

**PERBANDINGAN ALGORITMA XGBOOST DAN SVM PADA
CREDIT SCORING UNTUK MENENTUKAN KELAYAKAN
NASABAH
SKRIPSI**



OLEH:

KAISAR FAUZAN

20081010142

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2024**

LEMBAR PENGESAHAN SKRIPSI

Judul : PERBANDINGAN ALGORITMA XGBOOST DAN SVM PADA CREDIT SCORING UNTUK MENENTUKAN KELAYAKAN NASABAH

Nama: Kaisar Fauzan

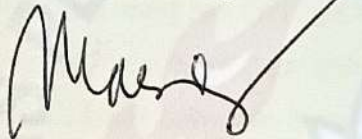
NPM : 20081010142

Telah Diseminarkan Dalam Ujian Skripsi Pada :

Mengetahui:

Dosen Pembimbing

1.



Dr. I Gede Susrama Mas Diyasa, ST. MT

NIP : 19700619 202121 1 009

2.

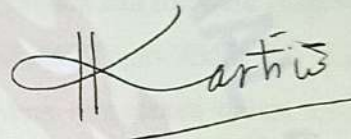


Eva Yulia Puspaningrum, S.Kom, M.Kom

NIP : 19890705 202121 2 002

Dosen penguji

1.



Dr. Ir. Kartini, S.Kom. MT

NIP. 19611110 199103 2 001

2.



Eka Prakarsa Mandvartha, S.T., M.Kom

NIP : 19880525 201803 1 001

Menyetujui:

Dekan

Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.

NIP : 19681126 199403 2 001

Koordinator Program Studi

Teknik Informatika



Fetty Tri Anggraeny, S.Kom, M.Kom

NIP : 19820211 202121 2 005

SURAT PERNYATAAN BEBAS DARI PLAGIASI

Saya, mahasiswa Program Studi Sarjana Informatika Universitas Pembangunan Nasional "Veteran" Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Kaisar Fauzan

NPM : 20081010142

Menyatakan dengan sesungguhnya bahwa Skripsi/Tugas Akhir yang saya kerjakan berjudul:

"PERBANDINGAN ALGORITMA XGBOOST DAN SVM PADA CREDIT SCORING UNTUK MENENTUKAN KELAYAKAN NASABAH"

bukan merupakan plagiasi sebagian atau keseluruhan dari Skripsi/Tugas Akhir/Penelitian orang lain dari juga bukan merupakan produk dan software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/Tugas Akhir ini secara keseluruhan adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di Institut Pendidikan lain. Bukti hasil pengecekan plagiasi dokumen ini dapat ditelusuri melalui QR Code di bawah.

Apabila di kemudian hari terbukti bahwa dokumen ini merupakan plagiasi karya orang lain, saya sanggup menerima sanksi sesuai aturan yang berlaku.

Demikian atas perhatiannya disampaikan terima kasih.

Surabaya, 14 July 2024

Hormat saya,



Kaisar Fauzan

NPM. 20081010142



PERBANDINGAN ALGORITMA XGBOOST DAN SVM PADA CREDIT SCORING UNTUK MENENTUKAN KELAYAKAN NASABAH

Nama Mahasiswa : Kaisar Fauzan

NPM : 20081010142

Program Studi : Informatika

Dosen Pembimbing 1 : Dr. I Gede Susrama Mas Diyasa, ST. MT

Dosen Pembimbing 2 : Eva Yulia Puspaningrum, S.Kom, M.Kom

Abstrak

Credit Scoring digunakan untuk menentukan kelayakan nasabah dalam mengajukan pinjaman pada lembaga atau instansi yang menyediakannya. *Credit Scoring* akan menentukan apakah nasabah layak untuk melakukan pinjaman

Beberapa penelitian telah dilakukan guna mengetahui kelayakan nasabah yang melakukan pinjaman dengan menggunakan sistem credit scoring. Penelitian ini bertujuan untuk mengetahui persentase nasabah yang layak dan tidak layak diberikan pinjaman oleh bank. Sejalan dengan itu, penelitian ini menggunakan dataset yang berasal dari perusahaan Home Credit Indonesia, dataset ini diperoleh dari situs *open source* yang bernama Kaggle.

Pada kesempatan ini, peneliti menggunakan beberapa metode, yakni CRISP-DM, XGBoost, dan SVM. Penggunaan CRISP-DM pada penelitian ini berguna untuk membantu peneliti dalam melakukan analisis data dengan lebih efektif dan teratur. Sedangkan, penggunaan XGBoost berguna untuk menghasilkan insight dengan akurasi yang tinggi dan pengolahan data dengan jumlah yang besar namun waktu yang dibutuhkan relatif cepat. Selain itu, penggunaan SVM berguna untuk menggeneralisasi suatu data dengan baik serta menghindari masalah overfitting pada saat klasifikasinya. Selain itu, pada penelitian ini menghasilkan beberapa insight yang berkaitan dengan credit scoring serta menghasilkan kelayakan dari seorang nasabah yang melakukan peminjaman.

Selain itu, untuk evaluasi terhadap hasil klasifikasi beserta tren menggunakan Confusion Matrix dan ROC AUC. Untuk akurasi dengan menggunakan algoritma SVM sebesar 70.88%, precision 63.91%, recall 95.89%, ROC-AUC 51.21%. untuk akurasi menggunakan XGBoost sebesar 93.01%, precision 97.88%, recall 91.08%, ROC_AUC 96.71%

Kata Kunci: *Credit Scoring, XGBoost, SVM, Confussion Matrix, ROC_AUC*

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Pebandingan Algoritma XGBoost Dan SVM Pada Credit Scoring Untuk Menentukan Kelayakan Nasabah".

Penulisan skripsi ini dimaksudkan untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan program Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur. Dalam proses penulisan skripsi ini, penulis mendapatkan banyak bantuan dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Ir. Akhmad Fauzi, MMT.,IPU selaku Rektor UPN "Veteran" Jawa Timur yang telah memberikan fasilitas dan kesempatan untuk menyelesaikan studi di UPN "Veteran" Jawa Timur
2. Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer, yang telah memberikan fasilitas dan kesempatan untuk menyelesaikan studi di Fakultas Ilmu Komputer.
3. Fetty Tri Anggraeny, S.Kom. M.Kom., selaku Ketua Program Studi Informatika, yang telah memberikan kemudahan dalam berbagai urusan akademik.
4. Dr. I Gede Susrama Mas Diyasa, ST. MT., selaku dosen pembimbing utama yang telah memberikan bimbingan, arahan, serta motivasi kepada penulis dalam menyusun skripsi ini.
5. Eva Yulia Puspaningrum, S.Kom, M.Kom., selaku dosen pembimbing kedua yang telah memberikan masukan berharga dan dukungan selama proses penulisan.

6. Dr. Ir. Kartini, S.Kom. MT., selaku dosen penguji pertama yang telah memberikan saran dan kritik yang membangun.
7. Eka Prakarsa Mandyartha, S.T., M.Kom., selaku dosen penguji kedua yang telah memberikan masukan berharga.
8. Orang tua dan keluarga tercinta yang selalu memberikan dukungan moral, spiritual, dan material kepada penulis.
9. Teman-teman dan rekan-rekan seperjuangan di Program Studi Informatika, yang telah memberikan semangat dan kerjasama selama masa studi.
10. Semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan bantuan dan dukungan selama penulisan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari berbagai pihak untuk penyempurnaan skripsi ini. Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi semua pihak yang berkepentingan.

Surabaya 08 juli 2024



penyusun

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	2
SURAT PERNYATAAN BEBAS DARI PLAGIASI	3
KATA PENGANTAR.....	6
DAFTAR ISI.....	8
DAFTAR GAMBAR.....	13
DAFTAR TABEL	14
BAB I PENDAHULUAN.....	16
1.1 Latar Belakang.....	16
1.2 Rumusan Masalah.....	18
1.3 Tujuan.....	18
1.4 Manfaat	19
1.5 Batasan Masalah.....	19
BAB II TINJAUAN PUSTAKA	20
2.1 Algoritma XGBoost.....	20
2.2 Algoritma SVM.....	20
2.3 Python.....	21
2.4 Power BI.....	21
2.5 Credit Scoring	22

2.6 CRISP-DM.....	22
2.6.1 Business Understanding	23
2.6.2 Data Understanding	24
2.6.2.1 application_{train test}.csv.....	24
2.6.2.2 bureau balance.csv.....	24
2.6.2.3 previous_application.csv.....	24
2.6.2.4 bureau.csv	25
2.6.2.5 POS_CASH_balance.csv.....	25
2.6.2.6 credit_card_balance.csv	25
2.6.2.7 installments_payments.csv	25
2.6.3 Data Preparation	25
2.6.4 Modelling.....	26
2.6.4.1 XGBoost.....	26
2.6.4.2 SVM.....	28
2.6.5 Evaluation.....	29
2.6.5.1 Confusion matrix	29
2.6.5.2 ROC-AUC.....	29
2.6.5.3 Cross-validation.....	30
2.6.5.4 T-TEST.....	30
2.6.6 Deployment.....	30
BAB III METODE PENELITIAN.....	31
3.1 Metode.....	31

3.2 Data akuisi.....	32
3.3 CRISP DM	35
3.3.1 Business Understanding.....	36
3.3.2 Data Understanding.....	36
3.3.3 Data Preparation	36
3.3.3.1 Data Selection.....	36
3.3.3.2 Data Cleaning.....	37
3.3.3.3 Data Transforming.....	37
3.3.4 Modelling	37
3.3.4.1 SVM	37
3.3.4.2 XGBoost.....	41
3.3.5 Evaluation.....	44
3.3.5.1 Confusion Matrix	46
3.3.5.2 ROC Curve	47
3.3.5.3 Cross -Validation	48
3.3.5.4 T-TEST.....	49
BAB IV HASIL DAN PEMBAHASAN.....	51
4.1 CRISP-DM.....	51
4.1.1 Business Understanding.....	51
4.1.2 Data Understanding.....	51
4.1.2.1 Distribusi TARGET.....	51
4.1.2.2 Pengaruh Gender Terhadap Kelayakan Nasasbah	52

4.1.2.3 Korelasi Antara Usia Nasabah Dan Kelayakan Nasabah	53
4.1.2.4 Pengaruh Jumlah Anak Terhadap Kelayakan Nasabah	54
4.1.2.5 Korelasi Status Kepemilikan Kendaraan Terhadap Kelayakan Nasabah	55
4.1.2.6 Korelasi Status Keluarga Terhadap Kelayakan Nasabah	56
4.1.3 Data Preparation	57
4.1.3.1 Data cleaning	57
4.1.3.2 Data Selection.....	58
4.1.3.3 Data Transformation	59
4.1.4 Modelling	60
4.1.4.1 Pengujian.....	60
4.1.4.1.1 Pengujian SVM	60
4.1.4.1.2 Pengujian XGBoost.....	62
4.1.4.2 Hasil	65
4.1.4.2.1 Hasil Klasifikasi oleh SVM	65
4.1.4.2.2 Hasil Klasifikasi oleh XGBoost.....	67
4.1.5 Evaluation.....	69
4.1.5.1 SVM	69
4.1.5.2 XGBoost.....	71
4.1.5.3 Cross-Validation.....	73
4.1.5.4 T-TEST.....	74
4.1.6 Deployment.....	74

BAB V KESIMPULAN DAN SARAN	76
5.1 Kesimpulan	76
5.2 Saran	76
DAFTAR PUSTAKA	77

DAFTAR GAMBAR

Gambar 2.1 Tahapan CRISP-DM.....	23
Gambar 2.2 Algoritma XGBoost.....	27
Gambar 2.3 Algoritma SVM.....	28
Gambar 3.1 flowchart	31
Gambar 3.2 Dataset Home Credit Default Risk	33
Gambar 3.3 Tipe Data pada Atribut.....	35
Gambar 3.3 Algoritma SVM.....	38
Gambar 3.5 Algoritma XGBoost.....	41
Gambar 3.7 confusion matrix.....	46
Gambar 3.9 koordinat ROC-curve.....	47
Gambar 4.1 Pie-chart distribusi Kelayakan Nasabah	52
Gambar 4.2 Pengaruh gender terhadap kelayakan nasabah	53
Gambar 4.3 Korelasi Antara Usia Nasabah Dan Kelayakan Nasabah.....	54
Gambar 4.4 Pengaruh Jumlah anak Terhadap Kelayakan Nasabah.....	55
Gambar 4.5 Status Kepemilikan Kendaraan Terhadap Kelayakan Nasabah	56
Gambar 4.6 Pengaruh Status Keluarga Terhadap Kelayakan Nasabah	57
Gambar 4.7 sebelum dan sesudah cleaning.....	58
Gambar 4.8 data selection	58
Gambar 4.9 data transformation	59
Gambar 4.10 hasil T-test dan p-value pada model SVM dan XGBoost.....	74
Gambar 4.11 Dashboard.....	75

DAFTAR TABEL

Tabel 2.1 confusion matrix.....	29
Tabel 3.1 Contoh Atribut pada dataset	33
Tabel 3.2 nilai untuk perhitungan rumus	39
Tabel 3.3 Skema Pengujian.....	44
Tabel 4.1 Tingkat akurasi model SVM dan XGBoost.....	60
Tabel 4.2 hasil skema uji SVM (70% train dan 30% test).....	61
Tabel 4.3 hasil skema uji SVM (80% train dan 20% test).....	61
Tabel 4.4 hasil skema uji SVM (90% train dan 10% test).....	62
Tabel 4.5 hasil skema uji XGBoost (70% train dan 30% test).....	62
Tabel 4.6 hasil skema uji XGBoost (80% train dan 20% test).....	63
Tabel 4.7 hasil skema uji XGBoost (90% train dan 10% test).....	64
Tabel 4.8 hasil akhir pengujian	64
Tabel 4.9 Hasil Klasifikasi SVM (70% train dan 30% test)	65
Tabel 4.10 Hasil Klasifikasi SVM (80% train dan 20% test)	66
Tabel 4.11 Hasil Klasifikasi SVM (90% train dan 10% test)	66
Tabel 4.12 Hasil Klasifikasi XGBoost (70% train dan 30% test)	67
Tabel 4.13 Hasil Klasifikasi XGBoost (80% train dan 20% test)	67
Tabel 4.14 Hasil Klasifikasi XGBoost (90% train dan 10% test)	68
Tabel 4.15 Hasil Evaluasi Model SVM (70% train dan 30% test).....	69
Tabel 4.16 Hasil Evaluasi Model SVM (80% train dan 20% test).....	69
Tabel 4.17 Hasil Evaluasi Model SVM (90% train dan 10% test).....	70