LAPORAN AKHIR

STUDI INDEPENDEN BERSERTIFIKAT

Pembuatan Machine Learning berbasis Sentence Classification di perusahaan Ford yang berbasis Custumer Relationship Management (CRM)

Diajukan untuk memenuhi persyaratan kelulusan

Bootcamp Data Science Rakamin batch 48

Disusun oleh :

Muhammad Fariz Shadiqin / Project Manager

Prima Praditya / Business / Data Analyst

Evira Marnolia / Data Scientist

Yosafat Angga Gilang Pratama / Data Engineer



**FOURTASTIC**

**TAHUN 2024**

Abstrak

Ford, sebuah perusahaan teknologi berbasis CRM, saat ini menghadapi kesulitan dalam merekrut karyawan karena deskripsi pekerjaan yang terlalu panjang. Untuk meningkatkan efisiensi, Ford ingin memiliki sistem yang dapat memberikan rekomendasi pekerjaan sesuai dengan keterampilan dan pengalaman kandidat, sehingga mereka lebih mudah menemukan posisi yang cocok. Data Science Lifecycle adalah panduan tahap demi tahap yang menggambarkan bagaimana machine learning dan teknik analitik lainnya dapat digunakan untuk menghasilkan insight dan prediksi dari data untuk mencapai tujuan bisnis. Dari beberapa model yang dilatih, model SVM dipilih karena menunjukkan hasil kinerja yang paling maksimal berdasarkan metric yang dipilih. Selain itu, model memiliki penanganan overfitting yang dinilai lebih efisien. Dengan memanfaatkan Feature Importance dari model SVM, kita bisa mendapatkan wawasan yang sangat berharga terkait proses rekrutmen. Beberapa business insights yang dapat diperoleh adalah tentang pengaruh kualifikasi, keterampilan, waktu respon, dan demografi terhadap penerimaan pelamar. Dengan cara ini, proses rekrutmen akan lebih efisien dan dapat menarik lebih banyak kandidat yang tepat.

Kata Pengantar

Puji syukur kehadirat Allah SWT atas segala rahmat dan hidayah-Nya sehingga kelompok yang berisikan 4 orang ini dapat menyelesaikan Final Project Bootcamp Rakamin ini dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad SAW, keluarga, sahabat, dan seluruh umatnya.

Laporan akhir ini disusun sebagai bentuk pertanggungjawaban atas tugas akhir yang telah diberikan. Pada kesempatan ini juga, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kak Dzikrul, selaku kakak pembimbing pada batch ke 48 Data Science ini yang telah memberikan bimbingan, arahan, dan motivasi selama proses pengerjaan project ini.
2. Tim Rakamin Academy yang telah menyelenggarakan Bootcamp ini dan menyediakan fasilitas serta materi yang sangat bermanfaat.
3. Rekan-rekan satu tim yang telah bekerja sama dengan baik dan saling mendukung selama proses pengerjaan project.
4. Seluruh pihak yang telah membantu dan mendukung penulis dalam menyelesaikan project ini.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna, oleh karena itu penulis sangat mengharapkan kritik dan saran yang bersifat membangun demi perbaikan di masa yang 1 akan datang. Penulis berharap laporan akhir ini dapat memberikan manfaat bagi semua pihak.

Jakarta, 8 Januari 2025

Muhammad Fariz Shadiqin

Project Manager

Daftar Isi

Abstrak ………………………………………………………………………...…… i

Kata Pengantar…………………………………………………………..………….. ii

Daftar Isi ………………………………………………………………..………….. iii

Daftar Gambar ……………………………………………………………………… iv

[Referensi ...………………………………………………………………………..… xix](#_4f1mdlm)

[**Bab I Pendahuluan**](#_2s8eyo1) ……………………………………………………………...... **5**

[I.1 Latar B](#_4f1mdlm)elakang ……………………………………………………………… 5

[I.2 Lingkup ……………………………………………………………………...](#_4f1mdlm) 6

I.3 Tujuan ……………………………………………………………………….. 6

[**Bab II Kelompok ‘FOURTASTIC’**](#_3rdcrjn) ………………………………………………. 7

[II.1 Struktur Organisasi ………………………………………………………… 7](#_4f1mdlm)

[II.2 Lingkup Pekerjaan …………………………………………………………](#_4f1mdlm)  8

[II.3 Deskripsi Pekerjaan …………………………………………………………](#_4f1mdlm)  10

[II.4 Jadwal Kerja ………………………………………………………………..](#_4f1mdlm)  11

[**Bab III Pembuatan Machine Learning berbasis Sentence Classification di Perusahaan Ford yang Berbasis Custumer Relationship Management (CRM)**](#_3whwml4) **..** [1](#_3whwml4)3

[III.1 Deskripsi Permasalahan ……………………………………………………](#_4f1mdlm) 13

[III.2 Proses Pelaksanaan Project …………………………………………………](#_4f1mdlm) 13

[III.3](#_4f1mdlm) Pencapaian Hasil [………………………………………………………….. 1](#_4f1mdlm)5

[**Bab IV Penutup** ……………………………………………………………………](#_1hmsyys) 17

[IV.1 Kesimpulan ………………………………………………………………… 1](#_4f1mdlm)7

[IV.2 Saran ……………………………………………………………………….. 1](#_4f1mdlm)7

**Bab V Lampiran A. PPT Final Project Presentation** [**…………………………**](#_4f1mdlm) **A - 1**

[**Bab VI Lampiran B. Notulensi Mentoring ……………………………………**](#_1v1yuxt)  **B - 1**

[**Bab VII Lampiran C. Dokumen Teknik ………………………………………**](#_2u6wntf) **C - 1**

Daftar Gambar

**Gambar 1** Jadwal kerja dari Rakamin Academy …………………………… 11

**Gambar 2** Judul dan metode final project. …………………………………. A - 1

**Gambar 3** Objectives dan Business metrics. ……………………………….. A - 1

**Gambar 4** Dataset dan data overview. ……………………………………… A - 1

**Gambar 5** Hasil EDA. …………….…………….…………….……………. A - 2

**Gambar 6** Feature selection. …………….…………….…………………… A - 2

**Gambar 7** Hasil modelling dan kesimpulan akhir. …………….………….. A - 2

# 

# Pendahuluan

## Latar Belakang

Ford, sebuah perusahaan teknologi berbasis Customer Relationship Management (CRM), saat ini menghadapi tantangan dalam proses rekrutmen karyawan. Salah satu permasalahan utama adalah deskripsi pekerjaan yang terlalu panjang dan kompleks, sehingga menyulitkan calon kandidat untuk memahami persyaratan yang diperlukan. Hal ini tidak hanya menghambat proses perekrutan, tetapi juga mengurangi minat kandidat potensial untuk melamar karena mereka merasa kesulitan mengidentifikasi kecocokan antara kemampuan mereka dan kebutuhan pekerjaan yang tertera (Pratama A D, Putri A N 2020).

Proses rekrutmen yang tidak efisien ini berpotensi menghambat pertumbuhan perusahaan karena kekurangan tenaga kerja yang kompeten. Sebagai perusahaan yang mengutamakan pengalaman pelanggan, Ford menyadari bahwa pengalaman pertama kandidat terhadap perusahaan sering kali dimulai dari proses rekrutmen. Oleh karena itu, sangat penting bagi Ford untuk memastikan bahwa proses rekrutmen memberikan kesan positif kepada kandidat, sekaligus memenuhi kebutuhan perusahaan akan tenaga kerja berkualitas.

Untuk mengatasi masalah ini, Ford berencana mengembangkan sistem berbasis Machine Learning yang dirancang untuk mengklasifikasikan kalimat dalam deskripsi pekerjaan. Sistem ini diharapkan mampu menyederhanakan informasi yang ada, sehingga lebih mudah dipahami oleh kandidat. Selain itu, sistem ini akan membantu mencocokkan keterampilan dan pengalaman kandidat dengan kebutuhan pekerjaan yang relevan. Dengan adanya teknologi ini, Ford berharap dapat meningkatkan efisiensi proses rekrutmen, mengurangi tingkat kesalahan dalam seleksi kandidat, dan menarik lebih banyak pelamar yang sesuai dengan kebutuhan perusahaan.

## Lingkup

Lingkup dari project ini adalah pengembangan sistem klasifikasi berbasis Machine Learning untuk mendukung proses rekrutmen di Ford. Sistem ini dirancang untuk:

1. Memproses dan menganalisis deskripsi pekerjaan yang ada.
2. Mengelompokkan kalimat berdasarkan relevansi dengan keterampilan dan pengalaman kandidat.
3. Memberikan rekomendasi posisi pekerjaan yang sesuai kepada kandidat.

Untuk mempermudah pelaksanaan project, kami membagi tim menjadi 4 *role*, yaitu Project Manager, Data Scientist, Data Engineer, dan Data Analyst. Penjelasan lebih lengkap dari masing-masing *role* akan dijelaskan pada bab selanjutnya.

## Tujuan

Tujuan dari project ini adalah:

1. Meningkatkan efisiensi proses rekrutmen di Ford dengan memanfaatkan Machine Learning untuk mengklasifikasikan deskripsi pekerjaan.
2. Mengasah dan menerapkan keterampilan yang diperoleh selama mengikuti Bootcamp Data Science Rakamin, termasuk penguasaan dalam pengembangan model Machine Learning, pengolahan data, dan evaluasi kinerja model.

# 

# Kelompok ‘FOURTASTIC’

## Struktur Organisasi

1. Project Manager (Muhammad Fariz Shadiqin):

Berperan untuk mengatur timeline sebuah project, memastikan bahwa pekerjaan akan selesai dengan tepat waktu, dan sebagai seorang koordinator untuk menghubungkan komunikasi dengan anggota tim lain. Selain itu, wajib memastikan bahwa kualitas project, business goals, performa model yang dibuat, dan pelaporan hasil project berjalan dengan baik.

1. Data Scientist (Evira Marnolia):

Berperan untuk mengembangkan model machine learning berdasarkan kebutuhan project dan juga mengevaluasi performa model menggunakan metric yang tepat. Selain itu, data scientist terus bereksperimen dengan berbagai teknik dan algoritma, sambil memanfaatkan insight yang didapat untuk membantu perusahaan membuat keputusan berbasis data yang lebih baik.

1. Data Engineer (Yosafat Angga Gilang Pratama):

Berperan untuk memastikan bahwa data yang dibutuhkan oleh Data Scientist dan Data Analyst tersedia dan siap untuk digunakan. Mereka bertanggung jawab untuk mendesain, membangun, dan memelihara infrastruktur data yang memungkinkan data dapat dikumpulkan, disimpan, dan diakses dengan efisien. Data Engineer juga memastikan bahwa data yang diolah bersih, terstruktur dengan baik, dan dapat digunakan untuk analisis lebih lanjut.

1. Business/Data Analyst (Prima Praditya):

Berperan dalam menginterpretasikan data yang telah diolah oleh Data Scientist, lalu menganalisisnya untuk memberikan insight yang relevan bagi pengembangan strategi bisnis. Sebagai Data Analyst, harus memahami kebutuhan bisnis dan mampu menerjemahkan data menjadi rekomendasi yang actionable. Dan

juga bertugas membuat laporan dan visualisasi data yang mudah dipahami oleh tim manajemen atau Mentor untuk mendukung pengambilan keputusan berbasis data.

## Lingkup Pekerjaan

1. Project Manager
2. Lingkup Pekerjaan:

* Membuat rencana project (timeline, milestone, resource allocation).
* Mengkoordinasikan tim dan memastikan semua tugas diselesaikan tepat waktu.
* Mengelola risiko project.
* Melakukan komunikasi dengan Mentor.

1. Kaitan dengan project:

* Menentukan target akurasi model.
* Mengatur jadwal pengumpulan dan preprocessing data.
* Mengatur jadwal pelatihan dan evaluasi model.
* Membuat laporan kemajuan project.

1. Data Scientist
2. Lingkup Pekerjaan:

* Menganalisis data untuk menemukan pola dan insight.
* Membangun fitur-fitur baru yang relevan.
* Visualisasi data.
* Menginterpretasi hasil model.

1. Kaitan dengan project:

* Menganalisis distribusi kelas dalam dataset.
* Membangun fitur-fitur seperti TF-IDF, word embeddings.
* Visualisasi hasil klasifikasi untuk memahami kekuatan dan kelemahan model.
* Menjelaskan hasil klasifikasi kepada Mentor.

1. Data Engineer
2. Lingkup Pekerjaan:

* Mengumpulkan informasi dari berbagai sumber.
* Membersihkan data (data cleaning).
* Membangun pipeline data.
* Menyimpan data dalam format yang sesuai.

1. Kaitan dengan project:

* Mengumpulkan dataset kalimat yang sudah dikategorikan.
* Membersihkan data dari noise atau inconsistency.
* Membangun pipeline untuk preprocessing data (tokenisasi, stop word removal, stemming/lemmatization).
* Menyimpan data dalam format yang dapat dibaca oleh model.

1. Business/Data Analyst
2. Lingkup Pekerjaan:

* Memahami tujuan bisnis dari project klasifikasi kalimat dan mendefinisikan metrik keberhasilan yang relevan.
* Mengeksplorasi data untuk menemukan insight bisnis yang berharga.
* Berkomunikasi dengan Mentor untuk memastikan model yang dikembangkan sesuai dengan kebutuhan bisnis.
* Evaluasi model dari perspektif bisnis.

1. Kaitan dengan project:

* Menentukan tujuan bisnis dari project klasifikasi kalimat.
* Memantau kinerja model secara berkala dan memberikan rekomendasi perbaikan.
* Mengevaluasi dampak bisnis dari model yang telah diimplementasikan.

## Deskripsi Pekerjaan

1. Project Manager

* Membuat rencana project yang detail, termasuk timeline, milestones, dan resource allocation.
* Mengadakan rapat kelompok secara berkala untuk membahas progress, kendala, dan solusi.
* Memastikan semua anggota kelompok memiliki tugas yang jelas dan bekerja sama dengan baik.
* Menyelesaikan konflik yang mungkin timbul dalam kelompok.
* Membuat laporan kemajuan project secara berkala kepada Mentor dan menyajikan hasil akhir project kepada Mentor.

1. Data Scientist

* Memilih algoritma machine learning yang sesuai untuk masalah klasifikasi.
* Melakukan eksperimen dengan berbagai algoritma.
* Melatih model dengan data yang telah disiapkan.
* Melakukan tuning hyperparameter untuk meningkatkan kinerja model.
* Mengevaluasi kinerja model menggunakan metrik yang sesuai (accuracy, precision, recall, F1-score).
* Menganalisis fitur penting yang digunakan oleh model untuk membuat prediksi.

1. Data Engineer

* Membersihkan data dari noise dan missing values.
* Melakukan tokenisasi, stop word removal, stemming atau lemmatization.
* Mengubah data teks menjadi format yang dapat diproses oleh model.
* Membangun pipeline data untuk memproses data secara otomatis dan efisien.
* Menyimpan data yang sudah bersih ke dalam database.

1. Business/Data Analyst

* Mendefinisikan metrik keberhasilan yang relevan dengan bisnis.
* Menganalisis data untuk menemukan pola, tren, dan insight yang berharga.
* Menganalisis data untuk mengidentifikasi jenis kalimat yang paling sulit diklasifikasi dan memberikan rekomendasi untuk meningkatkan kinerja model pada kelas tersebut.

## Jadwal Kerja



Gambar 1. Jadwal kerja dari Rakamin Academy

**Minggu 10:** Machine Learning (ML) Preparation

Fokus pada persiapan dasar untuk mempelajari Machine Learning, yang mencakup pengenalan konsep dasar ML, library yang akan digunakan (misal: Python, scikit-learn), dan lingkungan kerja.

**Minggu 11-20:** Tahapan Pembelajaran

1. Stage 0: Pengenalan awal pada project yang akan dikerjakan.
2. Stage 1-4: Pembelajaran bertahap mengenai berbagai topik ML, seperti:

* Git: Pengelolaan versi kode.
* Business Understanding: Memahami konteks bisnis dari project.
* ML Evaluation: Evaluasi model ML.
* Supervised Learning: Pembelajaran dengan data berlabel.
* Unsupervised Learning: Pembelajaran tanpa data berlabel.
* Deep Learning: Penggunaan neural network untuk tugas yang kompleks.
* Mentoring: Sesi mentoring rutin dengan mentor untuk membahas materi, kesulitan, dan progres project.
* Final Project: Jadwal mentoring project akhir akan disesuaikan dengan kesepakatan masing-masing kelompok.
* Mentoring: Terdapat sesi mentoring secara rutin setiap minggu, baik untuk materi maupun project akhir.

Berdasarkan timeline yang tertera di atas, program bootcamp Rakamin dirancang untuk memberikan pemahaman yang komprehensif mengenai Machine Learning. Masing-masing peserta dan kelompok akan diajak untuk mempelajari berbagai konsep dan teknik ML secara bertahap, sambil mengerjakan project nyata. Sesi mentoring yang rutin juga membantu peserta dan kelompok dalam mengatasi kesulitan dan mencapai tujuan pembelajaran.

# 

# Pembuatan Machine Learning berbasis Sentence Classification di perusahaan Ford yang berbasis Custumer Relationship Management (CRM)

## Deskripsi Permasalahan

## Ford adalah perusahaan berbasis Customer Relationship Management (CRM), menghadapi tantangan dalam proses rekrutmen yang tidak efisien karena deskripsi pekerjaan yang terlalu panjang dan kompleks. Hal ini membuat calon kandidat sulit memahami informasi penting, sehingga menurunkan minat mereka untuk melamar. Akibatnya, proses seleksi memakan waktu yang lama dan menghasilkan jumlah pelamar yang tidak sesuai kebutuhan perusahaan.

## Solusi yang diusulkan adalah pengembangan sistem berbasis Machine Learning untuk mengklasifikasikan kalimat dalam deskripsi pekerjaan menjadi beberapa kategori, seperti kualifikasi, tanggung jawab, dan persyaratan. Sistem ini dirancang untuk:

1. Menyederhanakan deskripsi pekerjaan agar lebih mudah dipahami.
2. Memberikan rekomendasi posisi pekerjaan kepada kandidat berdasarkan kecocokan keterampilan.
3. Meningkatkan efisiensi dan efektivitas rekrutmen, sehingga mengurangi waktu proses seleksi dan meningkatkan kualitas pelamar.

## Proses Pelaksanaan Project

## Pelaksanaan project Final Project Bootcamp Rakamin dimulai dengan pemahaman mendalam mengenai permasalahan yang dihadapi oleh Ford. Masalah utama adalah deskripsi pekerjaan yang panjang dan kompleks, yang menghambat calon kandidat dalam memahami informasi penting dan memengaruhi efisiensi proses rekrutmen. Tujuan dari project ini adalah mengembangkan sistem berbasis

## Machine Learning untuk mengklasifikasikan kalimat dalam deskripsi pekerjaan agar lebih terstruktur dan relevan.

## Tahapan pertama dalam project ini adalah pengumpulan data. Dataset yang terdiri dari lebih dari 60.000 baris data teks disediakan oleh Rakamin dalam format CSV. Data ini dieksplorasi untuk memahami pola umum, seperti distribusi kategori kalimat dan panjang kalimat rata-rata. Selama eksplorasi, ditemukan beberapa tantangan, seperti adanya nilai kosong dan data duplikat. Masalah ini diselesaikan dengan menghapus nilai kosong yang tidak signifikan dan menghilangkan data duplikat untuk memastikan integritas dataset.

## Langkah berikutnya adalah preprocessing data, yang melibatkan pembersihan teks dari noise seperti stopwords, simbol, dan karakter yang tidak relevan. Teknik tambahan seperti tokenisasi, stemming, dan lemmatization diterapkan untuk meningkatkan konsistensi data. Selain itu, dilakukan transformasi data dengan menambahkan fitur baru, seperti analisis sentimen (Sentiment) dan panjang kalimat (Sentence\_Length). Proses ini dirancang untuk memastikan data siap digunakan dalam pengembangan model Machine Learning.

## Pada tahap pengembangan model, Logistic Regression digunakan sebagai baseline. Selanjutnya, tim menguji algoritma yang lebih kompleks, seperti Support Vector Machine (SVM) dan Long Short-Term Memory (LSTM). Model SVM menunjukkan performa terbaik dengan akurasi mencapai 86%, sedangkan LSTM memberikan hasil yang lebih tinggi tetapi memerlukan waktu komputasi yang lebih lama. Akhirnya, model SVM dipilih untuk diimplementasikan karena memberikan hasil yang lebih stabil dan sesuai dengan kebutuhan bisnis.

## project ini menghadapi beberapa hambatan, termasuk ketidakseimbangan data dan overfitting pada model. Untuk mengatasi ketidakseimbangan data, teknik oversampling dengan SMOTE digunakan, sementara overfitting ditangani melalui regularisasi dan optimasi hyperparameter. Hambatan lainnya adalah perbedaan pemahaman awal antara anggota tim, yang diselesaikan dengan rapat rutin untuk menyamakan persepsi dan memastikan koordinasi yang baik.

## Sebagai hasil akhir, model SVM diintegrasikan ke dalam sistem CRM Ford dengan pipeline otomatis untuk memproses dan mengklasifikasikan deskripsi pekerjaan. project ini tidak hanya memberikan solusi teknis yang efektif tetapi juga meningkatkan efisiensi rekrutmen Ford secara signifikan.

## Pencapaian Hasil

## Efisiensi Proses Rekrutmen

## Project ini berhasil mengurangi durasi proses rekrutmen dari waktu semula, yaitu sekitar satu minggu, menjadi hanya 2–3 hari. Pencapaian ini diwujudkan melalui pengembangan sistem klasifikasi kalimat berbasis Machine Learning yang secara otomatis mengelompokkan deskripsi pekerjaan menjadi kategori yang lebih mudah dipahami oleh kandidat.

## Peningkatan Kualitas Pelamar

## Dengan implementasi sistem ini, jumlah pelamar yang memenuhi kriteria meningkat sebesar 20%. Sistem membantu menyelaraskan informasi dalam deskripsi pekerjaan dengan keterampilan yang dimiliki kandidat, sehingga kandidat lebih mudah menemukan posisi yang sesuai.

#### Akurasi dan Stabilitas Model

#### Model Support Vector Machine (SVM) yang dipilih untuk implementasi menunjukkan akurasi sebesar 86% pada data pengujian. Model ini lebih stabil dibandingkan dengan model lainnya, seperti Long Short-Term Memory (LSTM), yang meskipun memiliki akurasi lebih tinggi, memerlukan waktu komputasi yang lebih lama.

#### Integrasi Sistem

#### Model yang diimplementasikan telah diintegrasikan ke dalam sistem CRM Ford, memungkinkan proses otomatisasi mulai dari preprocessing data hingga klasifikasi kalimat. Ini memberikan efisiensi tambahan dalam pengelolaan proses rekrutmen.

Selain pencapaian tersebut, project ini menghadapi beberapa kesulitan, termasuk perbedaan pemahaman awal antar anggota tim terkait peran dan tanggung jawab. Hal ini diselesaikan melalui rapat yang dipimpin oleh Project Manager untuk menyamakan persepsi dan memastikan koordinasi yang baik. Kualitas dataset yang relatif bersih dan kolaborasi tim yang baik setelah penyesuaian peran masing-masing menjadi kemudahan yang memperlancar pelaksanaan project.

Keberhasilan ini tidak lepas dari pemahaman mendalam tim terhadap siklus data, kemampuan dalam menggunakan pustaka Python seperti Pandas, Scikit-learn, dan TensorFlow, serta bimbingan mentor yang memastikan setiap langkah project tetap sesuai dengan kebutuhan bisnis Ford.

# Bab IV

# Penutup

## Kesimpulan

Dengan memanfaatkan Feature Importance dari model SVM, kita bisa mendapatkan wawasan yang sangat berharga terkait proses rekrutmen. Beberapa business insights yang dapat diperoleh adalah tentang pengaruh kualifikasi, keterampilan, waktu respon, dan demografi terhadap penerimaan pelamar. Dengan implementasi strategi ini, proses rekrutmen menjadi lebih efisien dan dapat menghasilkan hasil yang lebih baik sesuai dengan metrik yang diinginkan.

## Saran

1. **Number of Applicants Accepted (Jumlah Pelamar yang Diterima):**

**Optimalisasi Proses Seleksi:** Dengan mengidentifikasi faktor-faktor seperti kualifikasi dan keterampilan sebagai fitur terpenting, fokuskan pada automasi seleksi berdasarkan kriteria ini untuk meminimalkan human error dan memastikan hanya pelamar yang memenuhi standar yang diproses lebih lanjut.

**Penyempurnaan Profil Kandidat:** Gunakan data kualifikasi untuk menyesuaikan lowongan pekerjaan agar dapat menarik pelamar yang lebih berkualitas dan memiliki peluang lebih tinggi untuk diterima.

1. **Key Performance Indicator (KPI):**

**Peningkatan KPI Rekrutmen:** Fokus pada waktu pemrosesan dan kecepatan rekrutmen. Jika proses rekrutmen masih memakan waktu lama, gunakan analisis sentimen untuk memahami apakah pelamar merasa prosesnya memakan waktu terlalu lama dan otomatisasi tahap-tahap tertentu.

**Set KPI untuk Automasi:** Tentukan KPI untuk pengurangan waktu evaluasi pelamar (misalnya, mengurangi waktu wawancara dan tes) dan peningkatan tingkat penerimaan berdasarkan data kualifikasi yang lebih preskriptif.

1. **Efisiensi Rekrutmen Karyawan**

**Implementasi Teknologi Rekrutmen:** Gunakan alat otomatisasi seperti video interview otomatis, AI screening tools, dan penjadwalan otomatis untuk mengurangi waktu yang dibutuhkan untuk menyeleksi pelamar.

**Penggunaan Tes Kemampuan Otomatis:** Implementasikan tes yang otomatis memberikan hasil dalam beberapa menit, dibandingkan dengan metode tradisional yang memakan waktu lebih lama.

**Review dan Penyederhanaan Proses Rekrutmen:** Identifikasi setiap tahapan dalam proses rekrutmen yang bisa dipercepat tanpa mengorbankan kualitas seleksi.

Referensi

*[1] Badan Pusat Statistik. (2022). Survei sosial ekonomi nasional. Jakarta: BPS.*

*[2] Pratama, A. D., & Putri, A. N. (2020). Pengaruh kualitas layanan terhadap kepuasan pelanggan pada UMKM berbasis online. Jurnal Manajemen Bisnis, 15(2), 123-135.*

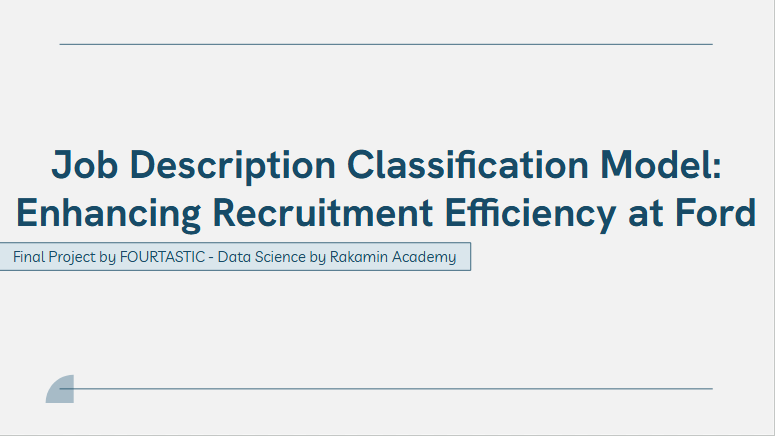
*[3] Rakamin Academy. (2023). Panduan belajar data science.*

*[4] Santoso, B. (2021, 15 Mei). Strategi pemasaran digital untuk UMKM di era pandemi. Jurnal Bisnis Digital, 3(1), 45-58.*

*[5] Sugiyono, P. (2017). Metode penelitian kuantitatif, kualitatif, dan R&D. Alfabet*

# Bab V

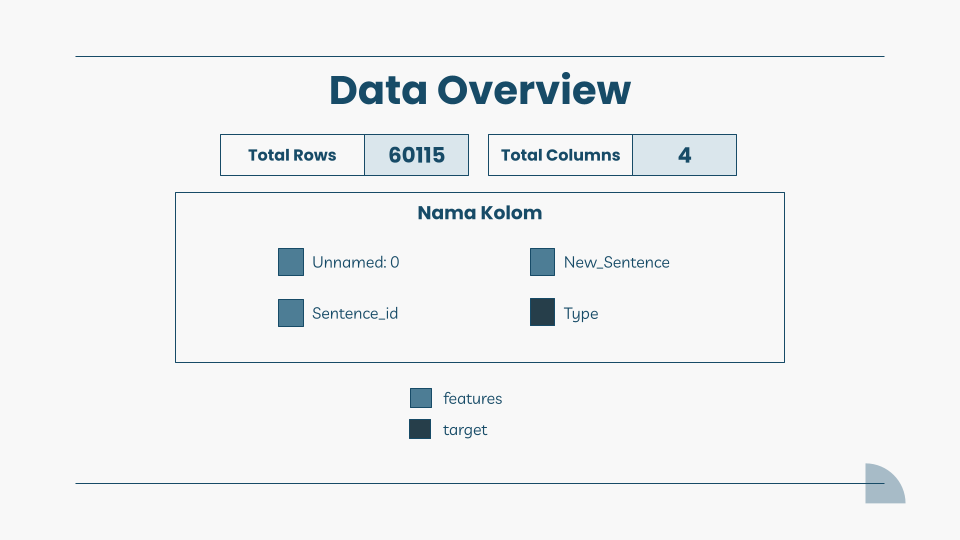
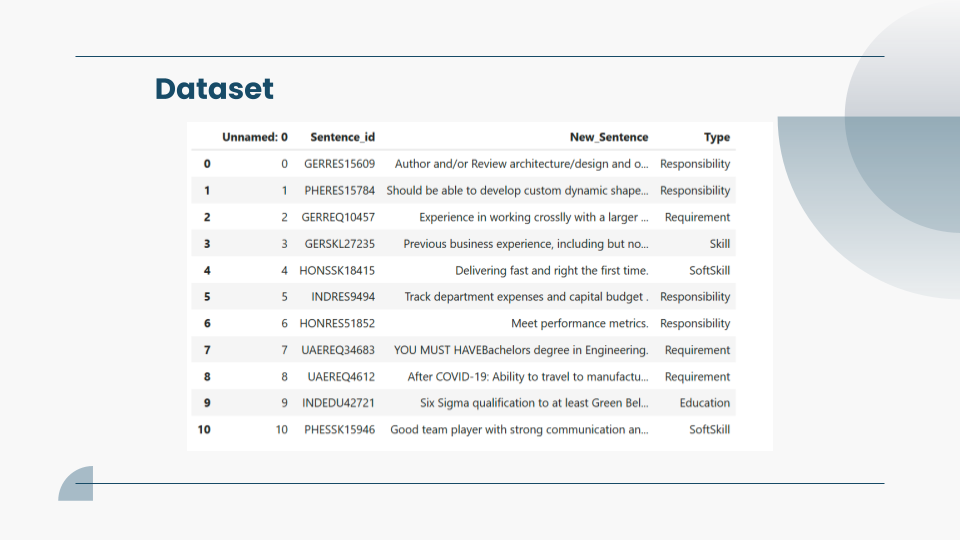
# Lampiran A. PPT Final Project Presentation

**

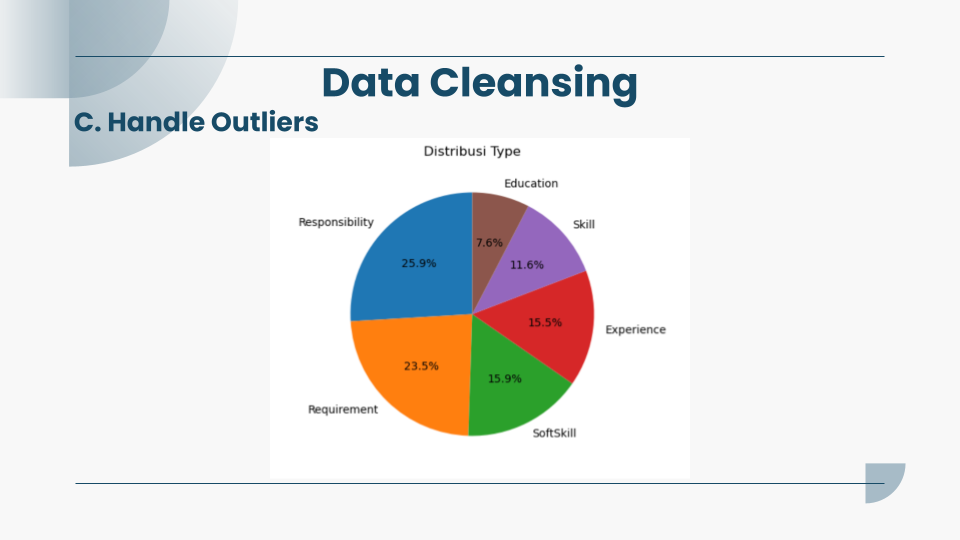
Gambar 2. Judul dan metode final project.

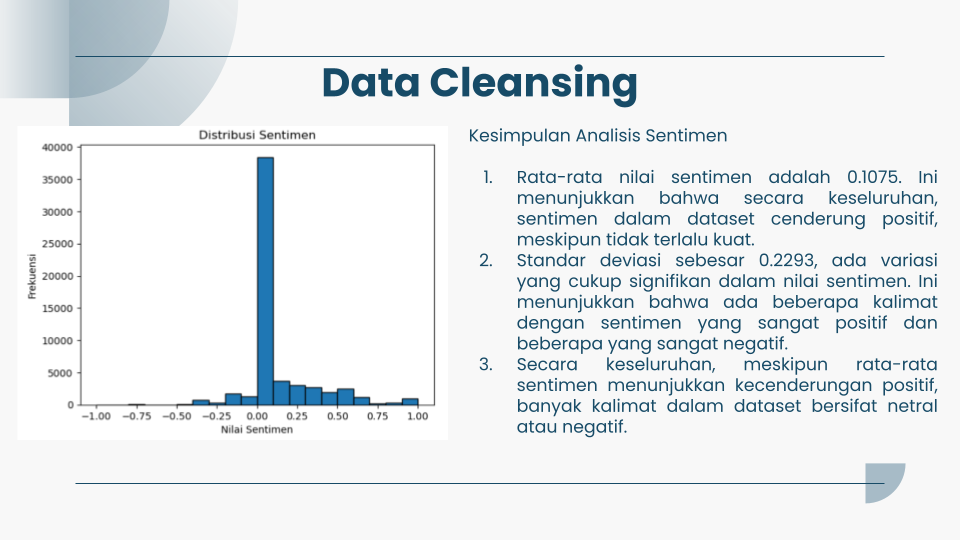
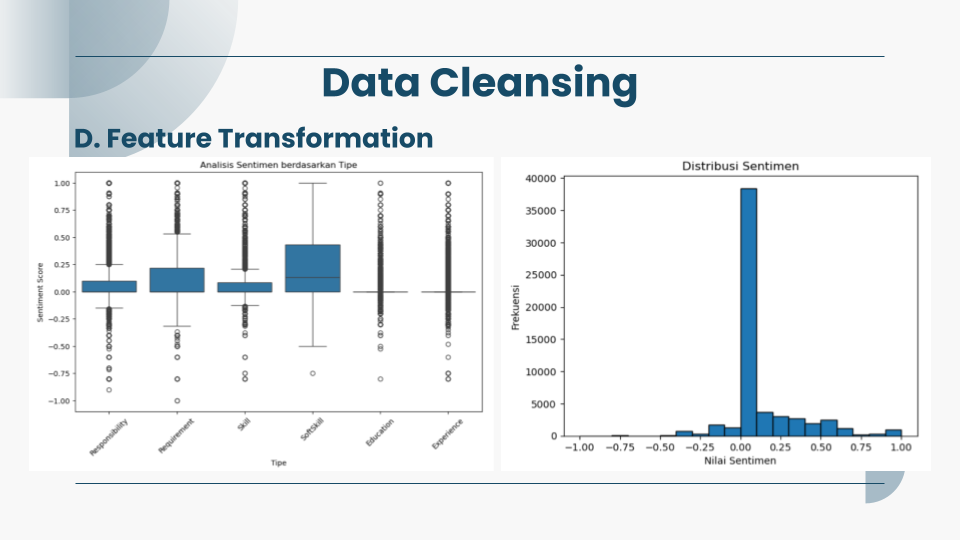
**

Gambar 3. Objectives dan Business metrics.

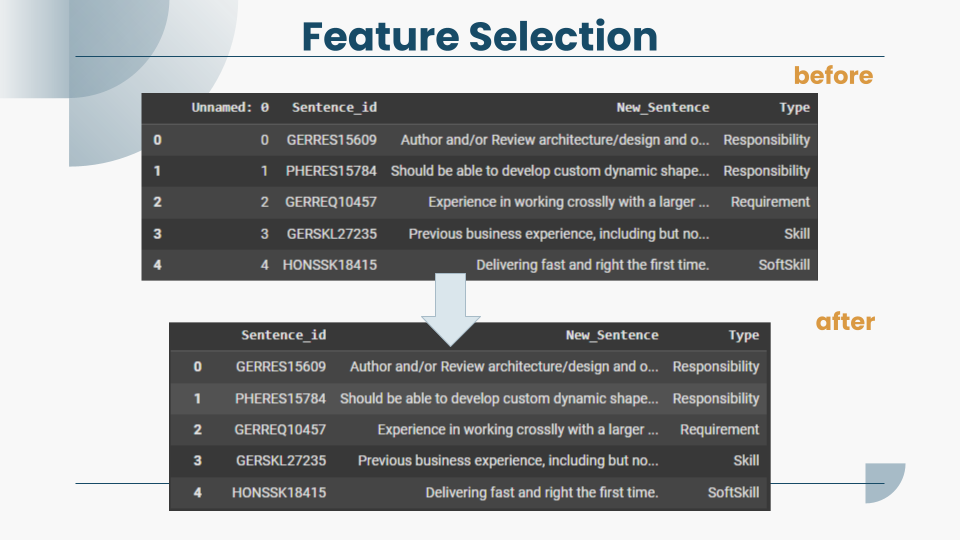
**

Gambar 4. Dataset dan data overview.

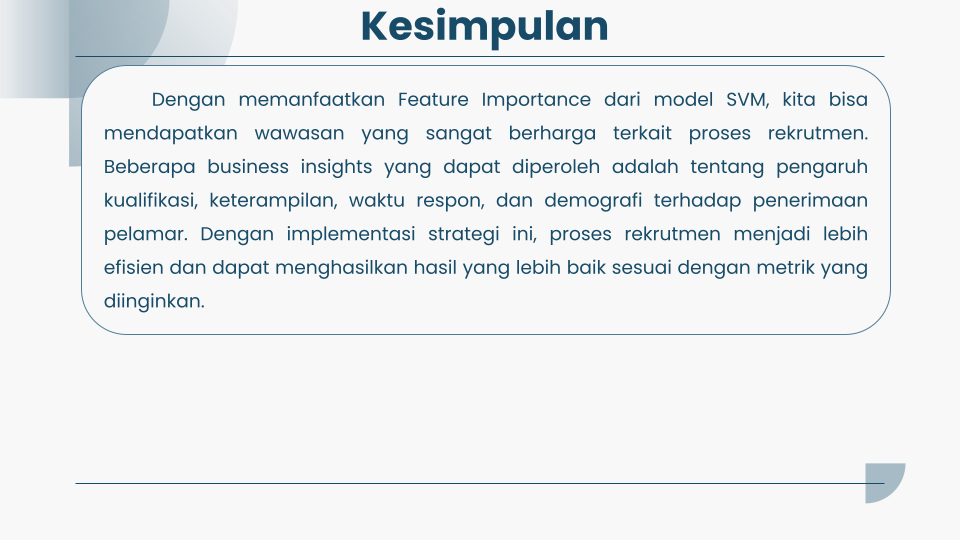
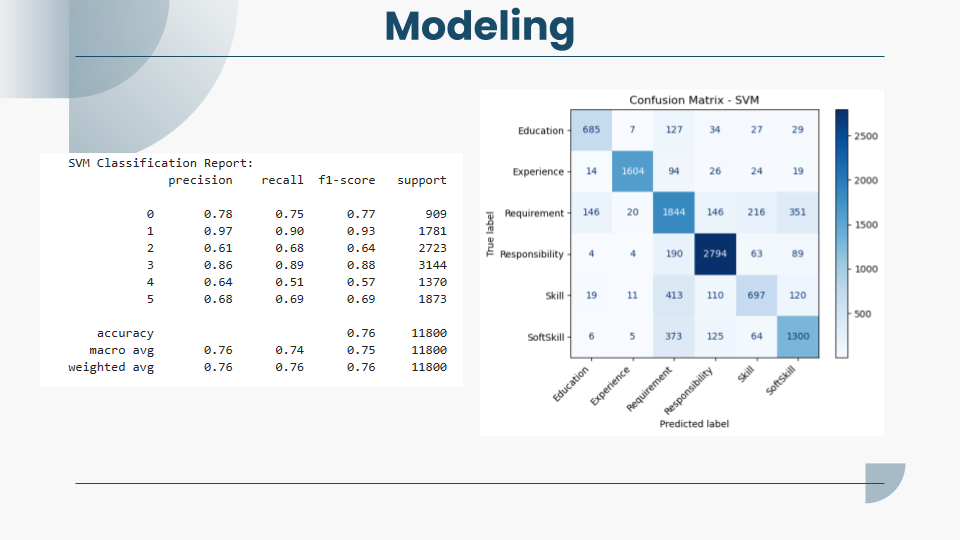
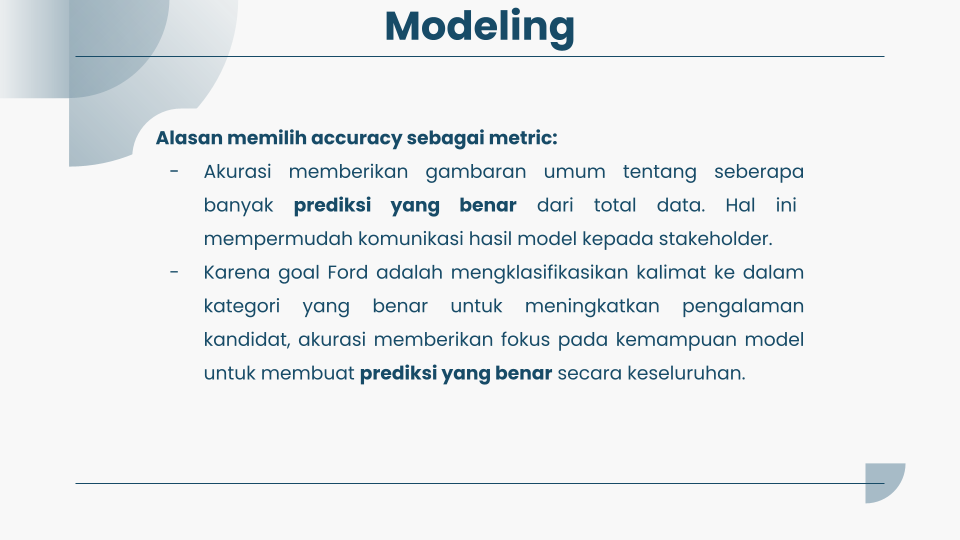
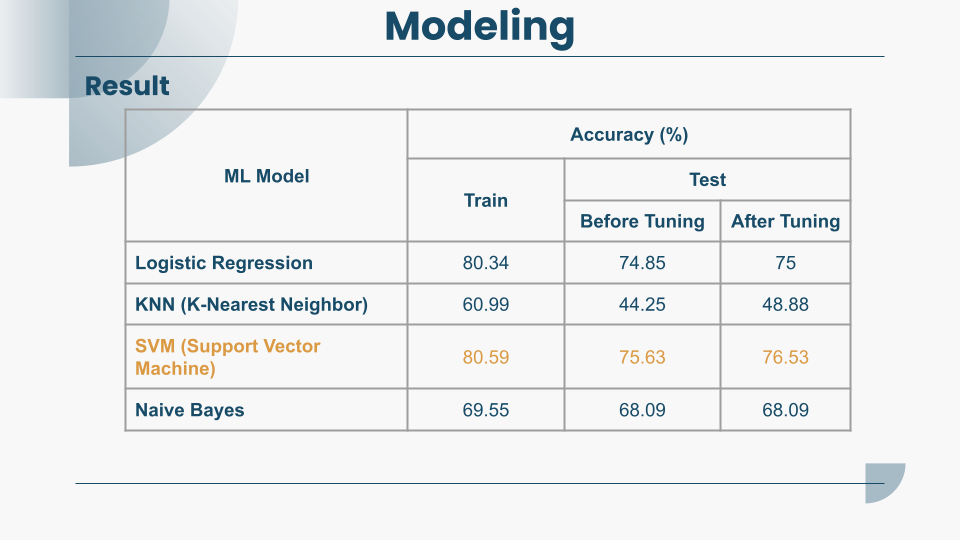
**

**

Gambar 5. Hasil EDA.

**

Gambar 6. Feature selection.

**

Gambar 7. Hasil modelling dan kesimpulan akhir.

# Bab VI

# Lampiran B. Notulensi Mentoring

| Minggu/Tgl | Kegiatan | Hasil |
| --- | --- | --- |
| 19 Oktober 2024 | Mentoring Stage 0  Dengan pembahasan :  1. Problem yang ingin diselesaikan dari dataset  2. Pembagian role dan tanggung jawab  3. Goals yang ingin dicapai  4. Penentuan objective yang sesuai  5. Business metric | Membahas mengenai problem dari Ford (perusahaan berbasis CRM), yang mengalami kesulitan dalam proses rekrutmen karena deskripsi pekerjaan yang panjang. Sehingga, dibutuhkan sistem yang  memberikan rekomendasi pekerjaan sesuai keterampilan dan pengalaman kandidat, sehingga proses rekrutmen lebih efisien dan menarik lebih banyak kandidat yang tepat.  Lalu, menentukan objective (turunan dari goals yang ingin dicapai), dan menentukan business metric yang sesuai berdasarkan impact yang  diharapkan dari model machine learning yang dibuat. |
| 02 November 2024 | Mentoring Stage 1  Dengan pembahasan :  1. Descriptive Statistic  2. Univariate & Multivariate Analysis  3. Business Insight  4. Data preparation untuk dataset non-numeric | Pada diskusi kali ini, kami membahas mengenai EDA dari dataset perusahaan Ford. Untuk Descriptive statistic, semua tipe data sudah sesuai dan tidak ada summary yang aneh. Missing  value boleh di drop saja karena jumlahnya tidak terlalu signifikan. Lalu, perlu dilakukan analisis lebih lanjut untuk duplicate data dengan dua kolom, yaitu Sentence\_id dan Type. Untuk  penambahan data insight, bisa dicari menggunakan word cloud. Kemudian, mentor menyarakan kelompok menggunakan API Gemini LLM untuk koreksi teks dari New\_Setence. Untuk di  pdf, bisa menambahkan data science life cycle dan jelaskan juga mengenai feature selection, feature mana yang akan kelompok gunakan. |
| 23 November 2024 | Mentoring Stage 2  Dengan pembahasan :  1. Data Cleansing (missing value, duplicated data, outliers, feature transformation, feature encoding, handle class imbalance)  2. Feature Engineering (feature selection, feature extraction, mencari feature tambahan) | Pada diskusi kali ini, kami membahas mengenai data preprocessing dari dataset perusahaan Ford. Untuk duplicated data sudah di-handle dan missing value di-drop karena jumlahnya  tidak signifikan. Tidak dilakukan handle outlier karena tidak terdapat data outlier. Dilakukan juga label encoding untuk column ‘Type’ dan menghapus stopwords pada column  ‘New\_Sentence’. Lalu, dilakukan analisis skor sentiment (positif, negatif, netral) untuk mengetahui sikap yang terkandung dalam kalimat. Mentor merekomendasikan untuk mencari  algoritma klasifikasi di sentiment, seperti Naïve Bayes, SVM, dll. Lakukan hyperparameter, lalu cari yang terbaik mana yang dipilih untuk sentiment. Berikan inovasi pada LLM untuk  permasalahan sentiment tsb. Kemudian, untuk handle class imbalance, kelompok melakukan oversampling dengan SMOTE. Mentor juga memberi saran untuk melampirkan flowchart  data science pada proses yang dilakukan di stage ke 2. Pada saat upload ke GitHub, pastikan dependency jelas, foldernya rapi, serta cara menjalankan kodenya juga sudah ada |
| 30 November 2024 | Mentoring Stage 3  Dengan pembahasan :  1. Model evaluation  2. Feature Importance | Pada diskusi kali ini, kami membahas mengenai data modeling yang telah dilakukan oleh kelompok dari dataset perusahaan Ford. Kelompok melakukan 2 ML model, yaitu Logistic  Regression dan LSTM. Akurasi yang diperoleh dengan menggunakan Logistic Regression untuk data training adalah 80,34% dan data testing 74,85% perbedaan yang signifikan ini  mengindikasi terjadinya overfitting sehingga kelompok melakukan hyperparameter tuning. Setelah dituning akurasi hanya menjadi 75%. Kelompok juga mencoba deep learning  menggunakan LSTM, diperoleh hasil pada data train 92,89% dan pada data test sebesar 74,96%. Terlihat hasil yang diperoleh juga terindikasi overfitting. Namun, setelah dilakukan  hyperparamete tuning dengan beberapa parameter pilihan, didapat hasil yang meningkat yaitu 89.51% pada data train dan 89,15% pada data test. Mentor menyarankan kepada kelompok  untuk lebih mengeksplor ML model lain agar hasil yang didapat bisa lebih representatif untuk dikomparasi dan menyimpan ML terbaik ke dalam metadata |
| 21 Desember 2024 | Mentoring Stage 4  Poin pembahasan:   1. Executive summary. 2. Perkuat di sistemnya ada fitur apa. 3. Gemini API dijadikan API. 4. Inovasi berupa API yg bisa merespon setiap input yg diberikan - ini bisa dari Gemini API atau model yang kalian gunakan. | Pertemuan kali ini membahas evaluasi setelah simulasi presentasi, perlu adanya beberapa revisi di dalam materi presentasinya. Serta cukup banyak membahas terkait dengan inovasi yang akan dikerjakan yaitu model API. Disarankan oleh mentor dengan banyak membaca-baca terkait Gemini API. |
|  |  |  |

# Bab VII

# Lampiran C. Dokumen Teknik

Stage 1 - EDA <https://drive.google.com/file/d/1Q5ztL_KXtnu6FfZJ2oETouDrvwUqUUDx/view?usp=sharing>

Stage 2

<https://drive.google.com/file/d/1yA1-fd_NjsP7x7_S4ocLTOeWL9W32FiL/view?usp=sharing>

Stage 3

<https://drive.google.com/file/d/1wGSV8IEUiVC-9vatFPioUGSiVI9vi8-5/view?usp=sharing>

Stage 4

<https://github.com/anggagilang11/Fourtastic/tree/stage4>

Modelling KNN-SVM-NB Stage 3

<https://drive.google.com/file/d/1kXN8slyXsd59NRRVnyNLFDQ9AC19gYRo/view?usp=sharing>

Modelling Logistic Regression Stage 3

<https://drive.google.com/file/d/1_ZFJg9w3ptjHL9BS4hQwHtf6bgNRCyFJ/view?usp=sharing>

GitHub

<https://github.com/anggagilang11/Fourtastic>