

BAB 1

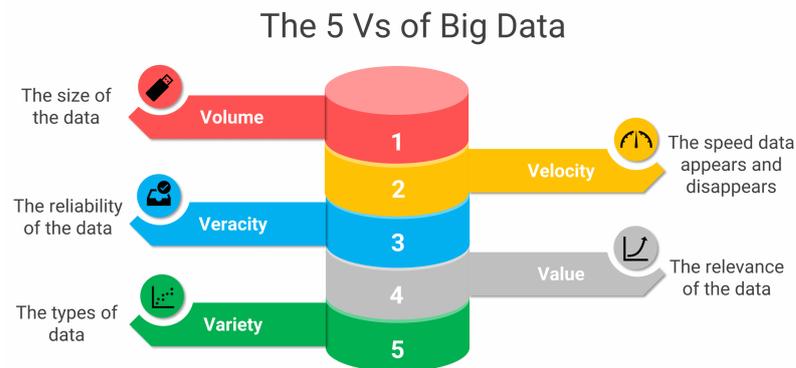
PENDAHULUAN

1.1 Latar Belakang Masalah



Gambar 1.1. Divisi Kompas Gramedia, Grup Retail & Publishing

Kompas Gramedia, melalui Divisi Retail & Publishing (contoh dari nama anak perusahaan terdapat di gambar 1.1), menjadi lokasi magang penulis karena perannya sebagai salah satu jaringan ritel buku terbesar di Indonesia dengan lebih dari 115 toko dan 6 penerbit internal [1][2]. Dalam divisi ini, tugas utama penulis adalah mengembangkan prototipe sistem *gender classification* dengan fokus utama pada pemanfaatan nama pelanggan sebagai satu-satunya fitur input, khususnya nama-nama khas Indonesia, sehingga nantinya perusahaan dapat memetakan dan menganalisis profil demografis per kawasan. Pembuatan sistem ini dikarenakan langkanya data-data *gender* yang ada di Kompas Gramedia, sehingga nantinya perusahaan dapat memetakan profil demografis tiap wilayah. Karena data *gender* yang tersedia cukup terbatas, proses ini menjadi krusial untuk mendukung personalisasi rekomendasi produk dan efektivitas kampanye pemasaran yang disesuaikan dengan distribusi *gender* di masing-masing daerah.

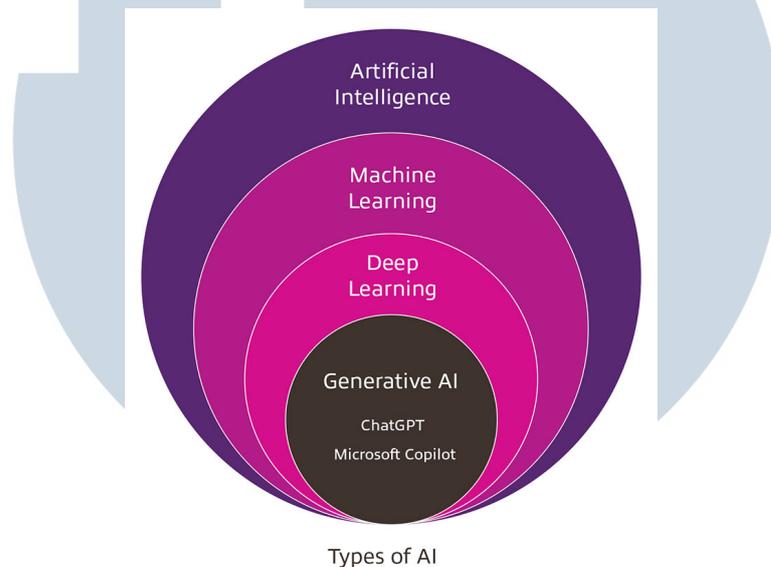


Gambar 1.2. The 5 vs of *Big Data*

Big Data merupakan kumpulan data dalam jumlah besar dan kompleks yang tidak dapat diproses menggunakan metode tradisional [3]. Lima karakteristik utama yang menentukan apakah suatu database termasuk dalam kategori *Big Data* berdasarkan Gambar 1.2 adalah:

1. **Volume:** Ukuran data yang sangat besar atau tidak terstruktur sehingga tidak dapat disimpan dalam tabel database tradisional dengan baris dan kolom. Oleh karena itu, *Big Data* disimpan dalam sistem berkas modern atau basis data berorientasi objek yang dirancang untuk menangani informasi dalam jumlah masif [4].
2. **Velocity:** Data dalam *Big Data* dihasilkan dan berubah dengan sangat cepat. Misalnya, sebuah tweet terbaru dapat segera tenggelam oleh konten yang lebih baru. Data yang berasal dari mesin dan perangkat juga bersifat sementara dan dapat hilang jika tidak segera ditangkap dan diproses [4].
3. **Veracity:** Tidak semua data dalam *Big Data* dapat dipercaya. Bot yang menghasilkan ulasan palsu atau akun spam yang menyebarkan promosi dapat mengaburkan hasil analisis. Selain itu, dataset yang tidak lengkap atau keterbatasan API dapat menyebabkan ketidakakuratan, sehingga sulit untuk menerapkan *Big Data* secara efektif [4].
4. **Value:** Nilai dari data tergantung pada kemampuan organisasi untuk mengekstrak insight yang relevan darinya. Data tanpa konteks atau interpretasi yang benar tidak memberikan manfaat strategis [4].

5. **Variety:** *Big Data* hadir dalam berbagai bentuk, seperti teks, audio, video, dan data sensor. Struktur dan maknanya juga bervariasi. Contohnya, kata "Washington" bisa merujuk pada sebuah negara bagian, ibu kota AS, atau nama seseorang, tergantung pada konteksnya. Bahkan sarkasme dalam media sosial, seperti "Bagus sekali, laptop baruku rusak di hari pertama," dapat memperumit analisis dengan menambah lapisan ambiguitas [4].



Gambar 1.3. Hierarki metode dalam kecerdasan buatan: Artificial Intelligence → Machine Learning → Pembelajaran Fitur (Representation Learning) → Deep Learning → Generative AI.

Untuk mengolah dan menganalisis *Big Data*, diperlukan teknik serta teknologi yang lebih maju. Salah satu solusi terbaik adalah Artificial Intelligence (AI). Gambar 1.3 menggambarkan bagian-bagian utama dalam AI, termasuk machine learning, deep learning, dan Generative AI. Secara umum, AI adalah teknologi yang memungkinkan komputer menjalankan tugas-tugas yang biasanya dilakukan oleh manusia. Umumnya, tugas-tugas tersebut melibatkan perhitungan kompleks, seperti optimasi, prediksi, dan klasifikasi.

Salah satu cabang dari AI adalah machine learning, yang memungkinkan komputer untuk belajar dari data dan membuat prediksi. Machine learning sendiri merupakan bagian dari AI, di mana konsep-konsep kecerdasan buatan digunakan untuk membangun model pembelajaran mesin. Di dalam machine learning terdapat deep learning, yang bekerja dengan prinsip menyerupai jaringan neuron di otak manusia untuk menyelesaikan tugas-tugasnya. Selanjutnya, ada Generative AI,

yang berbeda karena kemampuannya dalam menciptakan konten baru, seperti teks, gambar, dan suara [5].

Dengan bantuan AI, *Big Data* dapat dimanfaatkan dalam pengambilan keputusan, peningkatan pengalaman pelanggan, continuous intelligence, pengelolaan risiko yang lebih baik, efisiensi operasional, serta peningkatan kelincahan dan inovasi [4]. Tak heran jika banyak perusahaan berlomba-lomba untuk menguasai dan menganalisis *Big Data* mereka. Namun, untuk mencapai hal tersebut, dibutuhkan tenaga kerja yang ahli dan berpengalaman, atau minimal memiliki pemahaman dasar dalam bidang tersebut. Oleh karena itu, perusahaan sering merekrut profesional berpengalaman atau membuka peluang magang bagi mahasiswa untuk berkontribusi.

Pembuatan model klasifikasi *gender* sebelumnya telah dilakukan menggunakan algoritma Random Forest. Namun, saat proses evaluasi proyek dan diskusi bersama supervisor, disarankan untuk mengganti model tersebut dengan pendekatan baru. Hal ini dikarenakan model awal tidak menghasilkan performa yang memuaskan, dengan akurasi dan F1-score hanya berkisar antara 70% hingga 72%. Oleh karena itu, pengembangan model baru perlu segera dilakukan untuk memperoleh hasil yang lebih optimal dibandingkan dengan model sebelumnya.

Dengan konklusi, model *gender prediction* berbasis machine learning dan deep learning telah terbukti akurat memproyeksikan atribut *gender*. Ini bisa dari data transaksi dan ulasan pelanggan, misalnya dengan SVM, Random Forest, atau arsitektur CNN/VGG-16 pada wajah pelanggan di toko fisik[6][7]. Pekerjaan ini memerlukan fondasi teori pemasaran modern—sebagaimana diuraikan Kotler dalam *Marketing 5.0*—serta kapabilitas infrastruktur *Big Data* (5 V's: Volume, Velocity, Variety, Veracity, Value) untuk menangani data heterogen dalam skala masif [8][9]. Dengan tahapan kerja yang mencakup: praproses data nama, ekstraksi fitur (seperti bag-of-names dan embeddings), pelatihan model (SVM, Random Forest, LSTM), evaluasi metrik (*accuracy*, *precision*, *recall*, dan *F1-score*), prototipe ini diharapkan dapat menjadi fondasi bagi penerapan strategi data-driven decision making yang konsisten di seluruh unit Divisi Retail & Publishing Kompas Gramedia, sekaligus melampaui performa model yang digunakan sebelumnya [10][11].

1.2 Maksud dan Tujuan Kerja Magang

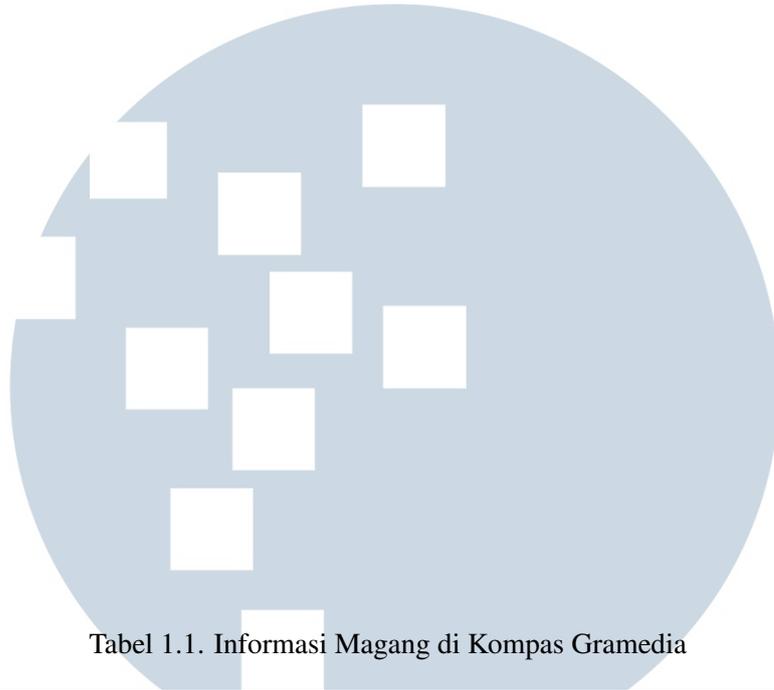
Pemilihan tempat magang dan ingin melakukan aktivitas tersebut ada dengan tujuan dan maksud, yaitu:

1. Pengembangan model machine learning dan deep learning sesuai dengan standar profesional.
2. Pembuatan model klasifikasi *gender* untuk menyusun *customer profile* bagi perusahaan Kompas Gramedia.
3. Pengembangan model klasifikasi alternatif yang memiliki performa lebih baik dibandingkan model sebelumnya.
4. Penyusunan *customer profile* untuk mendukung dan memperkuat strategi promosi yang dijalankan oleh Kompas Gramedia.

1.3 Waktu dan Prosedur Pelaksanaan Kerja Magang

Mengikuti struktur dari MBKM magang UMN, total waktu yang harus ditempuh selama magang bertotal 640 jam. Mahasiswa akan bekerja dan belajar selama di perusahaan dan di sarankan untuk bekerja bertotal 8 jam atau dari jam 08.00 WIB - 17.00 WIB. Mahasiswa juga memiliki waktu fleksibilitas yang jika mahasiswa datang lebih awal, maka mereka bisa pulang lebih awal dengan ketentuan harus bekerja total 8 jam. Mahasiswa dibolehkan, walau tidak sarankan, bekerja lebih dari 8 jam, namun tidak kurang dari 8 jam. Mahasiswa cuman bisa telat maksimal 8.30 WIB dan jika mahasiswa kelebihan, maka Mahasiswa sudah siap untuk mendapatkan sanksi dari perusahaan.

Magang ini akan juga dilaksanakan dengan menganut Work From Office (WFO). Mahasiswa yang mengikuti magang harus bekerja langsung di kantor perusahaan selama pelaksanaan dan tidak diijinkan untuk bekerja di rumah. Mahasiswa diberikan kesempatan untuk izin kepada supervisor, setengah hari atau sehabian, jika ada berhalangan, seperti sakit, keperluan kampus, acara mendadak, dll. Berikut adalah informasi lokasi dan alamat perusahaan magang pada tabel 1.1:



Tabel 1.1. Informasi Magang di Kompas Gramedia

Jenis Informasi	Detail
Nama Perusahaan	Kompas Gramedia
Alamat	Jl. Palmerah Barat No.22-26, RT.1/RW.2, Gelora, Tanah Abang, Jakarta Pusat
Periode Magang	1 Februari 2025 – 30 Juni 2025
Jumlah Jam	640 jam
Sistem Kerja	Work From Office (WFO)
Jam Kerja	08.00 WIB – 17.00 WIB (8 jam/hari)

UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA