

EXPLORASI ALGORITMA C4.5 DAN FORWARD FEATURE SELECTION UNTUK MENENTUKAN DEBITUR BAIK DAN DEBITUR BERMASALAH PADA PRODUK KREDIT TANPA AGUNAN (KTA)

Putu Gede Surya Cipta Nugraha¹, Gede Surya Mahendra²

¹ Prodi Teknik Informatika, STMIK STIKOM Indonesia, Denpasar, Indonesia

² Prodi Teknik Informatika, STMIK STIKOM Indonesia, Denpasar, Indonesia

e-mail: surya.cipta@stiki-indonesia.ac.id, gede.mahendra@stiki-indonesia.ac.id

Abstrak

Produk kredit Bank Umum yang sangat diminati oleh Badan Usaha atau Organisasi dan masyarakat salah satunya yaitu Kredit Tanpa Angunan (KTA), hal ini dikarenakan sistem kredit tidak membutuhkan jaminan dari debitur. Tetapi dalam jangka waktu proses kredit KTA tidak menutup kemungkinan debitur melakukan keterlambatan dalam melakukan pembayaran angsuran (menunggak) yang dikarenakan mengalami kegagalan dalam bisnis, kehilangan pekerjaan, uang digunakan untuk memenuhi kebutuhan lain serta berbagi macam alasan lainnya. Pada Bank ABC setiap nasabah yang terlambat melakukan pembayaran dapat dikelompokkan menjadi *Non Performance Loan (NPL)* atau yang sering disebut dengan kredit macet. Untuk mengatasi permasalahan tersebut diterapkan bidang Ilmu Komputer yaitu *Data Mining* untuk memprediksi kriteria debitur yang baik dan debitur bermasalah. Adapun metode atau *algoritma Data Mining* yang digunakan adalah kombinasi dari *algoritma C4.5* dan *Forward Feature Selection*. Pengujian *algoritma C4.5* dalam memprediksi menghasilkan tingkat *accuracy* sebesar 92.00%, *recall* sebesar 92.00% dan *precision* sebesar 92.00%. *Forward Feature Selection* berbasis *algoritma C4.5* lebih akurat dan efektif dalam memprediksi debitur yang baik dan debitur bermasalah dengan hasil *accuracy* sebesar 93.60%, *recall* sebesar 93.60%, *precision* sebesar 93.60% dan memperoleh atribut yang berpengaruh yaitu jangka waktu, maksimum kredit dan pekerjaan.

Kata kunci: Data Mining, *Algoritma C4.5*, *Forward Feature Selection*

Abstract

Commercial bank credit products that are in great demand by Business Entities or Organizations and the public, one of which is Unsecured Credit (KTA), this is because the credit system does not require collateral from the debtor. But in the term of the KTA credit process does not rule out the possibility of debtors making payments in arrears (delinquent) due to experience failure in business, lost work, money used to meet other needs and share various other reasons. At ABC Bank, every customer who is late in making a payment can be classified as Non Performance Loan (NPL) or what is often referred to as bad credit. To overcome these problems applied the field of Computer Science namely Data Mining to predict criteria for good debtors and problematic debtors. The Data Mining method or algorithm used is a combination of C4.5 algorithm and Forward Feature Selection. Testing the C4.5 algorithm in predicting produces an accuracy rate of 92.00%, recall of 92.00% and precision of 92.00%. Forward Feature Selection based on C4.5 algorithm is more accurate and effective in predicting good debtors and problem debtors with an accuracy of 93.60%, recall of 93.60%, precision of 93.60% and obtaining the influential attributes namely, time period, maximum credit and work.

Keywords : Data Mining, C4.5 Algorithm, Forward Feature Selection

PENDAHULUAN

Menurut Otoritas Jasa Keuangan (2017), Bank adalah badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkan kepada masyarakat dalam bentuk kredit dan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup masyarakat. Menurut Undang-Undang Nomor 10 Tahun 1998, Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga.

Produk kredit Bank Umum yang sangat diminati oleh Badan Usaha atau Organisasi dan masyarakat salah satunya yaitu Kredit Tanpa Angunan (KTA), hal ini dikarenakan sistem kredit tidak membutuhkan jaminan dari debitur. Kredit KTA dapat diberikan kepada debitur jika debitur memiliki syarat utama yaitu memiliki *track record* yang bagus dalam proses perkreditan. Tetapi dalam jangka waktu proses kredit KTA tidak menutup kemungkinan debitur melakukan keterlambatan dalam melakukan pembayaran angsuran (menunggak) yang dikarenakan mengalami kegagalan dalam bisnis, kehilangan pekerjaan, uang digunakan untuk memenuhi kebutuhan lain serta berbagai macam alasan lainnya. Pada Bank ABC setiap nasabah yang terlambat melakukan pembayaran dapat dikelompokkan menjadi *Non Performance Loan (NPL)* atau yang sering disebut dengan kredit macet.

Permasalahan NPL ini, jika tidak segera mendapatkan solusi pasti memberikan dampak yang buruk pada pihak Bank, hal ini bisa menyebabkan keuntungan yang diperoleh pihak Bank menjadi berkurang karena setiap adanya debitur yang terlambat melakukan pembayaran angsuran, maka pihak Bank harus menaruh sebagian dananya ke Bank Indonesia (BI) atau yang sering disebut dengan istilah CKPN (Cadangan Kerugian Penurunan Nilai). Untuk mengatasi permasalahan tersebut, peneliti menggunakan bidang Ilmu Komputer yaitu *Data Mining*. Menurut Larose (2005),

dalam Nuraeni (2017) *Data Mining* adalah proses menemukan korelasi baru yang bermakna, pola dan tren dengan memilah-milah sejumlah besar data yang tersimpan dalam repositori, menggunakan teknologi penalaran pola serta teknik-teknik statistik dan matematika. Dalam penelitian ini metode atau algoritma *Data Mining* yang digunakan adalah *algoritma C4.5* dan *forward feature selection*. *Algoritma C4.5* adalah pohon keputusan di mana setiap simpul cabang mewakili pilihan antara a sejumlah alternatif, dan setiap simpul daun mewakili keputusan. Hasil dari proses klasifikasi yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari *record* yang baru sedangkan *forward feature selection* merupakan salah satu metode pemodelan untuk menemukan kombinasi peubah yang terbaik dari suatu gugus peubah

Algoritma adalah metode efektif yang diekspresikan sebagai rangkaian terbatas. Algoritma juga merupakan kumpulan perintah untuk menyelesaikan suatu masalah. Perintah-perintah ini dapat diterjemahkan secara bertahap dari awal hingga akhir. Masalah tersebut dapat berupa apa saja, dengan syarat untuk setiap permasalahan memiliki kriteria kondisi awal yang harus dipenuhi sebelum menjalankan sebuah algoritma. Algoritma juga memiliki pengulangan proses (iterasi), dan juga memiliki keputusan hingga keputusan selesai (Maulana, 2017).

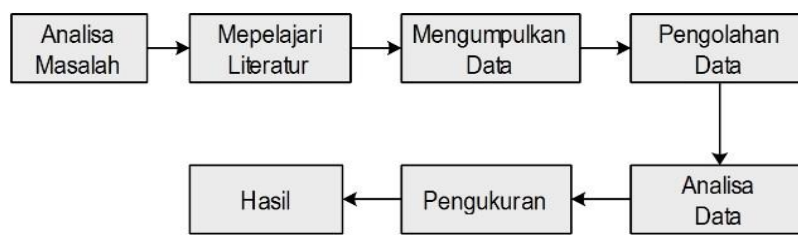
Beberapa peneliti telah melakukan penelitian menggunakan *algoritma C4.5* antara lain oleh Jafar Hamid, A., & Ahmed, T. M. (2016) dengan judul *Developing Prediction Model of Loan Risk in Banks Using Data Mining*, dengan hasil menemukan bahwa algoritma terbaik untuk klasifikasi pinjaman adalah algoritma J48 (C4.5) karena algoritma J48 (C4.5) memiliki akurasi tinggi sebesar 78.3784% dan kesalahan absolut rata-rata rendah. Melihat hasil yang sangat baik dari algoritma C4.5, penulis kembali menggunakan algoritma C4.5 dengan di kombinasikan dengan seleksi fitur dari *forward feature selection* untuk pemilihan parameter yang berpengaruh. Peneliti lainnya dilakukan oleh Sambani & Nuraeni (2018) dengan judul Penerapan *Algoritma C4.5* Untuk Klasifikasi Pola Penjurusan di

Sekolah Menengah Kejuruan (SMK) Kota Tasikmalaya, dengan hasil penelitian tingkat akurasi algoritma C4.5 sebesar 92.30%. Melihat hasil yang sangat baik dari algoritma C4.5, penulis kembali menggunakan algoritma C4.5 kedalam objek yang berbeda yaitu produk kredit KTA pada Bank ABC. Penelitian lainnya dilakukan oleh Saleh (2017) dengan judul Prediksi Kebangkrutan Perusahaan Menggunakan Algoritma C4.5 Berbasis *forward selection*, dengan hasil penelitian penerapan algoritma C4.5 berbasis *forward selection* menghasilkan tingkat akurasi sebesar 99.60% dan algoritma C4.5 berbasis *forward selection* dengan akurasi sebesar 99.61%. Melihat hasil yang sangat baik dari algoritma C4.5

berbasis *forward selection*, penulis menjadikan penelitian ini sebagai acuan dasar dalam melakukan kombinasi algoritma C4.5 dengan *forward feature selection*.

Dengan permasalahan diatas dan hasil penelitian sebelumnya, penulis tertarik untuk melakukan penelitian terhadap debitur di Bank ABC menggunakan algoritma C4.5 dan *forward feature selection* untuk memprediksi nasabah baik dan nasabah bermasalah. Sehingga tujuan dari penelitian ini adalah mengetahui hasil pengujian algoritma C4.5 dan metode *forward feature selection* dalam memprediksi debitur baik dan debitur bermasalah pada KTA di Bank ABC.

METODE



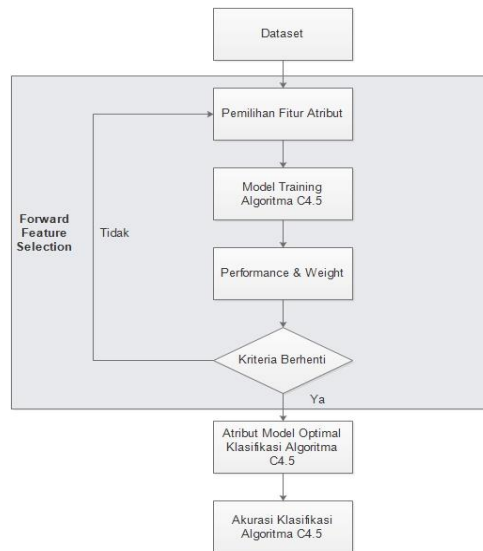
Gambar 1. Kerangka Berpikir

Kerangka pemikiran pada penelitian ini berdasarkan klasifikasi data mining dari dataset Bank ABC. Algoritma C4.5 diketahui bisa menangani masalah dataset yang besar. Sedangkan proses fitur seleksi yaitu *Forward Feature Selection* digunakan untuk menentukan atribut yang paling berpengaruh dan dapat membantu

meningkatkan hasil akurasi klasifikasi algoritma C4.5.

A. Desain Penelitian Algoritma C4.5 dan Forward Selection

Pada Penelitian kali ini menggunakan metode penelitian eksperimen, dengan desain penelitian sebagai berikut :

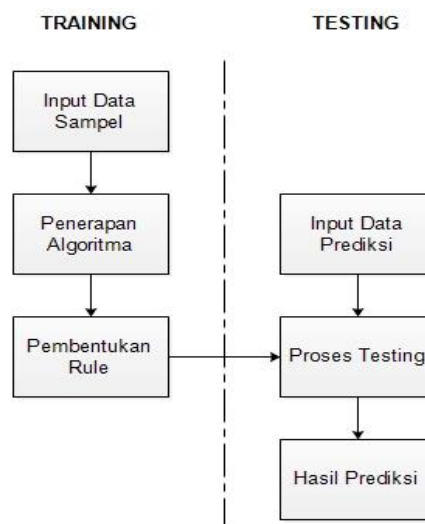


Gambar 2. Desain Penelitian *Algoritma C4.5 dan Forward Selection*

Pada gambar 2 merupakan langkah-langkah *algoritma C4.5* dengan metode *forward feature selection*. Dimana langkah-langkahnya dijelaskan sebagai berikut :

1. Dataset kredit diseleksi fitur menggunakan *forward feature selection*, metode *forward feature selection* adalah pemodelan dimulai dari nol peubah (empty model).
2. Pemilihan fitur atribut menggunakan metode *forward feature selection* untuk diuji menggunakan data *training algoritma C4.5*.
3. Dari proses uji atribut pada data *training algoritma C4.5* akan mendapatkan hasil dan pembobotan.
4. Apabila proses tersebut lolos maka akan mendapatkan suatu atribut/model yang optimal dari klasifikasi *algoritma C4.5*.
5. Sedangkan bila proses tersebut tidak berlanjut pada kriteria berhenti maka dilakukan proses pengulangan dari awal (pemilihan fitur seleksi *forward feature selection*) sampai mendapatkan atribut/model optimal.

B. Desain Penelitian *Algoritma C4.5*



Gambar 4. Desain Penelitian *Algoritma C4.5*

Pada gambar 4 dapat dijelaskan bahwasanya *algoritma C4.5* dimulai dengan dua langkah :

1. Proses *Training*.

Pada proses *training* yaitu memasukkan data sampel kedalam tabel yang disiapkan untuk proses perhitungan. Tabel tersebut meliputi atribut, jumlah data keseluruhan, jumlah data yang sudah terklasifikasi berdasarkan target yang ditentukan, dalam kasus ini yaitu debitur baik dan debitur bermasalah, serta kolom nilai *Entropy* dan *Gain*. Tahapan selanjutnya yaitu penerapan *algoritma C4.5* yaitu menghitung nilai *Entropy* dan *Gain* pada tiap-tiap atribut untuk dijadikan bentuk pohon. Bentuk pohon merupakan bentuk aturan klasifikasi yang akan diterapkan pada proses *testing*.

2. Proses *Testing*

Pada proses *testing* ini langkah yang dilakukan yaitu memasukkan data uji atau data prediksi. Atribut yang digunakan pada proses testing ini harus sesuai dengan atribut pada proses *training*. Setiap data atribut akan dibandingkan dengan aturan yang sudah terbentuk pada perhitungan data *training* sebelumnya. Selanjutnya data tersebut akan diklasifikasikan berdasarkan target yang ingin diketahui yaitu data kredit tanpa anggungan dengan kondisi atribut ini debitur baik atau debitur bermasalah.

C. Parameter Penelitian

Pada penelitian ini, parameter penelitian yang digunakan sebagai berikut :

Tabel 1. Parameter Penelitian

Parameter	Katagori
Jangka Waktu Kredit	a. Pendek b. Menengah c. Panjang
Maksimum Kredit	a. Kecil b. Sedang c. Besar
Umur	a. Remaja b. Dewasa c. Lansia
Angsuran	a. Kecil b. Sedang c. Besar
Lama Bekerja	a. Pendek b. Menengah c. Panjang
Jenis Kelamin	a. Laki-Laki b. Perempuan
Pekerjaan	a. Karyawan Swasta b. PNS c. Dosen d. Dokter e. BUMN
Status	a. Menikah b. Belum Menikah

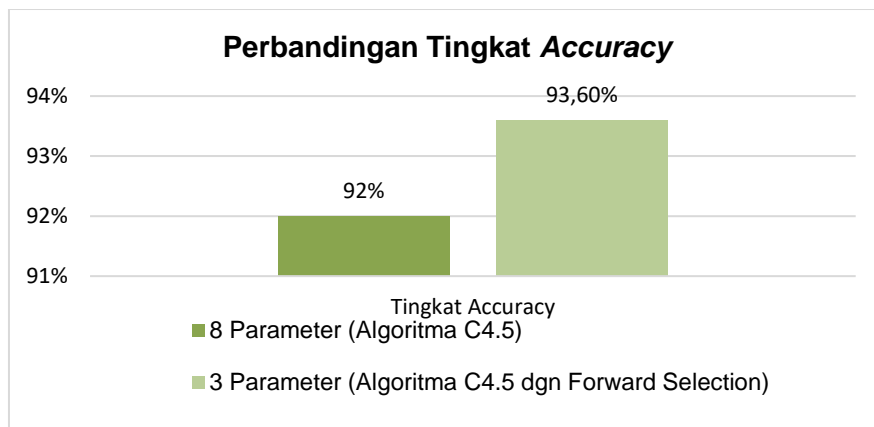
Dalam penelitian ini, menggunakan 8 parameter yaitu : jangka waktu kredit, maksimum kredit, umur, angsuran, lama bekerja, jenis kelamin, pekerjaan dan status.

HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dengan mengumpulkan data pada Bank ABC. Pada penelitian ini, peneliti menggunakan 8 atribut yaitu jangka waktu kredit, maksimum kredit, umur, angsuran, lama

bekerja, jenis kelamin, pekerjaan dan status dengan data *training* sebanyak 500 data dan data *testing* sebanyak 250 data.

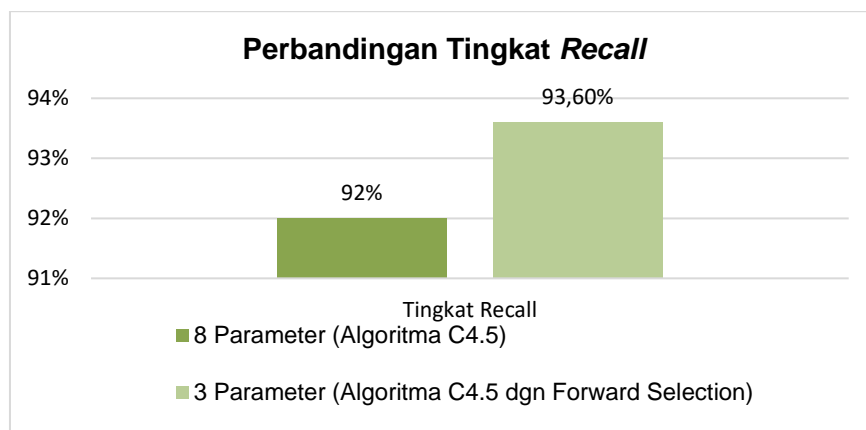
Perbandingan tingkat akurasi *algoritma C4.5* dan *forward feature selection* dapat dilihat pada gambar dibawah ini.



Gambar 5. Perbandingan Tingkat *Accuracy*

Pada gambar 5 memperlihatkan persentase *accuracy* dari masing-masing parameter yang digunakan. Dapat dilihat *accuracy* dengan 8 parameter menghasilkan persentase sebesar 92,00% sedangkan dengan 3 parameter menghasilkan persentase sebesar

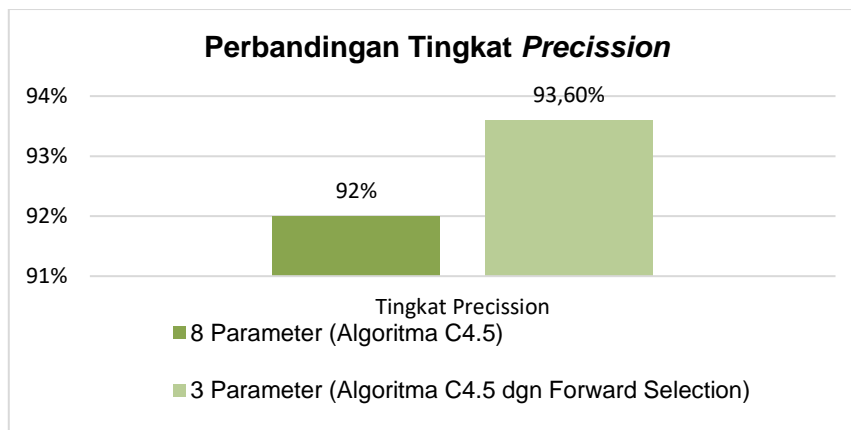
93,60%. Hasil ini diartikan tidak perlu mencari menggunakan 8 parameter untuk memprediksi kriteria nasabah cukup dengan 3 parameter yaitu jangka waktu kredit, maksimum kredit dan pekerjaan sudah bisa memprediksi kriteria nasabah baik dan bermasalah.



Gambar 6. Perbandingan Tingkat *Recall*

Pada gambar 6 memperlihatkan persentase *recall* dari masing-masing parameter yang digunakan. Dapat dilihat *recall* dengan 8 parameter menghasilkan persentase 92,00% sedangkan dengan 3 parameter menghasilkan persentase

sebesar 93,60% hasil ini didapatkan karena dengan 3 parameter tingkat keberhasilan algoritma untuk memprediksi data yang relevan hasilnya lebih baik dibandingkan dengan 8 parameter.



Gambar 7. Perbandingan Tingkat *Precision*

Pada gambar 7 memperlihatkan persentase *precision* dari masing-masing parameter yang digunakan. Dapat dilihat *precision* dengan 8 parameter menghasilkan persentase 92,00% sedangkan dengan 3 parameter menghasilkan persentase sebesar

Untuk membuktikan metode yang diusulkan dapat bekerja secara optimal, dilakukan serangkaian pengujian seperti mencari nilai *accuracy*, *recall*, *precision* serta membandingkan hasil tingkat *accuracy* metode C4.5 dengan penelitian terdahulu. Berdasarkan hasil pengujian *accuracy* algoritma C4.5 yang digunakan pada penelitian ini memberikan hasil *accuracy* yang masuk dalam katagori baik, begitu pula dengan penelitian yang dilakukan oleh Jafar Hamid, A., & Ahmed, T. M yang masuk dalam katagori baik dikarenakan hasilnya diatas 75%.

Pada penelitian selanjutnya yang dilakukan oleh Sambani & Nuraeni dengan menerapkan algoritma C4.5 dalam bidang penjurusan sekolah juga memberikan tingkat *accuracy* yang baik yaitu sebesar 92.30%. Pada penelitian ini juga membuktikan bahwa penerapan algoritma C4.5 dalam bidang perkreditan khususnya KTA dapat bekerja secara optimal karena memberikan tingkat *accuracy* yang baik itu sebesar 92.00%.

Pada penelitian selanjutnya yang dilakukan oleh Saleh dengan menerapkan kombinasi algoritma C4.5 dengan *forward selection* dalam bidang kebangkrutan perusahaan juga memberikan hasil yang baik sebesar 99.60% dan ketika dikombinasikan dengan *forward selection*

93,60%. Hasil ini didapatkan karena dengan 2 parameter tingkat ketepatan dalam memprediksi dengan benar pada sebuah data yang positif benar pada data yang sebenarnya lebih baik dibandingkan dengan 8 parameter.

tingkat *accuracy* meningkat menjadi 99.61%. Pada penelitian ini juga membuktikan bahwa tingkat *accuracy* algoritma C4.5 memberikan hasil yang baik yaitu sebesar 92.00% dan ketika dikombinasikan dengan *forward feature selection* tingkat *accuracy* meningkat menjadi 93.60% serta mendapatkan parameter yang paling berpengaruh yaitu jangka waktu, maksimum kredit dan pekerjaan.

SIMPULAN

Penelitian Dalam penelitian menentukan debitur baik dan debitur bermasalah pada produk Kredit Tanpa Agunan (KTA) di Bank ABC. Algoritma C4.5 menghasilkan tingkat *accuracy* yang baik yaitu sebesar 92.00%, *recall* sebesar 92.00% dan *precision* sebesar 92.00%. Dalam hal ini *algoritma C4.5* memanfaatkan fungsi seleksi fitur dari *forward feature selection* untuk pemilihan parameter data sehingga dapat membantu *algoritma C4.5* meningkatkan *accuracy* sebesar 93.60%, *recall* sebesar 93.60%, *precision* sebesar 93.60%. Dengan memperoleh parameter yang berpengaruh yaitu jangka waktu, maksimum kredit dan pekerjaan

DAFTAR PUSTAKA

- Dinesh Kumar, A., Pandi Selvam, R., & Sathesh Kumar, K. (2018). *Review on prediction algorithms in educational data mining. International Journal of Pure and Applied Mathematics*, 118(Special Issue 8), 531–537.
- Jafar Hamid, A., & Ahmed, T. M. (2016). *Developing Prediction Model of Loan Risk in Banks Using Data Mining. Machine Learning and Applications: An International Journal*, 3(1), 1–9. <https://doi.org/10.5121/mlaj.2016.3101>
- Larose, D.T. (2005). *Discovering Knowledge in Data*. Canada: Wiley-Interscience.
- Maulana, Gun Gun. 2017. Pembelajaran Dasar Algoritma Dan Pemrograman Menggunakan El-Goritma Berbasis Web. *Jurnal Teknik Mesin (JTM)*: Vol. 06, No. 2, Maret 2017.
- Nuraeni, N. (2017). Penentuan Kelayakan Kredit Dengan Algoritma Naïve Bayes Classifier: Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC. *Jurnal Teknik Komputer AMIK BSI*. 3 (1), ISSN. 2442-2436.
- Otoritas Jasa Keuangan. (2017). Bank Umum. [Online]. Available: <https://www.ojk.go.id/id/kanal/perbankan/Pages/Bank-Umum.aspx>, 28 Juni 2019.
- Sambani, E. B., & Nuraeni, F. (2018). Penerapan Algoritma C4.5 Untuk Klasifikasi Pola Penjurusan di Sekolah Menengah Kejuruan (SMK) Kota Tasikmalaya. *CSRID (Computer Science Research and Its Development Journal)*, 9 (3), 144.
- Saleh, A. (2015). Penerapan Data Mining Dengan Metode Klasifikasi Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa Dalam Mengikuti English Proficiency Test (Studi Kasus : Universitas Potensi Utama). ResearchGate
- Saleh, H. (2017). Prediksi Kebangkrutan Perusahaan Menggunakan Algoritma C4.5 Berbasis Forward Selection. *ILKOM Jurnal Ilmiah*, 9 (2), 173. ISSN 2087-1716. doi : 10.33096/ilkom.v9i2.97.173-180.
- Salim, M. (2016). Klasifikasi Tutupan Lahan Perkotaan Menggunakan Naive Bayes Berbasis Forward Selection. *Jurnal Teknosains*, 10 (2), 165-182.
- Undang-Undang Republik Indonesia Nomor 10 Tahun 1998 tentang perubahan atas Undang-Undang Nomor 7 Tahun 1992 tentang perbankan.