

BAB I

PENDAHULUAN

1.1. Latar Belakang

Di era transformasi digital yang semakin pesat, pertumbuhan data meningkat dengan cepat dan akan selalu meroket mencapai titik tertinggi dari tahun ke tahun. Pertumbuhan data ini berbanding lurus dengan penetrasi pengguna internet selama dekade terakhir dan diprediksi akan masih terus bertambah. Dimana hingga tahun 2024, pengguna yang telah mengakses internet mencapai 67,1 persen populasi global atau sekitar 5,44 miliar pengguna. Persentase ini meningkat secara signifikan jika dibandingkan pada tahun 2018 yang memperkirakan pengguna internet hanya sebesar 53 persen populasi global [1]. Perangkat yang digunakan pengguna internet terus menghasilkan data, saat ini diperkirakan data dibuat setiap hari mencapai 328,77 juta terabyte pada tahun 2024 dan di tahun 2025 diproyeksikan akan mencapai 181 zettabyte data akan dihasilkan [2].

Pertumbuhan data yang signifikan ini menawarkan peluang besar bagi pelaku diberbagai industri untuk mendapatkan wawasan yang lebih mendalam, meningkatkan efisiensi, dan mendukung pengambilan keputusan yang lebih tepat. Namun, manfaat ini hanya dapat dirasakan apabila data tersebut dapat dimanfaatkan dan dikelola dengan baik. Data yang terus bertambah dalam volume besar ini dirujuk dengan istilah *big data*, yang menuntut penerapan teknologi dan pendekatan analitis yang sesuai [3]. Secara umum big data sendiri mula-mula memiliki tiga karakteristik utama yang dikenal sebagai 3V. Volume yang merujuk pada jumlah informasi yang sangat besar, kecepatan (*velocity*) yang mencerminkan kecepatan data dihasilkan dan diproses secara *real-time*, serta variasi (*variety*) yaitu data memiliki karakteristik beragam bentuk dan variasi sumber data, mulai dari data terstruktur hingga data tidak terstruktur [4].

Data terstruktur adalah data yang memiliki format atau model yang terorganisir dan relatif mudah untuk dianalisis, seperti entri dalam basis data. Sementara itu, data tidak terstruktur mencakup berbagai jenis data yang kaya akan informasi,

namun kompleksitas dan ketidakteraturan menghadirkan tantangan, seperti gambar, video, suara, serta dokumen teks dalam format digital atau fisik [5]. Salah satu tantangan utama yang dihadapi oleh perusahaan saat ini adalah menangani data tidak terstruktur yang berasal dari dokumen-dokumen fisik, seperti faktur (*invoice*) dan tanda terima (*receipt*), yang penting dalam pencatatan keuangan perusahaan. Proses pengolahan dokumen ini dilakukan secara manual, yang memakan waktu dan rentan terhadap kesalahan. Oleh karena itu, teknologi yang dapat mengekstraksi informasi dari dokumen fisik secara otomatis sangat diperlukan guna meningkatkan efisiensi, serta mengoptimalkan waktu, tenaga, dan biaya.

Dalam menghadapi tantangan tersebut, penerapan teknologi menjadi solusi yang efektif. Salah satu inovasi teknologi yang memiliki potensi besar dalam mengatasi masalah pengolahan data tidak terstruktur adalah melalui penerapan *Machine Learning* (ML) dan *Artificial Intelligence* (AI) dengan menggunakan metode *Optical Character Recognition* (OCR). OCR mampu mengekstrak informasi teks dari gambar dan menerjemahkannya ke dalam komputer sehingga memudahkan dalam memproses dan analisis lebih lanjut [6]. Teknologi ini memberikan jalan keluar yang efisien dalam mengotomatiskan proses entri data dan mencegah kesalahan yang umum terjadi dalam metode manual. Seiring dengan perkembangan teknologi, kemampuan OCR juga semakin meningkat, memungkinkan ekstraksi data yang lebih kompleks dan beragam. Hal ini menjadikan OCR sebagai komponen penting dalam transformasi digital, terutama dalam industri yang bergantung pada data dokumen fisik.

Sebagai salah satu perusahaan *Software-as-a-Service* (SaaS) yang menyediakan berbagai layanan digital terbesar di Indonesia, PT Mid Solusi Nusantara (Mekari) atau yang dikenal dengan Mekari menyadari bahwa penting memanfaatkan teknologi OCR dalam memproses dokumen-dokumen bisnis. Perusahaan bergantung pada dokumen seperti faktur (*invoice*) dan tanda terima untuk kegiatan operasional sehari-hari, namun proses manual pengelolaan dokumen ini akan menguras waktu dan sumber daya. Melihat peluang ini, Mekari ingin memanfaatkan teknologi OCR dan berbagai pendekatan cerdas lainnya untuk membantu bisnis

dalam memproses dokumen dengan lebih cepat dan akurat, sehingga memungkinkan efisiensi operasional yang lebih baik.

Kesempatan untuk magang di Mekari sebagai *Machine Learning Intern* memberikan peluang berharga untuk memahami lebih dalam bagaimana proses bisnis dan inovasi produk dikembangkan. Melalui penggunaan teknologi terkini, Mekari ingin dapat benar-benar membantu dan menyelesaikan masalah yang dihadapi oleh pelanggan. Fokus utama dari setiap inovasi produk adalah bagaimana produk tersebut dapat mengatasi *pain-point* yang dirasakan oleh pengguna, seperti masalah dalam memproses dokumen secara manual yang lambat dan tidak efisien. Melalui pendekatan ini, perusahaan dapat memastikan bahwa produk yang dikembangkan tidak hanya relevan, tetapi juga memberikan nilai tambah bagi pelanggan.

Dalam peran sebagai *Machine Learning Intern* dalam proyek memiliki tanggung jawab utama sehari-hari adalah melakukan anotasi data gambar untuk memperkuat proses pembelajaran model yang digunakan dalam produk tersebut. Anotasi ini melibatkan penandaan label-label penting dalam dokumen untuk membantu model mengenali dan mengekstraksi informasi dengan lebih akurat. Selain itu, inisiatif juga diambil untuk mengembangkan model ML melalui pendekatan OCR, *Named Entity Recognition* (NER), dan *Large Language Model* (LLM) seperti GPT guna menguji apakah hasil ekstraksi dapat lebih fleksibel dibandingkan dengan model yang saat ini digunakan. Uji coba ini didasarkan pada pengetahuan yang diperoleh dari perkuliahan maupun pengalaman diluar perkuliahan, dengan tujuan untuk berkontribusi dalam peningkatan performa model secara keseluruhan.

Pengalaman magang di Mekari tidak hanya memberikan pemahaman yang lebih mendalam tentang dunia teknologi, tetapi juga memberikan wawasan mengenai permasalahan nyata (*real-world scenarios*) yang dihadapi oleh bisnis. Tantangan dalam memproses data tidak terstruktur menjadi salah satu contoh dari masalah yang dihadapi perusahaan-perusahaan saat ini. Selain itu, magang ini juga menyoroti betapa komunikasi yang efektif sangat penting dalam sebuah tim untuk

dapat mengembangkan produk yang maksimal dan memuaskan. Koordinasi yang baik dalam tim adalah kunci utama dalam memastikan setiap aspek dari produk yang dikembangkan mampu memenuhi kebutuhan pelanggan dengan baik.

1.2. Maksud dan Tujuan Kerja Magang

Program magang merupakan salah satu opsi kegiatan yang dapat dipilih oleh mahasiswa semester 7 di program studi Sistem Informasi (SI) Universitas Multimedia Nusantara (UMN) sebagai bagian dari pemenuhan persyaratan kelulusan. Kegiatan ini dirancang untuk memberikan pengalaman kerja nyata yang relevan dengan disiplin ilmu yang dipelajari. Pelaksanaan magang harus dilakukan dengan total minimal 640 jam kerja agar dapat dikonversikan menjadi 20 SKS. Pada periode ini, kegiatan magang dilaksanakan di Mekari, perusahaan di bidang teknologi dan solusi bisnis digital, dengan fokus pada produk yang memanfaatkan ML/AI.

Adapun maksud dari pelaksanaan magang di semester ini adalah sebagai berikut:

1. Mengaplikasikan pengetahuan dan keterampilan yang telah diperoleh selama masa perkuliahan, terutama di bidang ML/AI dan *Data Analytics* ke dalam dunia kerja nyata.
2. Memperoleh wawasan baru, baik dari segi *hard skill*, seperti kemampuan teknis dalam pengembangan model ML dan penerapan teknologi OCR, maupun *soft skill*, seperti kemampuan bekerja dalam tim, berpikir kritis, serta berkomunikasi secara efektif dalam lingkungan kerja profesional.

Tujuan pelaksanaan magang di Mekari sebagai *Machine Learning Intern*, antara lain:

1. Mendukung pengembangan salah satu proyek di Mekari dengan melakukan anotasi data gambar guna memperkuat kemampuan model dalam mengenali dan mengekstraksi informasi dari dokumen keuangan seperti faktur (*invoice*) dan tanda terima (*receipt*).

2. Berkontribusi dalam berbagai uji coba dan eksperimen pengembangan model dengan pemanfaatan OCR, NER, dan LLM, dengan tujuan untuk meningkatkan fleksibilitas terhadap tata letak dokumen yang ingin diproses.
3. Memperoleh pengalaman langsung dalam proses pengembangan produk teknologi yang berbasis ML, serta memahami bagaimana teknologi tersebut dapat menyelesaikan permasalahan nyata yang dihadapi oleh pengguna produk.
4. Meningkatkan kemampuan dalam hal *problem solving*, berkoordinasi dengan tim, dan beradaptasi dengan kebutuhan bisnis, melalui kolaborasi dengan karyawan Mekari dalam proses inovasi dan pengembangan produk.

1.3. Waktu dan Prosedur Pelaksanaan Kerja Magang

Pelaksanaan magang di Mekari sebagai *Machine Learning Intern* dapat dijabarkan menjadi 2 poin utama, yaitu:

1.3.1. Waktu Pelaksanaan Kerja Magang

Mekari menerapkan sistem kerja *hybrid*, di mana karyawan maupun peserta magang diharuskan untuk masuk ke kantor 1 hingga 2 kali dalam seminggu atau 15 kali ke kantor dalam satu kuartar, sedangkan hari kerja lain dapat dilakukan dari rumah (*work from home*). Penerapan sistem ini didukung oleh berbagai sarana teknologi yang memungkinkan koordinasi dan kolaborasi dilakukan secara efektif, meskipun bekerja dari jarak jauh. Salah satu alat yang digunakan untuk berkomunikasi dan berkolaborasi antar kolega adalah Google Workspace, yang memungkinkan diskusi, pembagian dokumen, dan rapat dilakukan secara online. Selain itu, Mekari menggunakan platform Talenta untuk keperluan absensi dan pelacakan kehadiran karyawan dan peserta magang. Adapun kegiatan magang di kantor Mekari berlokasi di Mid Plaza 2, Jl. Jenderal Sudirman No.4, Kel. Karet Tengsin, Kec. Tanah Abang, Kota Jakarta Pusat, Jakarta 10220.

Pelaksanaan magang dimulai pada tanggal 29 Juli 2024 dan berlangsung selama periode yang telah ditentukan sesuai kontrak. Jam kerja mengikuti

jadwal umum perusahaan, yaitu dari Senin hingga Jumat, berlaku *flexible hours* dengan waktu kerja yang disarankan adalah pukul 09.00 hingga 18.00 (8 jam + 1 jam). *Flexible working hour* paling telat dimulai pada pukul 10.00 dan waktu istirahat dimulai pada pukul 12.00 hingga 13.00 yang berikan waktu yang cukup untuk istirahat makan. Jadwal ini memungkinkan fleksibilitas dalam menyelesaikan tugas, baik di kantor maupun dari rumah, sesuai dengan kebijakan *hybrid* yang berlaku. Untuk hari libur dan cuti, perusahaan mengikuti kalender hari libur nasional dan cuti bersama yang ditetapkan oleh pemerintah Indonesia. Gambar 1.1 di bawah merupakan *ganttt chart* yang akan memberikan gambaran besar kegiatan pelaksanaan magang di Mekari.

| No | Deskripsi Pekerjaan | Juli | | | | Agustus | | | | September | | | | Oktober | | | |
|----|---|------|---|---|---|---------|---|---|---|-----------|---|---|---|---------|--|--|--|
| | | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | | | |
| 1 | Onboarding, pengenalan lingkungan kerja, penjelasan tugas, penempatan tim proyek. | | | | | | | | | | | | | | | | |
| 2 | Anotasi/pemberian label pada gambar dan dokumen melalui platform Document AI. | | | | | | | | | | | | | | | | |
| 3 | Uji coba pembuatan model menggunakan metode OCR, NER, dan LLM. | | | | | | | | | | | | | | | | |
| 4 | Melakukan POC bersama tim data untuk pemrosesan dokumen yang lebih <i>general</i> . | | | | | | | | | | | | | | | | |

Gambar 1. 1 *Timeline* Magang

1.3.2. Prosedur Pelaksanaan Kerja Magang

Prosedur pelaksanaan kerja magang dapat dibagi menjadi tiga tahapan, yaitu pra-magang, saat magang, dan pasca-magang. Berikut adalah penjabaran terhadap masing-masing tahapan:

1.3.2.1. Pra-magang

Tahapan persiapan magang dimulai pada pertengahan bulan Mei 2024, dengan pembaruan *Curriculum Vitae* (CV) dan portofolio di berbagai platform seperti LinkedIn dan GitHub. Setelah itu, dilakukan pencarian

peluang magang melalui berbagai portal pekerjaan dan situs web resmi perusahaan yang membuka lowongan. Pada bulan Juni, setelah meninjau situs karier Mekari, terbuka kesempatan magang sebagai *Machine Learning Intern*, dan proses aplikasi dilakukan. Sekitar satu hingga dua minggu kemudian, Mekari menghubungi untuk mengikuti tahapan rekrutmen yang meliputi psikotes dan wawancara. Setelah berhasil melalui tahapan seleksi, peserta magang diundang untuk bergabung dalam WhatsApp *Group* (WAG) khusus *new joiners* dan diminta untuk mengambil perangkat kerja sehari sebelum *onboarding* dimulai. Selain itu, sebagai bukti diterima sebagai peserta magang, diperlukan *Letter of Acceptance* (LoA) untuk diserahkan kepada dosen koordinator magang agar disetujui. Setelah LoA sesuai dan telah disetujui, mahasiswa mendapatkan akses ke halaman web merdeka.umn.ac.id untuk melakukan registrasi hingga proses input *daily task*. Dengan demikian, tahapan pra-magang ini memastikan bahwa seluruh persyaratan administratif dan teknis telah terpenuhi sebelum pelaksanaan magang dimulai. Hal ini juga menjadi langkah penting dalam mempersiapkan mahasiswa untuk memasuki lingkungan kerja yang lebih profesional dan terstruktur.

1.3.2.2. Saat Magang

Pelaksanaan magang dimulai pada tanggal 29 Juli 2024, dengan penempatan di bawah *Business and Finance Manager—Innovation* yang menangani salah satu proyek di Mekari untuk proses ekstraksi teks dari gambar atau dokumen seperti faktur (*invoice*) dan *receipt*. Adapun jam dan sistem kerja yang berlaku selama pelaksanaan magang dilakukan sesuai dengan yang telah dijelaskan sebelumnya. Penjelasan lebih detail terkait sistem kerja ini disampaikan pada awal masuk kantor pada hari *onboarding*. Selama *onboarding*, anggota baru mendapatkan gambaran umum mengenai berbagai aspek penting di kantor, mulai dari sejarah perusahaan, struktur organisasi, nilai-nilai yang dijunjung, hingga benefit yang diterima sebagai karyawan.

Selama kegiatan magang, tugas harian meliputi melakukan anotasi gambar menggunakan platform Google Cloud Platform—Document AI berdasarkan skema atau label yang telah ditentukan. Selain itu, sebagai inisiatif pribadi, serangkaian uji coba dilakukan untuk mengembangkan model ML yang mengombinasikan OCR, NER, dan LLM. Tujuan dari uji coba ini adalah untuk meningkatkan fleksibilitas dalam pemrosesan dokumen bisnis, mengingat variasi tata letak (*layout*) yang dimiliki oleh dokumen seperti faktur (*invoice*) dan tanda terima seperti *receipt*.

Koordinasi dengan tim dilakukan secara *hybrid*, dengan pertemuan tatap muka di kantor satu hingga dua kali per minggu, dan juga dilakukan secara *remote* menggunakan Google Workspace untuk rapat dan kolaborasi. Setiap peserta magang diwajibkan melakukan absensi harian melalui platform Talenta. Mekari juga menyediakan berbagai alat-alat produktivitas seperti software Jira yang membantu proses kerja dan manajemen proyek, memastikan bahwa tugas-tugas dapat diselesaikan secara efisien, baik saat bekerja dari kantor maupun dari rumah.

1.3.2.3. Pasca-Magang

Setelah kegiatan magang selesai, tahapan berikutnya adalah finalisasi penyusunan laporan magang berdasarkan hasil kerja magang yang telah dilakukan. Pada periode semester 7 ini, bimbingan laporan magang dilakukan bersama Dr. Santo Fernandi S.Kom., M.Kom., sebagai dosen pembimbing. Proses bimbingan dilakukan dengan minimal delapan kali, mengikuti prosedur yang ditetapkan oleh Universitas Multimedia Nusantara (UMN), melalui sistem *progress bar* yang harus dipenuhi di situs web merdeka.umn.ac.id. Sidang magang dijadwalkan berlangsung pada bulan Januari 2025 sebagai tahapan akhir untuk menyelesaikan seluruh proses magang.