4.5

Desi Marlina¹, Muhammad Bakri²

Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia¹
Teknik Komputer, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia²

desi_marlina@teknokrat.ac.id¹, muhammadbakri@teknokrat.ac.id²

Received: (8 Maret 2021) **Accepted:** (15 Maret 2021) **Published:** (29 Maret 2021)

Abstract

Lending activities are one of the activities carried out by savings and loan cooperatives every day, in which there are customer transactions. This research is devoted to classifying in predicting customer transactions that will be in arrears or smoothly in installment payments at the BMT Sahabat Rawajitu Selatan cooperative. The problem faced in this study is the cooperative BMT Sahabat Rawajitu Selatan in conducting analyzes of each customer transaction is very lacking. This can be seen from the existence of customers who are in arrears for their installment payments, even when there are customers who find it difficult to make their installment payments. The method used in this research is data classification is C4.5 using primary data obtained from the cooperative BMT Sahabat Rawajitu Selatan for the classification process. This determination is classified as a classification process using Rapid Miner for classification and using cross validation as validation of data accuracy. The results of this study are the prediction of customer transactions of the cooperative BMT Sahabat Rawajitu Selatan to produce information, which is used to assist the financing (loan) division in determining which customers are likely to be in arrears or smoothly in making payments in order to make decisions related to customer transactions to be carried out. by customers for the future, that is, they will be in arrears or withdrawals in making payments.

Keywords: Cooperatives, Savings and Loans, Data Mining, C4.5 Algorithm, and Cross Validation.

Abstrak

Kegiatan peminjaman merupakan salah satu kegiatan yang dilakukan koperasi simpan pinjam setiap harinya, dimana didalam nya terdapat transaksi-transaksi nasabah. Penelitian ini di khususnya untuk mengklasifikasi dalam memprediksi transaksi nasabah yang akan nunggak atau lancar dalam pembayaran angsuran pada koperasi BMT Sahabat Rawajitu Selatan. Adapun masalah yang di hadapi pada penelitian ini adalah koperasi BMT Sahabat Rawajitu Selatan dalam melakukan analisis-analisis setiap transaksi nasabahnya sangat lah kurang. Hal ini dapat dilihat dari adanya nasabah-nasabah yang nunggak pembayaran angsuran bahkan sampai ada nasabah yang sulit dalam melakukan pembayaran angsurannya. Metode yang digunakan dalam penelitian ini adalah klasifikasi data adalah C4.5 dengan menggunakan data primer yang didapat dari koperasi BMT Sahabat Rawajitu Selatan untuk proses klasifikasi. Penentuan tersebut digolongkan sebagai proses pengklasifikasian menggunakan Rapid Miner untuk klasifikasi dan menggunakan cross validation sebagai validasi akurasi data. Hasil penelitian ini adalah prediksi transaksi nasabah koperasi BMT Sahabat Rawajitu Selatan untuk menghasilkan suatu informasi, yang digunakan untuk membantu bagian pembiayaan (pinjaman) dalam menentukan nasabah mana yang kemungkinan akan nunggak ataupun lancar dalam melakukan pembayaran guna mengambil keputusan yang berkaitan dengan transaksi nasabah yang akan dilakukan oleh nasabah untuk masa yang akan datang yaitu akan nunggak ataupun lancar dalam melakukan pembayaran.

Kata Kunci: Koprasi, Simpan Pinjam, Data Mining, Algoritma C4.5, dan Cross Validation.

To cite this article:

Marlina, Bakri. (2021). Penerapan Data Mining Untuk Memprediksi Transaksi Nasabah Dengan Algoritma C4.5. Jurnal Teknologi dan Sistem Informasi, Vol (2) No.1 , 23 - 28

1. Pendahuluan

Salah satu jenis koperasi syariah yang ada pada saat ini adalah koperasi Baitul Maal Wattamwil (BMT) Sahabat Rawajitu selatan. Koperasi BMT Sahabat Rawajitu Selatan adalah koperasi yang bergerak di unit simpan pinjam. Koperasi ini berada di Jalan Poros Gedung Karya Jitu, Kec. Rawajitu Selatan, Kab. Tulang Bawang. Koperasi ini merupakan salah satu koperasi jasa keuangan syariah (KJKS) yang berada di Kecamatan Rawajitu Selatan dan kegiatan atau fungsinya nya pun sama dengan koperasi umum lainnya yaitu anggota koperasi yang kelebihan dana dan menyimpan dana nya di koperasi, akan disalurkan kembali pada anggotanya yang membutuhkan dana dan jika memungkinkan pihak koperasi juga dapat meminjamkan dananya kepada masyarakat yang membutuhkan, namun yang membedakannya adalah menggunakan sistem bagi hasil.

Saat ini Koperasi BMT Sahabat Rawajitu Selatan dalam melakukan analisis-analisis setiap transaksi nasabahnya sangat lah kurang. Hal ini dapat dilihat dari nasabah-nasabah yang banyak sekali nunggak pembayaran angsuran bahkan sampai ada nasabah yang sulit dalam melakukan pembayaran angsurannya. Serta analisa yang dilakukan manager koperasi selama ini adalah dengan melakukan pendekatan personal pada semua nasabah kredit. Hal ini dilakukan karena keterbatasan tenaga analisis yang cukup mahal. Hal demikian lah yang memicu terjadinya nasabah nunggak dalam pembayaran. Dengan ada nya masalah tersebut cara atau metode yang dapat dilakukan yaitu dengan menggunakan metode klasifikasi. Metode klasifikasi ini untuk mengelompokan data mining dan menentukan sebuah record data baru [1], [2], [3], [4], [5]. Pada data mining pengelompokan data atau yang sering disebut dengan klasifikasi data ini sangat berguna untuk memprediksi transaksi nasabah mana yang akan nunggak ataupun lancar dalam pembayaran. Salah satu metode yang digunakan untuk klasifikasi adalah C4.5 dimana algoritma ini adalah salah satu teknik decision tree yang sering digunakan dan menghasilkan sebuah pohon keputusan yang lebih mudah di pahami [6], [7], [8], [9], [10].

Tujuan penelitian ini adalah untuk mengetahui klasifikasi dalam memprediksi transaksi nasabah Koperasi BMT Sahabat Rawajitu Selatan menggunakan algoritma C4.5. Diharapkan dengan diterapkannya algoritma C4.5 ini dapat membantu karyawan koperasi bagian pembiayaan untuk dapat memprediksi nasabah mana yang nantinya akan lancar atau pun nunggak dalam pembayaran. Dan hasil dari penelitian ini pun diharapkan dapat digunakan untuk membantu menentukan strategi baru dalam analisis transaksi nasabah.

2. Tinjauan Pustaka

2.1. Data Mining

Perkembangan data mining yang pesat tidak dapat lepas dari perkembangan teknologi informasi yang memungkinkan data dalam jumlah yang besar terakumulasi. Tetapi pertumbuhan yang pesat dari akumulasi data telah menciptakan suatu kondisi yang disebut dengan “rich of data but poor of information” karena data yang terkumpul itu tidak dapat digunakan dalam suatu aplikasi yang berguna. Bahkan tidak jarang kumpulan data tersebut dibiarkan begitu saja sehingga tercipta “data tombs” (kuburan data) [11], [12].

2.1.1. Tahapan – Tahapan Data Mining

Karena data mining adalah suatu rangkaian proses maka dibagi menjadi beberapa tahap antara lain [13], [14] :

- Pembersihan data: untuk membuang data yang tidak konsisten dan *noise*.
- Integrasi data: untuk menggabungkan data dari beberapa sumber.
- Transformasi data : untuk mengubah data menjadi bentuk yang sesuai untuk di *mining*.
- Aplikasi teknik *data mining*.
- Evaluasi pola yang ditemukan : untuk menemukan informasi yang menarik atau pun bernilai.
- Presentasi pengetahuan dengan teknik visualisasi.

2.2. Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan “jika maka”, decision tree ataupun formula matematis. Decision tree merupakan salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasikan oleh manusia. Serta setiap percabangan menyatakan kondisi yang harus dipenuhi dan tiap ujung pohon menyatakan kelas data atau atribut data [15],[16].

2.3. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang sangat populer yang digunakan oleh banyak peneliti di dunia, Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang di ciptakan oleh J. Rose Quinlan. Adapun tahapan-tahapan untuk melakukan perhitungan algoritma C4.5 diantaranya, menyiapkan data training, menentukan akar dari pohon dengan menghitung entropy, kemudian menghitung nilai gain, setelah itu menentukan tupel yang ingin dipartisi [6], [12].

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan :

S : Himpunan Kasus

A : Atribut

S_i : Jumlah Kasus pada Partisi ke-i

N : Jumlah Partisi

Atribut A | S_i : Jumlah Kasus dalam S

Adapun untuk mencari nilai Entropy, digunakan rumus sebagai berikut :

$$Entropy(S) = \sum_{i=1} - p_i * log_2 p_i$$

Keterangan :

S : Himpunan Kasus

A : Fitur

N : Jumlah Partisi S

p_i : Proporsi dari S_i thdp S

2.4. Cross Validation

Validasi dan pengujian adalah Pengujian yang dilakukan untuk mengetahui semua fungsi bekerja dengan baik atau tidak. Validasi dilakukan dengan Ten-fold Cross Validation. Tenfold Cross Validation adalah validasi yang dilakukan dengan cara membagi suatu set data menjadi sepuluh segmen yang berukuran sama besar dengan cara melakukan pengacakan data. Validasi dan pengujian dilakukan untuk mengetahui tingkat akurasi, presisi, dan recall dari hasil prediksi klasifikasi. Akurasi adalah persentase dari catatan yang diklasifikasikan dengan benar dalam pengujian dataset. Presisi adalah persentase data yang diklasifikasikan sebagai model baik yang sebenarnya juga baik. Recall adalah pengukuran tingkat pengenalan positif sebenarnya [17].

3. Metode Penelitian

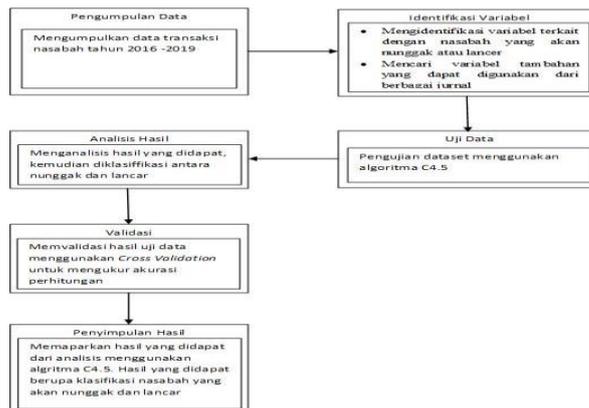
3.1. Tahapan Penelitian

Metode penelitian merupakan tata cara dalam tahapan penelitian pada dasarnya adalah tahapan-tahapan antara konsep-konsep yang ingin diamati atau diukur melalui penelitian yang akan dilakukan. Tahap Penelitian dapat dilihat pada Gambar 1 berikut:

3.2. Identifikasi Variabel

Identifikasi ini terdiri dari 4 variabel atau fitur yaitu:

1. Gaji Bersih,
2. Jumlah Kredit,
3. Jumlah Tanggungan,
4. Inflasi,



Gambar 1. Tahapan Penelitian

4. Hasil dan Pembahasan

4.1. Hasil Perhitungan Manual C4.5

Data yang menjadi variabel penentu dalam pembentukan pohon keputusan pada penelitian ini adalah gaji bersih, jumlah kredit, inflasi, dan jumlah tanggungan. Setelah itu, sebelum kita menghitung nilai again dan entropy dari masing-masing variabel, hal pertama yang harus dilakukan adalah mencari nilai kategori setiap variabelnya sebagai berikut:

Tabel 1 Kategori Nilai Jumlah Tanggungan

Atribut	Nilai
Jumlah Tanggungan	1-3
	4-6
	>6

Sumber: (Purwanto & Taftazani, 2018)

Tabel 2 Kategori Nilai Gaji Bersih

Atribut	Nilai
Gaji Bersih	<1.5jt
	1.5jt-2.4jt
	>=2.5jt

Sumber: (Dewi & Latrini, 2020)

Tabel 3 Kategori Nilai Jumlah Kredit

Atribut	Nilai
Jumlah Kredit	1jt - 5jt
	6jt - 10jt
	>=11jt

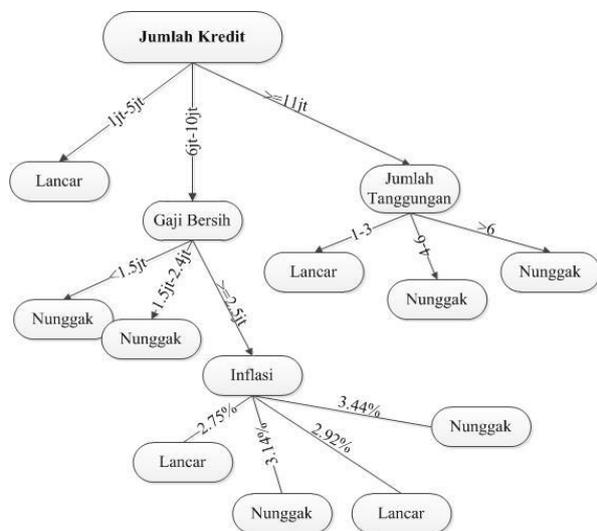
Sumber: (Anggota, 2016)

Tabel 4 Kategori Nilai Inflasi

Atribut	Nilai
Data Inflasi	2,75%
	3,14%
	2,92%
	3,44%

Sumber: (Statistik, 2019)

Setelah itu, langkah selanjutnya adalah menentukan nilai entropy dan gain sesuai rumus yang ada dan menghasilkan bahwa atribut yang memiliki nilai gain tertinggi dan dijadikan sebagai node akar pohon keputusan adalah jumlah kredit. Dengan demikian terbentuklah pohon keputusan akhir sebagai berikut:



Gambar 2 Pohon Keputusan

Dari pohon keputusan yang didapat maka aturan atau rule yang terbentuk adalah sebagai berikut :

1. *If* Jumlah Kredit = 1-5jt *Then* keputusan=Lancar
2. *If* Jumlah Kredit = 6jt-10jt, *And* Gaji Bersih = <1.5jt *Then* keputusan=Nunggak
3. *If* Jumlah Kredit = 6jt-10jt, *And* Gaji Bersih = 1.5jt-2.4jt, *Then* keputusan=Nunggak
4. *If* Jumlah Kredit = 6jt-10jt, *And* Gaji Bersih = >=2.5jt, *And* inflasi = 2.75% *Then* keputusan = Lancar
5. *If* Jumlah Kredit = 6jt-10jt, *And* Gaji Bersih = >=2.5jt, *And* inflasi = 3.14% *Then* keputusan = Nunggak
6. *If* Jumlah Kredit = 6jt-10jt, *And* Gaji Bersih = >=2.5jt, *And* inflasi = 2.92% *Then* keputusan = Lancar
7. *If* Jumlah Kredit = 6jt-10jt, *And* Gaji Bersih = >=2.5jt, *And* inflasi = 3.44% *Then* keputusan = Nunggak
8. *If* Jumlah Kredit = >=11jt, *And* Jumlah Tanggungan = 1-3 *Then* keputusan=Lancar
9. *If* Jumlah Kredit = >=11jt, *And* Jumlah Tanggungan = 4-6 *Then* keputusan=Nunggak
10. *If* Jumlah Kredit = >=11jt, *And* Jumlah Tanggungan = >6 *Then* keputusan=Nunggak

4.2. Pembahasan Hasil Pengujian

Berdasarkan hasil pengujian yang akan dilakukan dengan tahapan data mining yaitu :

1. Pembersihan Data
2. Integrasi Data
3. Transformasi Data
4. Pengujian
5. Evaluasi Pola Nilai
6. Teknik Presentasi

Maka menghasilkan suatu tabel prediksi sebagai berikut:

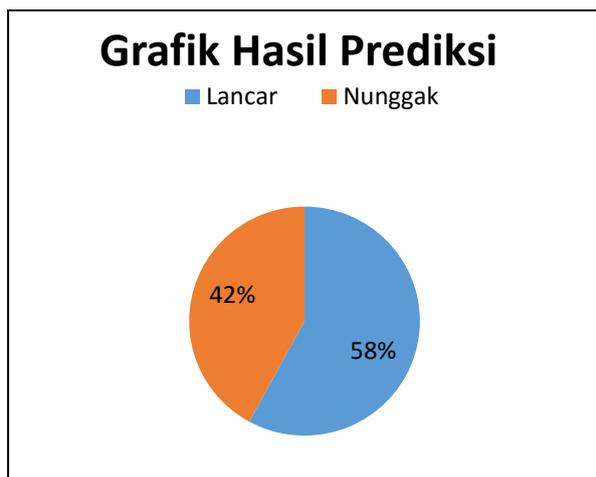
Tabel 5 Hasil Prediksi

Data Trining:				
Keputusan	True Lancar	True Nunggak	Class Precision	Akurasi
Pred. Lancar	587	202	74.40%	
Pred. Nunggak	168	100	37.31%	65.00%
Class Recall	77.75%	33.11%		
Data Testing:				
Keputusan	True Lancar	True Nunggak	Class Precision	Akurasi
Pred. Lancar	138	17	89.03%	
Pred. Nunggak	15	94	86.24%	87.88%
Class Recall	90.20%	84.68%		

Berdasarkan hasil prediksi yang sudah tertera pada tabel di atas yang menghasilkan nilai sebagai berikut:

1. **Akurasi**
Merupakan hasil perhitungan dari klasifikasi data benar di bagi dengan jumlah keseluruhan data.
2. **Class Precision**
Merupakan hasil dari jumlah prediksi positif benar dibagi dengan keseluruhan dari hasil yang prediksi benar.
3. **Class Recall**
Merupakan hasil dari jumlah prediksi positif benar di bagi dengan keseluruhan prediksi positif benar.
4. **Classification Error**
Merupakan hasil penjumlahan dari klasifikasi yang salah kemudian di bagi dengan keseluruhan data.

Sehingga dari hasil klasifikasi di atas, maka dapat menghasilkan sebuah grafik hasil Prediksi sebagai berikut:



Gambar 12 Grafik Hasil Prediksi

4.3. Analisis Hasil

Diketahui berdasarkan analisis perhitungan entropy dan nilai gain variabel yang nilai gain tertinggi dan yang menjadi node akar pohon keputusan adalah jumlah kredit. Selanjutnya, setelah melakukan pengujian dataset, diketahui bahwa hasil prediksi data training maupun data testing, menunjukkan bahwa nilai confidence itu adalah yang menentukan untuk memprediksi lancar dan nunggak nya setiap nasabah. Data set yang telah di ujikan di atas baik data training maupun data testing, menunjukkan bahwa klasifikasi algoritma C4.5 memiliki nilai akurasi sebesar 65.00% , mean precision sebesar 55.86%, classification error sebesar 35%, dan recall sebesar 55.43% untuk data training sedangkan data testing memiliki tingkat akurasi sebesar 87.88%, precision sebesar 87.66%, dan recall sebesar 87.44%. Dan adapun hasil prediksinya adalah 58% nasabah membayar dengan lancar dan 42% nasabah membayar nunggak, hal ini berdasarkan nilai confidence dari masing-masing nasabah dilihat dari munculnya prediksi kredit macet dikarenakan data inflasi kecil dan memiliki jumlah kredit besar pada setiap nasabah.

5. Kesimpulan dan Saran

5.1. Kesimpulan

Berdasarkan pembahasan dari penulisan laporan skripsi dapat diambil kesimpulan yaitu :

1. Klasifikasi dalam memprediksi transaksi nasabah yang akan nunggak atau lancar dalam pembayaran menggunakan algoritma C4.5 pada Koperasi BMT Sahabat Rawajitu Selatan menggunakan validasi cross validation.
2. Hasil klasifikasi terhadap transaksi nasabah yang akan nunggak ataupun lancar pada Koperasi BMT Sahabat Rawajitu Selatan dari algoritma C4.5 menggunakan data uji sebanyak 1321 yang terdiri dari 1057 data training dan 264 data testing. Dimana pengujian ini menghasilkan tingkat

akurasi sebesar 65.00% , mean precision sebesar 55.86%, classification error sebesar 35%, dan recall sebesar 55.43% untuk data training sedangkan data testing memiliki tingkat akurasi sebesar 87.88%, precision sebesar 87.66%, dan recall sebesar 87.44%.

3. Adapun hasil prediksinya adalah 58% nasabah membayar dengan lancar dan 42% nasabah membayar nunggak, hal ini berdasarkan nilai confidence dari masing-masing nasabah.
4. Nilai gain tertinggi di dapatkan oleh variabel jumlah kredit dan gaji bersih yang bernilai 0.1166 dan 0.0645 sedangkan variabel jumlah tanggungan dan inflasi memiliki nilai gain terendah yaitu 0.0054 dan 0.
5. Hasil prediksi dibagi menjadi 2 bagian, yaitu "Lancar" dan "Nunggak".

5.2. Saran

Adapun beberapa saran dari peneliti untuk penelitian lebih lanjut yaitu:

1. Mengkombinasikan Algoritma C4.5 dengan teknik lain agar hasil nya dapat lebih akurat.
2. Untuk pengujiannya bisa menggunakan tools lain seperti weka dan lain sebagainya.
3. Menggunakan dataset yang lebih banyak agar hasilnya lebih baik dan akurat.
4. Menambahkan atribut atau variabel baru yang sesuai dengan kebutuhan agar hasil akurasi yang di dapat bisa lebih baik lagi.

Daftar Pustaka

- [1] H. Sulistiani And A. Tjahyanto, "Heterogeneous Feature Selection For Classification Of Customer Loyalty Fast Moving Consumer Goods (Case Study: Instant Noodle).," *J. Theor. Appl. Inf. Technol.*, Vol. 94, No. 1, 2016.
- [2] H. Sulistiani, K. Muludi, And A. Syarif, "Implementation Of Dynamic Mutual Information And Support Vector Machine For Customer Loyalty Classification," In *Journal Of Physics: Conference Series*, 2019, Vol. 1338, No. 1, P. 12050.
- [3] H. Sulistiani, "Penerapan Algoritma Klasifikasi Sebagai Pendukung Keputusan Pemberian Beasiswa Mahasiswa," 2018.
- [4] A. A. Aldino, D. Darwis, A. T. Prastowo, And C. Sujana, "Implementation Of K-Means Algorithm For Clustering Corn Planting Feasibility Area In South Lampung Regency," In *Journal Of Physics: Conference Series*, 2021, Vol. 1751, No. 1, P. 12038.

- [5] M. Bakri, "Penerapan Data Mining Untuk Clustering Kualitas Batu Bara Dalam Proses Pembakaran Di Pltu Sebalang Menggunakan Metode K-Means," *J. Teknoinfo*, Vol. 11, No. 1, P. 6, 2017.
- [6] A. A. Aldino And H. Sulistiani, "Decision Tree C4. 5 Algorithm For Tuition Aid Grant Program Classification (Case Study: Department Of Information System, Universitas Teknokrat Indonesia)," *Eduatic-Scientific J. Informatics Educ.*, Vol. 7, No. 1, 2020.
- [7] A. Nurkholis And I. S. Sitanggang, "Optimalisasi Model Prediksi Kesesuaian Lahan Kelapa Sawit Menggunakan Algoritme Pohon Keputusan Spasial," *J. Teknol. Dan Sist. Komput.*, Vol. 8, No. 3, Pp. 192–200, 2020.
- [8] A. Nurkholis And T. Susanto, "Algoritme Spatial Decision Tree Untuk Evaluasi Kesesuaian Lahan Padi Sawah Irigasi," *J. Resti (Rekayasa Sist. Dan Teknol. Informasi)*, Vol. 4, No. 5, Pp. 978–987, 2020.
- [9] N. Neneng And Y. Fernando, "Klasifikasi Jenis Daging Berdasarkan Analisis Citra Tekstur Gray Level Co-Occurrence Matrices (GlcM) Dan Warna," *Pros. Semnastek*, 2017.
- [10] N. Neneng, N. U. Putri, And E. R. Susanto, "Klasifikasi Jenis Kayu Menggunakan Support Vector Machine Berdasarkan Ciri Tekstur Local Binary Pattern," *Cybernetics*, Vol. 4, No. 02, Pp. 93–100, 2021.
- [11] S. Styawati And K. Mustofa, "A Support Vector Machine-Firefly Algorithm For Movie Opinion Data Classification," *Ijccs (Indonesian J. Comput. Cybern. Syst.*, Vol. 13, No. 3, Pp. 219–230, 2019.
- [12] D. Alita, I. Sari, A. R. Isnain, And S. Styawati, "Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa," *J. Data Min. Dan Sist. Inf.*, Vol. 2, No. 1, Pp. 17–23, 2021.
- [13] E. Prasetyo, *Data Mining Mengubah Data Menjadi Informasi*. 2014.
- [14] H. Sulistiani And D. Darwis, "Penerapan Metode Agile Untuk Pengembangan Online Analytical Processing (Olap) Pada Data Penjualan (Studi Kasus : Cv Adilia Lestari)," *J. Coreit*, Vol. 6, No. 1, Pp. 50–56, 2020.
- [15] D. Darwis, E. S. Pratiwi, And A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Eduatic-Scientific J. Informatics Educ.*, Vol. 7, No. 1, 2020.
- [16] Z. Abidin, A. Wijaya, And D. Pasha, "Aplikasi Stemming Kata Bahasa Lampung Dialek Api Menggunakan Pendekatan Brute-Force Dan Pemograman C," *J. Media Inform. Budidarma*, Vol. 5, No. 1, Pp. 1–8, 2021.
- [17] A. A. R. I. Aldino, "Dimensi Metrik Hasil Operasi Tertentugraf Petersen Diperumum." Universitas Lampung, 2019.