

Memprediksi Nasabah Kredit yang Berisiko dengan Menerapkan Model CRISP-DM

Bank merupakan salah satu sumber modal yang dipakai untuk mengembangkan usaha oleh pengusaha dan pegiat Usaha Mikro Kecil dan Menengah (UMKM). Hampir 46% pelaku usaha mengalami kesulitan membayar tagihan atau hutang (Azizah, 2022) yang menyebabkan bank harus selektif untuk menyetujui permintaan hutang dari pelaku usaha yang digunakan untuk menambah modal bisnisnya. Oleh sebab itu, bank akan mengalami banyak kesulitan dalam perputaran uang jika kejadian hutang macet terus berulang maka perlu dilakukan analisis terjadinya hutang macet dengan menggali data (mining) histori pelaku usaha yang akan berhutang.

Penggalan data nasabah atau peminjam untuk modal/kredit tetap menarik untuk diteliti, terutama mengingat inflasi ekonomi saat ini yang sulit diprediksi. Hal ini menjadi perhatian utama bagi industri keuangan atau lembaga pemberi pinjaman, termasuk lembaga perbankan ketika mengidentifikasi dan memprediksi data nasabah yang berisiko sebagai dasar pengambilan keputusan operasional.

Ada beberapa algoritma atau metode klasifikasi data mining yang dapat digunakan sebagai strategi pemasaran dan periklanan, antara lain Support Vector Machines (SVM), Naïve Bayes (NB), dan Decision Tree (DT). Metode SVM merupakan salah satu metode pembelajaran mesin (supervised learning) untuk mengklasifikasikan data historis dengan mencari bidang pemisah atau hyperplane terbaik yang dapat memisahkan data dimensional secara sempurna ke dalam kelas-kelas (Zainuddin dan Selamat, 2014).

Prinsip kerja dari metode SVM ini adalah mencari ruang pemisah yang paling optimal dari sekumpulan data dalam kelas yang berbeda. Hyperplane dapat ditemukan dengan memaksimalkan margin atau jarak antara titik kelas terdekat (support vector) dan hyperplane. Data sampel biasanya tidak dapat dipisahkan secara linier, tetapi SVM memperkenalkan gagasan untuk meningkatkan dimensi data. Secara umum menggunakan dimensi ruang yang lebih tinggi akan menyebabkan masalah mesin dan overfitting. Masalah ini dapat diselesaikan dengan menggunakan dot-product dalam ruang (Boswell, 2003).

Penelitian ini mengusulkan untuk mengklasifikasikan nasabah perbankan berisiko dan tidak berisiko untuk pemberian pinjaman modal menggunakan algoritma SVM yang mengacu pada

sebuah standar proses data mining standar CRISP-DM. Dataset penelitian terdiri dari data publik berskala besar yang diolah melalui langkah-langkah data mining untuk menemukan pola yang menjadi dasar untuk mengklasifikasikan nasabah kredit bank yang berisiko atau tidak berisiko.

Langkah penelitian ini dilakukan dengan cara mengadopsi sebuah standar proses data mining yang dikenal sebagai CrossIndustry Standard Process for Data Mining atau CRISP-DM. Standar proses tersebut terdiri dari lima fase yaitu fase pemahaman bisnis, fase pemahaman data, fase pengolahan data, fase pemodelan, dan fase evaluasi & validasi.



Gambar 1. Standar Proses Model CRISPDM (Sumber: <https://www.datascience-pm.com/crisp-dm-2/>)

1. Fase Pemahaman Bisnis

Dataset penelitian merupakan data publik yang diunduh dari kaggle.com tentang prediksi nasabah berisiko.

2. Fase Pemahaman Data

Dataset tersebut terdiri atas 11 atribut prediktor dan 1 label dengan penjelasan yang terdapat dalam Tabel 1.

Tabel 1. Keterangan Atribut Prediktor

Variabel	Keterangan
Income	Pendapatan
Age	Usia
Experience	Pengalaman Kerja Keseluruhan

Married/Single	Status Pernikahan
House_Ownership	Status Kepemilikan Rumah
Car_Ownership	Status Kepemilikan Mobil
Profession	Profesi Saat Ini
City	Kota Tempat Tinggal
State	Provinsi Tempat Tinggal
Current_Job_Years	Lama Bekerja di Pekerjaan Saat Ini
Current_House_Years	Lama Tinggal di Rumah Saat Ini

Tabel 2. Keterangan Atribut Label/Kelas

Variabel	Keterangan
Risk Flag	Keputusan Peminjaman (label)

3. Fase Pengolahan Data

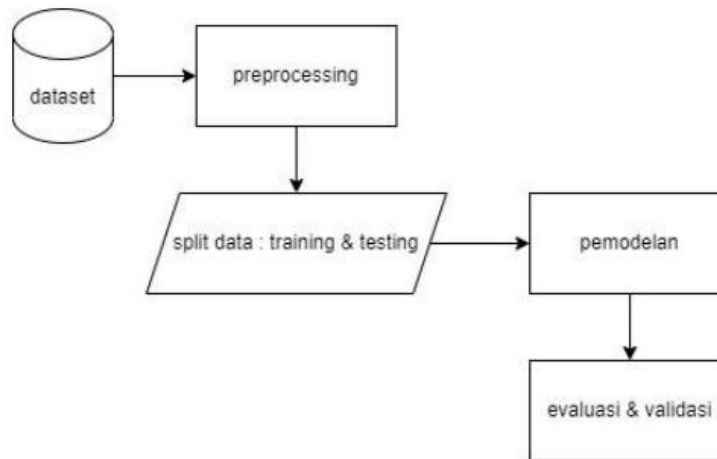
Pada tahap ini dilakukan seleksi atribut, pembersihan data, dan membagi data menjadi data training dan data testing.

4. Fase Pemodelan (Modelling Phase)

Penelitian prediksi ini menggunakan metode SVM.

5. Fase Evaluasi dan Validasi

Pada fase ini dilakukan pengukuran performa model menggunakan teknik Confussion Matrix, serta 10-fold Cross Validation untuk memvalidasi model.



Gambar 2. Tahapan Pemodelan Prediksi Support Vector Machine (Sumber : <https://univ45sby.ac.id/ejournal/index.php/informatika/article/view/291>)

Algoritma ini digunakan untuk mengklasifikasi data linier dan tidak linier. Menghitung lebar margin secara maksimal perlu dilakukan untuk mendapatkan hyperplane yang baik di antara satu kelas dengan kelas lain. Margin adalah jarak dari hyperplane atau bidang pemisah optimal terhadap point terdekat yang berada di masing-masing kelas. Point paling dekat disebut support vector (Nugroho, dkk., 2003).

Proses pelaksanaan penelitian ini dilakukan dengan tahapan berikut.

Sample Data Set

Pengambilan dataset awal menggunakan variabel kriteria yang diambil dari kaggle.com. Atribut yang dipakai sebagai variabel/parameter yang mempengaruhi kelayakan nasabah dalam melakukan peminjaman seperti pada Tabel 1.

Id	Income	Age	Experienc	Married/S	House_Ov	Car_Own	Profession	CITY	STATE	CURRENT_	CURRENT_	Risk_Flag
1	1303834	23	3	single	rented	no	Mechanic	Rewa	Madhya_P	3	13	0
2	7574516	40	10	single	rented	no	Software_	Parbhani	Maharash	9	13	0
3	3991815	66	4	married	rented	no	Technical	Alappuzha	Kerala	4	10	0
4	6256451	41	2	single	rented	yes	Software_	Bhubanes	Odisha	2	12	1
5	5768871	47	11	single	rented	no	Civil_serv	Tiruchirap	Tamil_Na	3	14	1
6	6915937	64	0	single	rented	no	Civil_serv	Jalgaon	Maharash	0	12	0
7	3954973	58	14	married	rented	no	Librarian	Tiruppur	Tamil_Na	8	12	0
8	1706172	33	2	single	rented	no	Economist	Jamnagar	Gujarat	2	14	0
9	7566849	24	17	single	rented	yes	Flight_att	Kota[6]	Rajasthan	11	11	0
10	8964846	23	12	single	rented	no	Architect	Karimnag	Telangana	5	13	0
11	4634680	78	7	single	rented	no	Flight_att	Hajipur[3]	Bihar	7	12	0
12	6623263	22	4	single	rented	no	Designer	Adoni	Andhra_P	4	14	0
13	9120988	28	9	single	rented	no	Physician	Erode[17]	Tamil_Na	9	12	0
14	8043880	57	12	single	rented	no	Financial	Kollam	Kerala	8	10	0
15	9420838	48	6	single	rented	no	Technical	Madurai	Tamil_Na	6	10	1
16	5694236	39	2	married	rented	yes	Economist	Anantapu	Andhra_P	2	10	0
17	7315840	71	8	married	rented	no	Air_traffic	Kamarhati	West_Ben	8	14	0
18	3666346	56	12	single	rented	no	Politician	Bhusawal	Maharash	12	11	1
19	2241112	28	8	single	rented	no	Police_off	Sirsa	Haryana	6	14	0
20	5431918	40	1	single	rented	no	Artist	Amaravati	Andhra_P	1	14	0

Gambar 3. Tahapan Pemodelan Prediksi Support Vector Machine (Sumber : <https://www.kaggle.com/datasets/subhamjain/loan-prediction-based-on-customer-behavior?select=Training+Data.csv>)

Dengan menyesuaikan kondisi di Indonesia maka peneliti menghilangkan atribut profession, city, state karena berbeda dengan karakter nasabah di Indonesia. Dilanjutkan dengan pemilihan atribut yang dijadikan label/kelas sebagai variabel prediksi. Pada Gambar 3 ditampilkan atribut dari kaggle.com.

Data tersebut dilakukan preprocessing data dengan melakukan proses pada Rapidminer menggunakan Operator Read CSV untuk membaca data. Mengubah tipe data untuk atribut yang terpilih yang mempunyai tipe nominal ke tipe numerik dengan menggunakan Operator nominal to numerical. Kemudian data dibagi menjadi data latih dan data tes dengan rasio 70% dan 30% dengan operator Split data. Ketika data siap diolah maka dengan menggunakan operator SVM yang diberi input data latih, menghasilkan output modul (mod). Kemudian output SVM dijadikan input pada operator apply model dan diberi input data tes yang merupakan output dari split data. . Untuk mengetahui performance dari hasil klasifikasi.

Hasil dari performance yang dicapai didapatkan tingkat akurasi mencapai 80% dengan nilai perbandingan antara data latih dan data tes adalah 70% dan 30%, dan mencapai tingkat akurasi 80% dengan nilai data perbandingan antara data latih dan data tes adalah 80% dan 20%.

Penelitian ini menemukan tingkat akurasi yang dihasilkan oleh Algoritma SVM yaitu 80%, Hasil pengukuran performa model menunjukkan true positive sebanyak 24 nasabah dari 30 nasabah. Model prediksi yang diusulkan telah tervalidasi sehingga dapat dimanfaatkan untuk bahan pengetahuan pengambilan kebijakan.

Dashboard development:

