

IMPLEMENTASI METODE C4.5 DAN NAIVE BAYES BERBASIS ADABOOST UNTUK MEMPREDIKSI KELAYAKAN PEMBERIAN KREDIT

Putu Gede Surya Cipta Nugraha^{1,*}, Gede Rasben Dantes², Kadek Yota Ernanda Aryanto³.

¹ Universitas Pendidikan Ganesha

² Universitas Pendidikan Ganesha

³ Universitas Pendidikan Ganesha

Abstrak

At PT. BPR XYZ credit problems is a very vital issue, where if many debtors are delinquent in payment it will increase the NPL value of the bank itself. Increasing the NPL value above 5% indicates that the bank is not healthy. From the above problems, then in this study aims to perform the implementation process of data mining methods to determine the accuracy level of prediction of creditworthiness at PT. BPR XYZ, so that the future of credit problems can be overcome. Data mining methods used in the prediction process are C4.5 and Naive Bayes methods, where both methods are implemented and the accuracy level comparison process is used to see which method is more accurate in predicting creditworthiness. Both methods are also embedded AdaBoost method with the aim of increasing the accuracy in the process of prediction of creditworthiness feasibility. The result obtained from the comparison of method accuracy level, stated that the better accuracy is C4.5 method that is 90.00% with the precision level of 86.67%. As for the accuracy of Naive Bayes method that is equal to 70.00% with the precision level of 79.71%. Then with the addition of AdaBoost method in predicting creditworthiness proved to increase the higher accuracy value of 91.54% in method C4.5 and by 78.13% in Naive Bayes method. From the description above, with the implementation of AdaBoost method on the method of C4.5 and Naive Bayes can improve the accuracy of the prediction of creditworthiness of PT. BPR XYZ. In addition, the implementation of the AdaBoost-based C4.5 method can be a recommendation for PT. BPR XYZ in conducting predictive process of credit worthiness in the future.

Keywords:

Data Mining, Accuracy, C4.5, Naive Bayes, AdaBoost

PENDAHULUAN

PT. BPR XYZ merupakan salah satu lembaga perbankan resmi yang berfungsi melayani proses simpan pinjam untuk golongan pengusaha mikro, kecil dan menengah. PT. BPR XYZ memiliki beberapa produk-produk, dimana produk-produk yang ditawarkan berfungsi membantu masyarakat dalam sektor ekonomi. Produk-produk yang ditawarkan dilembaga ini adalah tabungan, deposito dan kredit. Yang mana salah satu produk yang paling berkembang dilembaga ini adalah produk kredit. Menurut pasal 1 angka 11 undang-undang nomor 10 tahun 1998, kredit adalah proses memberikan dana yang berdasarkan persetujuan atau kesepakatan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak yang melakukan peminjaman untuk melunasi utangnya sesuai dengan jangka waktu yang telah disepakati. Didalam kegiatan perkreditan, kita juga mengenal istilah NPL (*Non Performing Loan*). NPL adalah salah satu indikator kesehatan kualitas aset bank.

Permasalahan kredit bermasalah pada PT. BPR XYZ sering terjadi akibat dari petugas analisis yang tidak berhati-hati atau kurang cermat dalam proses melakukan analisa pemberian kredit. Selain itu dipengaruhi juga oleh karakter debitur yang kurang baik yaitu debitur yang melanggar surat perjanjian dalam melakukan proses pembayaran pada waktu yang telah disepakati. Jika kredit bermasalah tersebut tidak diatasi maka dapat memberikan tingkat NPL yang cukup tinggi (diatas 5%), maka PT. BPR XYZ tersebut masuk kedalam kondisi BPR yang tidak sehat. Untuk mengetahui kelayakan pemberian kredit

* Corresponding author.

E-mail Addresses: surya.ciptanugraha@gmail.com (Putu Gede Surya Cipta Nugraha)

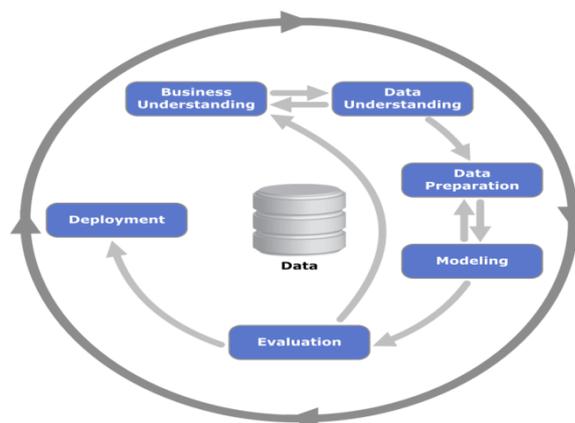
dimasa mendatang di PT. BPR XYZ, perlu adanya prediksi yang akurat yang salah satunya memanfaatkan teknologi dibidang data mining.

Menurut Hermawati (2013), Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisa dan mengekstraksi pengetahuan (knowledge) secara otomatis. Dalam data mining terdapat beberapa algoritma klasifikasi yang dapat digunakan untuk analisis kredit seperti algoritma C4.5 dan Naïve Bayes. Klasifikasi C4.5 dan Naïve Bayes berjalan dengan baik dalam melakukan klasifikasi dalam pemberian kredit, kejadian ini dibuktikan dengan penelitian yang dilakukan oleh Masripah (2015) yang berjudul "Evaluasi Penentuan Kelayakan Pemberian Kredit Koperasi Syariah Menggunakan Algoritma Klasifikasi C4.5", dalam penelitian tersebut mendapatkan kesimpulan bahwa akurasi algoritma C4.5 sebesar 88.00%. Penelitian lain yang dilakukan oleh Masripah (2016) yang berjudul "Komparasi Algoritma Klasifikasi Data Mining Untuk Evaluasi Pemberian Kredit", dalam penelitian tersebut mendapatkan kesimpulan bahwa akurasi algoritma C4.5 sebesar 88.90% sedangkan algoritma Naïve Bayes sebesar 88.00%. Jadi algoritma C4.5 memprediksi lebih akurat dibandingkan algoritma Naïve Bayes.

Berdasarkan latar belakang diatas, penulis tertarik melakukan penelitian kelayakan pemberian kredit di PT. BPR XYZ dengan menggunakan algoritma C4.5 dan Naïve Bayes. Selanjutnya dilakukan penerapan algoritma C4.5 dan Naïve Bayes menggunakan metode boosting yaitu AdaBoost. Metode AdaBoost digunakan dengan tujuan metode ini dapat meningkatkan akurasi dalam proses klasifikasi dan prediksi.

METODE PENELITIAN

Dalam melakukan penelitian ini, penulis menggunakan metode penelitian eksperimen. Design *eksperimen* yang digunakan adalah CRISP-DM (*Cross Standard Industry For Data Mining*). Model proses CRISP-DM memiliki enam tahapan sebagai berikut :



Gambar 1 Model Proses CRISP-DM

Tahap pertama yaitu *Business Understanding*. Pada tahap ini dilakukan kegiatan survey lapangan untuk melihat dan mengetahui secara langsung kondisi serta permasalahan yang terjadi. Dari hasil survey lapangan diperoleh hasil pada tiap bulannya tingkat NPL selalu tinggi yaitu diatas 5%. Tahap ke dua yaitu *Data Understanding*. Pada tahap ini dilakukan proses pengumpulan data, kemudian melakukan analisa data serta melakukan evaluasi kualitas data yang digunakan dalam penelitian ini. Dari proses perijinan data yang diajukan, diperoleh data debitur pada tahun 2014 sebanyak 162 record data, tahun 2015 sebanyak 240 record data dan pada tahun 2016 sebanyak 320 record data. Dari masing-masing data memiliki 37 atribut dengan nilai yang dimiliki oleh atribut merupakan nilai kategorikal dan nilai angka.

Tahap ke tiga yaitu *Data Preparation*. Pada tahap ini dilakukan persiapan data mentah selanjutnya menentukan atribut yang digunakan menganalisa masalah. Dalam *Data Preparation* ada beberapa teknik *preprocessing* yang digunakan untuk mengolah data mentah yaitu: a) *Data Cleaning*, pada tahap ini dilakukan pembersihan nilai-nilai dalam atribut yang kosong dan menghapus data yang kosong (*missing values* dan *noisy*), b) *Data Integration*, pada tahap ini dilakukan penyatuan data yang berbeda kedalam satu data. Dalam hal ini, ada dua data yang diambil sebagai data *warehouse* yaitu database debitur dan data nominatif kredit, c) *Data Reduction*, jumlah atribut dan data yang digunakan untuk data training dan data uji mungkin terlalu banyak dan hanya beberapa atribut dan data yang diperlukan dalam melakukan penelitian sehingga atribut dan data yang tidak diperlukan akan dihapus. Dalam tahap ini juga dilakukan proses penghapusan data yang terindetifikasi duplikasi. Karena seperti yang kita tahu bahwa debitur yang

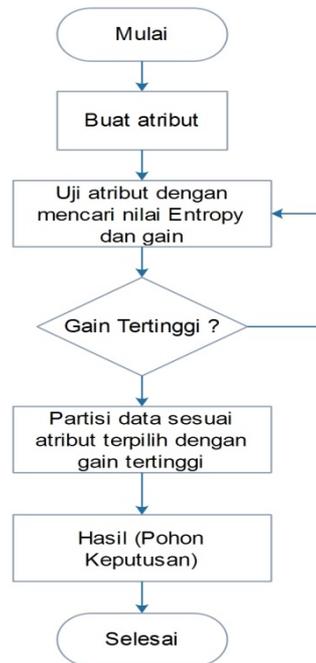
melakukan kredit dalam kurun waktu tertentu, pasti akan tercatat pada pelaporan tahun kedepannya. d) *Data conversion*, data yang masih memiliki nilai berupa angka harus dilakukan proses konversi data dengan klasifikasi data dari atribut. Data klasifikasi atribut terdapat pada Tabel 1 dibawah ini.

Tabel 1 Klasifikasi Atribut

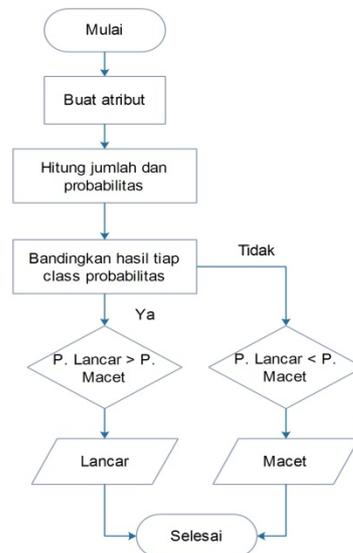
Atribut	Klasifikasi	Keterangan
Umur	17-25 tahun	Umur Nasabah
	26-45 tahun	
	>45 tahun	
Jumlah Pinjaman	1.000.000 - 50.000.000	Jumlah Pinjaman Nasabah
	50.000.001 - 500.000.000	
	>500.000.000	
Jangka Waktu	1 - 12 Bln : Pendek	Jangka Waktu Kredit
	13 - 36 Bln : Menengah	
	>36 : Panjang	
Jenis Kredit	Modal Kerja Konsumsi Elektronik & Motor Mingguan Karyawan	Jenis Kredit Berdasarkan Penggunaan
Sektor Ekonomi	Petani Nelayan Wiraswasta Pedagang Pegawai Negeri Sipil	Golongan Pekerjaan Debitur
Jaminan	BPKB Sepeda Motor	Jenis Jaminan yang digunakan Debitur
	BPKB Mobil	
	Sertifikat Tanah	
	Sertifikat Bangunan	
Bunga	Gaji	Bunga kredit yang ditawarkan untuk Debitur
	Flat	
	Menurun Murni Anuitas	
Tunggakan Pokok	Tidak Memiliki Memiliki	Tunggakan Pokok Debitur
Tunggakan Bunga	Tidak Memiliki Memiliki	Tunggakan Bunga Debitur

Tahap ke empat yaitu *Modelling*. Pada tahap ini dilakukan pemilihan teknik pemodelan yang tepat. Pada penelitian ini, menggunakan model atau algoritma C4.5 dan Naïve Bayes berbasis Ada Boost. Setelah melewati tahap *Data Preparation*, sehingga pada penelitian ini digunakan data training sebanyak 162 record sedangkan untuk data uji 1 digunakan sebanyak 80 record dan untuk data uji 2 digunakan sebanyak 80 record. Tahap ke lima yaitu *Evaluation*. Pada tahap ini dilakukan evaluasi dari model yang sudah diciptakan dari masing-masing algoritma. Pengukuran tingkat akurasi algoritma C4.5 dan Naïve Bayes berbasis AdaBoost untuk memprediksi kelayakan pemberian kredit menggunakan *confusion matrix*. Tahap ke enam atau tahap terakhir yaitu *Deployment*. Pada tahap ini dilakukan pemanfaatan model yang

telah dibuat untuk membangun sistem. Pada penelitian ini, hanya dilakukan sampai pada tahap ke lima yaitu *evaluation*. Flowchart Algoritma C4.5 dan Flowchart Algoritma Naïve Bayes tersedia pada Gambar 2 dan 3 di bawah ini.



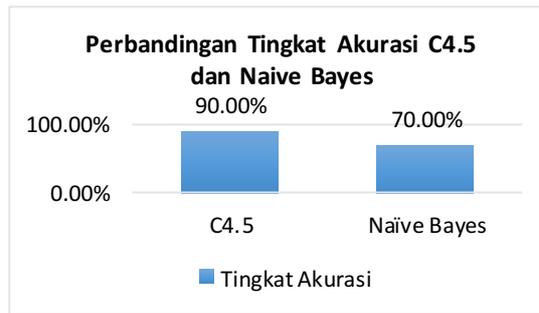
Gambar 2 Flowchart Algoritma C4.5



Gambar 3 Flowchart Algoritma Naïve Bayes

ANALISIS DAN PEMBAHASAN

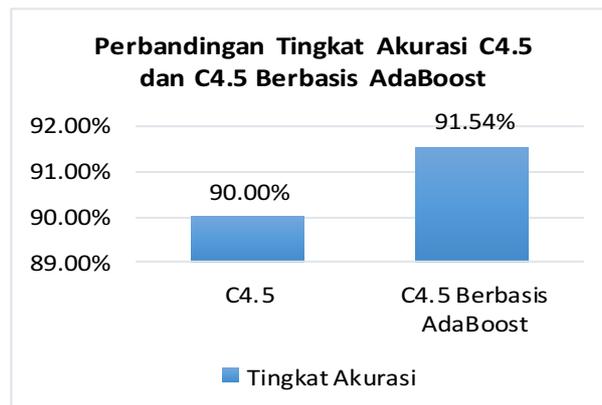
Dalam melakukan penelitian ini, data yang digunakan adalah data kredit dari tahun 2014 sampai tahun 2016. Dimana masing-masing data tiap tahunnya, akan dijadikan data uji dan data training. Data training diperoleh dari tahun 2014 sedangkan untuk data uji 1 diperoleh dari tahun 2015 dan data uji 2 diperoleh dari tahun 2016. Masing - masing data memiliki 9 kriteria atribut seperti *Tunggakan_Bunga*, *Tunggakan_Pokok*, *Umur*, *Jenis_Jaminan* *Jumlah_Pinjaman*, *Type Bunga*, *Jangka Waktu*, *Sektor Ekonomi*, dan *Jenis_Kredit*. Perbandingan Tingkat Akurasi Algoritma C4.5 dan Naïve Bayes dapat dilihat pada Gambar 4 dibawah ini.



Gambar 4 Tingkat akurasi Algoritma C4.5 dan Naive Bayes

Pada Gambar diagram 4 dapat dilihat tingkat akurasi algoritma C4.5 sebesar 90.00%, nilai tersebut lebih besar dibandingkan dengan akurasi algoritma Naive Bayes yang hanya memperoleh tingkat akurasi sebesar 70.00%.

Tingkat akurasi pada algoritma C4.5 lebih besar dalam melakukan proses prediksi kelayakan pemberian kredit dikarenakan pada algoritma C4.5 mampu melakukan prediksi dengan benar pada sebuah data yang juga positif benar pada data yang sebenarnya, hal ini dibuktikan dengan dinilai *precision* dari algoritma C4.5 sebesar 86.67%, nilai tersebut lebih besar jika dibandingkan dengan nilai *precision* dari algoritma Naive Bayes yang hanya sebesar 79.71%. Selain itu kemampuan pada algoritma C4.5 dalam memprediksi data yang relevan juga sangat baik, hal itu dibuktikan dengan perolehan nilai *recall* sebesar 100.00%. Perbandingan Tingkat Akurasi C4.5 dengan C4.5 Berbasis AdaBoost dapat dilihat pada gambar 5 dibawah ini.

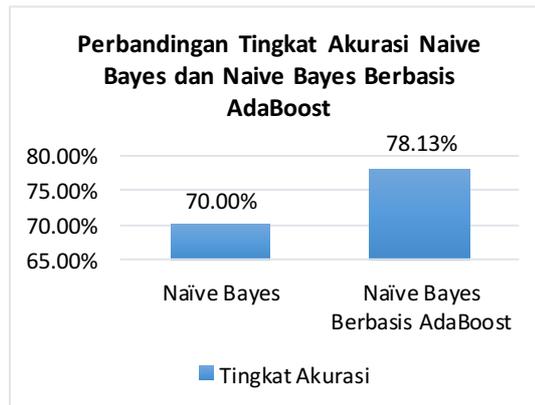


Gambar 5 Tingkat Akurasi C4.5 dengan C4.5 Berbasis AdaBoost

Pada Gambar diagram 5 dapat dilihat tingkat akurasi algoritma C4.5 sebesar 90.00%, nilai tersebut lebih kecil dibandingkan dengan akurasi algoritma C4.5 berbasis AdaBoost yang memperoleh tingkat akurasi sebesar 91.54%. Tingkat akurasi pada algoritma C4.5 berbasis AdaBoost lebih besar dalam melakukan proses prediksi kelayakan pemberian kredit dikarenakan pada algoritma C4.5 berbasis AdaBoost mampu melakukan prediksi dengan benar pada sebuah data yang juga positif benar pada data yang sebenarnya, hal ini dibuktikan dengan dinilai *precision* dari algoritma C4.5 berbasis AdaBoost sebesar 88.55%, nilai tersebut lebih besar jika dibandingkan dengan nilai *precision* dari algoritma C4.5 yang hanya sebesar 86.67%.

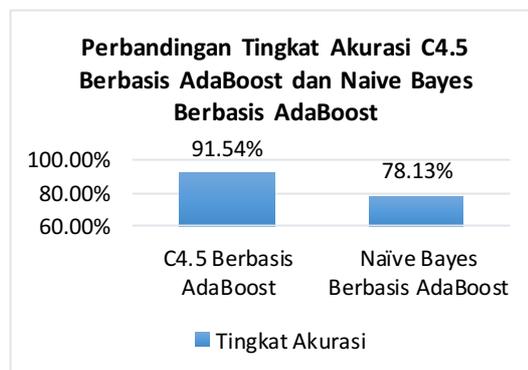
Kemampuan pada algoritma C4.5 berbasis AdaBoost dalam memprediksi data yang relevan juga sangat baik, hal itu dibuktikan dengan perolehan nilai *recall* sebesar 100.00%. Selain itu, tingkat akurasi juga dipengaruhi oleh metode AdaBoost karena metode AdaBoost dapat menyesuaikan diri dengan data dan algoritma C4.5. Metode AdaBoost juga dapat mengurangi kesalahan dari *weak classifier* dan meningkatkan akurasi dari algoritma C4.5. Hal ini dikarenakan pada dasarnya metode AdaBoost dimunculkan untuk mampu mengurangi kesalahan dalam proses pembelajaran.

Perbandingan Tingkat Akurasi C4.5 dengan C4.5 Berbasis AdaBoost dapat dilihat pada Gambar 6 dibawah ini.



Gambar 6 Tingkat Akurasi Naive Bayes dengan Naive Bayes Berbasis AdaBoost

Pada Gambar diagram 6 dapat dilihat tingkat akurasi algoritma Naive Bayes sebesar 70.00%, nilai tersebut lebih kecil dibandingkan dengan akurasi algoritma Naive Bayes berbasis AdaBoost yang memperoleh tingkat akurasi sebesar 78.13%. Tingkat akurasi pada algoritma Naive Bayes berbasis AdaBoost lebih besar dalam melakukan proses prediksi kelayakan pemberian kredit dikarenakan pada algoritma Naive Bayes berbasis AdaBoost mampu melakukan prediksi dengan benar pada sebuah data yang juga positif benar pada data yang sebenarnya, hal ini dibuktikan dengan dinilai *precision* dari algoritma Naive Bayes berbasis AdaBoost sebesar 86.94%, nilai tersebut lebih besar jika dibandingkan dengan nilai *precision* dari algoritma Naive Bayes yang hanya sebesar 79.71%. Kemampuan pada algoritma Naive Bayes berbasis AdaBoost dalam memprediksi data yang relevan juga sangat baik, hal itu dibuktikan dengan perolehan nilai *recall* sebesar 77.74%. Selain itu, tingkat akurasi juga dipengaruhi oleh metode AdaBoost karena metode AdaBoost dapat menyesuaikan diri dengan data dan algoritma Naive Bayes. Metode AdaBoost juga dapat mengurangi kesalahan dari *weak classifier* dan meningkatkan akurasi dari algoritma Naive Bayes. Hal ini dikarenakan pada dasarnya metode AdaBoost dimunculkan untuk mampu mengurangi kesalahan dalam proses pembelajaran.



Gambar 7 Tingkat Akurasi C4.5 Berbasis AdaBoost dan Naive Bayes Berbasis AdaBoost

Pada gambar diagram 7 dapat dilihat tingkat akurasi algoritma C4.5 berbasis AdaBoost sebesar 91.54%, nilai tersebut lebih besar dibandingkan dengan akurasi algoritma Naive Bayes berbasis AdaBoost yang memperoleh tingkat akurasi sebesar 78.13%. Tingkat akurasi pada algoritma C4.5 berbasis AdaBoost lebih besar dalam melakukan proses prediksi kelayakan pemberian kredit dikarenakan pada algoritma C4.5 berbasis AdaBoost mampu melakukan prediksi dengan benar pada sebuah data yang juga positif benar pada data yang sebenarnya, hal ini dibuktikan dengan dinilai *precision* dari algoritma C4.5 berbasis AdaBoost sebesar 88.55%, nilai tersebut lebih besar jika dibandingkan dengan nilai *precision* dari algoritma Naive Bayes berbasis AdaBoost yang hanya sebesar 86.94%. Selain itu kemampuan pada

algoritma C4.5 berbasis AdaBoost dalam memprediksi data yang relevan juga sangat baik, hal itu dibuktikan dengan perolehan nilai *recall* sebesar 100.00%.

KESIMPULAN

Kesimpulan dalam penelitian ini sebagai berikut; 1) Penerapan algoritma data mining menggunakan C4.5 dan Naïve Bayes Berbasis AdaBoost dapat dilakukan untuk melakukan proses prediksi kelayakan pemberian kredit, 2) Algoritma C4.5 dalam melakukan prediksi kelayakan pemberian kredit memiliki tingkat akurasi yang lebih besar daripada algoritma Naïve Bayes dengan tingkat akurasi sebesar 90.00% dan tingkat *precision* sebesar 86.67%, 3) Penambahan metode AdaBoost dalam memprediksi kelayakan pemberian kredit terbukti memberikan nilai akurasi yang lebih tinggi yakni sebesar 91.54% pada algoritma C4.5 dan sebesar 78.13% pada algoritma Naïve Bayes. Jadi nyata dengan penerapan metode AdaBoost pada algoritma C4.5 dan Naïve Bayes dapat meningkatkan akurasi prediksi kelayakan pemberian kredit.

DAFTAR PUSTAKA

- Bank Indonesia, Peraturan Bank Indonesia Nomor 6/10/PBI/2004 tanggal 12 April 2004 perihal Sistem Penilaian Tingkat Kesehatan Bank Umum
- C. Ruan, Q. Ruan, and X. Li, "Real Adaboost feature selection for Object Recognition," in IEEE 10th International Conference On Signal Processing Proceeding , pp. 1402–1405, 2010
- Fajar Astuti Hermawati, Data Mining. Yogyakarta: Andi, 2013
- Larose, Daniel T. 2005, Discovering Knowledge Ind Data: An Introduction To Data Mining. John Willey & Sons. Inc
- Larose D, T., 2006, Data Mining Methods and Models, Jhon Wiley & Sons, Inc. Hoboken New Jersey
- Siti Masripah. (2015). Evaluasi Penentuan Kelayakan Pemberian Kredit Koperasi Syariah Menggunakan Algoritma Klasifikasi C4.5. Jurnal Pilar Nusa Mandiri Vol XI. ISSN 1978-1946
- Siti Masripah. (2016). Komparasi Algoritma Klasifikasi Data Mining Untuk Evaluasi Pemberian Kredit. BINA INSANI ICT JOURNAL Vol 3. 187-193. ISSN 2537-9777 (Online)
- Undang-undang Nomor 10 Tahun 1998 tentang Perubahan Atas Undang-undang Nomor 7 Tahun 1992 tentang Perbankan.