

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Magang merupakan suatu kegiatan / program yang tercantum dalam kurikulum Politeknik Negeri Jember yang dilaksanakan pada semester VII untuk program D-IV, kegiatan magang termasuk salah satu syarat kelulusan mahasiswa Politeknik Negeri Jember.

PT Universal Big Data (UBIG) adalah perusahaan berbasis teknologi yang bergerak di bidang pengembangan perangkat lunak, analisis big data, dan kecerdasan buatan (AI). Mereka menawarkan solusi seperti **Nobox.AI** untuk pemasaran dan layanan pelanggan berbasis AI, menyediakan data lintas kategori untuk pengguna global, serta mendukung startup melalui inkubasi dan konsultasi manajemen. Selain itu, UBIG aktif dalam kemitraan akademik dengan menyelenggarakan pelatihan dan magang berbasis proyek untuk mempersiapkan tenaga kerja terampil yang siap masuk ke dunia industri .(“Soft Launching Nobox.AI – PT. Universal Big Data (UBIG) | event.mcc.or.id,” n.d.)

Dalam kegiatan magang pihak yang bersangkutan memberikan penulis sebuah tugas dengan melanjutkan pengembangan Web Scraping yang terintegrasi dengan **Nobox.AI** berupa sebuah aplikasi desktop yang bertujuan mempermudah pengguna dalam pengambilan data dalam jumlah besar. Data yang terambil dapat digunakan untuk berbagai keperluan, seperti analisis bisnis, riset pasar, pemantauan harga, atau pengembangan aplikasi. data yang tersedia secara publik di situs web dapat diubah menjadi format terstruktur seperti database atau file spreadsheet, sehingga lebih mudah dianalisis dan digunakan untuk pengambilan keputusan atau pengembangan solusi berbasis data.

Dalam web scraping, berbagai fungsi memiliki peran penting untuk memastikan pengambilan data berjalan efektif dan sesuai dengan kebutuhan. Get List digunakan untuk mengambil daftar elemen atau item dari halaman utama yang relevan dengan data yang diinginkan. Sementara itu, Get Detail berfungsi untuk menggali informasi lebih mendalam dari setiap elemen yang ditemukan dalam daftar tersebut. Pagination memungkinkan navigasi antar halaman untuk memastikan semua data yang tersebar dapat diakses. Fitur Search membantu dalam memfilter hasil berdasarkan kata kunci atau parameter tertentu. Dalam konteks pengambilan data kontak, fungsi seperti No hp format WhatsApp digunakan untuk memformat nomor HP agar sesuai dengan standar WhatsApp. Apabila tidak memerlukan

login, kode harus dapat menonaktifkan kolom pengguna dan memeriksa jika input null melalui kode validasi. Fungsi Phone Format dan Filter phone sesuai negara memastikan nomor HP diformat dengan benar dan disaring berdasarkan lokasi negara yang ditentukan. Selanjutnya, fitur seperti Cek no result found memastikan skrip berhenti atau memberikan logika berdasarkan tidak adanya hasil. Terakhir, Cek semua kategori memastikan semua kategori dari sumber data telah terakses melalui selector yang ditentukan. Semua operasi ini bergantung pada pemilihan selector dari web, yang berfungsi sebagai penunjuk elemen yang relevan dalam struktur HTML.

1.2 Tujuan Dan Manfaat

Dalam penulisan laporan akhir magang ini 2 (dua) tujuan, yaitu tujuan umum dan tujuan khusus yang dijelaskan sebagai berikut:

1.2.1 Tujuan Utama Magang

Tujuan umum kegiatan magang ini antara lain

1. mengintegrasikan teori yang diperoleh selama pendidikan formal dengan praktik langsung di lingkungan profesional, sehingga memberikan pengalaman kerja yang nyata kepada peserta.
2. Mengembangkan keterampilan dan kompetensi melalui pengalaman kerja di perusahaan atau instansi, peserta dapat mengembangkan keterampilan teknis dan non-teknis, seperti komunikasi, pemecahan masalah, kerja tim, serta kemampuan adaptasi dalam berbagai situasi kerja.
3. membangun koneksi dengan profesional dan pihak yang berpengalaman di industri yang relevan. Jaringan ini bisa menjadi peluang untuk mendapatkan referensi atau pekerjaan di masa depan.
4. memiliki kesempatan untuk memahami seluk-beluk dunia profesional dan membangun pengalaman kerja yang meningkatkan daya saing mereka dalam mencari pekerjaan di masa mendatang.

1.2.2 Tujuan Khusus Magang

Ada beberapa tujuan khusus dari magang antara lain:

1. Menguasai teknik pengembangan web scraping dengan bahasa C# dan framework pendukungnya seperti Selenium atau WebView.
2. Memahami pengelolaan data hasil scraping, termasuk filtering, penyimpanan, dan pemrosesan.
3. Mengembangkan solusi scraping yang efisien dan sesuai dengan kebijakan hukum dan etika.
4. Meningkatkan kemampuan debugging dan pengoptimalan kode untuk menangani berbagai struktur situs web.
5. Memahami integrasi hasil scraping dengan sistem atau database yang relevan.

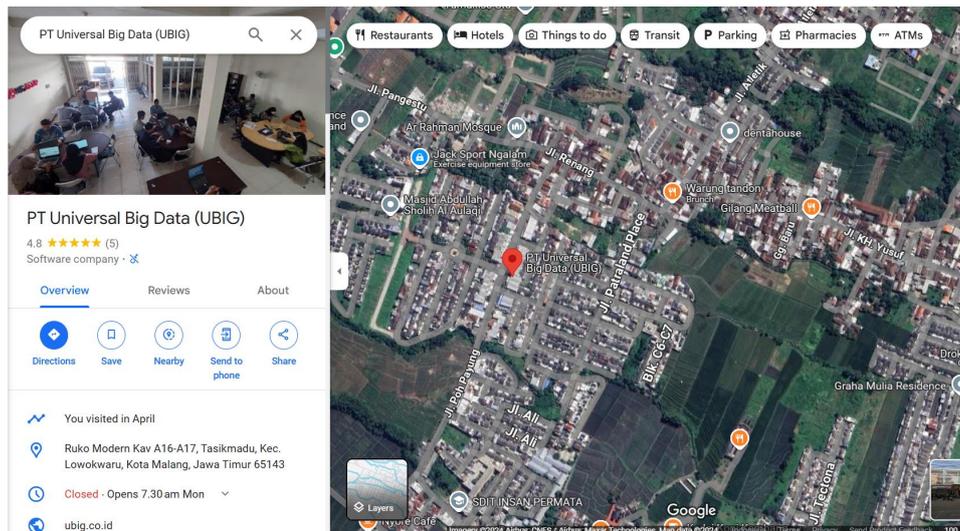
1.2.3 Manfaat Magang

Manfaat dari magang ialah sebagai berikut:

- 1) Memberikan kesempatan untuk menerapkan pengetahuan teoritis dalam situasi kerja nyata.
- 2) Meningkatkan keterampilan teknis, komunikasi, dan manajemen waktu melalui pekerjaan langsung.
- 3) Membantu memahami budaya kerja, etika profesional, dan proses bisnis di industri.
- 4) Memperluas jaringan profesional dengan rekan kerja, mentor, dan komunitas industri.
- 5) Memberikan nilai tambah pada CV dengan pengalaman kerja yang relevan.
- 6) Membantu mengevaluasi minat dan kesesuaian dengan bidang pekerjaan yang dipilih.
- 7) Membuka peluang untuk direkrut sebagai karyawan tetap jika kinerja memuaskan.

1.3 Lokasi Dan Waktu

Kegiatan magang dimulai tanggal 1 juli 2024 sampai dengan 31 desember 2024. Magang ini bertempat di PT. Universal Big Data Ruko Modern Kav A16-A17, Tasikmadu, Kec. Lowokwaru, Kota Malang, Jawa Timur 65143.



Gambar 1. 1 Peta Lokasi Perusahaan

1.4 Metode Pelaksanaan

Metode yang digunakan dalam pelaksanaan magang ini adalah sebagai berikut:

a. Observasi

Secara umum, metode observasi dapat didefinisikan sebagai proses sistematis untuk mengumpulkan informasi dengan cara mengamati suatu objek atau sistem secara langsung tanpa memengaruhi kondisinya. Metode ini melibatkan pengamatan langsung terhadap struktur, elemen, dan perilaku situs web yang akan diambil datanya. Dalam konteks ini, observasi dilakukan untuk memahami bagaimana data diatur pada halaman web, seperti lokasi elemen HTML, penggunaan atribut seperti kelas atau ID, serta pola yang muncul dalam struktur data. Tujuan utama metode observasi adalah untuk mengidentifikasi pola atau elemen yang dapat digunakan sebagai referensi dalam membangun kode atau skrip scraping. Misalnya, pengembang dapat mengamati apakah elemen data seperti judul, harga, atau deskripsi produk memiliki atribut unik atau berulang di seluruh halaman, sehingga dapat dibuat selector yang akurat.

b. Metode Studi Pustaka

Metode studi pustaka adalah pendekatan yang digunakan untuk mengumpulkan informasi atau data melalui kajian terhadap literatur, dokumen, buku, jurnal, artikel ilmiah, maupun sumber digital yang relevan dengan topik tertentu. Dalam studi pustaka, peneliti memanfaatkan referensi teoretis maupun hasil penelitian terdahulu untuk memahami konsep dasar, mengidentifikasi masalah, atau menemukan solusi yang mendukung

penelitian atau proyek yang sedang dilakukan. Metode ini tidak hanya membantu memperkuat argumen, tetapi juga memastikan bahwa penelitian atau laporan yang dibuat memiliki landasan akademis yang kuat dan relevan.

Dalam konteks laporan magang, penerapan metode studi pustaka dapat digunakan untuk mendukung analisis atau pembahasan terkait tugas yang dijalankan selama magang. Misalnya, jika magang dilakukan di bidang pengembangan web, studi pustaka dapat digunakan untuk memahami teknologi yang digunakan, seperti framework atau bahasa pemrograman tertentu. Selain itu, metode ini juga bisa membantu mempelajari praktik terbaik (best practices) dalam implementasi sistem yang relevan. Dalam laporan, penerapan studi pustaka biasanya terlihat pada bagian pendahuluan, tinjauan pustaka, dan pembahasan, di mana referensi digunakan untuk menjelaskan latar belakang, memperkuat argumentasi, atau menghubungkan hasil magang dengan teori yang telah ada. Hal ini menunjukkan bahwa laporan didasarkan pada kombinasi antara pengalaman praktis dan kajian teoritis, sehingga lebih komprehensif dan bernilai akademis.

c. Implementasi

Metode implementasi dalam pengembangan web scraping menggunakan XPath melibatkan proses identifikasi, seleksi, dan ekstraksi elemen-elemen tertentu dari halaman web berdasarkan path XML atau HTML. (Rizaldi and Putranto, 2017) XPath (XML Path Language) adalah bahasa query yang digunakan untuk menavigasi struktur dokumen HTML atau XML, memungkinkan pengembang untuk menargetkan elemen berdasarkan hierarki, atribut, atau teksnya. Implementasi ini dimulai dengan inspeksi elemen di browser untuk menentukan pola atau struktur elemen yang mengandung data yang diinginkan, seperti tabel, div, atau tag khusus. (Prastyo et al., n.d.)

Dalam praktiknya, XPath digunakan untuk membuat selector yang presisi, seperti memilih elemen berdasarkan ID (`//div[@id='example']`), kelas (`//div[@class='item']`), atau hierarki elemen (`//ul/li[1]/a`). Setelah XPath yang tepat ditemukan, skrip scraping dikembangkan menggunakan pustaka seperti Selenium, BeautifulSoup, atau lxml untuk mengeksekusi query XPath dan mengekstrak data. Metode ini sangat bermanfaat untuk menangani struktur halaman yang kompleks dan memastikan hanya data yang relevan yang diambil.