

Perbandingan Kinerja Algoritma Data Mining Prediksi Persetujuan Kartu Kredit

IPIN SUGIYARTO
WINDU GATA

Program Studi Magister Ilmu Komputer
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
Jl. Kramat Raya No. 18 Jakarta Pusat
Email: 14002082@nusamandiri.ac.id, wgata@nusamandiri.ac.id

Abstract. *Credit analysis needs to identify and assess factors that can influence customers in credit returns. Accurate measurement and good management capability in dealing with credit risk is an effort to save the economic operations unit and is beneficial for a stable and healthy financial system. Failure to identify credit risk results in loss of income and extends credit risk to a bad type of threat to profitability. Data mining prediction techniques are used to determine credit risk. Using the Cross-Industry Standard Process for data mining CRISP-DM. This study has tested the model using a neural network using PCA feature selection and optimized with the PSO algorithm to predict credit card approval. Several experiments were conducted to see the best results. The results of this study prove the use of a single Neural Net method produces an accuracy of 80.33%. while the use of the hybrid PCA+NN+PSO method has been proven to increase accuracy to 82.67%. Likewise, the AUC NN value of 0.706 increased to 0.749 when the NN was optimized using PSO and using the PCA. This study implements and compares PCA-based SVM, L. Regression and NN algorithms and optimize PSO to improve accuracy in credit card approval predictions.*

Keywords: *Credit analysis, predictions, accuracy, NN, PCA, PSO.*

Abstrak. Analisis kredit perlu mengidentifikasi dan menilai faktor yang dapat mempengaruhi nasabah dalam pengembalian kredit. Pengukuran yang akurat dan kemampuan manajemen yang baik dalam menghadapi risiko kredit merupakan upaya penyelamatan unit operasi ekonomi dan bermanfaat untuk sistem keuangan yang stabil dan sehat. Kegagalan mengidentifikasi risiko kredit berakibat hilangnya pendapatan dan memperluas risiko kredit bertipe buruk ancaman bagi profitabilitas. Metode yang digunakan Cross-Industry Standard Process for Data Mining CRISP-DM yang terdiri dari beberapa tahap yaitu Business Understanding, Data Processing, Model Algoritma, Evaluation. Penelitian ini telah melakukan pengujian model menggunakan neural network dengan menggunakan feature selection PCA dan dioptimasi dengan algoritma PSO untuk memprediksi persetujuan kartu kredit. Beberapa eksperimen dilakukan untuk melihat hasil terbaik. Hasil penelitian ini membuktikan penggunaan metode tunggal Neural Net menghasilkan akurasi sebesar 80.33%. sedangkan penggunaan hybrid method PCA+Neural Net + PSO terbukti meningkatkan akurasi menjadi 82.67%. Begitu juga nilai AUC NN sebesar 0.706 meningkat menjadi 0.749 ketika Neural Network dioptimasi menggunakan PSO dan menggunakan feature selection PCA. Penelitian ini mengimplementasikan dan membandingkan algoritma SVM, Log. Regression dan Neural Network berbasis PCA dan optimize PSO untuk meningkatkan akurasi pada prediksi persetujuan kartu kredit.

Kata kunci: Analisis kredit, prediksi, akurasi, neural network, PCA, PSO.

PENDAHULUAN

Dunia perbankan dan instansi keuangan berperan strategis dalam pembangunan nasional (Menarianti, 2015). Peraturan dan regulasi perbankan berubah dengan diterapkannya *internet banking*. Seluruh data nasabah terkait pinjaman yang lancar maupun bermasalah dapat dipantau dan diverifikasi melalui *internet banking*. Sehingga resiko kredit dapat ditekan. Pada umumnya pemberian kredit dipengaruhi beberapa faktor seperti kepercayaan, kesepakatan, jangka waktu, risiko dan balas jasa. Analisis kredit perlu mengidentifikasi dan menilai faktor-faktor yang dapat mempengaruhi nasabah dalam pengembalian kredit.

Pengukuran yang akurat dan kemampuan manajemen yang baik dalam menghadapi risiko kredit merupakan upaya penyelamatan unit operasi ekonomi dan bermanfaat untuk sistem keuangan yang stabil dan sehat secara keseluruhan dan pembangunan ekonomi yang berkesinambungan. Kegagalan mengidentifikasi risiko kredit mengarah pada hilangnya pendapatan dan memperluas kredit untuk risiko kredit yang bertipe buruk adalah ancaman bagi profitabilitas.

Komponen yang mempengaruhi risiko kredit, adalah debitur akan gagal membayar klaim yang akan ditanggung debitur jika tidak memenuhi kewajiban membayar dan nominal yang hilang akibat risiko atau gagal bayar (Menarianti, 2015).

Teknik klasifikasi *data mining* dapat digunakan untuk menentukan risiko kredit. *Data mining* adalah proses komputasi yang mengungkapkan pola dikumpulkan data dengan menggunakan metode seperti kecerdasan buatan, pembelajaran mesin, statistik dan lain-lain (Lee, Chiu, Lu, & Chen, 2002). Metode yang digunakan dalam *data mining* diselidiki dalam dua kelompok sebagai prediktif dan deskriptif. Dalam metode prediktif, model dibuat dengan menggunakan dataset yang hasilnya diketahui. Misalnya di Bank, sifat pelanggan yang membayar kembali kreditnya bisa terungkap dan model dapat di buat dengan menggunakan kumpulan data sebelumnya tentang pendanaan klient. Setelah itu model bisa digunakan pada yang baru. Pelanggan untuk menentukan kemungkinan membayar kredit mereka kembali. Dalam metode deskriptif, sebuah hubungan dapat di cari antara dua kumpulan data misalnya kebiasaan berbelanja dua budaya yang berbeda dapat diselidiki untuk kesamaan.

Berdasarkan beberapa penelitian sebelumnya seperti: Teknik klasifikasi *data mining* dalam menentukan peningkatan kualitas kredit dan penurunan resiko kredit dengan menggunakan *Logistic Regression, Discriminant Analysis, K-Nearest Neighbor, TAN Technique, Naives Bayes, Decision Tree (C.45), Associative Classification, Neural Network* dan *Support Vector Machine* menjadi beberapa relevansi terhadap penelitian ini.

Dari beberapa penelitian yang dilakukan bahwa jaringan saraf tiruan dapat melakukan pengolahan data dengan baik untuk memprediksi *default* dari klient kartu kredit dan menunjukkan akurasi tertinggi (Pasha, Fatima, Dogar, & Shahzad, 2017).

Jaringan syaraf tiruan mendekati dalam hal rata-rata tingkat klasifikasi yang benar dan dapat memperkirakan biaya klasifikasi yang salah (Akkoç, 2012). Prediksi dari jaringan saraf tiruan akan lebih baik lagi jika di optimasi dengan beberapa metode seperti model *hybird* mendapatkan bantuan dari analisis-analisis diskriminan dalam menyediakan dukungan statistik, mengurangi jumlah input variable dan memberikan solusi awal yang lebih baik (Lee et al., 2002). Penelitian tersebut memang mendukung hipotesis bahwa dua tahap pendekatan *scoring* kredit hibrida yang digunakan dalam penelitian sebelumnya akan memiliki akurasi penilaian kredit yang lebih baik dan lebih baik karakteristik konvergensi untuk model jaringan syaraf yang di rancang. Pada analisis komparasi penelitian sebelumnya dengan menggunakan *cross validation, confussion matrix, ROC curve* dan *T-Test* untuk beberapa algoritma klasifikasi data mining dapat disimpulkan bahwa algoritma Linear Regression memiliki akurasi tertinggi (Menarianti, 2015). Berdasarkan hasil

penelitian sebelumnya dapat di ambil kesimpulan bahwa metode *Neural Network*, Naives Bayes dan *Logistic Regression* memiliki tingkat akurasi yang baik sehingga untuk penelitian selanjutnya akan mencoba membandingkan antara algoritma *Logistic Regression*, *Support Vector Machine* dan *Neural Network* berbasis *feature selection* PCA dan *parameter optimize* PSO untuk mengetahui hasil tingkat akurasi yang terbaik dalam peningkatan kualitas kredit dan mengurangi risiko kredit.

Landasan dan Kerangka Pemikiran

Kredit merupakan penyediaan dana transaksi pinjam meminjam atas persetujuan dan kesepakatan antar pihak bank atau instansi keuangan dengan nasabahnya, serta mewajibkan peminjam untuk membayar dalam jangka waktu yang ditentukan (Menarianti, 2015). Pemberian kredit dilakukan dengan mengidentifikasi dan menilai faktor yang mempengaruhi risiko kredit. Hilangnya pendapatan dan ancaman profitabilitas merupakan hal yang perlu diwaspadai dalam pemberian kredit (Menarianti, 2015).

Industri perbankan dalam mengembangkan kartu kredit merupakan suatu kemajuan yang luar biasa. Pada tahun 1730 kartu kredit pertama kali dikeluarkan berbentuk kartu plastik. Pengguna kartu plastik telah meningkat secara besar-besaran untuk pembelian barang. Dengan menggunakan kartu plastik ini, setiap individu atau pelanggan dan perusahaan pengguna dijauhkan dari risiko (Pasha et al., 2017). Dalam beberapa tahun terakhir para kreditur di Singapura mengalami krisis utang kartu kredit, kartu kredit dan kecurangan diperkirakan akan mencapai puncaknya pada kuartal ketiga 2006 (Lee et al., 2002).

Dalam rangka meningkatkan pangsa pasar, mengeluarkan kartu bank di negara Taiwan mengeluarkan uang tunai dan kartu kredit ke pelamar yang tidak memenuhi syarat. Pada saat yang sama sebagian besar pemegang kartu terlepas dari mentabilitas pembayaran mereka, lebih dari kartu kredit yang digunakan untuk konsumsi dan akumulasi kredit besar dan uang tunai hutang kartu kredit. Krisis menyebabkan pukulan bagi keyakinan pembiayaan konsumen dan hal tersebut merupakan tantangan besar bagi keduanya yaitu bank dan pemegang kartu (Yeh & Lien, 2009).

Penggunaan kartu kredit yang ekstensif adalah penyebab persaingan di industry kredit. Oleh karena itu, ada kebutuhan untuk memperluas dan menerapkan mesin teknik belajar *Machine Learning* untuk dapat mengelola data. Hal tersebut akan menghemat waktu dan mengurangi kesalahan. Aktivitas penipuan juga meningkat karena pertumbuhan cepat dalam kartu kredit transaksi. Penipuan adalah istilah yang tersebar luas.

Penambangan data berisi berbagai teknik yang terlibat didalamnya menyelidiki data yang dapat di akses dan meringkasnya menjadi informasi berharga. Dalam ilmu komputer penambangan data digunakan untuk mendeteksi pola dan hubungan antara besar jumlah data dalam *data base* relasional raksasa. Untuk beberapa algortima pada *machine learning* pola ini digunakan sebagai *input* (Pasha et al., 2017). Algoritma penambangan data banyak digunakan untuk fitur seleksi, klasifikasi, pengelompokan dan aturan farming. Isi paragraf ketiga dari bab kedua.

Data Mining

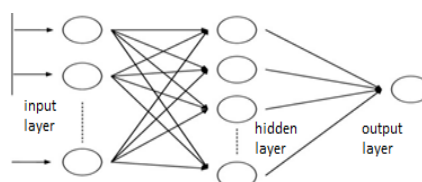
Data mining merupakan pengumpulan dan pemakaian data historis untuk menemukan pola atau hubungan dalam data berukuran besar. *Data mining* adalah proses menganalisis data dari perspektif yang berbeda dan menyimpulkannya menjadi pengetahuan baru dan penting yang dapat digunakan untuk meningkatkan profit (Angga Ginanjar Mabur, 2012). Data mining dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan *field* dari sebuah relasional *data base* yang besar.

Dalam aplikasinya, data *mining* sebenarnya merupakan salah satu bagian proses Cross Industry Standard Process for Datamining (CRISP-DM) yang dikembangkan oleh Daimler Chrysler (then Daimler –Benz), SPSS (then ISL), dan NCR di tahun 1999, CRISP-DM V 1.0 di publis dan diselesaikan serta didokumentasi. Motede ini memberikan kerangka kerja yang seragam dan sebagai pedoman untuk penambangan data (data mining). Proses terdiri dari enam fase atau tahap yang terstruktur dan didefinisikan dengan baik. Fase-fase ini diantaranya yaitu, Business Understanding, Data Understanding, Data Preparation, Model, Evaluation dan Deploy- ment.

Artificial Neural Network

Artificial Neural Network atau Jaringan Saraf Tiruan memiliki prinsip kerja yang mirip dengan cara kerja sel-sel saraf, di mana memiliki bagian yang berfungsi untuk melakukan pemrosesan informasi. Bagian ini akan dihubungkan dengan bagian-bagian serupa melalui jaringan yang memiliki fungsi sebagai rangkaian masukan (*input*) dan keluaran (*output*). Bagian yang memiliki fungsi input dan output ini mirip dengan fungsi dari *Neurit* dan *Dendrit* yang terdapat pada sel saraf manusia. Sehingga dapat dikatakan bahwa ANN adalah suatu struktur pemroses informasi yang terdistribusi dan bekerja secara paralel, yang terdiri dari elemen pemroses yang dihubungkan dengan suatu koneksi.

Setiap elemen pemroses memiliki koneksi keluaran tunggal yang bercabang (*fan out*) ke sejumlah koneksi kolateral yang diinginkan. Keluaran dari elemen pemroses tersebut dapat merupakan hasil pemrosesan data yang dapat berupa persamaan matematis, fungsi atau metode. Proses yang berlangsung pada setiap elemen pemroses harus benar-benar dilakukan secara lokal, yaitu keluaran hanya bergantung pada nilai masukan pada saat data itu diperoleh secara lokal, yaitu keluaran hanya bergantung pada nilai masukan pada saat itu yang diperoleh melalui koneksi dan nilai yang tersimpan dalam memori lokal. ANN merupakan sebuah proses yang terdistribusi secara paralel dan memiliki kemampuan untuk menyimpan pengetahuan yang didapatkannya dari pembelajaran dan membuatnya tetap tersedia untuk dapat digunakan.



Gambar 1. Artificial Neural Network Model

Neural network merupakan kategori ilmu *soft Computing* yang mengadopsi cara kerja otak manusia yang mampu memberikan stimulasi atau rangsangan, melakukan proses, dan memberikan *output*.

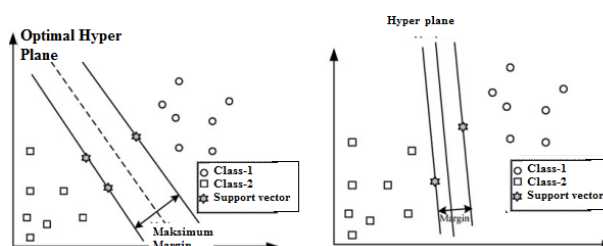
Output yang didapat dari neural network berupa variasi stimulasi dan proses yang terjadi di dalam otak manusia. Kemampuan manusia dalam memproses informasi yang didapat adalah hasil dari kompleksitas proses di dalam otak manusia. Jaringan antar saraf saling mengirimkan informasi satu sama lain dan menghasilkan sebuah informasi yang dapat direpresentasikan sebagai target atau tujuan. Algoritma *neural network* dapat diaplikasikan untuk keperluan *classification*, *pattern resognition*, *function approximation*, *control*, dan *datamining* untuk mencari bentuk tersendiri dari data atau biasa disebut sebagai *knowledge discovery*.

Karakteristik dari *artificial neural network* dapat dilihat dari pola hubungan antar neuron, metode penentuan bobot dari tiap koneksi dan fungsi aktifasinya. Secara umum

model *artificial neural network* terdiri dari, input layer, hidden layer, Output Layer dan fungsi aktivasi.

Support Vector Machine

Support Vector Machine (SVM) adalah metode *learning machine* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyper plane* terbaik yang memisahkan dua buah *class* pada input space. *Hyper plane* terbaik adalah *hyperplane* yang terletak ditengah-tengah antara dua set obyek dari dua *class*. *Hyperplane* pemisah terbaik antara kedua *class* dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya. Margin adalah jarak antara *hyperplane* tersebut dengan pattern terdekat dari masing-masing *class*. *Pattern* yang paling dekat ini disebut sebagai *support vector*.



Gambar 2. *Optimal hyper plan with maximum margin & Hyper plan with margin.*

Logistic regression

Model regresi adalah metode yang biasa digunakan pada prognosticates hubungan barang antara lebih dari dua kelompok. Kendala dalam hal ini model adalah sifat dari target yang seharusnya bernilai biner. *Analog* dengan regresi berganda, yang kuat dengan teknik yang ditawarkan olehnya. Model regresi logistic juga menyediakan ANOVA untuk jawaban tanpa gangguan. Fungsi independen variabel $y_1, y_2, y_3 \dots y_n$ dengan respons yang biner di alam, adalah bagian dari keluarga eksponensial dengan “ $\log \left(\frac{\prod_1}{1-\prod_1} \right), \dots \dots \log \left(\frac{\prod_n}{1-\prod_n} \right)$ ” sebagai parameter kanonik kondisi ilmiah korelasi antara kanonik parameter dan vektor variabel deskriptif x adalah dinyatakan sebagai persamaan $\text{Log} \left(\frac{\prod_i}{1-\prod_i} \right) = x\beta_i$.

Keanggotaan *linear* antara penjelas dan logaritma peluang menciptakan keanggotaan *non-linear* di antara kemungkinan y yang setara dengan 1 dan variabel penjelas dalam vektor. $\prod_i \exp(x\beta_i / (1 + \exp(x\beta_i)))$ Untuk menangani masalah klasifikasi, logistik regresi adalah algoritma yang tepat. Namun, hasil yang dihitung dapat ditunjukkan sebagai probabilitas

Personal Component Analysis

Personal Components Analysis (PCA) atau analisis komponen utama adalah teknik yang digunakan untuk menyederhanakan suatu data, dengan cara mentransformasikan data secara linier sehingga terbentuk koordinat baru dengan varian maksimum. Analisis komponen utama dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan (Putra, Pradina, & Hafidz, 2016).

Tujuan utama analisis komponen utama ialah untuk mengurangi dimensi peubah-peubah yang saling berhubungan dan cukup banyak variabelnya sehingga lebih mudah untuk menginterpretasikan data-data tersebut. Metode yang digunakan yaitu menentukan komponen utama dengan melakukan alih ragam *orthogonal* atau membentuk kombinasi linier $Y = A' X$. Dari proses tersebut akan dipilih beberapa komponen keragaman total data semula.

Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) adalah algoritma perulangan dengan pencarian secara acak. Setiap solusi individu partikel tanpa *volume* dan tanpa massa di ruang pencarian. Di setiap iterasi, partikel memperbarui dirinya sendiri sesuai dengan dua potong informasi ekstrem. Bagian pertama adalah individu nilai optimal *p-best* ditemukan oleh partikel individu dan lainnya adalah nilai optimal global yang ditemukan oleh populasi. Dengan cara ini, populasi partikel dapat mendekati partikel individu yang memiliki adaptasi nilai yang baik dan pada akhirnya menemukan solusi optimal.

Algoritma PSO dapat digambarkan sebagai berikut: misalkan dalam A ruang pencarian target D-dimensi, N partikel membentuk A populasi partikel, posisi dan kecepatan ke-I partikel dinyatakan sebagai: $X_j = (X_{j1}, X_{j2} \dots X_{jD})$ dan $V_j = (V_{j1}, V_{j2} \dots, V_{jD})$. Posisi individual optimal partikel individu yaitu: $P_j = (P_{j1}, P_{j2}, \dots, P_{jD})$ (*pbest*). Posisi optimal global dari seluruh partikel adalah $P_g = (P_{g1}, P_{g2}, \dots, P_{gD})$ (*gbest*). Setiap partikel memperbarui kecepatan dan posisi secara iteratif sesuai dengan rumus berikut:

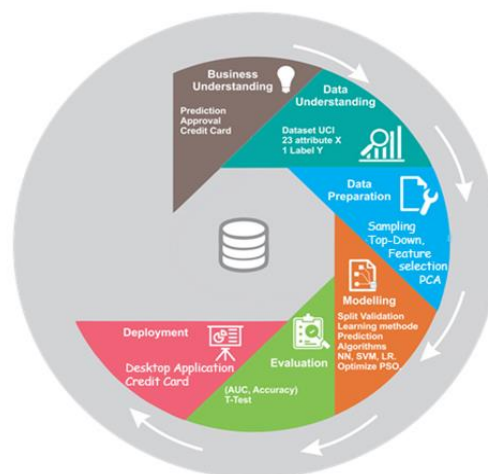
$$V_{jd}^{t+1} = \omega V_{jd}^t + C_1 r_1 + (P_{jd}^t - X_{jd}^t) + C_2 r_2 (P_{gd}^t - X_{jd}^t)$$

$$X_{jd}^{t+1} = X_{jd}^t + V_{jd}^{t+1}$$

Berikut keterangan rumus; $i = (1,2, \dots N)$, $d = (1,2, \dots D)$; ω adalah bobot kelembaman; C_1 dan C_2 merupakan korelasi konstanta; r_1 dan r_2 merupakan angka acak yang didistribusikan secara merata antara 0 dan 1; V_{jd}^t adalah kecepatan partikel saat ini i ; V_{jd}^{t+1} kecepatan partikel i yang diperbarui; X_{jd}^t adalah posisi saat ini dari partikel i ; X_{jd}^{t+1} adalah posisi yang diperbarui dari partikel i .

METODE

Metodologi penelitian yang digunakan penulis dalam penelitian eksperimen ini dengan menggunakan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) terdiri dari enam tahap yaitu *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, Deployment*.



Gambar 3. CRISP-DM Modifikasi.

Business Understanding

Pemahaman bisnis meliputi penetapan tujuan bisnis, penilaian situasi terkini, penetapan tujuan bisnis, penetapan tujuan penggalian data dan pengembangan rencana proyek.

Berdasarkan ini, pendekatan komparasi antara metode algoritma *Logistic Regression*, *Neural Network* dan *Support Vector Machine* berbasis PCA dan optimize dengan PSO akan mampu mendapatkan nilai tingkat akurasi yang terbaik untuk menentukan persetujuan kartu kredit pada 2 tipe kategori disetujui dan tidak disetujui.

Data Understanding

Tahap berikut setelah tujuan bisnis dan rencana proyek ditetapkan, pemahaman data mempertimbangkan data yang dibutuhkan. Langkah ini bisa meliputi pengumpulan data awal, deskripsi data, eksplorasi data dan verifikasi kualitas data. Eksplorasi data seperti peninjauan statistik rangkuman yang meliputi tampilan visual variabel-variabel kategorik bisa terjadi pada akhir tahap ini. Model-model seperti analisis pengelompokan cluster (*cluster analysis*) dapat pula diterapkan ke dalam tahap ini, dengan tujuan mengidentifikasi pola dalam data tersebut.

Dataset yang digunakan dalam penelitian ini adalah *dataset approval credit card UCI Repository* yang diambil 10% dengan model sampling statistik top-down sebanyak 3000 dari 30000 *record* dengan 23 atribut, 1 atribut digunakan sebagai label. Atribut-atribut dataset tersebut terdiri dari limit ball, sex, edu, marriage, age, *payment* 1 – 6, bill amount 1 – 6 dan payment amount 1-6. Kemudian dari artribut-atribut tersebut memiliki label target yaitu 1 berarti disetujui dan 0 tidak disetujui

Data Preparation

Data preparation adalah setelah sumber data yang tersedia diidentifikasi, sumber data tersebut perlu diseleksi, dibersihkan, dibangun ke dalam wujud yang dikehendaki dan di bentuk. Pembersihan dan transformasi data dalam persiapan model data perlu dilakukan pada tahap ini. Eksplorasi data secara lebih mendalam juga dapat diterapkan dalam tahap ini, dan penggunaan model-model tambahan sekali lagi memberikan peluang untuk melihat berbagai pola berdasarkan pemahaman bisnis.

Tahap ini untuk mengurangi dimensi data yang cukup besar maka digunakan feature selection dengan PCA (Principle Component Analysis) dan pengambilan sample data dengan cara top-down yaitu data diambil 1500 dari range teratas dan data diambil 1500 dari range terbawah.

Modelling

Metode penggalian data, seperti visualisasi penggambaran data dan penetapan hubungan serta analisis pengelompokan untuk mengidentifikasi variabel yang berhubungan satu sama lain bermanfaat bagi analisis awal. Alat bantu seperti induksi aturan yang digeneralisasi dapat mengembangkan aturan-aturan asosiasi awal. Begitu pemahaman data yang lebih luas diperoleh seringkali melalui pengenalan pola yang di picu dengan melihat *output* model, model-model lebih terinci yang sesuai dengan jenis data tersebut dapat diterapkan. Pembagian data ke dalam data latihan dan data uji juga diperlukan untuk pembuatan model.

Pada tahap ini dibuatkan model pendekatan yang akan digunakan menggunakan tools rapidminer untuk membandingkan kinerja algoritma antara pendekatan metode algoritma *Logistic Regression*, *Neural Network* dan *Support Vector Machine* berbasis PCA dan optimasi dengan PSO untuk masing-masing tipe persetujuan dalam *dataset credit card UCI Repository*.

Evaluation

Hasil model sebaiknya dievaluasi dalam konteks tujuan bisnis yang ditetapkan pada tahap awal pemahaman bisnis. Hal ini akan mengarahkan pada identifikasi kebutuhan lain kerap kali melalui pengenalan pola, sering kali kembali ke tahap-tahap CRISP-DM sebelumnya. Perolehan pemahaman bisnis merupakan prosedur berulang dalam penggalian data, di mana hasil dari beragam visualisasi, fakta statistik, dan metode kecerdasan buatan menunjukkan hubungan-hubungan baru kepada pengguna yang memberikan pemahaman yang lebih mendalam mengenai operasi perusahaan.

Penelitian ini akan menggunakan teknik *Split Validation* dengan perangkat lunak (*software*) Rapidminer untuk pengujian model pendekatan metode algoritma NN, NN berbasis PCA, Logistic Regression berbasis PCA dan Support Vector Machine berbasis PCA, sebelum membandingkan nilai tingkat akurasi masing-masing tipe kategori *approval* yang terdapat di dataset *credit card* UCI Repository. Penjelasan lebih detail untuk pengujian dari model pendekatan metode algoritma NN, NN berbasis PCA, Logistic Regression berbasis PCA dan SVM berbasis PCA dan optimize PSO akan dijelaskan pada bab IV penelitian ini.

Deployment

Penggalian data dapat digunakan baik untuk membuktikan hipotesis sebelumnya, ataupun untuk penemuan pengetahuan pengidentifikasian hubungan yang tidak terduga dan bermanfaat. Melalui pengetahuan yang ditemukan dalam tahap awal proses CRISP-DM, model yang kuat dapat diperoleh yang mungkin kemudian dapat diterapkan pada kegiatan bisnis untuk berbagai keperluan, termasuk memprediksi atau mengidentifikasi situasi-situasi penting. Model-model ini perlu dipantau untuk mengawasi adanya perubahan dalam operasi, karena apa yang mungkin tepat untuk saat ini mungkin tidak lagi tepat satu tahun ke depan. Jika perubahan besar benar-benar terjadi, model tersebut sebaiknya dibuat ulang. Merupakan hal yang bijaksana.

Pada tahap ini, membangun sebuah system interface berbasis aplikasi desktop berdasarkan dari penelitian setelah mendapatkan hasil nilai tingkat akurasi terbaik dari persetujuan kartu kredit yang terdapat di dalam dataset *credit card* UCI Repository menggunakan salah satu pendekatan metode algoritma prediksi terpilih yang mempunyai akurasi terbaik.

HASIL DAN PEMBAHASAN

Business Understanding

Pengukuran yang akurat dan kemampuan manajemen yang baik dalam menghadapi risiko kredit merupakan upaya penyelamatan unit operasi ekonomi dan bermanfaat untuk sistem keuangan yang stabil dan sehat secara keseluruhan dan pembangunan ekonomi yang berkesinambungan. Kegagalan mengidentifikasi risiko kredit mengarah pada hilangnya pendapatan dan memperluas kredit untuk risiko kredit yang bertipe buruk adalah ancaman bagi profitabilitas.

Kesalahan analisa kredit dapat menyebabkan risiko kredit, seperti menghilangkan nasabah, ketidakpastian pembayaran dana pinjaman bahkan ketidakmampuan nasabah dalam mengembalikan pinjaman dana kredit. Kriteria penilaian kredit seperti sifat atau watak seseorang, kemampuan membayar, penggunaan dana, kondisi sosial, ekonomi dan politik serta jaminan yang diajukan diperlukan untuk memberikan informasi mengenai itikad baik dan kemampuan membayar seorang nasabah.

Data Understanding

Penelitian ini menggunakan sumber dataset dari UCI repository tentang *approval credit card*. Dataset yang diolah berjumlah 30.000 rows data, 23 atribut prediktor dan dengan Label target 1 dan 0. Berikut ini atribut yang digunakan dalam penelitian yaitu:

Tabel 1. Tabel Input dan Output Approval Credit Card.

Keterangan	Variabel	Kode
<i>Input</i>	Limmit_Bal	X1
	Sex	X2
	Education	X3
	Marriage	X4
	Age	X5
	Pay_0	X6
	Pay_2	X7
	Pay_3	X8
	Pay_4	X9
	Pay_5	X10
	Pay_6	X11
	Bill_Amt1	X12
	Bill_Amt2	X13
	Bill_Amt3	X14
	Bill_Amt4	X15
	Bill_Amt5	X16
	Bill_Amt6	X17
	Pay_Amt1	X18
	Pay_Amt2	X19
	Pay_Amt3	X20
	Pay_Amt4	X21
	Pay_Amt5	X22
	Pay_Amt6	X23
<i>Out</i>	Default Credit Card	Y

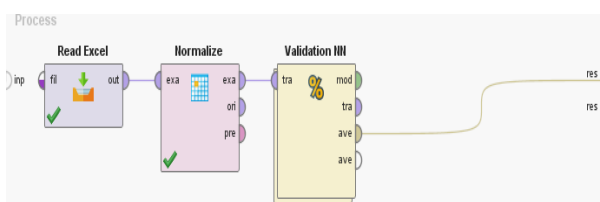
Data Preparation

Pada tahap ini data dinormalisasi untuk keseragaman range data min-0 sampai dengan maks-1 , kemudian dilakukan teknik feature selection PCA (*Principal Component Analysis*). Feature selection digunakan untuk memperoleh atribut-atribut yang paling berpengaruh menurut algoritma *feature selection* PCA.

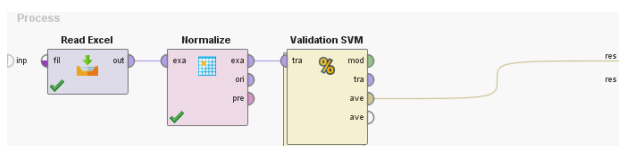
Setelah data terpilih maka didapat dari 23 atribut menjadi 15 atribut baru dan langkah selanjutnya mengurangi dimensi data dari 30.000 menjadi 3000 rows. Teknik ini disebut dengan dimension reduction tujuannya mengurangi beban proses dan tanpa mengurangi tingkat akurasi. Untuk pemilihan data secara random yaitu mengambil top data sejumlah 1500 sampel dan down data sejumlah 1500 sampel

Pembuatan Modeling Metode

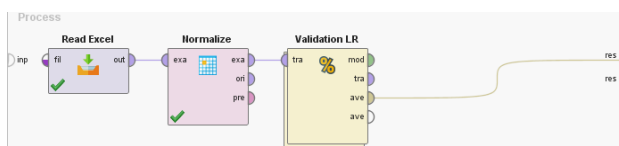
Pada tahap ini model untuk metode menggunakan CRISP-DM yang telah dimodifikasi, sebagai berikut:



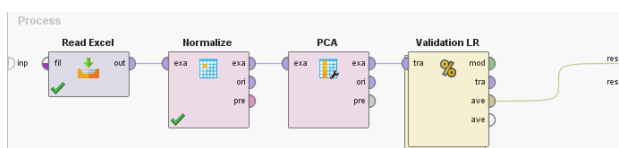
Gambar 4. Model Proses Metode Neural Network.



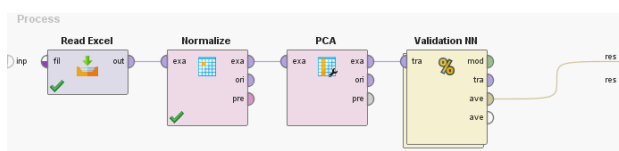
Gambar 5. Model Proses Metode Support Vector.



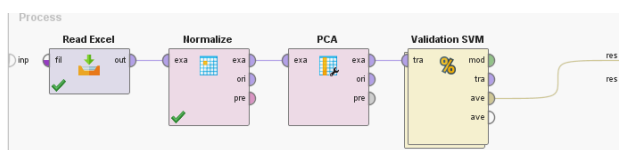
Gambar 6. Model Proses Metode Logistic Regression.



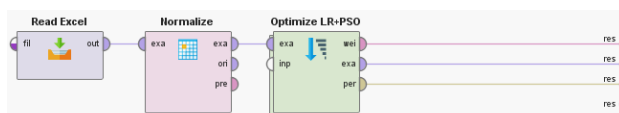
Gambar 7. Model Proses Metode Logistic Regression+PCA.



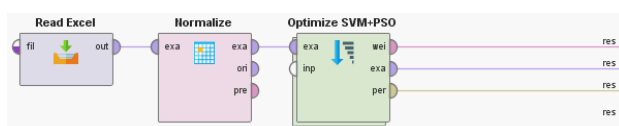
Gambar 8. Model Proses Metode NN+PCA.



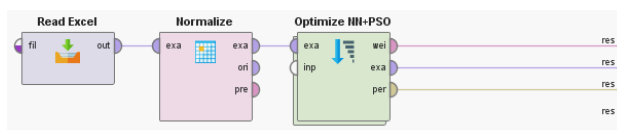
Gambar 9. Model Proses Metode SVM+PCA.



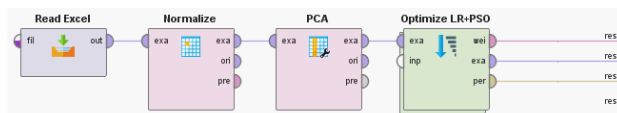
Gambar 10. Model Proses Metode Logistic Regression+PSO.



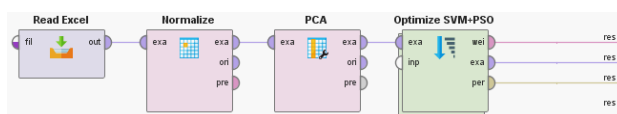
Gambar 11. Model Proses Metode SVM+PSO.



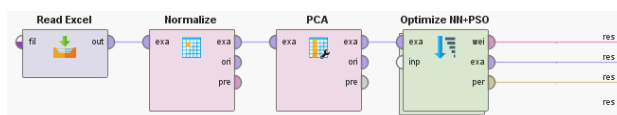
Gambar 12. Model Proses Metode NN+PSO.



Gambar 13. Model Proses Metode LR+PCA+PSO.



Gambar 14. Model Proses Metode SVM+PCA+PSO.



Gambar 15. Model Proses Metode NN+PCA+PSO.

Hasil Penelitian

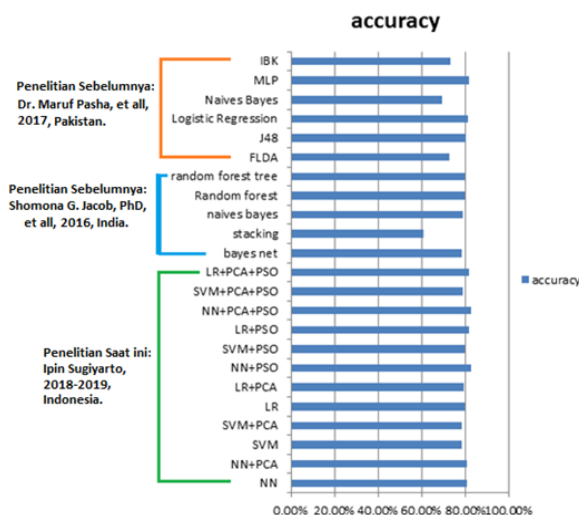
Tahap ini merupakan hasil dari pengolahan pada proses validasi untuk mengetahui akurasi komparasi prediktif dari ke-3 metode yang digunakan yaitu Algoritma *Neural Network*, *Support Vector Machine* dan *Logistic Regression* serta penambahan metode *feature selection* PCA dan *parameter Optimize* dengan PSO. Berikut ini adalah hasil penyajian dari data yang sudah di olah berupa nilai akurasi dan grafik perbandingan akurasi pada penelitian terkait di tahun sebelumnya dan penelitian saat ini.

Tabel 2. Hasil Komparasi Algoritma Datamining Pada Penelian Approval Credit Card.

Peneliti	Algoritma	Akurasi
Ajay, Ajay Vankates & Shomona Gracia Jacob, PhD, India, Tahun (2016).	Bayes Net	0.781
	Stacking	0.607
	Naives Bayes	0.786
	Random Forest	0.798
	Random Forest Tree	0.797
Dr. Maruf Pasha, Meherwar Fatima, Abdul Hanan, & Furrakh Shahzad, Pakistan, Tahun (2017).	FLDA	0.724
	J48	0.803
	Logistic Regression	0.81
	Naives Bayes	0.694
	MLP	0.817
Ipin Sugiyarto, Indonesia, Tahun 2018-2019.	IBK	0.729
	Neural Net.	0.803
	Support Vector Machine	0.782
	Logistic Regression	0.794

Peneliti	Algoritma	Akurasi
	NN+PCA	0.806
	SVM+PCA	0.782
	LR+PCA	0.793
	NN+PSO	0.824
	SVM+PSO	0.794
	LR+PSO	0.813
	NN+PCA+PSO	0.826
	SVM+PCA+PSO	0.787
	LR+PCA+PSO	0.816

Berdasarkan tabel 2. diperoleh beberapa hasil prediksi dengan algoritma yang beragam dan yang dilakukan pada penelitian tahun sebelumnya masing-masing memiliki nilai akurasi tertinggi seperti di tahun 2016 didapatkan hasil tertinggi pada algoritma random forest yaitu nilai akurasi 79.80 %, kemudian tahun 2017 algoritma tertinggi dengan MLP yaitu akurasi 81.70% dan penelitian saat ini tahun 2018-2019 algoritma terbaik dengan Neural Net + PCA + PSO dengan akurasi 82.60%.



Gambar 16. Grafik Perbandingan Akurasi Penelitian Pada *Approval Credit Card*.

PENUTUP

Simpulan

Penelitian ini telah melakukan pengujian model menggunakan *neural network* dengan menggunakan *feature selection* PCA dan dioptimasi dengan algoritma *Particle Swarm Optimize* (PSO) untuk memprediksi persetujuan kartu kredit. Beberapa eksperimen dilakukan untuk melihat hasil terbaik. Hasil penelitian ini membuktikan penggunaan metode *Neural Network* menghasilkan akurasi sebesar 80.33%. sedangkan penggunaan *feature selection* dan *parameter optimize* dengan method *Neural Network + PCA + PSO* terbukti meningkatkan akurasi menjadi 82.67%. Begitu juga nilai AUC NN sebesar 0.706 meningkat menjadi 0.749 ketika *Neural Network* dioptimasi menggunakan PSO dan menggunakan *feature selection* PCA.

Saran

Adapun model yang telah terbentuk selanjutnya dapat dikembangkan dan dapat diimplementasikan kedalam sebuah aplikasi. Sehingga dapat membantu dan memudahkan bagi para pemegang kepentingan dalam pengambilan sebuah keputusan untuk

memprediksi persetujuan kredit untuk nasabah diperbankan. Saran pada penelitian ini bahwa optimasi parameter pada *Neural Network* menggunakan *Particle Swarm Optimize* dapat meningkatkan akurasi dan *feature selection* PCA digunakan untuk mengurangi dimensi data agar proses pengolahan lebih efisien dari data yang banyak dan beragam. Namun ada beberapa faktor yang dapat di coba untuk penelitian selanjutnya, untuk mencari metode yang sama-sama memiliki hasil terbaik, yaitu untuk mengembangkan sistem informasi keputusan prediksi persetujuan kredit, menerapkan sistem informasi manajerial di perbankan dan melakukan penelitian lanjutan yang mempunyai pengaruh terhadap hasil prediksi nasabah kreditur.

DAFTAR PUSTAKA

- Akkoç, S. (2012). An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research*, 222(1), 168–178. <https://doi.org/10.1016/j.ejor.2012.04.009>
- Angga Ginanjar Mabur, R. L. (2012). PENERAPAN DATA MINING UNTUK MEMPREDIKSI Program Studi Teknik Informatika Jurnal Komputer dan Informatika (KOMPUTA). *Jurnal Komputer Dan Informatika (KOMPUTA)*, 1, 53–57.
- Lee, T. S., Chiu, C. C., Lu, C. J., & Chen, I. F. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications*, 23(3), 245–254. [https://doi.org/10.1016/S0957-4174\(02\)00044-1](https://doi.org/10.1016/S0957-4174(02)00044-1)
- Menarianti, I. (2015). Klasifikasi data mining dalam menentukan pemberian kredit bagi nasabah koperasi. *Jurnal Ilmiah Teknosains*, 1(1), 1–10.
- Pasha, M., Fatima, M., Dogar, A. M., & Shahzad, F. (2017). Performance Comparison of Data Mining Algorithms for the Predictive Accuracy of Credit Card Defaulters. *International Journal of Computer Science and Network Security*, 17(3), 178–183.
- Putra, S. F., Pradina, R., & Hafidz, I. (2016). Feature Selection pada Dataset Faktor Kesiapan Bencana pada Provinsi di Indonesia Menggunakan Metode PCA (Principal Component Analysis). *Jurnal Teknik Its*, 5(2), 5–9.
- Yeh, I. C., & Lien, C. hui. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 36(2 PART 1), 2473–2480. <https://doi.org/10.1016/j.eswa.2007.12.020>