

**IMPLEMENTASI *MACHINE LEARNING* MENGGUNAKAN
METODE REGRESI PADA DATASET UDEMY *FINANCE &
ACCOUNTING COURSE***

PRAKTEK KERJA LAPANGAN



Oleh :

HAIDAR TRIARI RESPATI

NPM : 21082010086

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

Universitas Pembangunan Nasional “Veteran” Jawa Timur

SURABAYA

2024

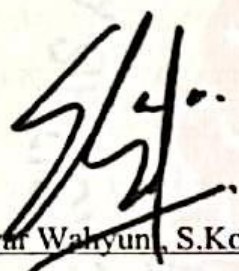
HALAMAN PENGESAHAN

Judul : IMPLEMENTASI *MACHINE LEARNING* PADA DATASET
UDEMY *FINANCE & ACCOUNTING COURSE* DENGAN
MENGUNAKAN METODE REGRESI PADA
PT MITRA TALENTA GRUB
Oleh : HAIDAR TRIARI RESPATI NPM. 21082010086

Menyetujui,

Pembimbing

Pembimbing Lapangan


(Eka Dyan Wahyuningsih, S.Kom, M.Kom)

NIP. 198412012021212005


(A. Asfar Jaya)

Mengetahui,

Dekan
Fakultas Ilmu Komputer

Koordinator Program Studi
Sistem Informasi


(Prof. Dr. Ir. Novirina Hendrasarie, M.T)
NIP. 196811261994032001


(Agung Brastama Putra, S.Kom, M.Kom)
NIP. 198511242021211003

ABSTRAK

Industri e-learning berkembang pesat dengan banyak orang mencari cara baru untuk belajar mandiri dan meningkatkan keterampilan. Udemy, sebagai salah satu platform pembelajaran online terkemuka, menawarkan berbagai kursus dalam berbagai bidang. Penelitian ini menggunakan dataset Finance & Accounting Course dari Udemy untuk memprediksi apakah kursus akan ditambahkan ke dalam Daftar Keinginan oleh pengguna. Untuk tujuan ini, kami mengaplikasikan tiga model Machine Learning: Logistic Regression, Random Forest, dan Support Vector Machine (SVM).

Ketiga model ini menunjukkan akurasi sempurna (1.0) dalam memprediksi variabel target, yang menunjukkan bahwa mereka sangat efektif dalam memahami pola dalam data. Random Forest mengidentifikasi panjang dan lebar kelopak sebagai fitur paling penting dalam prediksi, menunjukkan kemampuan model ini untuk menangkap pengaruh signifikan dari fitur tertentu. Logistic Regression, di sisi lain, memberikan pemahaman yang lebih mendetail tentang hubungan linear antara fitur dan probabilitas kursus ditambahkan ke Daftar Keinginan. Model ini memberikan wawasan yang jelas tentang bagaimana setiap fitur berkontribusi terhadap hasil prediksi. SVM, dengan kemampuannya untuk menghasilkan batas keputusan yang efektif dalam ruang fitur, juga menunjukkan performa yang sangat baik.

Penelitian ini memberikan wawasan berharga bagi pengajar dan pengembang kursus di Udemy untuk memahami faktor-faktor yang mempengaruhi keputusan pengguna dalam menambahkan kursus ke Daftar Keinginan mereka. Dengan memanfaatkan model Machine Learning yang tepat, mereka dapat mengidentifikasi fitur-fitur yang paling penting dan meningkatkan kualitas serta daya tarik kursus mereka. Hasil penelitian ini juga dapat digunakan sebagai dasar untuk pengembangan sistem rekomendasi yang lebih baik di platform e-learning seperti Udemy, sehingga pengguna dapat lebih mudah menemukan kursus yang sesuai dengan kebutuhan dan minat mereka.

Kata Kunci: *Machine Learning, Regresi Linear, Random Forest, SVM*

KATA PENGANTAR

Puji syukur kami panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga kami dapat menyelesaikan laporan PKL yang berjudul "*Implementasi Machine Learning Menggunakan Metode Regresi pada Dataset Udemy Finance & Accounting Course*". Laporan ini disusun untuk memenuhi tugas Praktek Kerja Lapangan (PKL) dan sebagai bagian dari pengembangan kompetensi mahasiswa di bidang Sistem Informasi.

Penyusunan laporan ini tidak lepas dari dukungan, semangat, dan bimbingan dari berbagai pihak. Ucapan terima kasih disampaikan kepada semua pihak yang telah memberikan dukungan, baik secara moral maupun material. Terima kasih kepada :

1. Tuhan Yang Maha Esa, atas petunjuk dan rahmat-Nya selama proses penyusunan laporan ini.
2. Agung Bramasta, S.Kom., M.Kom., selaku Kepala Program Studi Sistem Informasi di Universitas Pembangunan Nasional "Veteran" Jawa Timur, yang telah memberikan izin serta arahan yang berharga selama pelaksanaan Praktek Kerja Lapangan.
3. Eka Dyar Wahyuni, S.Kom, M.Kom selaku Dosen Pembimbing yang turut memberikan izin, arahan, serta bimbingan yang berharga sepanjang proses Praktek Kerja Lapangan berlangsung.
4. A. Asfar Jaya selaku Mentor yang membimbing banyak hal terkait materi selama kegiatan berlangsung.
5. Kedua orang tua yang selalu mendukung dalam pelaksanaan kegiatan praktik kerja lapangan.
6. Teman - teman Prodi Sistem Informasi yang telah memberi memberi semangat dan motivasi tiada henti.

Laporan ini telah disusun dengan penuh dedikasi dan upaya maksimal. Namun, penulis menyadari bahwa laporan ini masih jauh dari sempurna dan terdapat banyak kekurangan. Oleh karena itu, penulis mengharapkan adanya saran dan kritik yang membangun dari semua pihak untuk perbaikan di masa mendatang.

Semoga laporan ini bermanfaat bagi para pembaca dan dapat meningkatkan pengetahuan.

Dengan demikian, laporan ini menyajikan hasil Implementasi *Machine Learning* pada dataset *Udemy Finance & Accounting Course*. Laporan ini diharapkan dapat menjadi referensi bagi pengembangan model prediksi di bidang *e-learning* dan memberikan wawasan baru tentang pentingnya fitur dalam prediksi keberhasilan kursus.

Sidoarjo, 15 Juli 2024

Penyusun

DAFTAR ISI

SAMPUL DEPAN	i
HALAMAN JUDUL	ii
HALAMAN PENGESAHAN	Error! Bookmark not defined.
ABSTRAK	iii
KATA PENGANTAR.....	v
DAFTAR ISI.....	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN	xiii
BAB I.....	1
PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	2
1.3. Tujuan.....	2
1.4. Manfaat.....	2
BAB II	3
GAMBARAN UMUM TEMPAT PKL	3
2.1. Profil Organisasi	3
2.2. Tujuan Organisasi	4
2.3. Struktur Organisasi	4
2.4. Bidang Usaha Organisasi	5
BAB III.....	6
PELAKSANAAN PKL	6
3.1. Tinjauan Pustaka	6
3.1.1. <i>Machine Learning</i>	6
3.1.2. Dataset.....	6
3.1.3. <i>Deep Learning</i>	7
3.1.4. Regresi.....	8
3.1.5. Finance and Accounting	8
3.1.6. Udemy Course	9
3.2. Waktu dan Tempat Pelaksanaan PKL	10
3.2.1 Tempat dan Waktu Laporan	10

3.2.2	Pelaksanaan	10
BAB IV		13
HASIL DAN PEMBAHASAN		13
4.1.	Pengolahan Data	13
4.1.1.	Input Library	13
4.1.2.	Membaca dan Input Dataset	14
4.1.3.	Eksplorasi Dataset.....	15
4.1.4.	Pra-Pemrosesan Data	23
4.2.	<i>Modelling Machine Learning</i>	33
4.2.1.	Pembagian Data	33
4.2.2.	Pemodelan.....	35
4.2.3.	Optimasi Model.....	Error! Bookmark not defined.
4.3.	Analisis Hasil	Error! Bookmark not defined.
4.4.	Penjelasan Hasil	Error! Bookmark not defined.
4.4.1.	Akurasi Model.....	Error! Bookmark not defined.
4.4.2.	Pentingnya Fitur dan Koefisien	Error! Bookmark not defined.
BAB V		39
PENUTUP		39
5.1.	Kesimpulan	39
5.2.	Saran	40
DAFTAR PUSTAKA		41
LAMPIRAN		42

DAFTAR TABEL

Tabel 3.1 Daftar Kegiatan yang Dilakukan.....	12
Tabel 4.1 Akurasi Model.....	50
Tabel 4.2 Tabel Hasil Kalkulasi <i>Feature Importance & Coefficients</i>	50

DAFTAR GAMBAR

Gambar 2.1 Logo PT Mitra Talenta Grub	3
Gambar 4.1 Baris Kode Program Input Library	13
Gambar 4.2 Baris Kode Program Impor Data Google Drive.....	14
Gambar 4.3 Output Baris Kode Program Impor Data Google Drive.....	14
Gambar 4.4 Membaca File CSV Udemy Data FA Course	15
Gambar 4.5 Kode Program Memeriksa Data Frame.....	15
Gambar 4.6 Menghitung Distribusi Kolom	17
Gambar 4.7 Output Menghitung Distribusi Kolom Nilai False	18
Gambar 4.8 Memeriksa jumlah Baris dan Kolom	18
Gambar 4.9 Memeriksa Jumlah Baris dan Kolom.....	18
Gambar 4.10 Identifikasi Struktur dan Informasi Data Frame	19
Gambar 4.11 Output Struktur dan Informasi Data Frame.....	19
Gambar 4.12 Mengubah Nama Kolom Asli.....	19
Gambar 4.13 Mengubah Nama Kolom Asli.....	20
Gambar 4.14 Hitung Jumlah Duplikat Baris Data Frame	20
Gambar 4.15 Output Hitung Jumlah Duplikat Baris Data Frame	20
Gambar 4.16 Memeriksa Jumlah Nilai hilang	21
Gambar 4.17 Output Memeriksa Jumlah Nilai hilang	21
Gambar 4.18 Memeriksa dan Mengubah Nilai Hilang Kolom	21
Gambar 4.19 Memeriksa Nilai Hilang dan Hitung Nilai True.....	22
Gambar 4.20 Output Memeriksa Nilai Hilang dan Hitung Nilai True.....	23
Gambar 4.21 Statistik Deskriptif Data Frame.....	23
Gambar 4.22 Output Tabel Statistik Deskriptif Data Frame	23
Gambar 4.23 Visualisasi Diagram Batang	23
Gambar 4.24 Output Visualisasi Diagram Batang	24
Gambar 4.25 Box Plot Distribusi Nilai Kolom Numerik.....	24

Gambar 4.26 Output Box Plot Distribusi Nilai Kolom Numerik.....	25
Gambar 4.27 Akses Index Nama Kolom (Atribut)	25
Gambar 4.28 Output Akses Index Nama Kolom (Atribut)	25
Gambar 4.29 Mengubah Tipe Data	25
Gambar 4.30 Memeriksa Perubahan Tipe Data	26
Gambar 4.31 Memeriksa Perubahan Tipe Data	26
Gambar 4.32 Memeriksa Informasi Struktur Data Frame	27
Gambar 4.33 Memeriksa Informasi Struktur Data Frame Kiri	27
Gambar 4.34 Memeriksa Informasi Struktur Data Frame Kanan	27
Gambar 4.35 Encoding Data Object	27
Gambar 4.36 Menghapus Kolom	28
Gambar 4.37. Memeriksa Informasi Struktur Data Frame	29
Gambar 4.38 Tabel Hasil Struktur Data Frame Bagian Kiri	29
Gambar 4.39 Tabel Hasil Struktur Data Frame Bagian Kanan	29
Gambar 4.40 Scalling Data	29
Gambar 4.41 Tabel Output Scalling Data	30
Gambar 4.42 Tabel Output Scalling Data	30
Gambar 4.43 Dekomposisi dan Identifikasi Dataset menggunakan PCA.....	31
Gambar 4.44 Hasil Dekomposisi & Identifikasi Dataset menggunakan PCA.....	32
Gambar 4.45 Kalkulasi Heatmap dari Korelasi Matriks Data Frame	33
Gambar 4.46 Hasil Kalkulasi Heatmap dari Korelasi Matriks Data Frame	34
Gambar 4.47 Syntaks untuk Menampilkan Heatmap	34
Gambar 4.48 Tabel Heatmap.....	35
Gambar 4.49 Pemrosesan data	35
Gambar 4.50 Pembagian Dataset.....	36
Gambar 4.51 Mendistribusikan Dataset.....	36
Gambar 4.52 Output Distribusi.....	37
Gambar 4.53 Menstandarisasi Fitur Numerik.....	37

Gambar 4.54 Mengubah Kategorikal dalam Dataset	38
Gambar 4.55 Membuat Model	39
Gambar 4.56 Melatih Model.....	40
Gambar 4.57 Evaluasi Kinerja Model.....	41
Gambar 4.58 Menampilkan Hasil Evaluasi	41
Gambar 4.59 Persiapan Data.....	43
Gambar 4.60 Hiperparameter Tuning	44
Gambar 4.61 Mengevaluasi Beberapa Model.....	45
Gambar 4.62 Analisis Akurasi	47
Gambar 4.63 Menampilkan Pentingnya Fitur yang diperoleh dari Model	48

DAFTAR LAMPIRAN

Gambar Lampiran 1. Dokumentasi Diskusi Final Project	47
Gambar Lampiran 2. Dokumentasi Latihan Presentasi Final Project	47
Gambar Lampiran 3. Dokumentasi Presentasi Final Project	48