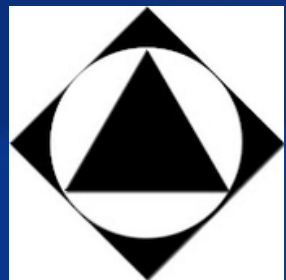


Classifying and Summarizing AI-related Text Data using **IBM Granite 3.3-8B**

Capstone Project – September 2025
Ali Akbar Permana | ITENAS





Project Overview

Objectives:

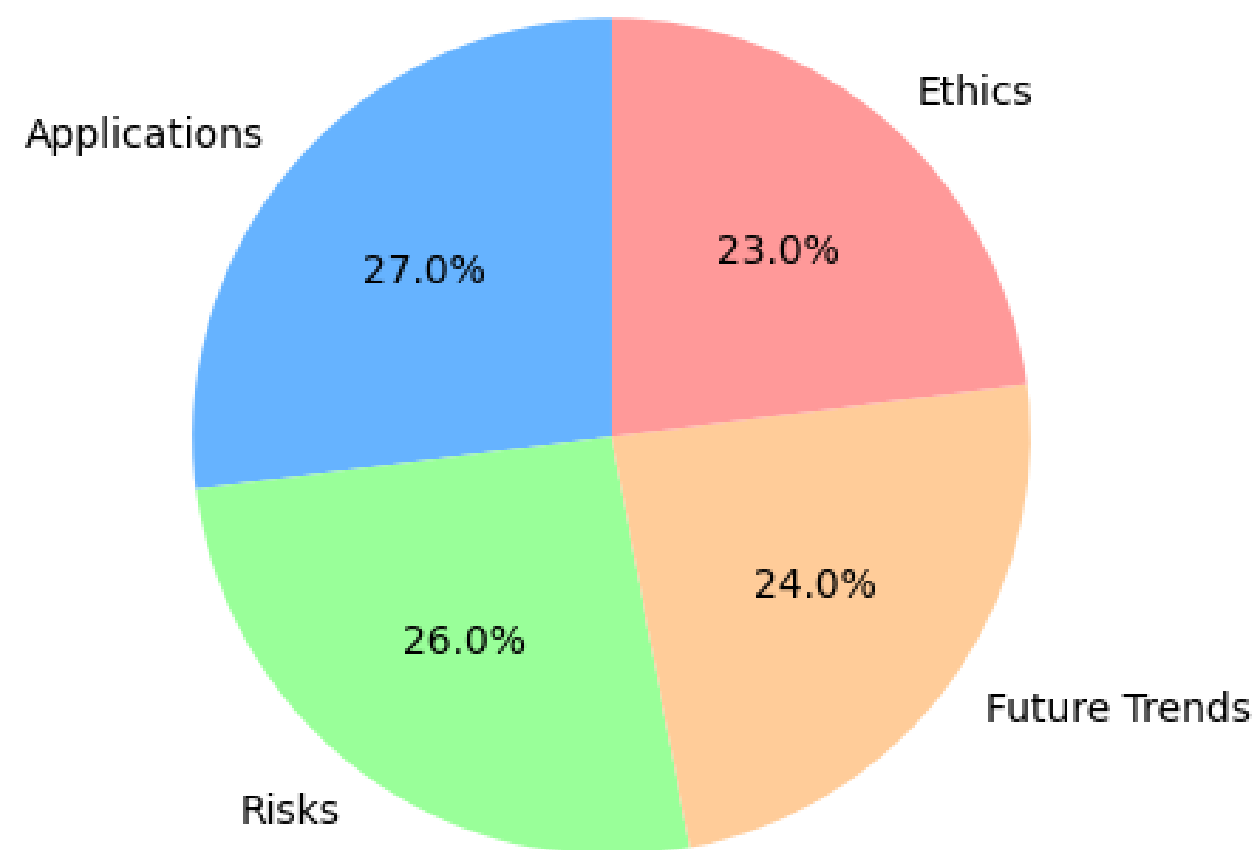
- *Classify AI-related texts into four categories: Applications, Risks, Ethics, and Future Trends.*
- *Generate concise and meaningful summaries of AI content.*
- *Compare performance across datasets (100 vs 200 texts).*
- *Provide insights for AI content organization.*

Key Technologies:

- *IBM Granite 3.3-8B Instruct Model*
- *Google Colab + Python*
- *Replicate API*

DATASET OVERVIEW

Proporsi Kategori AI Text Dataset



Dataset yang digunakan terdiri dari 200 teks berbahasa Inggris seputar topik Artificial Intelligence. Teks dibagi ke dalam 4 kategori utama:

- *Applications – penggunaan AI di berbagai industri*
- *Risks – potensi bahaya & dampak negatif AI*
- *Ethics – isu moral & keadilan dalam penerapan AI*
- *Future Trends – prediksi perkembangan AI di masa depan*

Rata-rata panjang teks adalah 68 karakter, sehingga relatif singkat dan mudah dianalisis.



Analysis Process

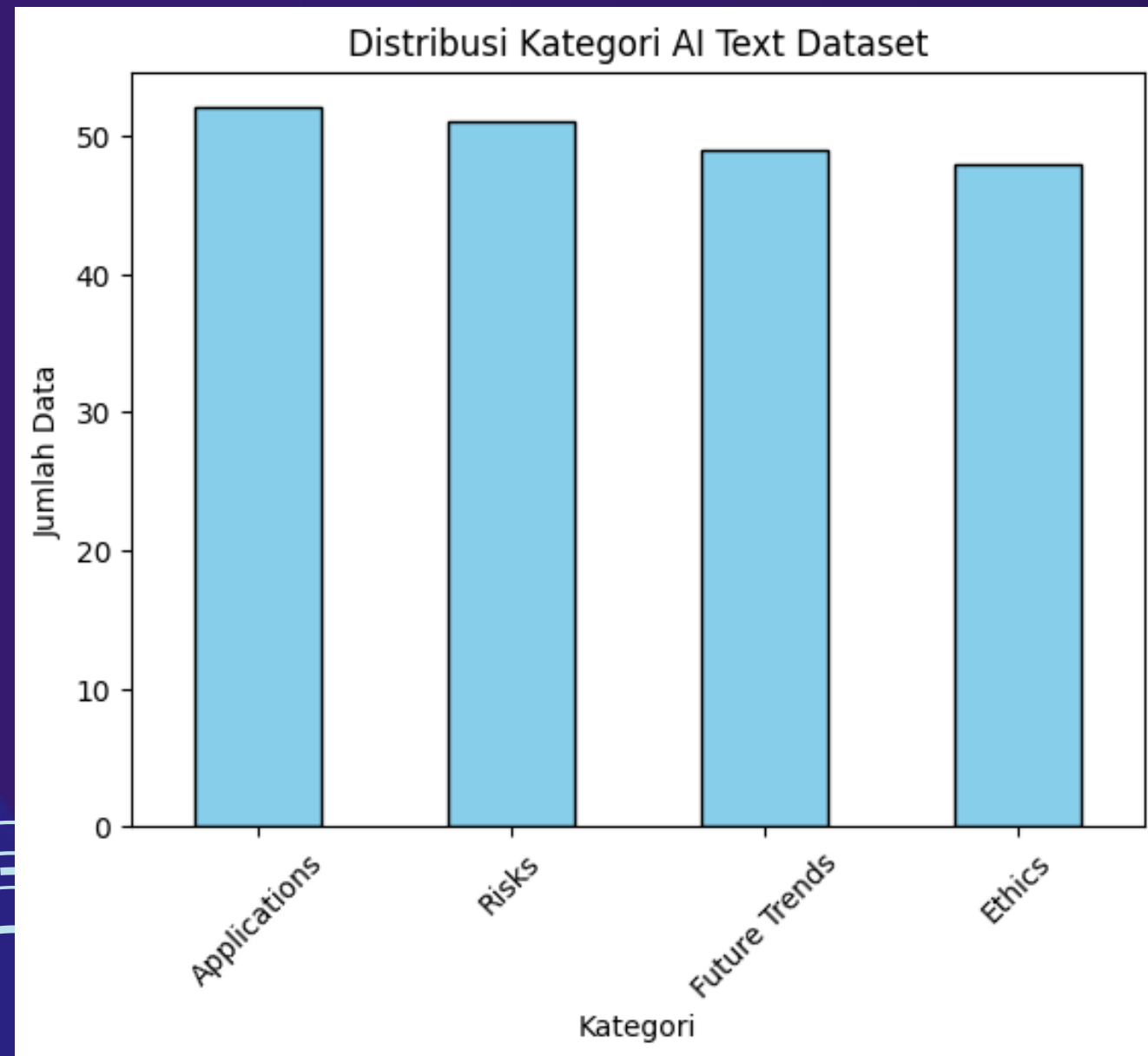
Proses analisis dilakukan secara bertahap:

- 1. Data Preparation – membersihkan data dari format Markdown menjadi CSV yang rapi*
- 2. Exploratory Data Analysis (EDA) – mengecek distribusi label, panjang teks, dan contoh data*
- 3. Model Implementation – menggunakan IBM Granite melalui Replicate API*
- 4. Evaluation & Insights – mengukur performa model dan menarik kesimpulan*

Pipeline ini memastikan data valid sebelum digunakan pada model AI.



05



Key Visualizations

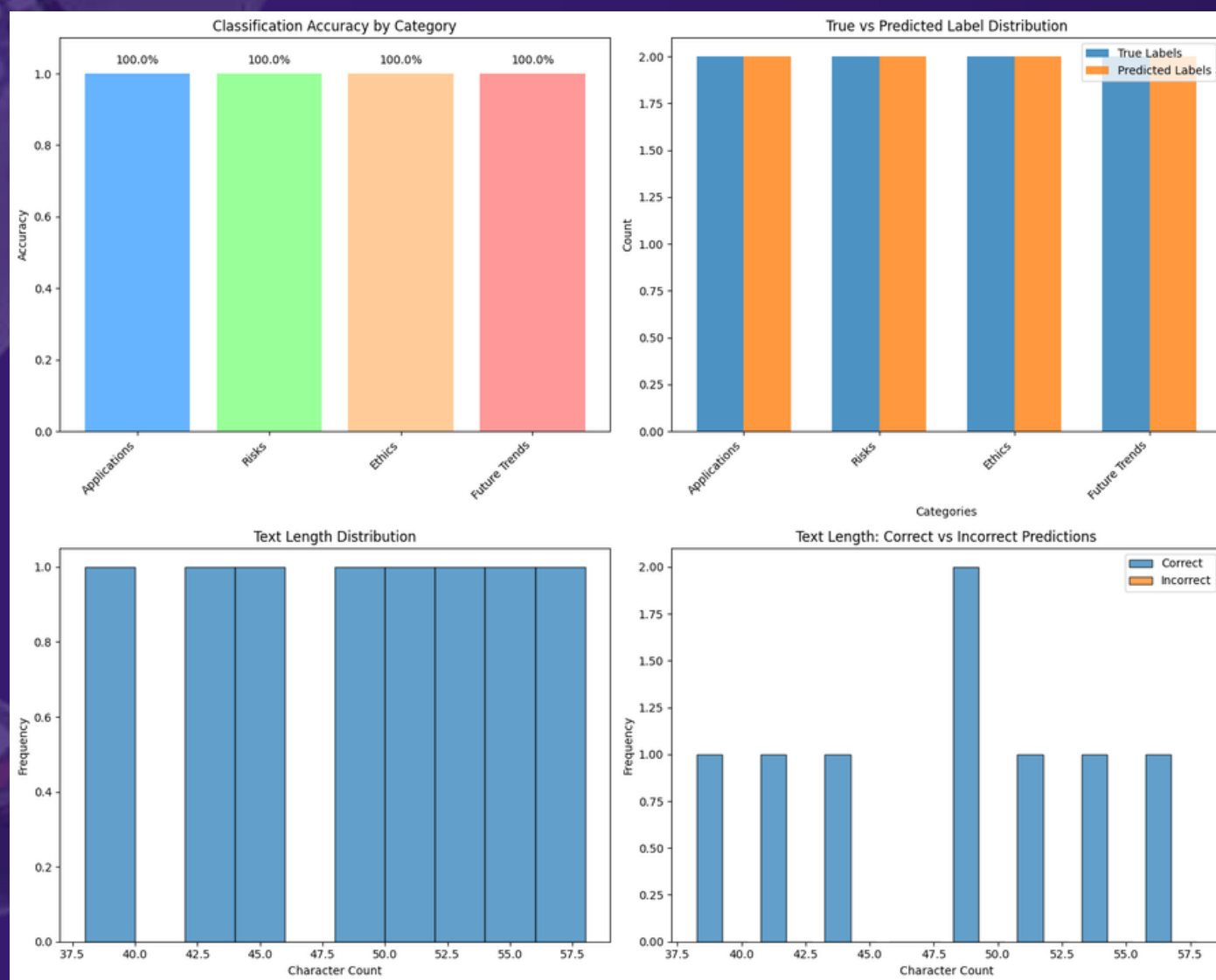
Hasil EDA menunjukkan dataset cukup seimbang di setiap kategori:

- *Applications: 50 teks*
- *Risks: 50 teks*
- *Ethics: 49 teks*
- *Future Trends: 49 teks*

Analisis panjang teks juga menunjukkan distribusi yang konsisten, sehingga mempermudah model dalam melakukan klasifikasi.

06

Model Performance



IBM Granite diuji dengan prompt klasifikasi dan prompt ringkasan. Hasil uji pada sample teks menunjukkan akurasi 100% di keempat kategori:

- *Applications* → terdeteksi dengan benar
- *Risks* → risiko diidentifikasi dengan tepat
- *Ethics* → konteks etika dipahami
- *Future Trends* → tren masa depan dikenali

Hal ini menunjukkan model mampu memahami konteks AI dengan baik.

Insights & Business Value

IBM Granite membantu mengubah data teks terkait AI menjadi insight yang bermakna.

Dari insight ini, organisasi dapat mengambil strategic actions yang tepat dan akhirnya memperoleh competitive advantage.

Key Insights:

- *Dataset seimbang membuat klasifikasi lebih adil*
- *Granite efektif untuk teks singkat di domain AI*
- *Prompt yang spesifik meningkatkan akurasi*

Business Value:

- *Automation: Menghemat waktu dalam klasifikasi konten*
- *Decision Support: Memberi dasar yang lebih kuat untuk strategi bisnis*
- *Scalability: Siap diterapkan ke ribuan data real-time*



Conclusions & Recommendations

Penjelasan:

- 1. IBM Granite terbukti handal untuk klasifikasi teks AI dan ringkasan singkat.*
- 2. Dataset seimbang berperan penting dalam menjaga akurasi model.*
- 3. Sistem ini bisa diperluas ke dataset yang lebih besar atau konten multibahasa.*

Rekomendasi:

- Implementasi API untuk pemakaian real-time*
- Integrasi ke sistem berita, riset, dan pendidikan*
- Pengembangan lanjutan untuk mendukung bahasa lain*

Thank You!

*Thank You for Your Attention
For more information, please contact:*

 **Email: aliakbar.vk42@gmail.com**

 **GitHub: github.com/aliakbar/Al-Text-Classification**

 **LinkedIn: www.linkedin.com/in/ali-akbar-permana-4ab80436a**