

**IMPLEMENTASI CRISP-DM (CROSS INDUSTRY STANDARD
PROCESS FOR DATA MINING) PADA DATA SET
HOME CREDIT DEFAULT RISK**

PRAKTEK KERJA LAPANGAN



Oleh :

ATHIYAH FITRIYANI BASUKI

20081010056

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2023**

**LEMBAR PENGESAHAN
PRAKTEK KERJA LAPANGAN**

Judul : IMPLEMENTASI CRISP-DM (CROSS INDUSTRY
STANDARD PROCESS FOR DATA MINING) PADA DATA
SET HOME CREDIT DEFAULT RISK
Oleh : ATHIYAH FITRIYANI BASUKI
NPM : 20081010056

Telah Diseminarkan Dalam Ujian PKL, pada :
Hari Jumat, Tanggal 07 Juli 2023

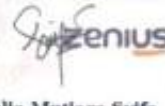
Menyetujui

Dosen Pembimbing



Eva Yulia Puspaningrum, S.Kom, M.Kom
NIP : 19890705 2021212 002

Pembimbing Lapangan



Aulia Mutiara Syifa
NIP : 20210227

Mengetahui



**Dekan
Fakultas Ilmu Komputer**
Dr. H. Noylian Hendrasari, M.T.
NIP 19681126 199403 2 001

**Koordinator Program Studi
Informatika**

Fetty Tri Anggrany, S.Kom., M.Kom
NIP 19820211 2021212 005

Judul : Implementasi Crisp - dm (Cross Industry Standard Process For Data Mining) Pada Data Set Home Credit Default Risk

Studi Kasus : PT Zona Edukasi Nusantara

Penulis : Athiyah Fitriyani Basuki NPM. 20081010056

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

ABSTRAK

Program Magang & Studi Independen Bersertifikat (SIB) batch 4 ini merupakan bagian dari program Kampus Merdeka yang bertujuan untuk memberikan kesempatan kepada mahasiswa untuk belajar dan mengembangkan diri melalui aktivitas di luar kelas perkuliahan, namun tetap diakui sebagai bagian dari perkuliahan. Program ini diperuntukkan bagi mahasiswa yang ingin memperlengkapi dirinya dengan menguasai kompetensi spesifik dan praktis yang juga dicari oleh dunia usaha dunia industri.

Zenius dengan program Data Analyst menjalankan program selama kurang lebih 5 bulan yang dimana menggunakan platform zoom meeting dengan expert atau mentor dan diisi dengan quiz, discussion, case study, serta terakhir yaitu final project. Final project akan menjadi puncak akhir dari pembelajaran. Project akhir penulis yaitu mengolah data set home credit default risk. Dalam bidang credit scoring, solusi Data Science dikembangkan menggunakan framework CRISP-DM. Solusi ini mengikuti serangkaian langkah yang meliputi business understanding, data understanding, data preparation, modeling, and evaluation. Tahap pemahaman bisnis melibatkan identifikasi tujuan bisnis dan kebutuhan yang ada. Data yang relevan kemudian dikumpulkan dan dieksplorasi pada tahap pemahaman data. Selanjutnya, data dipersiapkan melalui proses pembersihan dan transformasi. Pada tahap pemodelan, model credit scoring dibangun menggunakan berbagai algoritma Machine Learning. Akhirnya, performa model dievaluasi dengan mengujinya pada data yang tidak digunakan dalam pelatihan. Framework ini menyediakan pendekatan sistematis dalam

mengatasi permasalahan credit scoring dan dapat diterapkan untuk meningkatkan proses pengambilan keputusan dalam penilaian kredit.

Kata kunci : *Solusi Data Science, credit scoring, framework CRISP-DM, algoritma Machine Learning, pengambilan keputusan, penilaian kredit.*

ABSTRACT

The Certified Independent Study & Internship Program (SIB) batch 4 is part of the Merdeka Campus program which aims to provide opportunities for students to learn and develop themselves through activities outside the lecture class, but are still recognized as part of the lecture. This program is intended for students who wish to equip themselves by mastering specific and practical competencies that are also sought after by the business and industrial world.

Zenius with the Data Analyst program runs the program for approximately 5 months which uses a zoom meeting platform with experts or mentors and is filled with quizzes, discussions, case studies, and finally the final project. The final project will be the final pinnacle of learning. The author's final project is processing the home credit default risk data set. In the field of credit scoring, Data Science solutions are developed using the CRISP-DM framework. This solution follows a series of steps which include business understanding, data understanding, data preparation, modeling, and evaluation. The business understanding stage involves identifying business goals and existing needs. Relevant data is then collected and explored at the data understanding stage. Next, the data is prepared through a process of cleaning and transforming. At the modeling stage, the credit scoring model is built using various Machine Learning algorithms. Finally, the model's performance is evaluated by testing it on data not used in training. This framework provides a systematic approach to addressing credit scoring issues and can be applied to improve the decision-making process in credit scoring.

Keywords : *Data Science solutions, credit scoring, CRISP-DM framework, Machine Learning algorithms, decision making, credit scoring*

Kata Pengantar

Puji syukur kehadiran Allah SWT yang telah memberikan keberkahan dan kesihatan kepada penulis, sehingga penulis dapat menyelesaikan kegiatan pengerjaan laporan akhir ini dengan judul *“Implementasi Crisp - dm (Cross Industry Standard Process For Data Mining) Pada Data Set Home Credit Default Risk (Studi Kasus:PT Zona Edukasi Nusantara)”*. Laporan akhir ini dibuat untuk memenuhi mata kuliah praktek kerja lapangan sebagai salah satu syarat untuk menyelesaikan Strata Satu (S1) di Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur. Penulis berharap dengan penyusunan laporan akhir ini mampu menambah ilmu baru dan bermanfaat bagi semua pihak. Penulis menyadari bahwa dalam penyelesaian laporan akhir PKL ini tidak terlepas dari bantuan beberapa pihak yang telah memberikan bimbingan, motivasi dan bantuan sehingga penulis dapat menyelesaikan laporan akhir ini dengan sebaik mungkin. Tak lupa penulis sadari dalam penulisan laporan akhir ini masih jauh dari kata sempurna, masih banyak kekurangan mengingat keterbatasan pengetahuan dan kemampuan penulis. Atas segala kekurangan dan ketidaksempurnaan laporan akhir ini, maka penulis sangat mengharapkan masukan, kritik dan saran dari semua pihak yang bersifat membangun untuk penyempurnaan laporan akhir ini.

Surabaya, 28 Juli 2023



Athiyah Fitriyani Basuki

UCAPAN TERIMA KASIH

Segala puji dan syukur penulis ucapkan kepada Allah SWT karena atas berkah dan rahmat-Nya, penulis mampu menyelesaikan laporan kerja praktek di PT. Zona Edukasi Nusantara dengan sebaik - baiknya. Selama masa kerja praktik yang terhitung mulai tanggal 20 Februari 2023 s.d 30 juni 2023, penulis melakukan serangkaian kegiatan yang dirangkum ke dalam laporan ini sebagai syarat kelulusan pada jenjang pendidikan S1 di Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Pada kesempatan kali ini, penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dalam penyelesaian laporan kerja praktek ini, di antaranya:

1. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia selaku penyelenggara program Kampus Merdeka.
2. PT Zona Edukasi Nusantara - Zenius yang telah memberikan kesempatan kepada saya untuk belajar pada Data Analyst Path.
3. Prof. Dr. Ir. Akhmad Fauzi, M.MT. selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom selaku Ketua Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
6. Eva Yulia Puspaningrum, S.Kom, M.Kom, selaku Dosen Wali dan Dosen Pembimbing MBKM.
7. Kak Erwin selaku mentor yang telah memberikan pendampingan dan saran kepada penulis dalam memahami materi.
8. Seluruh anggota Tim Zenius yang telah sabar menanggapi berbagai keperluan dari penulis pada saat program ini berlangsung.
9. Kedua Orang tua tercinta yang senantiasa mendoakan saya dan selalu mendukung serta memberikan semangat.

10. Seluruh teman-teman, mahasiswa/i Studi Independen Zenius 2023 Data Analyst yang saling membantu, mendukung, dan bekerja sama selama program berlangsung.

Penulis menyadari bahwa laporan ini masih jauh dari kata sempurna sehingga diperlukan evaluasi untuk peningkatan kualitas yang berkelanjutan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari para pembaca. Penulis mengharapkan semoga laporan ini dapat menambah wawasan dan bermanfaat bagi para pembaca.

Daftar Isi

ABSTRAK.....	iii
KATA PENGANTAR.....	v
UCAPAN TERIMA KASIH.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Praktek kerja lapangan.....	2
1.4 Manfaat/ Kegunaan Praktek Kerja Lapangan.....	2
BAB II GAMBARAN UMUM TEMPAT PKL.....	3
2.1 Sejarah Zenius.....	3
2.2 Struktur Organisasi.....	4
2.3 Visi Misi Zenius.....	5
2.4 Lingkup Pekerjaan.....	5
2.5 Deskripsi Pekerjaan.....	8
BAB III PELAKSANAAN.....	9
3.1 Waktu dan Tempat MBKM.....	9
3.2 Pelaksanaan.....	10
3.3 Tinjauan pustaka.....	10
3.3.1 CRISP - DM.....	10
3.3.2 Data Missing value.....	11
3.3.3 Outlier.....	11
3.3.4 Logistic regression.....	12
3.3.5 XGBoost.....	12
3.3.6 Random Forest.....	12
3.3.7 Decision Tree.....	13
3.3.8 K - Nearest Neighbor.....	13
3.3.9 ROC.....	13
3.3.10 Kaggle.....	13
3.3.11 Google Colab.....	14
3.3.12 Looker studio.....	14
3.4 Pelaksanan PKL MBKM.....	14
3.4.1 Business Understanding : Home Credit.....	14
3.4.2 Describe Data.....	15
3.4.3 Data Preparation.....	16

3.4.4 EDA (Exploratory Data Analysis).....	19
3.4.5 Data splitting.....	35
3.4.6 Modeling.....	36
BAB IV HASIL DAN PEMBAHASAN.....	37
4.1 Hasil Perancangan Model.....	37
4.1.1 Logistic Regression.....	37
4.1.2 XGBoots.....	38
4.1.3 Random Forest.....	39
4.1.4 Decision Tree.....	41
4.1.5 K-Nearest Neighbors.....	42
4.2 Evaluation.....	43
4.3 Deployment.....	44
BAB V PENUTUP.....	45
5.1 Kesimpulan.....	45
5.2 Saran.....	45
Daftar Pustaka.....	46
Lampiran A. TOR.....	47
Lampiran B. Log Activity.....	51
Lampiran C. Dokumen Teknik.....	67

Daftar Gambar

Gambar 2.1 Struktur Organisasi Zenius.....	4
Gambar 3.1 Diagram process Crisp - DM.....	10
Gambar 3.2 Class Diagram Data set Home Credit.....	16
Gambar 3.3 Deteksi Data Duplikat.....	17
Gambar 3.4 Kolom dengan missing value dan persentase nya.....	17
Gambar 3.5 Kolom Missing Value dengan Feature value <40%.....	18
Gambar 3.6 Tampilan Outlier.....	18
Gambar 3.7 Handling Outlier.....	18
Gambar 3.8 Kolom Anomali sebelum Handling.....	19
Gambar 3.9 Kolom Anomali setelah Handling.....	19
Gambar 3.10 Kolom Label Encoder.....	19
Gambar 3.11 Pie Chart Customer Distribution by Target.....	20
Gambar 3.12 Crosstab dan Count plot Contract Type Vs Target.....	20
Gambar 3.13 Crosstab dan Count plot Gender Vs Target.....	21
Gambar 3.14 Crosstab dan Count plot Car Ownership Status Vs Target.....	22
Gambar 3.15 Crosstab dan Count plot Realty Ownership Status Vs Target.....	23
Gambar 3.16 Crosstab dan Count plot Suite Type Vs Target.....	23
Gambar 3.17 Crosstab dan Count plot Income Type Vs Target.....	24
Gambar 3.18 Crosstab dan Count plot Education Type Vs Target.....	25
Gambar 3.19 Crosstab dan Count Plot Family Status Vs Target.....	26
Gambar 3.20 Crosstab dan Count Plot Housing Type Vs Target.....	27
Gambar 3.21 Kolom Crosstab Occupation Type Vs Target.....	28
Gambar 3.22 Crosstab dan Count Plot Process Day Vs Target.....	29
Gambar 3.23 Kolom Crosstab Organization Type Vs Target.....	29
Gambar 3.24 Boxplot Amount Credit of the Loan Vs Target.....	30
Gambar 3.25 Boxplot Amount Income Vs Target.....	30
Gambar 3.26 Histogram plot Age Vs target.....	31
Gambar 3.27 Catplot Car Ownership Status, The Number of Children, Target, and House/Flat Ownership Status.....	32
Gambar 3.28 Catplot Income Type, Amount of Goods Price, Target, and House/Flat Ownership Status.....	32
Gambar 3.29 Catplot Family Status, Amount of Income, Target, and House/Flat Ownership Status.....	33
Gambar 3.30 Catplot Contract Type, The Number of Children, Target, and Car Ownership.....	33
Gambar 3.31 Catplot Income Type, Amount of Credit, Target, and Contract Type.....	34
Gambar 3.32 Catplot Housing Type, Amount Credit of Loan, Target, and Rating	

of Region where Client Lives.....	34
Gambar 3.33 Catplot Education Type, Amount Credit of Loan, Target, and Rating of Region where Client Lives.....	35
Gambar 3.34 Catplot Family Status, Amount Credit of Loan, Target, and Rating of Region where Client Lives.....	35
Gambar 3.35 Subplot Balance of Target Before and After SMOTE.....	36
Gambar 4.1 Test Accuracy Logistic Regression.....	38
Gambar 4.2 Train Accuracy Logistic Regression.....	38
Gambar 4.3 Test Accuracy XGBoots.....	39
Gambar 4.4 Train Accuracy XGBoots.....	39
Gambar 4.5 Test Accuracy Random Forest.....	40
Gambar 4.6 Train Accuracy Random Forest.....	41
Gambar 4.7 Test Accuracy Random Forest.....	41
Gambar 4.8 Train Accuracy Random Forest.....	42
Gambar 4.9 Test Accuracy K-Nearest Neighbors.....	43
Gambar 4.10 Train Accuracy K-Nearest Neighbors.....	43
Gambar 4.11 Dashboard Bagian Loan Application.....	44
Gambar 4.12 Dashboard Bagian Evaluation Metric.....	44

Daftar Tabel

Tabel 2.1 Ruang lingkup Universal Fundamental Skills.....	5
Tabel 2.2 Ruang Lingkup Modul Specific Fundamental and Technical Skills (Data Analytics).....	6
Tabel 3.1 Jadwal Kerja.....	9
Tabel 3.2 Application_train/application_test.....	16
Tabel 3.3 bureau.....	31
Tabel 3.4 bureau_balance.....	32
Tabel 3.5 previous_application.....	33
Tabel 3.6 POS_CASH_BALANCE.....	36
Tabel 3.7 credit_card_balance.....	37
Tabel 3.8 installments_payment.....	39
Tabel 3.9 Application_train/application_test final.....	40
Tabel 3.10 bureau final.....	42
Tabel 3.11 bureau_balance final.....	43
Tabel 3.12 POS_CASH_BALANCE final.....	43
Tabel 3.13 credit_card_balance final.....	44
Tabel 3.14 previous_application final.....	45
Tabel 3.15 installments_payment final.....	46
Tabel 4.1 Perbandingan before & after cross-val.....	74