

SKRIPSI
IMPLEMENTASI WEB SCRAPING PADA
MEDIA SOSIAL INSTAGRAM



Disusun Oleh:

N a m a : Rio Baskara

NIM : 18523254

PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA

2022

HALAMAN PENGESAHAN DOSEN PEMBIMBING

IMPLEMENTASI WEB SCRAPING PADA

MEDIA SOSIAL INSTAGRAM

TUGAS AKHIR



HALAMAN PENGESAHAN DOSEN PENGUJI

IMPLEMENTASI WEB SCRAPING PADA

MEDIA SOSIAL INSTAGRAM

TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 27 Mei 2022

Tim Penguji

Fayruz Rahma, S.T., M.Eng.

Anggota 1

Andhik Budi Cahyono, S.T., M.T.

Anggota 2

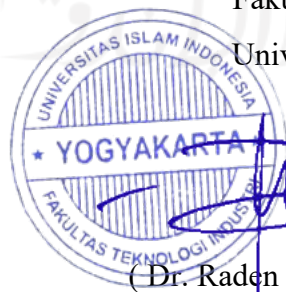
Moh. Idris, S.Kom., M.Kom.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia



(Dr. Raden Teduh Dirgahayu, S.T., M.Sc.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Rio Baskara

NIM : 18523254

Tugas akhir dengan judul:

IMPLEMENTASI WEB SCRAPING PADA MEDIA SOSIAL INSTAGRAM

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 28 April 2022



(Rio Baskara)

HALAMAN PERSEMBAHAN

Bismillahirrahmanirrahim

Dengan mengucap rasa syukur kepada Allah SWT, maha pengasih dan maha penyayang, Alhamdulillahilahi robbil'alamin, saya persembahkan laporan Tugas Akhir saya kepada kedua orang tua saya tercinta, mama dan papa, tanpa dukungan dan doa-doanya lah saya berada di posisi seperti sekarang ini.



HALAMAN MOTO

وَأَحْسِنُ كَمَا أَحْسَنَ اللَّهُ إِلَيْكَ

“dan berbuat baiklah (kepada orang lain) sebagaimana Allah telah berbuat baik,
kepadamu,”

(QS Al-Qashash : 77)



KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa karena atas rahmat dan anugerah-Nya, penulis dapat menyelesaikan skripsi ini dengan judul “Implementasi Web Scraping Pada Media Sosial Instagram” yang merupakan salah satu persyaratan untuk memperoleh derajat Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia.

Dalam penyusunan skripsi ini, penulis sangat menyadari bahwa penyusunan skripsi ini tidak akan selesai tanpa adanya doa, bimbingan, dorongan dan bantuan dari berbagai pihak, oleh karena itu pada kesempatan kali ini penulis ingin mengucapkan rasa terima kasih kepada:

1. Allah SWT Yang Maha Besar, Maha Baik, dan Maha Sempurna, hanya karena ridha dan izin-Nya segala proses penyusunan skripsi ini dapat berjalan hingga tahap ini.
2. Mama tercinta Nadia Irene B.BA., dan Papa tercinta Ir. Eddy Anwar yang senantiasa mendukung baik secara moral dan finansial, memberi semangat, cinta dan kasih sayang, serta doa sehingga penulis dapat menyelesaikan skripsi ini.
3. Kakak-kakak tercinta Indra Ardy Pratama, S.E., Tiara Ardiani, S.Psi., Indah Putri, Dicky Dwicahyo, S.Pd., adik tersayang Dhanapatih Pradana dan kemenakan tercinta Kalandra Ahmad Adhyaksa dan Kaffi Ahmad Alvarendra yang telah memberikan semangat, kasih sayang, dan doa sehingga penulis dapat menyelesaikan skripsi ini.
4. Prof. Dr. Ir. Hari Purnomo, M.T., IPU selaku dekan FTI UII yang telah menyediakan sarana untuk menyelesaikan pendidikan S1.
5. Dr. Raden Teduh Dirgahayu, S.T., M.Sc. selaku Kaprodi Informatika FTI UII – Program Sarjana yang telah menyediakan sarana untuk menyelesaikan pendidikan S1.
6. Ibu Fayruz Rahma, S.T., M.Eng selaku dosen pembimbing skripsi yang selalu meluangkan waktu, memberikan arahan dan semangat serta selalu mendampingi selama penyelesaian skripsi ini.
7. Azkia Raihani S.KG. yang telah sabar menemani, memberi semangat, dukungan, kasih sayang, serta doa sehingga penulis dapat menyelesaikan skripsi ini.
8. Sahabat terdekat penulis, Ahmad Irfani yang senantiasa menemani, memberikan semangat, dukungan dan doa kepada penulis untuk dapat menyelesaikan skripsi ini.

9. Sahabat terdekat penulis, Fauzi Dahlan dan Alfiansyah Indra yang telah memberikan semangat, dukungan, dan doa kepada penulis untuk dapat menyelesaikan skripsi ini.
10. Sahabat terdekat penulis, Akhmad Rifki Febrianto, Hanif, dan Ricky Nagata yang telah memberikan semangat, dukungan, dan doa kepada penulis untuk dapat menyelesaikan skripsi ini.
11. Keluarga HMIF UII yang sudah menjadi media terbaik penulis dalam berkembang dan meningkatkan potensi diri.
12. Seluruh staf FTI UII baik dari akademis maupun administratif yang sudah memberi banyak bantuan bagi penulis selama proses penyusunan skripsi.
13. Kepada seluruh pihak yang tidak dapat disebutkan satu persatu yang telah membantu kelancaran penyelesaian penyusunan skripsi ini.

Penulis berharap semoga Tuhan Yang Maha Esa akan membalas segala kebaikan semua pihak yang telah membantu. Akhir kata, penulis menyadari bahwa penelitian ini masih jauh dari sempurna, karena itu penulis mengharapkan saran untuk menyempurnakan penelitian ini. Penulis berharap agar hasil penelitian ini dapat memberikan informasi dan berguna bagi negara, masyarakat maupun bagi perkembangan ilmu informatika.

Yogyakarta, 28 April 2022



(Rio Baskara)

SARI

Dinas Komunikasi dan Informatika Daerah Istimewa Yogyakarta (Diskominfo DIY) memiliki sebuah sistem analitik berbasis *big data*, yang difokuskan pada pengembangan data analitik dan pendukung pengambilan keputusan, serta merujuk pada dimensi-dimensi *Jogja Smart Province* (JSP) yang diberi nama Jogja Center. Penelitian ini bertujuan untuk memenuhi kebutuhan data Jogja Center yang menjadi dasar kebutuhan analisis. Pengumpulan data yang akan digunakan menggunakan teknik *web scraping* dengan bahasa pemrograman *Python*. Metode yang akan digunakan terdiri dari *analysis*, *coding*, dan *testing*. Hasil dari penelitian ini berupa data dalam bentuk format dokumen *.CSV* yang telah dibersihkan. Dapat disimpulkan bahwa dengan menggunakan teknik *web scraping* dapat memudahkan pengguna dalam mengumpulkan data secara cepat.

Kata kunci: *Web Scraping*, *Python*.

GLOSARIUM

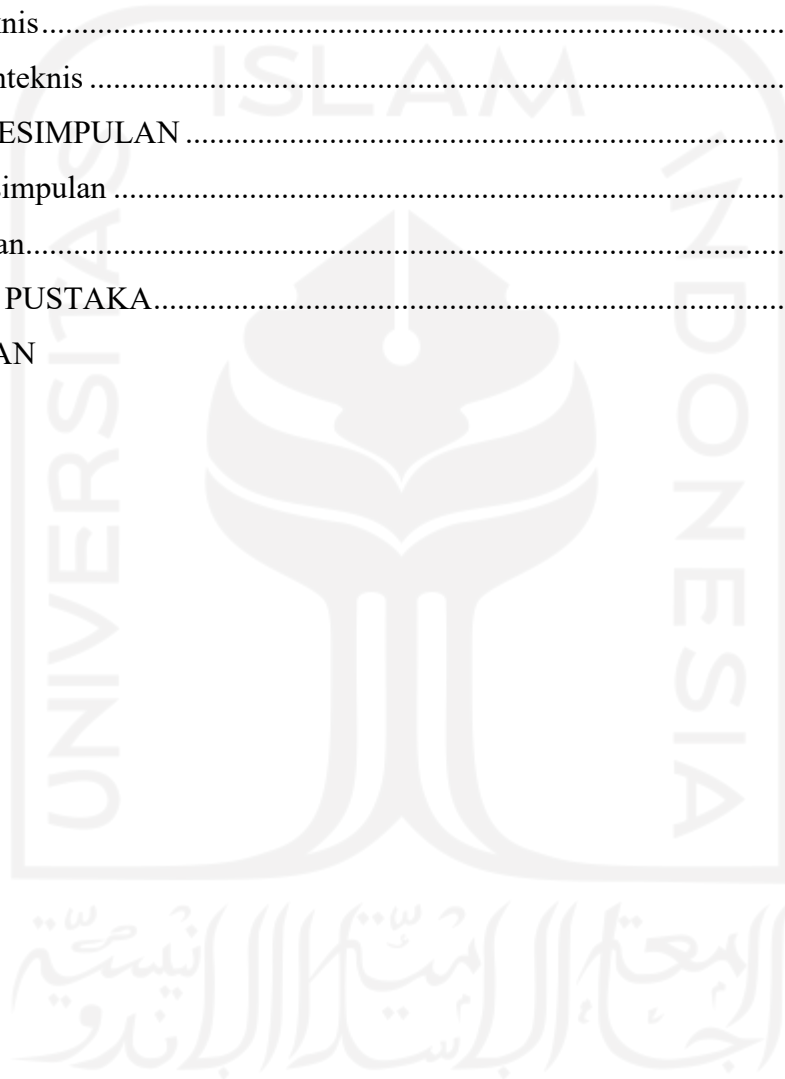
<i>Mounting</i>	proses penghubungan antara kode pemrograman dengan satu atau lebih dokumen yang berada di luar jaringan.
<i>Python</i>	bahasa pemrograman yang diciptakan oleh Guido Van Rosum pada awal tahun 1990 untuk keperluan umum di CWI.
<i>Sentimen</i>	pendapat atau pandangan yang didasarkan pada perasaan yang berlebih-lebihan.
<i>Spreadsheet</i>	aplikasi atau program komputer yang digunakan untuk memanipulasi, menangkap, dan menampilkan data yang disusun secara sistematis.
<i>Tweet</i>	Sebuah pesan yang diposting ke Twitter berisi teks, foto, GIF, dan/atau video.
<i>Waterfall</i>	metode pengembangan perangkat lunak yang urutan pengerjaannya dimulai dari proses perencanaan, analisa, desain, dan implementasi pada sistem.



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN DOSEN PEMBIMBING	ii
HALAMAN PENGESAHAN DOSEN PENGUJI	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTO	vi
KATA PENGANTAR	vii
SARI	ix
GLOSARIUM	x
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Ruang Lingkup Magang	2
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Sistematika Penulisan	3
BAB II DASAR TEORI	5
2.1 Web Scraping	5
2.1.1 Jenis-Jenis <i>Web Scraping</i>	6
2.1.2 Manfaat Web Scraping	6
2.1.3 Kendala Web Scraping	6
2.2 Extract, Transform, Load	7
2.3 Data Cleaning	7
2.4 Python	8
BAB III PELAKSANAAN MAGANG	10
3.1 Manajemen Aktivitas Magang	10
3.1.1 Pendataan Tempat Wisata di Daerah Istimewa Yogyakarta	10
3.1.2 Pendataan dan Rekapitulasi Direktori Putusan Pengadilan	15
3.1.3 Visualisasi Perbandingan Data Covid-19 dan Vaksinasi Dengan Data Mobilitas	20

3.1.4	Analisis Sentimen Publik Berdasarkan Kata Kunci.....	24
3.2	<i>Scraping</i> Media Sosial Instagram	30
3.2.1	Perencanaan Proyek	30
3.2.2	Pelaksanaan Proyek.....	31
3.2.3	Pemantauan dan Pengendalian Proyek	35
3.2.4	Penutupan Proyek	36
BAB IV REFLEKSI PELAKSANAAN MAGANG		37
4.1	Teknis.....	37
4.2	Nonteknis	38
BAB V KESIMPULAN		39
5.1	Kesimpulan	39
5.2	Saran.....	39
DAFTAR PUSTAKA.....		41
LAMPIRAN		



DAFTAR TABEL

Tabel 3.1 Pembagian waktu dalam pelaksanaan proyek pendataan tempat wisata	10
Tabel 3.2 Pembagian waktu dalam pelaksanaan proyek pendataan direktori	16
Tabel 3.3 Pembagian waktu dalam pelaksanaan proyek visualisasi data	20
Tabel 3.4 Pembagian waktu dalam pelaksanaan proyek analisis sentimen publik.....	25
Tabel 3.5 Pembagian waktu dalam pelaksanaan proyek scraping media sosial instagram	30



DAFTAR GAMBAR

Gambar 1.1 Kantor DISKOMINFO DIY	2
Gambar 2.1 Guido Van Rossum.	8
Gambar 3.1 Hasil pembangian kategori tugas.	11
Gambar 3.2 Pencarian menggunakan google.....	12
Gambar 3.3 Pengambilan titik koordinat menggunakan google maps	13
Gambar 3.4 Pengecekan akun media sosial dan tagar menggunakan Instagram.....	14
Gambar 3.5 Hasil pendataan yang dikumpulkan pada spreadsheet.....	14
Gambar 3.6 Kumpulan dokumen direktori putusan pengadilan	16
Gambar 3.7 <i>Library</i> yang digunakan	17
Gambar 3.8 Potongan kode untuk melakukan <i>mouting</i>	17
Gambar 3.9 Potongan kode ekstraksi data	18
Gambar 3.10 Hasil rekapitulasi direktori.....	19
Gambar 3.11 Data vaksinasi DIY	21
Gambar 3.12 Tampilan penggunaan <i>datawrapper</i>	22
Gambar 3.13 Hasil visualisasi data mobilitas dengan data Covid-19.....	22
Gambar 3.14 Hasil visualisasi data Covid-19 dengan data vaksinasi.....	23
Gambar 3.15 Potongan kode membaca dokumen yang akan digunakan.....	25
Gambar 3.16 Potongan kode perhitungan jenis <i>tweet</i>	26
Gambar 3.17 Potongan kode perhitungan jumlah <i>tweet</i> per-jam.....	27
Gambar 3.18 Potongan kode perhitungan jenis sentimen.....	27
Gambar 3.19 Tampilan proses analisis menggunakan spreadsheet.....	28
Gambar 3.20 Tampilan salah satu slide presentasi yang telah diberikan tweet pendukung	29
Gambar 3.21 Penyimpanan <i>username</i> dan <i>password</i> pada variable.....	32
Gambar 3.22 Potongan kode scraping dan penyimpanan format json.....	33
Gambar 3.23 Proses <i>running</i> menggunakan terminal.....	33
Gambar 3.24 Potongan kode mengubah json menjadi CSV	34
Gambar 3.25 Hasil scraping.....	35

BAB I

PENDAHULUAN

1.1. Latar Belakang

Dinas Komunikasi dan Informatika Daerah Istimewa Yogyakarta merupakan sebuah instansi pemerintah yang memiliki tugas membantu Gubernur melaksanakan urusan pemerintahan bidang komunikasi dan informatika dan urusan pemerintahan bidang persandian. Dinas Komunikasi dan Informatika Daerah Istimewa Yogyakarta memiliki beberapa proyek yang dikembangkan salah satunya adalah Jogja Center, sebuah sistem analitik berbasis *big data*, yang difokuskan pada pengembangan data analitik dan pendukung pengambilan keputusan, serta merujuk pada dimensi-dimensi *Jogja Smart Province (JSP)*.

Sistem analitik yang dikembangkan memerlukan data yang perlu diperbarui secara berkala untuk memberikan hasil analisis yang lebih baik. Sebelum ditampilkan di halaman Jogja Center, data-data tersebut perlu dikumpulkan menggunakan cara yang disebut dengan *web scraping*. Hal tersebut dilakukan dengan membuat suatu program yang dapat mengambil data dari berbagai sumber, salah satunya dapat diimplementasikan pada media sosial instagram yang kemudian data tersebut dapat disimpan dalam bentuk format dokumen *Comma Separated Values* atau CSV. Data-data tersebut berupa koordinat lokasi, *caption*, *hashtag* atau tagar, dan lain-lain yang kemudian dapat dianalisis untuk menentukan sebaran mobilitas masyarakat maupun sentimen masyarakat yang kemudian ditampilkan.

Web scraping merupakan teknik yang digunakan untuk mengumpulkan data yang dapat dilakukan secara manual yaitu dengan melakukan *copy paste* data yang diinginkan maupun secara otomatis yaitu dengan membuat sebuah program atau *code* yang dapat melakukan proses pengambilan data dari sebuah halaman web. Dalam melakukan *web scraping* terdapat beberapa metode yang dapat dilakukan yaitu menyalin data secara manual, *parsing HTML*, *regular expression*, dan lain-lain. Tidak dipungkiri teknik *web scraping* memiliki kekurangan yaitu sampai saat ini belum ada teknik *web scraping* yang 100% efektif. Selain itu hasil yang didapatkan tidak selalu rapih, maka perlu juga untuk memahami struktur halaman *website* yang dituju, karena tidak semua data dapat diekstrak dengan mudah, sering kali program harus dijalankan berulang kali yang mengakibatkan akses terhadap halaman web tersebut terblokir.

Namun ada pula manfaat dari melakukan *web scraping* yaitu data-data yang didapatkan akan lebih terfokus yang dapat memudahkan dalam pencarian sesuatu.

Instagram merupakan sebuah platform media sosial di mana penggunanya dapat membagikan foto maupun video kepada pengguna lainnya. Selain itu penggunanya juga dapat menambahkan hal-hal lainnya seperti keterangan dan juga lokasi. Data-data tersebut dapat dikumpulkan menggunakan teknik *web scraping* dengan membuat program yang dapat melakukan hal tersebut secara otomatis.

Dalam proyek ini data yang dikumpulkan digunakan untuk keperluan analisis salah satunya pada sektor pariwisata yaitu untuk mengetahui *insight* mengenai tempat wisata. Pengumpulan data dilakukan menggunakan teknik *web scraping* pada media sosial instagram dikarenakan instagram merupakan salah satu platform media sosial yang paling banyak digunakan. Per bulan Oktober tahun 2021, di Indonesia terdapat lebih dari 90 juta pengguna (Annur, 2021). Dengan banyaknya pengguna aktif media sosial instagram di Indonesia dapat memberikan hasil data seperti tempat wisata baru, *update* mengenai kegiatan dengan tingkat pembaruan lebih tinggi dibandingkan dengan platform media sosial lainnya.

1.2. Ruang Lingkup Magang



Gambar 1.1 Kantor DISKOMINFO DIY

Pelaksanaan magang dilakukan di Dinas Komunikasi dan Informatika Daerah Istimewa Yogyakarta, ditampilkan pada Gambar 1.1. Magang dilaksanakan selama 6 bulan, yaitu dimulai pada bulan Maret 2021 dan berakhir pada bulan September 2021. Kegiatan magang dilakukan dengan tanggung jawab kerja untuk melakukan *web scraping* pada sosial media

untuk mengumpulkan data yang akan ditampilkan pada sistem. Selain itu juga terdapat penugasan kolaborasi yang dikerjakan Bersama peserta magang lainnya yaitu melakukan pendataan tempat wisata di Daerah Istimewa Yogyakarta, pendataan dan rekapitulasi direktori putusan pengadilan, visualisasi perbandingan data Covid-19 dan vaksinasi dengan data mobilitas, dan analisis sentiment public berdasarkan kata kunci.

Kegiatan magang dilakukan di dalam tim proyek Jogja Center yang terdiri dari beberapa bagian yaitu terdapat *Project Manager* sebagai posisi yang bertanggung jawab atas berjalannya proyek, *Data Engineer* sebagai posisi yang mengumpulkan data, *Data Analyst* sebagai posisi yang melakukan analisis dan melakukan *labeling*, dan *UI* sebagai posisi yang bertanggung jawab atas tampilan dari halaman web Jogja Center. Kegiatan magang dilaksanakan di bawah Bidang Keamanan Informasi dan Persandian.

1.3. Tujuan

Tujuan dari implementasi *web scraping* untuk mengumpulkan data pada media sosial instagram adalah untuk memenuhi kebutuhan data yang menjadi dasar dari hasil analisis. Setiap hasil dari pengumpulan data yang telah dianalisis akan ditampilkan pada halaman Jogja Center.

1.4. Manfaat

Manfaat dari implementasi *web scraping* untuk mengumpulkan data pada media sosial instagram adalah terpenuhinya kebutuhan data yang menjadi dasar dari kebutuhan analisis yang kemudian akan ditampilkan pada halaman Jogja Center.

1.5. Sistematika Penulisan

Untuk memahami lebih jelas laporan ini, materi-materi yang tertulis dikelompokkan menjadi beberapa subbab dengan sistematika penyampaian sebagai berikut:

SARI

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang, ruang lingkup magang, tujuan, manfaat dan sistematika penulisan.

BAB II DASAR TEORI

Bab ini berisi teori-teori yang berupa pengertian dan penjelasan yang diambil dari kutipan kajian literatur dan beberapa web yang berkaitan dengan penyusunan laporan.

BAB III PELAKSANAAN MAGANG

Bab ini berisi dokumentasi dari kegiatan dan tugas yang diberikan selama kegiatan magang berlangsung.

BAB IV REFLEKSI PELAKSANAAN MAGANG

Bab ini berisi penjelasan hasil dan manfaat yang didapatkan setelah melakukan kegiatan magang.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran dari hal-hal yang telah dijelaskan pada bab-bab sebelumnya.

DAFTAR PUSTAKA

LAMPIRAN



BAB II

DASAR TEORI

2.1. *Web Scraping*

Web scraping adalah salah satu metode mengumpulkan data yang digunakan untuk mengekstraksi data dari suatu halaman web. Halaman web dibangun menggunakan bahasa berbasis teks (HTML dan XHTML), sering kali berisi banyak data berguna dalam bentuk teks. Sebagian besar halaman web dirancang untuk kemudahan penggunaan oleh manusia, bukan untuk kemudahan penggunaan otomatis. Akibatnya, alat dan perangkat lunak khusus telah dikembangkan untuk melakukan *web scraping*. (Sahria, 2022)

Web scraping secara hukum merupakan hal yang sah untuk dilakukan selama data yang dikumpulkan digunakan untuk kepentingan pribadi dan dalam tidak melanggar undang-undang yang berlaku. Jika data akan dipublikasikan atau jika konten memiliki hak cipta dan melanggar persyaratan layanan maka ada beberapa preseden hukum yang perlu diperhatikan.

Di *Feist Publications, Inc. v. Rural Telephone Service Co.*, Mahkamah Agung Amerika Serikat memutuskan bahwa *scraping* dan penerbitan ulang data, seperti daftar telepon, diizinkan. Hal ini serupa dengan Australia, *Telstra Corporation Limited v. Phone Directories Company Pty Ltd*, menunjukkan bahwa hanya data dengan penulis yang dapat diidentifikasi dilindungi dengan hak cipta. (Jarmul, 2017)

Kasus lainnya di Amerika Serikat, mengevaluasi penggunaan kembali berita *Associated Press* untuk produk berita gabungan, dinyatakan sebagai pelanggaran hak cipta di *Associated Press v. Meltwater*. Hal ini menunjukkan bahwa ketika data yang didapatkan merupakan data publik seperti lokasi bisnis dan daftar telepon, data tersebut dapat diterbitkan ulang dengan mengikuti aturan yang berlaku. Namun, jika data tersebut merupakan opini, ulasan atau data pengguna pribadi, kemungkinan besar tidak dapat dipublikasikan ulang karena alasan hak cipta. (Jarmul, 2017)

Web scraping biasanya dibuat untuk menargetkan situs web atau situs tertentu dengan tujuan mengumpulkan informasi spesifik di situs tersebut. *Scraper* dibuat untuk mengakses halaman spesifik dan perlu dimodifikasi jika situs berubah atau jika informasi yang ada di situs diubah.

Sebaliknya, *web crawling* biasanya dibuat dengan cara umum, yaitu menargetkan situs web dari serangkaian domain tingkat atas atau seluruh web. *Crawler* dapat dibuat untuk

mengumpulkan informasi yang lebih spesifik, tetapi umumnya digunakan untuk mengambil sedikit informasi umum dari berbagai situs atau halaman yang berbeda dan mengikuti tautan ke halaman lain.

2.1.1 Jenis-Jenis *Web Scraping*

Dalam melakukan *web scraping* terdapat beberapa teknik yang umumnya digunakan. Yang pertama, menyalin data secara manual. Melakukan *web scraping* secara manual merupakan bentuk paling sederhana yang dapat dilakukan. Terkadang teknologi untuk melakukan *web scraping* terbaik sekalipun tidak dapat menggantikan pemeriksaan manual yang dilakukan oleh manusia. Teknik yang kedua merupakan *parsing HTML*. Dengan teknik ini, *web scraping* dapat dilakukan tidak hanya pada halaman *website* yang bersifat statis, tetapi juga dinamis. *Parsing HTML* juga memungkinkan untuk menyalin data dalam jumlah yang besar dalam waktu singkat.

2.1.2 Manfaat *Web Scraping*

Dalam penggunaannya *web scraping* memiliki beberapa manfaat. Yang pertama, dapat membandingkan ulasan dalam skala besar karena *web scraping* pada dasarnya adalah mengumpulkan data. Yang kedua, dapat digunakan untuk mencari informasi. Sebagai contoh ketika ingin memulai suatu bisnis, pencarian informasi menggunakan *web scraping* dapat menjadi salah satu cara untuk mengetahui apakah produk atau jasa yang akan dikembangkan merupakan sesuatu yang dibutuhkan dan diminati oleh masyarakat.

Dengan menggunakan *web scraping* proses mengumpulkan data dapat dilakukan dengan lebih cepat. Selain itu, jika data yang dikumpulkan jumlahnya sangat besar, proses tersebut juga dapat dilakukan secara otomatis. *Web scraping* dapat membantu untuk menganalisa data, karena mampu mengumpulkan data secara detail dan efisien. (Josi, 2014)

2.1.3 Kendala *Web Scraping*

Dalam melakukan proses *web scraping* terdapat beberapa hal yang menjadi kendala. Yang pertama adalah belum adanya teknik *web scraping* yang efektif. Yang kedua, data yang didapatkan tidak terstruktur dengan baik. Hal tersebut dapat terjadi dikarenakan adanya sisa-sisa teks yang tidak diinginkan seperti contoh tag HTML. Yang ketiga adalah dapat terjadi

pemblokiran terhadap alamat IP baik secara manual maupun berdasarkan *geolocation* dari alamat IP tersebut.

Terdapat metode lain yang digunakan oleh beberapa situs web untuk mencegah dilakukannya *web scraping*, seperti mendeteksi dan melarang *bot* melihat halaman mereka dan juga menonaktifkan API layanan web yang mungkin terekspos oleh sistem.

2.2. *Extract, Transform, Load*

Extract, transform, load atau dapat disingkat menjadi ETL merupakan prosedur umum dalam menyalin data dari suatu sumber ke sistem yang dituju. Proses ini dibagi menjadi tiga tahap. Tahap pertama adalah *extract* atau ekstraksi. Proses ekstraksi merupakan proses yang paling penting karena dapat menjadi sebuah acuan keberhasilan untuk tahap selanjutnya. Sebagian besar dari proyek database menggabungkan data dari sumber yang berbeda. Setiap sumber juga dapat menggunakan format data yang berbeda. Format data pada umumnya berbentuk JSON atau XML.

Proses tahap kedua yaitu *transform*. Pada tahap ini data disiapkan untuk dimuat di target akhir dengan melakukan serangkaian fungsi pada data yang telah diekstrak. Proses *transform* ini berfungsi untuk membersihkan data atau memisahkan data-data yang tepat untuk digunakan. Namun terdapat tantangan ketika sistem yang berbeda melakukan interaksi yaitu antarmuka dan komunikasi sistem yang relevan. Kumpulan karakter yang mungkin tersedia di satu sistem mungkin tidak tersedia di sistem lainnya.

Tahapan terakhir adalah *load*, di mana data dimasukkan ke target akhir. Proses *load* dapat dibagi menjadi menjadi dua, yaitu *full load* dan *incremental load*. *Full load* merupakan metode untuk memuat seluruh data secara bersamaan untuk menjadi catatan baru pada *database*. Metode ini berguna untuk menghasilkan data yang tumbuh secara eksponensial namun sulit untuk diatur. Sedangkan *incremental load* merupakan metode untuk memuat data secara interval terjadwal. Metode *incremental load* membandingkan data yang masuk dengan data yang sudah ada, oleh karena itu, metode ini menghasilkan data tambahan jika ditemukan data yang baru dan memudahkan untuk diatur. (Bansal, 2015)

2.3. *Data Cleaning*

Data cleaning atau pembersihan data adalah adalah proses mengoreksi data yang tidak sesuai dari kumpulan catatan, tabel, atau basis data, mengacu pada pengidentifikasian bagian data yang tidak lengkap, tidak benar, tidak akurat, atau tidak relevan, kemudian mengganti,

memodifikasi, atau menghapus data yang tidak sesuai. Untuk mendapatkan data hasil yang baik, diperlukan untuk memenuhi beberapa kriteria yaitu validitas, akurasi, kelengkapan, konsistensi, dan keseragaman.

Dalam melakukan pembersihan data terdapat proses yang harus dilakukan. Yang pertama yaitu menghapus data yang tidak penting atau kurang relevan. Kemudian mengedit kesalahan pada struktur seperti adanya kesalahan penulisan, kesalahan penggunaan huruf besar dan kecil, dan lain-lain. Hal tersebut dilakukan untuk memudahkan dalam pengolahan data.

2.4. Python

Python merupakan bahasa pemrograman yang diciptakan oleh Guido Van Rossum, dapat dilihat pada Gambar 2.1. Guido Van Rossum menciptakan python pada awal tahun 1990 untuk keperluan umum di Centrum voor Wiskunde and Informatica (CWI).

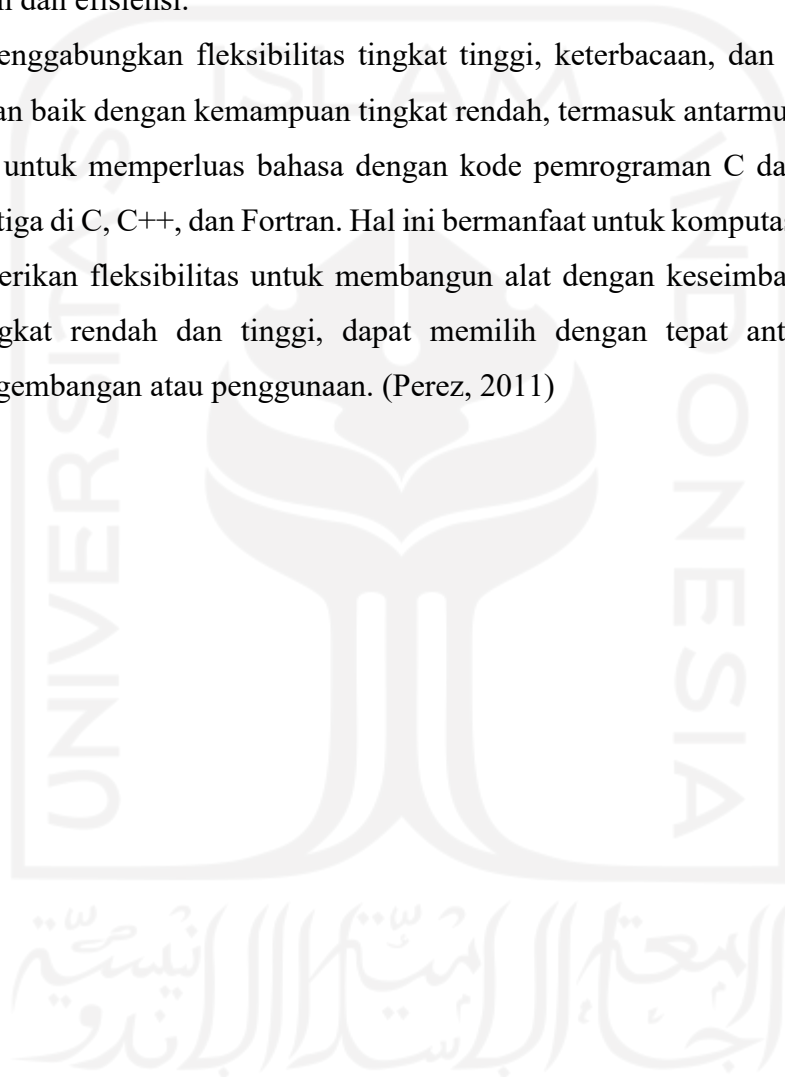


Gambar 2.1 Guido Van Rossum

Sebelum menciptakan bahasa pemrograman Python, Guido Van Rossum telah mengerjakan bahasa pemrograman lainnya, yaitu bahasa pemrograman ABC, yang dikembangkan di CWI sebagai bahasa pengajaran yang menekankan kejelasan. Meskipun proyek ABC pada akhirnya ditutup, van Rossum mengambil banyak pelajaran ketika ia mulai menulis Python sebagai alat untuk digunakan dalam proyek penelitian sistem operasi dan multimedia.

Python adalah bahasa pemrograman yang dinamis. Hirarki nomornya mencakup bilangan bulat asli dengan panjang arbitrer, titik mengambang presisi perangkat keras, bilangan kompleks, dan dukungan *library* untuk bilangan rasional dan titik mengambang presisi arbitrer. Python juga memiliki string yang kuat, daftar ukuran variabel, set, dan array asosiatif yang sangat fleksibel disebut sebagai *dictionaries* dalam Python. Jenis-jenis ini memberi Python kosakata yang beragam untuk mengekspresikan banyak pertanyaan algoritmik kompleks dengan kejelasan dan efisiensi.

Python menggabungkan fleksibilitas tingkat tinggi, keterbacaan, dan antarmuka yang terdefinisi dengan baik dengan kemampuan tingkat rendah, termasuk antarmuka C resmi yang memungkinkan untuk memperluas bahasa dengan kode pemrograman C dan menautkan ke *library* pihak ketiga di C, C++, dan Fortran. Hal ini bermanfaat untuk komputasi ilmiah modern sehingga memberikan fleksibilitas untuk membangun alat dengan keseimbangan yang tepat antara fitur tingkat rendah dan tinggi, dapat memilih dengan tepat antara kinerja dan kemudahan pengembangan atau penggunaan. (Perez, 2011)



BAB III

PELAKSANAAN MAGANG

3.1 Manajemen Aktivitas Magang

3.1.1 Pendataan Tempat Wisata di Daerah Istimewa Yogyakarta

Daerah Istimewa Yogyakarta sebagai salah satu tujuan yang digemari oleh masyarakat untuk menghabiskan waktu dengan berwisata di berbagai kategori tempat wisata. Kategori yang dimaksud adalah wisata museum, wisata *heritage* atau warisan budaya, wisata pantai, wisata alam, wisata belanja, wisata religi, wisata kuliner, wisata agro, wisata gua, dan wisata budaya. Dengan banyaknya kategori tempat wisata di Daerah Istimewa Yogyakarta yang tersebar di kabupaten dan kota, dilakukan pendataan di mana data yang dikumpulkan adalah nama tempat wisata tersebut, lokasi berupa koordinat dalam peta, akun media sosial resmi dari tempat wisata tersebut, akun yang memiliki kaitan dengan tempat tersebut, serta tagar yang menandai tempat wisata tersebut.

A. Perencanaan Proyek

Pembagian waktu dalam pengerjaan proyek pendataan tempat wisata di Daerah Istimewa Yogyakarta dibagi dibagi menjadi 5 poin seperti yang dapat dilihat pada Tabel 3.1.

Tabel 3.1 Pembagian waktu dalam pelaksanaan proyek pendataan tempat wisata

No	Aktivitas	Durasi
1	Pembagian tugas	1 hari
2	Pencarian tempat wisata belanja	1 minggu
3	Pencarian tempat wisata religi	1 minggu
4	Pencarian tempat wisata kuliner	1 minggu
5	Pencarian tempat wisata agro	1 minggu

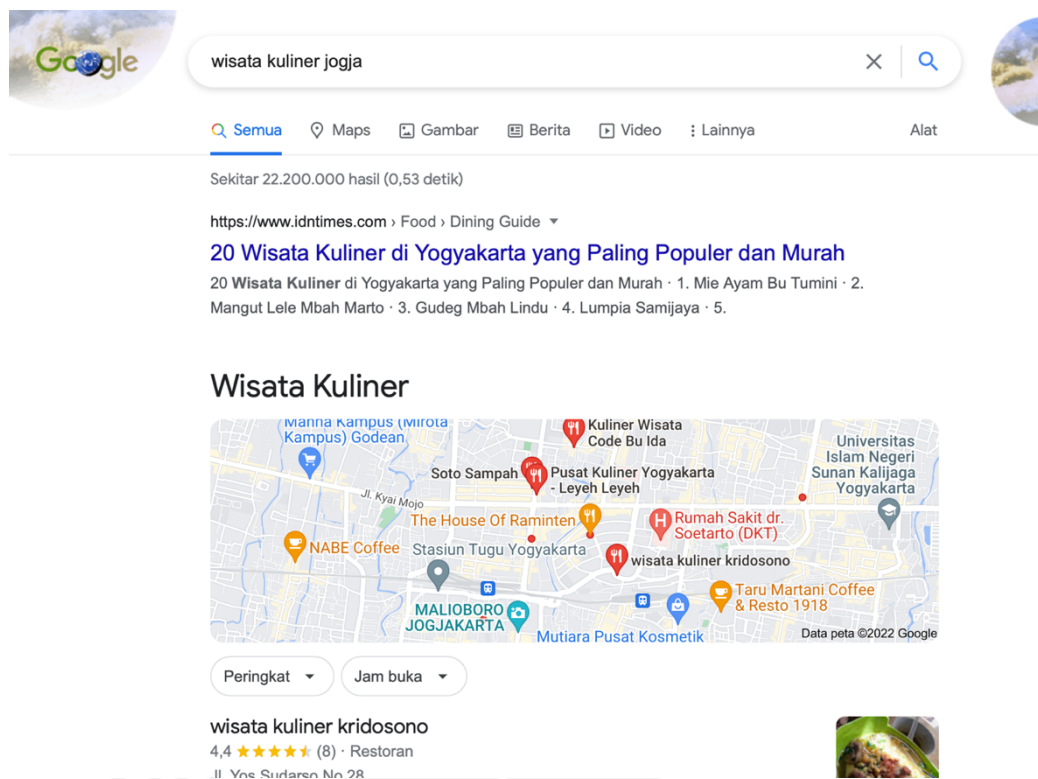
B. Pelaksanaan Proyek

Proyek ini dimulai dengan pembagian tugas dengan sesama peserta magang. Dari 10 kategori yang ada, yaitu museum, *heritage*, pantai, alam, belanja, religi, kuliner, agro, gua, dan budaya dibagi menjadi 3 kategori kepada 2 orang dan 4 kategori kepada 1 orang seperti yang ditampilkan pada Gambar 3.1.

3	Auliya	wisata museum (kata kunci museum, ada sekitar
4	Auliya	wisata heritage (kraton jogja dan pakualaman, be
5	Husnul	wisata pantai
6	Husnul	wisata alam
7	Rio	wisata belanja
8	Rio	wisata religius
9	Rio	wisata kuliner
10	Rio	wisata agro
11	Husnul	wisata gua
12	Auliya	wisata budaya

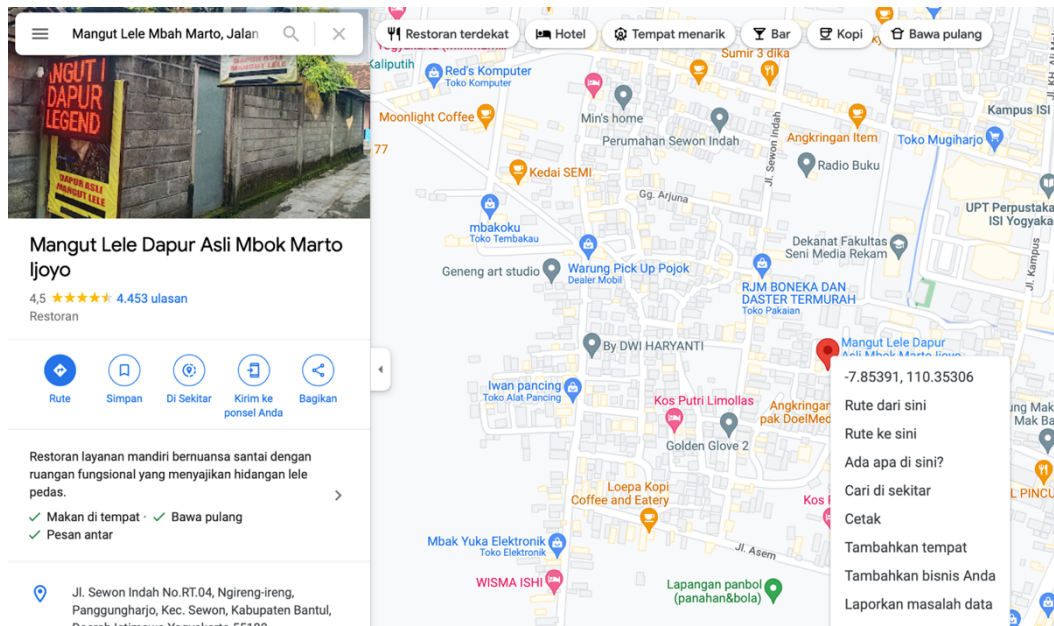
Gambar 3.1 Hasil pembagian kategori tugas

Proses pencarian dilakukan secara manual dengan melakukan pencarian pada *Google* dan memanfaatkan aplikasi Instagram. Pencarian dilakukan dengan menggunakan kata kunci seperti “wisata kuliner jogja” atau “wisata kuliner yogyakarta” seperti yang ditampilkan pada Gambar 3.2.



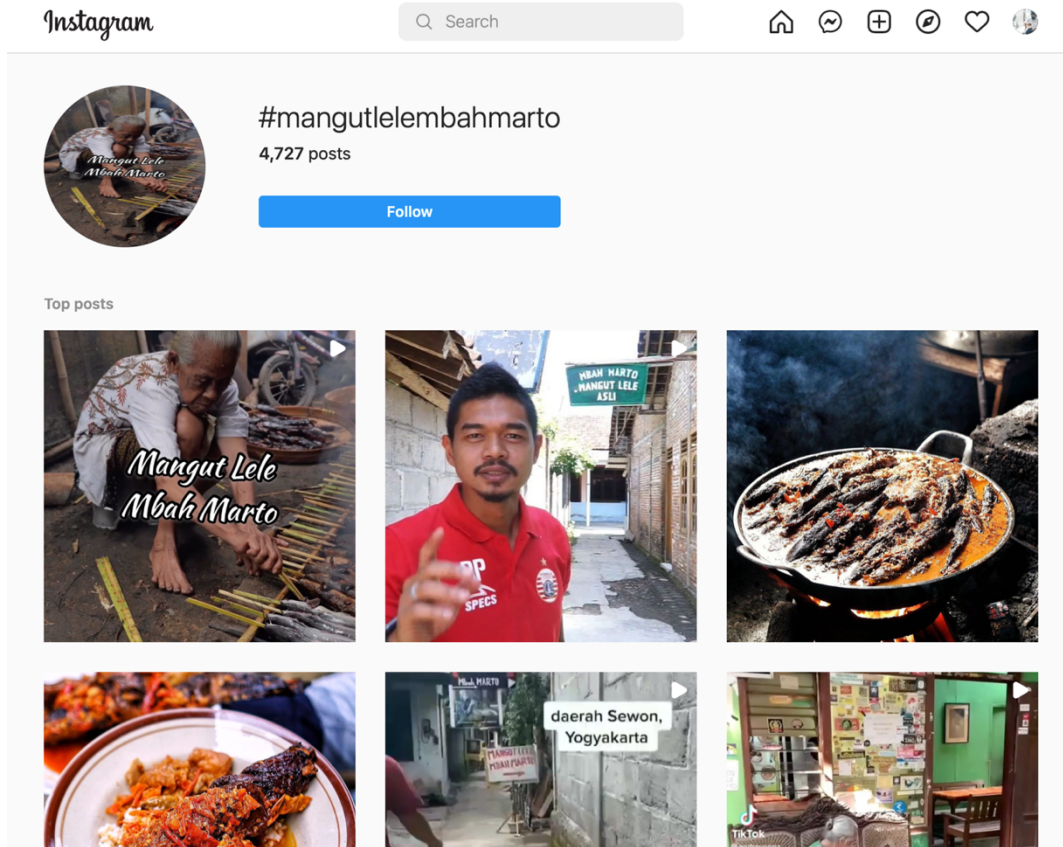
Gambar 3.2 Pencarian menggunakan google

Dari nama-nama tempat wisata yang telah dicatat, kemudian dilakukan pencarian lokasi tempat tersebut menggunakan *google maps*. Lokasi dari tempat tersebut dilakukan pengecekan titik koordinat dalam peta seperti pada Gambar 3.3.



Gambar 3.3 Pengambilan titik koordinat menggunakan google maps

Tahap terakhir yaitu dengan melakukan pencarian menggunakan aplikasi *Instagram* untuk mencari apakah terdapat akun resmi dari tempat tersebut dan tagar apa yang digunakan oleh masyarakat untuk mendeskripsikan tempat tersebut seperti pada Gambar 3.4.



Gambar 3.4 Pengecekan akun media sosial dan tagar menggunakan Instagram

Seluruh data yang telah dikumpulkan, dicatat dalam *spreadsheet* seperti pada Gambar 3.5 yang kemudian dilaporkan kepada mentor mengenai hasil yang telah ditemukan. Pendataan tempat wisata akan digunakan untuk menjadi acuan sebaran tempat wisata berdasarkan kategori pada situs Jogja Center.

1	No	Nama	Akun resmi	Akun terkait 1	Akun Terkait 2	Latitude	Longitude	Hashtag 1	Hashtag 2	Hashtag
2	1	Gudeg Yu Djum Wijilan	gudegyudjum			-7.804937315	110.3667291	#gudegyudjum	#gudegyudjumpusat	#gudegyuj
3	2	Bakmi Jawa Jogokariyan	bakmijawajogokariyan			-7.822965065	110.364042	#bakmijawajogokariyan	-	-
4	3	Warung Kopi Klotok	kopiklotok			-7.670322429	110.4213278	#warungkopiklotok	#warungkopiklotokpakem	-
5	4	Warung Kopi Merapi	kopimerapi_			-7.602594435	110.4458456	#kopimerapi	#kopimerapijogja	#warungkopir
6	5	Sate Klathak Pak Pong	satepakpong			-7.871203883	110.3874707	#satepakpong	#sateklathakpakpong	#sateklatakpa
7	6	House Of Raminten	houseoframinten			-7.785180157	110.3713911	#houseoframinten	#ramintenjogja	#raminte
8	7	Mangut Lele Mbah Marto	-			-7.854025556	110.3529177	#mangutlelembahmarto	#mangutlelembahmartojogja	-
9	8	Sop Ayam Pak Min	-			-7.79505213	110.3690658	#sopayampakminklaten	#sopayampakmin	#sopakmin
10	9	Angkringan Kopi Joss Pak Agus	-			-7.788811397	110.3665458	#angkringankopijoss	#angkringankopijos	-
11	10	Soto Bathok Mbah Katro	-			-7.760930499	110.4470468	#sotobathokmbahkatro	-	-
12	11	Gudeg Bromo Bu Tekluk	gudegbromobutekluk			-7.773722972	110.389591	#gudegbromo	#gudegbromobutekluk	#gudegbrom
13	12	Gudeg Permata Bu Narti	-			-7.801209955	110.3731956	#gudegpermata	#gudegpermatajogja	-
14	13	Oseng Mercon Bu Narti	-			-7.801298625	110.3574771	#osengmerconbunarti	#osengmerconjogja	-
15	14	Tempo Gelato Kaliurang	tempogelato.official			-7.762717005	110.3797015	#tempogelato	#tempogelatojogja	#tempogelatoyo
16	15	Bakpia Pathok 25	bakpia25official			-7.796937424	110.3580701	#bakpiapathok25	#bakpiapathok25original	-
17	16	Bakpia Kukus Tugu Jogja Kaliurang	bakpiatugujogja			-7.758846904	110.3814964	#bakpiakukusjogja	#bakpiakukustugujogja	-
18	17	Bakpia Kurnia Sari	bakpiakurniasari.official			-7.808520581	110.3879417	#bakpiakurniasari	#bakpiakurniasarijogja	-
19	18	Bakpia Patuk 75	bakpiapatuk75			-7.793636639	110.3531803	#bakpia75	-	-
20	19	Bakpia Kencana	bakpiakencana			-7.783005741	110.4062724	#bakpiakencana	#bakpiakencanajogja	-

Gambar 3.5 Hasil pendataan yang dikumpulkan pada spreadsheet

C. Pemantauan dan Pengendalian Proyek

Pemantauan dan pengendalian proyek dilakukan secara daring bersama mentor dan peserta magang lainnya. *Project manager* dan mentor memberikan masukan dalam setiap rapat pemantauan. Pada rapat yang dilaksanakan terdapat permasalahan dimana adanya kendala dari peserta magang yaitu tempat wisata yang didata mencakup dalam 2 kategori yang berbeda yang kemudian diberikan saran untuk melakukan diskusi untuk memilih kategori mana yang dirasa lebih tepat. Pemantauan dilaksanakan sebanyak satu kali dalam satu minggu yang dilaksanakan setiap hari Jum'at menggunakan aplikasi pihak ketiga seperti *google meet* atau *zoom*. Peserta magang akan melakukan *share screen* dan memaparkan perkembangan pekerjaan dalam proyek yang telah diberikan. Baik mentor maupun peserta magang lainnya dapat memberikan masukan jika terdapat hal-hal yang kurang sesuai.

D. Penutupan Proyek

Proyek pendataan diakhiri dengan terkumpulnya data tempat-tempat wisata di Daerah Istimewa Yogyakarta berdasarkan kategori yang telah ditentukan. Dokumen pendataan berupa *spreadsheet* diberikan kepada mentor untuk diolah dan digunakan.

3.1.2 Pendataan dan Rekapitulasi Direktori Putusan Pengadilan

Daerah Istimewa Yogyakarta memiliki daerah yang luas, terdiri dari 4 kabupaten dan 1 kota. Jogja Center sebagai sistem analitik memberikan analisis perkara yang mencakup 4 kabupaten dan 1 kota di Daerah Istimewa Yogyakarta. Analisis perkara yang disajikan dibagai menjadi dua bagian, yaitu jenis perkara perdata dan jenis perkara pidana.

A. Perencanaan Proyek

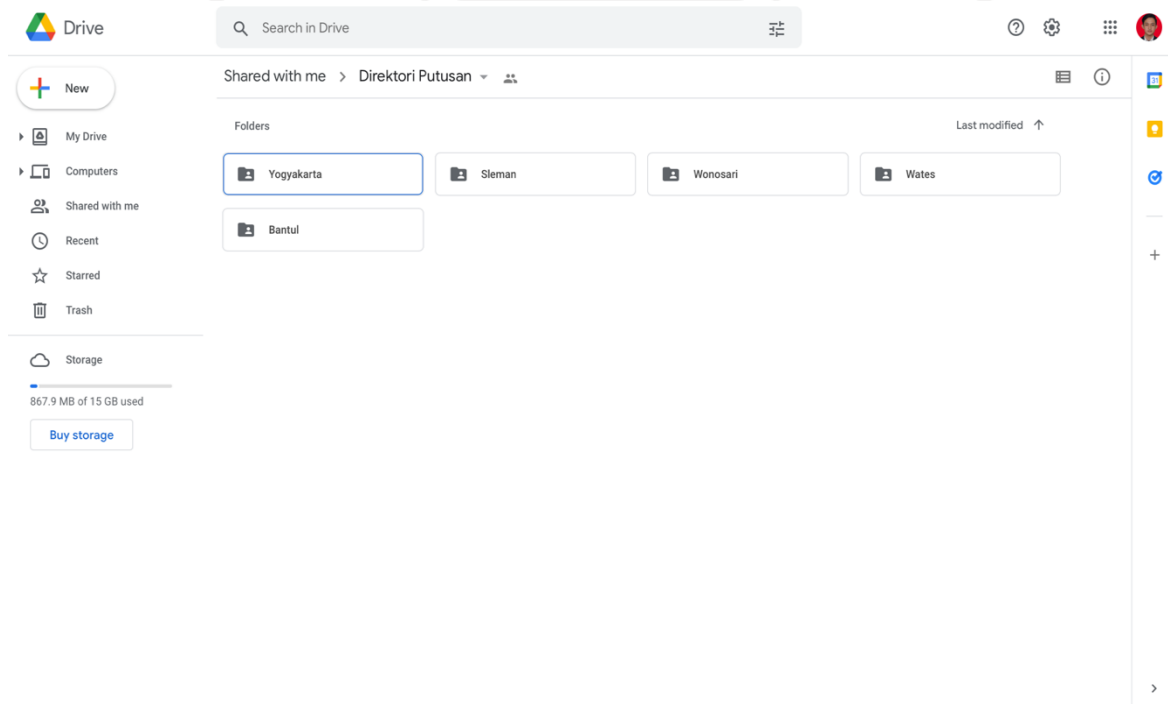
Pembagian waktu dalam pengerjaan proyek pendataan dan rekapitulasi direktori putusan pengadilan dibagi menjadi 3 poin seperti yang dapat dilihat pada Tabel 3.2.

Tabel 3.2 Pembagian waktu dalam pelaksanaan proyek pendataan direktori

No	Aktivitas	Durasi
1	Pembagian tugas	2 hari
2	Pendataan direktori Pengadilan Negeri	1 minggu
3	Pendataan direktori Pengadilan Agama	1 minggu

B. Pelaksanaan Proyek

Proyek diawali dengan melakukan pembagian tugas rekapitulasi per kabupaten dan kota dengan peserta magang lain. Pembagian dilakukan seadil-adilnya dengan mempertimbangkan jumlah data yang akan dikerjakan. Dokumen dikumpulkan menggunakan *google drive* untuk memudahkan pembagian data seperti yang ditampilkan pada Gambar 3.6.



Gambar 3.6 Kumpulan dokumen direktori putusan pengadilan

Selanjutnya dilakukan proses pendataan yang dilakukan berdasarkan jenis perkara. Karena jumlah data yang cukup banyak dan berkas yang diberikan dalam bentuk ekstensi *file* PDF, maka perlu dilakukan ekstraksi data yang dilakukan menggunakan bahasa pemrograman Python. Tahap pertama yang perlu dilakukan yaitu melakukan *import library*. *Library* yang digunakan yaitu *pdfminer* seperti pada Gambar 3.7.

```
1 import fitz
2 import glob
3 import pdfminer
4 import os
5 import nltk
6 from nltk.tokenize import sent_tokenize
7 nltk.download('punkt')
```

Gambar 3.7 Library yang digunakan

Kemudian dilakukan *mounting* atau menghubungkan *google drive* ke *google colab* sehingga *google drive* memberikan izin kepada *google colab* untuk mengakses dokumen yang terdapat dalam *google drive* seperti pada Gambar 3.8.

```
9 from google.colab import drive
10 drive.mount ('/content/gdrive')
```

Gambar 3.8 Potongan kode untuk melakukan mounting

Selanjutnya dilakukan proses ekstraksi data. Pada proses ini, kode program membaca apakah nama dokumen sesuai dengan kata kunci yang telah ditentukan. Kemudian dilakukan perulangan untuk tiap dokumen di mana ketika program menemukan terdapat kata kunci yang telah ditentukan, program akan menyimpan kalimat di mana kata kunci tersebut berada secara utuh seperti pada potongan kode pemrograman di baris 20–32 pada Gambar 3.9. Kata kunci yang ditentukan yaitu untuk mendapatkan nama dokumen, tanggal, jenis perkara, pasal yang mengatur, dan perkara.

```

3 for fl in files:
4     if 'pdt' in fl.lower():
5         jenis_perkara = "perdata"
6     elif 'pid' in fl.lower():
7         jenis_perkara = "pidana"
8
9     doc = fitz.open(fl)
10    all_text = ""
11    for page in doc:
12        text = page.get_text()
13        all_text += text
14
15    sentences = sent_tokenize(all_text)
16
17    kuhap = []
18    perkara = []
19    tanggal = []
20    for sen in sentences:
21        if "kuhap" in sen.lower():
22            kuhap.append(sen)
23        if "melakukan tindak pidana" in sen.lower():
24            perkara.append(sen)
25        if "dilaksanakan pada hari" in sen.lower():
26            tanggal.append(sen)
27    print(str(count))
28    print(fl)
29    print(tanggal)
30    #print(jenis_perkara)
31    #print(kuhap)
32    #print(perkara)
33
34    count = count + 1

```

Gambar 3.9 Potongan kode ekstraksi data

Hasil ekstraksi data kemudian dipindahkan ke dalam dokumen *spreadsheet* untuk kemudian diberikan kembali kepada mentor seperti pada Gambar 3.10.

putusan pengadilan ☆ 📄 ☁

File Edit View Insert Format Data Tools Extensions Help Last edit was made on September 5, 2021 by Cut Husnul

100% \$ % .0 .00 123 Default (Ari... 10 B I S A

A1 fx file_name

	A	B	C	D	E
1	file_name	tanggal	jenis_perkara	pasal	duduk_perkara
2	1-Pdt.P-2021-PN_Smn.pdf	18 Januari 2021	Perdata	UU No 23 tahun 2006 jo PP No 25 tahun 2008	Perbaikan Kesalahan Dalam Akta Kelahiran
3	1-Pid.C-2021-PN_Smn.pdf	5 Februari 2021	Pidana	Pasal 364 KUH Pidana Jo Pasal 53 KUH Pidana dan Peraturan Mahkamah Agung Republik Indonesia Nomor 02 tahun 2012 serta UU No. 8 tahun 1981	Pencurian
4	1-Pid.S-2021-PN_Smn.pdf	13 Januari 2021	Pidana	Pasal 37 Perda Kabupaten Sleman Nomor 08 Tahun 2019, Pasal 55 ayat ke-1 KUHP	Melakukan Peredaran dan Penjualan Minuman Beralkohol Tanpa Izin Secara Bersama-sama
5	10-Pdt.G-2021-PN_Smn.pdf	23 Februari 2021	Perdata	Pasal 271 jo 272 RV	Perbuatan Melawan Hukum
6	10-Pdt.P-2021-PN_Smn.pdf	28 Januari 2021	Perdata	UU No 24 tahun 2013	Permohonan Ganti Nama
7	11-Pdt.P-2021-PN_Smn.pdf	28 Januari 2021	Perdata		Permohonan Ganti Nama
8	12-Pdt.P-2021-PN_Smn.pdf	21 Januari 2021	Perdata		Permohonan Ganti Nama
9	121-Pdt.G-2020-PN_Smn.pdf	7 Januari 2021	Perdata	Pasal 271 jo 272 RV	Perbuatan Melawan Hukum
10	125-Pdt.G-2020-PN_Smn.pdf	4 Januari 2021	Perdata		Perbuatan Melawan Hukum
11	13-Pdt.P-2021-PN_Smn.pdf	3 Februari 2021	Perdata	UU No 23 tahun 2006	Permohonan Ganti Nama
12	13-Pid.B-2021-PN_Smn.pdf	1 Maret 2021	Pidana	Pasal 362 KUHP	Pencurian
13	14-Pdt.G.S-2020-PN_Smn.pdf	21 Januari 2021	Perdata		Perbuatan Melawan Hukum
14	14-Pdt.P-2021-PN_Smn.pdf	2 Februari 2021	Perdata	Pasal 52 UU No 23 tahun 2006 Jo UU No 24 tahun 2013	Permohonan Ganti Nama
15	15-Pdt.P-2021-PN_Smn.pdf	16 Februari 2021	Perdata	Pedoman Pelaksanaan Tugas dan Administrasi Pengadilan Dalam Empat Lingkungan Peradilan Buku II Edisi 2007 Mahkamah Agung RI tahun 2009 hal 47	Penerimaan/Penolakan Warisan

Gambar 3.10 Hasil rekapitulasi direktori

C. Pemantauan dan Pengendalian Proyek

Proyek pendataan dan rekapitulasi direktori putusan pengadilan dilakukan selama 2 minggu. Untuk memastikan proyek yang dilakukan sesuai dengan arahan, dilakukan pemantauan dan pengendalian proyek secara rutin bersama mentor dan peserta magang lainnya. Pada rapat yang dilaksanakan terdapat kendala yaitu ada beberapa dokumen putusan pengadilan yang telah dilakukan ekstraksi tidak terdapat keterangan tanggal yang jelas yang kemudia diberikan solusi oleh mentor yaitu dengan melakukan pengecekan kembali pada situs pengadilan negeri atau pengadilan agama pada kabupaten yang bersangkutan. Pemantauan dilaksanakan sebanyak satu kali dalam satu minggu yang dilaksanakan setiap Jum'at menggunakan aplikasi pihak ketiga seperti *google meet* maupun *zoom*. Peserta magang akan melakukan *share screen* dan memaparkan perkembangan pekerjaan dalam proyek yang telah diberikan. Baik mentor maupun peserta magang lainnya dapat memberikan masukan jika terdapat hal-hal yang kurang sesuai.

D. Penutupan Proyek

Proyek rekapitulasi diakhiri dengan pengumpulan dokumen hasil rekapitulasi direktori putusan pengadilan berupa *link spreadsheet* yang kemudian akan digunakan sebagai data analisis sosial bagian analisis perkara menurut data pengadilan pada situs Jogja Center.

3.1.3 Visualisasi Perbandingan Data Covid-19 dan Vaksinasi Dengan Data Mobilitas

Di tahun 2021, angka kasus Covid-19 di Indonesia mencapai angka lebih dari 6 juta kasus dengan 156 ribu meninggal dunia. Kasus Covid-19 melonjak pada bulan Juli 2021 bertepatan dengan hari raya Idul Adha dan adanya varian Covid-19 baru yaitu varian delta. Data Covid-19 dan vaksinasi dilakukan perbandingan dengan data mobilitas, untuk memberikan gambaran apakah lonjakan kasus Covid-19 dipengaruhi dengan dilakukannya vaksinasi dan tingkat mobilitas masyarakat.

A. Perencanaan Proyek

Pembagian waktu dalam pengerjaan proyek visualisasi perbandingan data covid-19 dan vaksinasi dengan data mobilitas dibagi dibagi menjadi 4 poin seperti yang dapat dilihat pada Tabel 3.3.

Tabel 3.3 Pembagian waktu dalam pelaksanaan proyek visualisasi data

No	Aktivitas	Durasi
1	Pencarian data mobilitas	1 minggu
2	Pencarian data vaksin	1 minggu
3	Pencarian data Covid-19	1 minggu
4	Analisis dan visualisasi	1 minggu

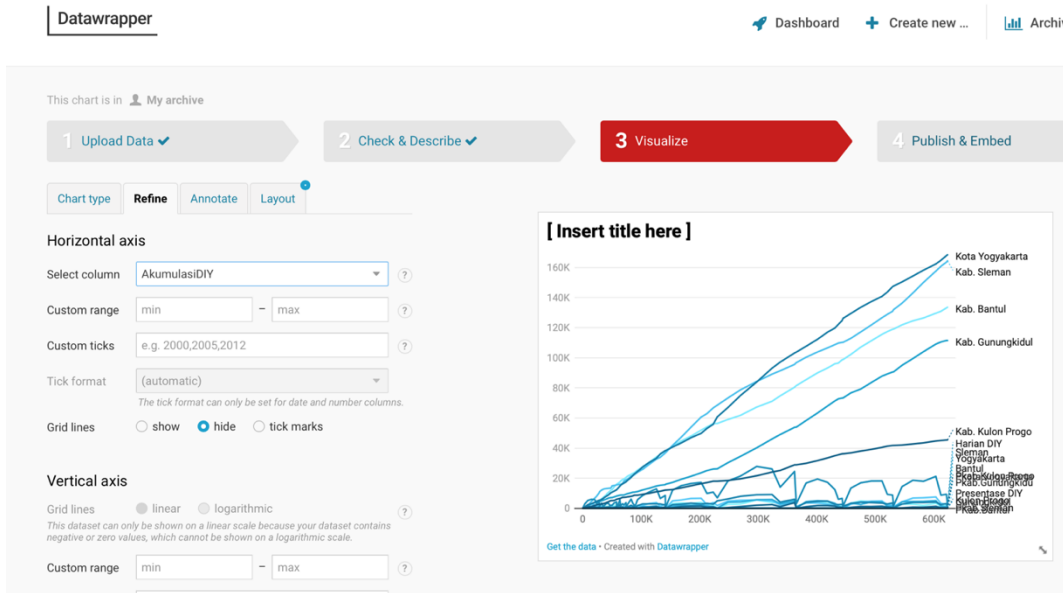
B. Pelaksanaan Proyek

Proyek diawali dengan pencarian data mobilitas, data Covid-19, dan data vaksinasi di Daerah Istimewa Yogyakarta salah satunya dari situs resmi Kemenkes yang telah diunduh seperti pada Gambar 3.11.

	A	B	C	D	E	F	G	H	I	J	
1	Provinsi	Tanggal	Bantul	Gunungkidul	Kulon Progo	Sleman	Yogyakarta	Kab. Bantul	Kab. Gunung	Kab. Kulon Pr	Kab. S
2	Daerah Istim	01/05/21	2	0	63	1	0	2	0	63	
3	Daerah Istim	02/05/21	0	0	0	0	398	2	0	63	
4	Daerah Istim	03/05/21	774	779	247	653	541	776	779	310	
5	Daerah Istim	04/05/21	1139	1087	304	1194	1510	1915	1866	614	
6	Daerah Istim	05/05/21	1048	1109	561	2684	563	2963	2975	1175	
7	Daerah Istim	06/05/21	1052	808	729	2247	938	4015	3783	1904	
8	Daerah Istim	07/05/21	612	404	292	1126	566	4627	4187	2196	
9	Daerah Istim	08/05/21	542	247	456	2302	846	5169	4434	2652	
10	Daerah Istim	09/05/21	0	5	0	779	0	5169	4439	2652	
11	Daerah Istim	10/05/21	1105	559	176	1074	505	6274	4998	2828	
12	Daerah Istim	11/05/21	628	286	213	355	462	6902	5284	3041	
13	Daerah Istim	12/05/21	2	3	22	178	34	6904	5287	3063	
14	Daerah Istim	13/05/21	0	0	0	0	0	6904	5287	3063	
15	Daerah Istim	14/05/21	0	0	0	0	0	6904	5287	3063	
16	Daerah Istim	15/05/21	4	8	0	0	14	6908	5295	3063	
17	Daerah Istim	16/05/21	0	9	0	0	2	6908	5304	3063	
18	Daerah Istim	17/05/21	534	101	80	156	367	7442	5405	3143	
19	Daerah Istim	18/05/21	712	372	273	226	692	8154	5777	3416	
20	Daerah Istim	19/05/21	1361	491	498	160	714	9515	6268	3914	
21	Daerah Istim	20/05/21	1199	563	470	164	835	10714	6831	4384	
22	Daerah Istim	21/05/21	547	112	384	53	685	11261	6943	4768	
23	Daerah Istim	22/05/21	790	166	325	96	642	12051	7109	5093	
24	Daerah Istim	23/05/21	154	0	0	1406	359	12205	7109	5093	
25	Daerah Istim	24/05/21	1997	428	87	312	710	14202	7537	5180	
26	Daerah Istim	25/05/21	1727	552	767	350	874	15929	8089	5947	
27	Daerah Istim	26/05/21	0	2	14	1	1	15929	8091	5961	
28	Daerah Istim	27/05/21	1499	612	590	377	2478	17428	8703	6551	
29	Daerah Istim	28/05/21	655	103	155	66	578	18083	8806	6706	
30	Daerah Istim	29/05/21	1040	24	564	463	940	19123	8830	7270	
31	Daerah Istim	30/05/21	184	0	0	1564	481	19307	8830	7270	
32	Daerah Istim	31/05/21	1218	532	130	1009	861	20525	9362	7400	

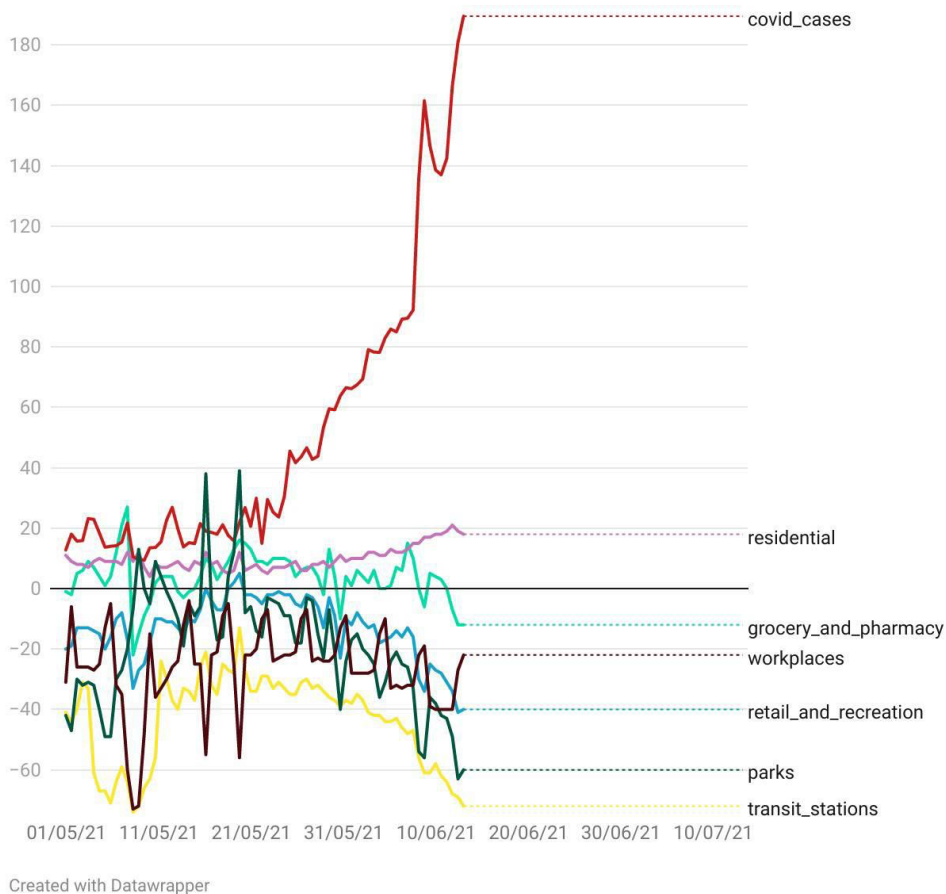
Gambar 3.11 Data vaksinasi DIY

Kemudian dilakukan proses visualisasi data menggunakan data mobilitas, data Covid-19, dan data vaksinasi di Daerah Istimewa Yogyakarta yang telah dikumpulkan sebelumnya. Data yang akan dianalisis akan dipilih dan dijadikan dalam satu dokumen untuk mempermudah dalam proses analisis dan visualisasi. Visualisasi data dilakukan menggunakan situs pemvisualisasian data yang dapat digunakan secara gratis yaitu situs datawrapper yang dapat dilihat pada Gambar 3.12. Hasil dari visualisasi data dapat dilihat pada Gambar 3.13 dan Gambar 3.14.



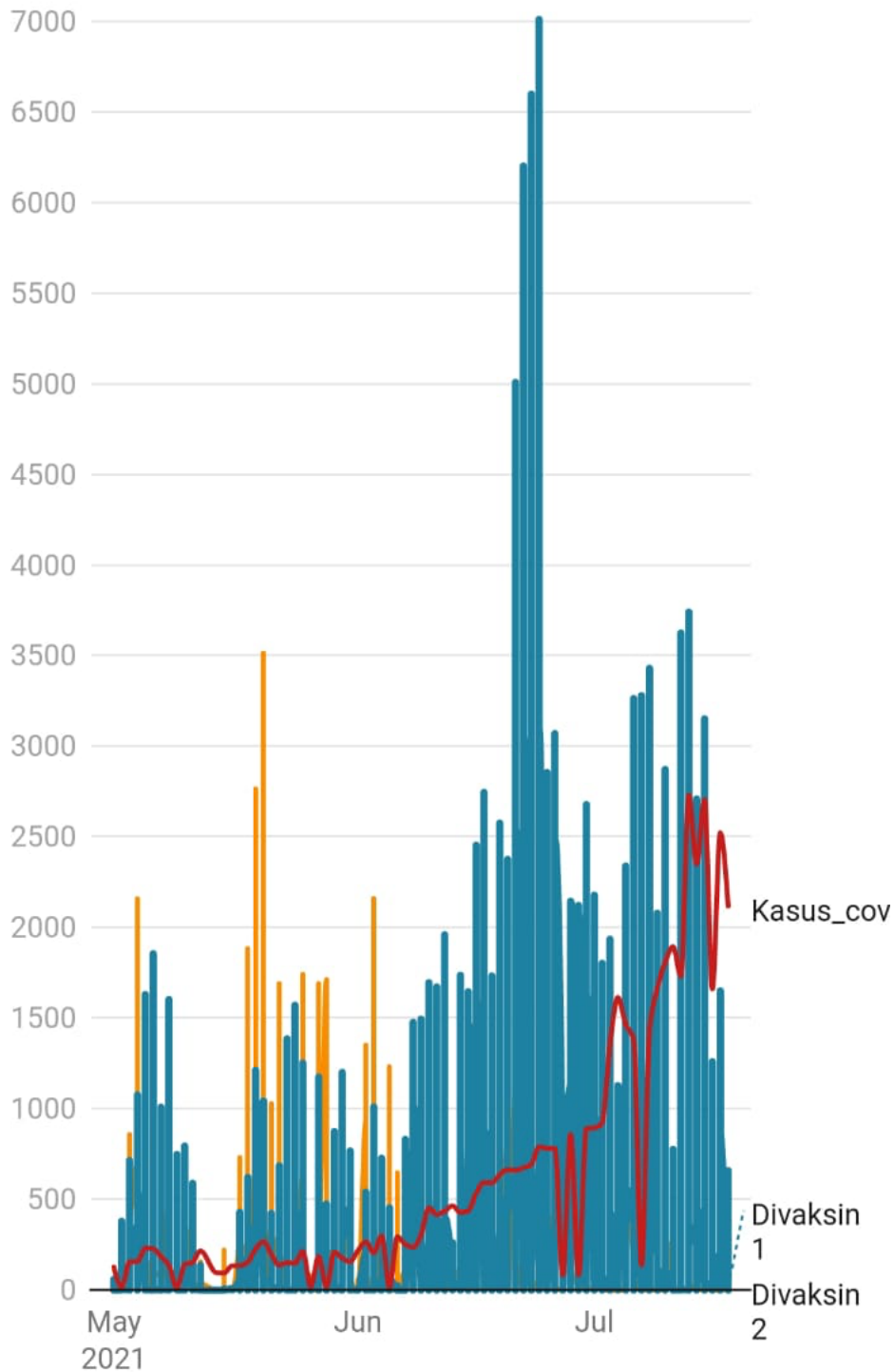
Gambar 3.12 Tampilan penggunaan datawrapper

Comparing between Mobility and Covid Cases in Yogyakarta



Gambar 3.13 Hasil visualisasi data mobilitas dengan data Covid-19

Data Monitoring Vaksinasi dengan kasus covid



Created with [Datawrapper](#)

Gambar 3.14 Hasil visualisas data Covid-19 dengan data vaksinasi

C. Pemantauan dan Pengendalian Proyek

Pemantauan dan pengendalian proyek dilakukan secara rutin bersama mentor dan peserta magang lainnya. Pada rapat yang dilaksanakan terdapat kendala dimana data yang didapatkan tidak sesuai dengan jangka waktu yang diinginkan. *Project manager* dan mentor kemudian memberikan solusi dengan memberikan rekomendasi situs lainnya untuk melakukan pencarian data dengan jangka waktu yang sesuai. Pemantauan dilaksanakan paling sedikit sebanyak satu kali dalam satu minggu yang dilaksanakan setiap Jum'at menggunakan aplikasi pihak ketiga seperti *google meet* maupun *zoom*. Peserta magang akan melakukan *share screen* dan memaparkan perkembangan pekerjaan dalam proyek yang telah diberikan. Baik mentor maupun peserta magang lainnya dapat memberikan masukan jika terdapat hal-hal yang kurang sesuai.

D. Penutupan Proyek

Pada situs Jogja Center, terdapat direktori analisis yang memberikan analisis yang berhubungan dengan Daerah Istimewa Yogyakarta. Visualisasi perbandingan data Covid-19 dan vaksinasi dengan data mobilitas menjadi salah satu analisis yang dilakukan. Proyek ini diakhiri dengan memberikan hasil visualisasi kepada mentor sehingga dapat digunakan dalam pembuatan artikel analisis pada situs Jogja Center.

3.1.4 Analisis Sentimen Publik Berdasarkan Kata Kunci

Di masa pandemi, banyak masyarakat yang memiliki pandangan tersendiri mengenai problematika yang ada hingga kinerja pemerintah dalam menanggulangi hal tersebut. Proyek analisis sentimen publik berdasarkan kata kunci menjadi sebuah sarana untuk mengetahui apakah sentimen yang diberikan masyarakat berupa sentimen positif, negatif atau netral.

A. Perencanaan Proyek

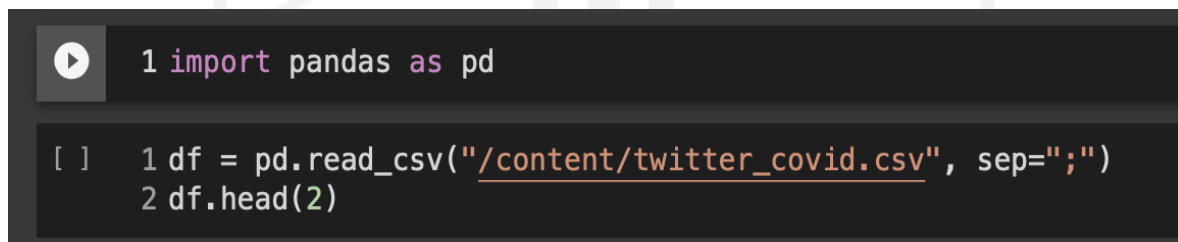
Pembagian waktu dalam pengerjaan proyek analisis sentimen publik berdasarkan kata kunci dibagi dibagi menjadi 5 poin seperti yang dapat dilihat pada Tabel 3.4.

Tabel 3.4 Pembagian waktu dalam pelaksanaan proyek analisis sentimen publik

No	Aktivitas	Durasi
1	Analisis sentimen minggu pertama	1 minggu
2	Analisis sentimen minggu kedua	1 minggu
3	Analisis sentimen minggu ketiga	1 minggu
4	Analisis sentimen minggu keempat	1 minggu

B. Pelaksanaan Proyek

Proyek diawali dengan mempersiapkan data hasil sentimen publik di media sosial *Twitter* yang diberikan oleh mentor terhadap 7 tema kata kunci yaitu bansos, covid, malioboro, PPKM, Sri Sultan, Vaksin Jogja, dan Isolasi. Dari 7 tema kata kunci yang ada, pembagian tugas kepada pemegang yaitu 2 orang mengerjakan analisis 2 tema kata kunci dan 1 orang mengerjakan analisis 3 tema kata kunci. Mengerjakan analisis diawali dengan melakukan *import library* yaitu *library Pandas*. Kemudian membuka dokumen sentimen menggunakan nama dokumen yang dituju seperti pada Gambar 3.15.



```

1 import pandas as pd

[ ] 1 df = pd.read_csv("/content/twitter_covid.csv", sep=";")
     2 df.head(2)

```

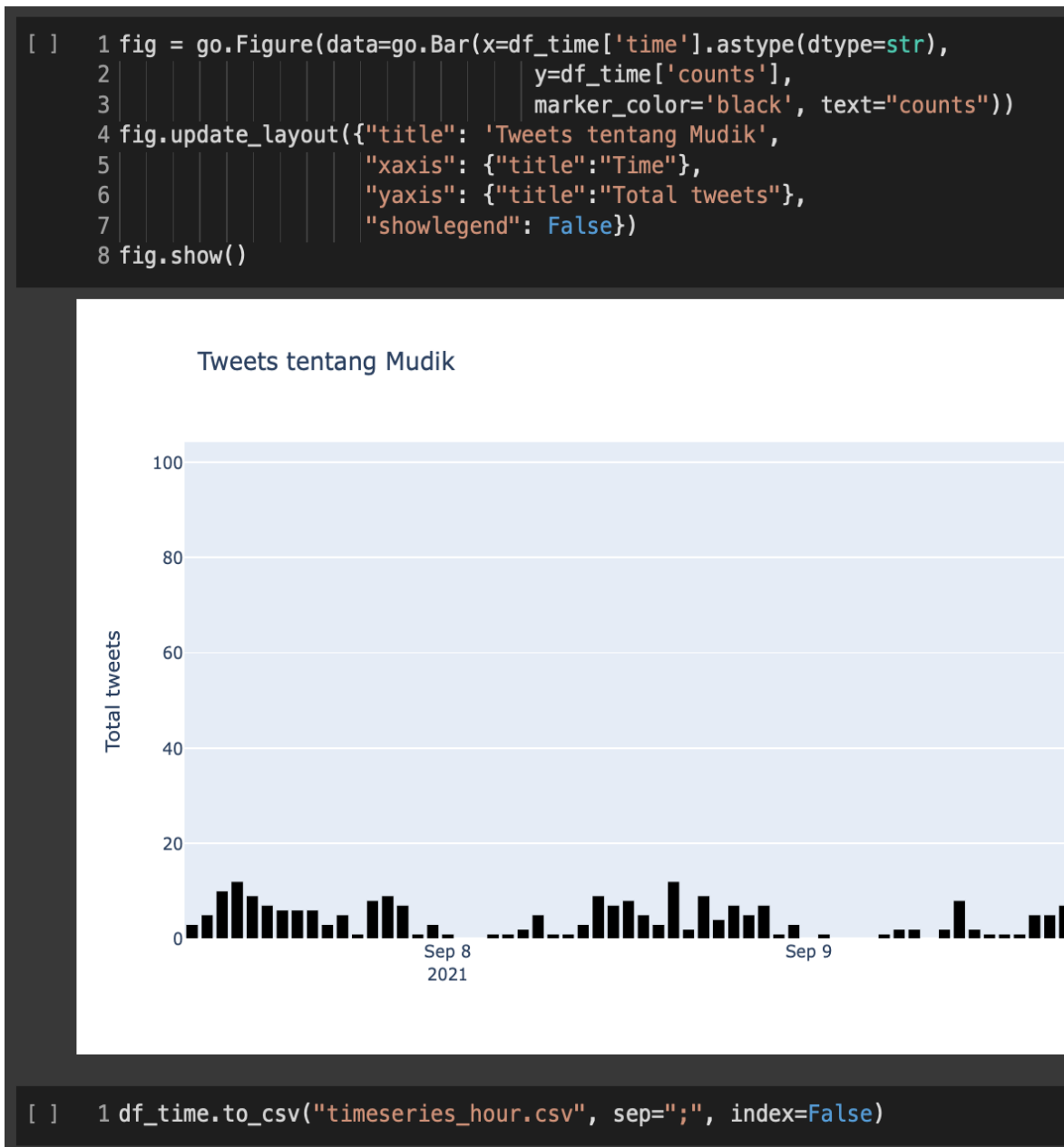
Gambar 3.15 Potongan kode membaca dokumen yang akan digunakan

Dilanjutkan dengan penghitungan jumlah dari jenis *tweet* yang ditemukan. Selain menghitung jumlah *tweet* secara keseluruhan seperti pada Gambar 3.16, penghitungan juga dilakukan berdasarkan jumlah *tweet* per-jamnya seperti pada Gambar 3.17. Selanjutnya dilakukan penghitungan jenis sentimen yang ditemukan apakah berupa sentimen positif, sentimen negatif atau bukan berupa sentimen positif atau sentimen negatif tetapi netral yang dapat dilihat pada Gambar 3.18.

```
[ ] 1 df['type'].value_counts()
RT          688
tweet       411
mention     19
Name: type, dtype: int64
```

Gambar 3.16 Potongan kode perhitungan jenis tweet



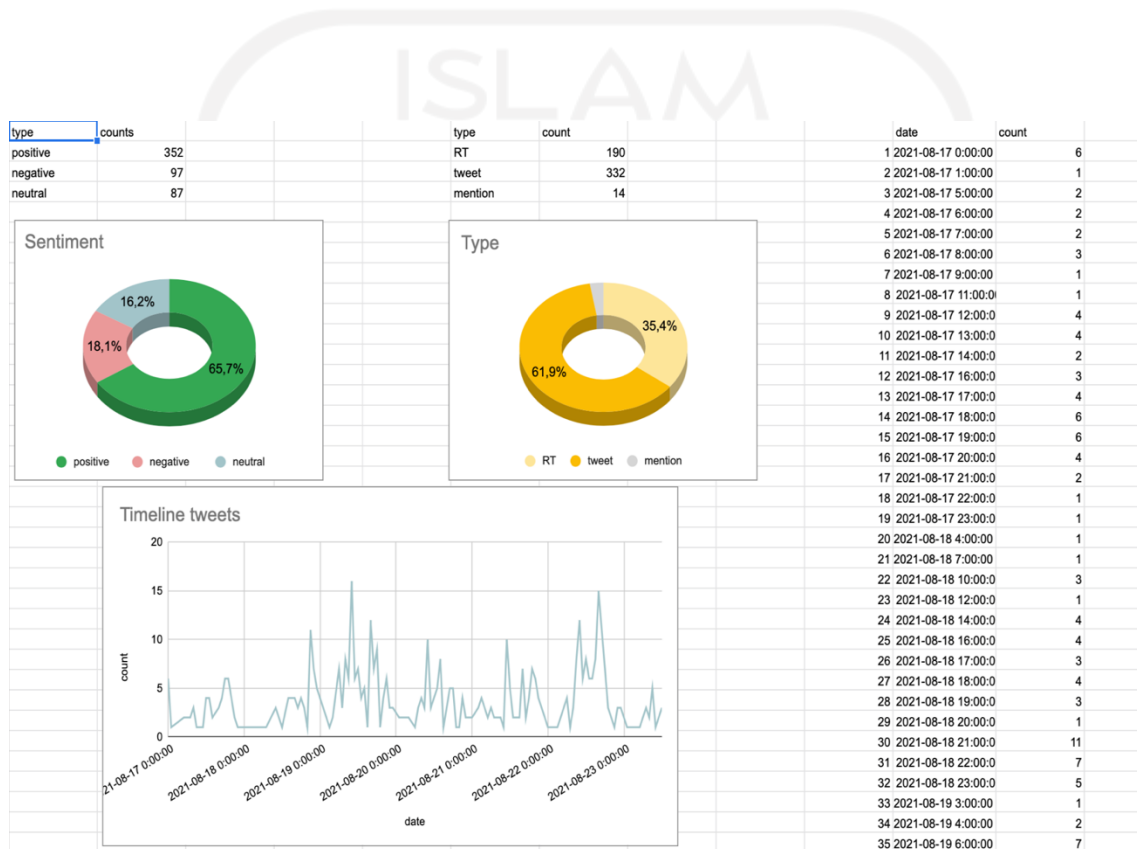


Gambar 3.17 Potongan kode perhitungan jumlah tweet per-jam

```
▼ Sentiments
[ ] 1 sentiments = df['sentiment'].value_counts().rename_axis('type').reset_index(name='counts')
2
3 sentiments.to_csv("sentiments.csv", sep=";", index=False)
```

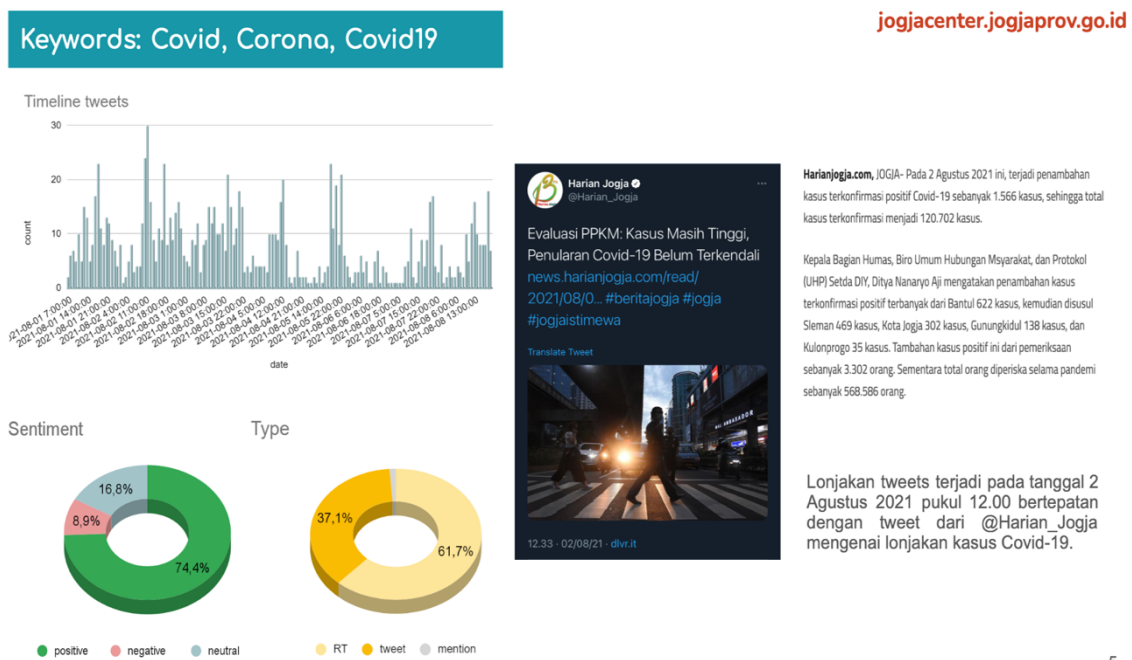
Gambar 3.18 Potongan kode perhitungan jenis sentimen

Setelah data pendukung analisis telah dikumpulkan, proses analisis mulai dikerjakan. Data dikumpulkan pada *google spreadsheet*, kemudian dibuat diagram mengenai sentimen publik dan diagram jenis *tweet* yang ditemukan. Selain diagram, juga dibuat grafik sebaran *tweet* per-jam dalam bentuk diagram garis seperti yang dapat dilihat pada Gambar 3.19 untuk melihat sebaran *tweet* yang ditemukan.



Gambar 3.19 Tampilan proses analisis menggunakan spreadsheet

Hasil dari analisis yang dilakukan diberikan kembali kepada mentor dalam bentuk *file power point*. Untuk mendukung analisis yang telah dilakukan, juga diberikan data pendukung berupa *tweet* yang berhubungan dengan tema kata kunci tersebut seperti yang dapat dilihat pada Gambar 3.20 dapat dilihat bahwa terjadi lonjakan *tweet* pada tanggal 2 Agustus 2021 pukul 12 yang bertepatan dengan *tweet* mengenai lonjakan kasus Covid-19. Selain itu, dalam *file power point* yang dibuat juga ditambahkan mengenai berita hoaks yang ditemukan di minggu tersebut.



5

Gambar 3.20 Tampilan salah satu slide presentasi yang telah diberikan tweet pendukung

C. Pemantauan dan Pengendalian Proyek

Pemantauan dan pengendalian proyek kolaborasi analisis sentimen publik berdasarkan kata kunci dilakukan secara rutin bersama mentor dan peserta magang lainnya. Pemantauan dilaksanakan paling sedikit sebanyak satu kali dalam satu minggu yang dilaksanakan setiap hari Selasa menggunakan aplikasi pihak ketiga seperti *google meet* maupun *zoom*. Pada rapat yang dilaksanakan terdapat kendala dimana terdapat data pendukung dan hasil analisis yang kurang sesuai. Mentor memberikan saran untuk melakukan pencarian lebih lanjut mengenai data pendukung agar hasil analisis yang diberikan dapat lebih sesuai. Peserta magang akan melakukan *share screen* dan memaparkan perkembangan pekerjaan dalam proyek yang telah diberikan. Baik mentor maupun peserta magang lainnya dapat memberikan masukan jika terdapat hal-hal yang kurang sesuai.

D. Penutupan Proyek

Proyek analisis sentimen diakhiri dengan memberikan dokumen *spreadsheet* yang digunakan untuk melakukan analisis serta *slide* presentasi kepada mentor sehingga nantinya dapat digunakan kembali untuk keperluan analisis mendatang.

3.2 Scraping Media Sosial Instagram

Jogja Center merupakan sebuah proyek yang didirikan oleh Dinas Komunikasi dan Informatika Daerah Istimewa Yogyakarta pada tahun 2020. Jogja Center merupakan sebuah sistem analitik berbasis *big data*, yang difokuskan pada pengembangan data analitik dan pendukung pengambilan keputusan, serta merujuk pada dimensi-dimensi *Jogja Smart Province* (JSP). Dalam proyek Jogja Center, yang merupakan sistem informasi berbasis analitik, dibutuhkan data-data sebagai sumber yang akan dianalisis.

3.2.1 Perencanaan Proyek

Dalam pengerjaan proyek dibutuhkan perencanaan waktu pengerjaan yang dapat dilihat pada Tabel 3.5. Pembagian waktu tersebut tidak termasuk dengan adanya aktivitas kecil tambahan yang diberikan untuk dikerjakan selama waktu pengerjaan proyek. Pengerjaan proyek ini menggunakan metode *Extract, Transform, load* (ETL) dimana data diekstraksi oleh penulis, kemudian dilakukan transformasi data, yaitu merubah data dengan format JSON menjadi format CSV. Dilanjutkan dengan *load* yaitu menyimpan data ke media penyimpanan. Proses *load* tidak dilakukan oleh penulis melainkan oleh mentor.

Tabel 3.5 Pembagian waktu dalam pelaksanaan proyek *scraping* media sosial Instagram

No	Aktivitas	Durasi
1	Mempelajari proyek	2 Minggu
2	Mendesain kode pemrograman	2 minggu
3	Pelaksanaan proyek	5 bulan

Tahap awal yang dilakukan untuk memulai proyek yaitu *research* atau dilakukan pencarian informasi serta memahami mengenai tugas yang diberikan yaitu *scraping* dimulai dari pemahaman mengenai *scraping* hingga proses pembuatan kode. Setelah dilakukan *research*, tahapan selanjutnya yaitu menentukan bahasa pemrograman serta *library, module*,

dan *package* yang akan digunakan. Pada proyek ini bahasa pemrograman yang digunakan adalah Python.

Bahasa pemrograman Python dipilih dengan alasan yaitu Python merupakan salah satu bahasa pemrograman dengan struktur kode yang sederhana. Penggunaan bahasa pemrograman Python juga dapat memudahkan dalam pemahaman ketika proses pembuatan kode.

3.2.2 Pelaksanaan Proyek

Pelaksanaan proyek *scraping* media sosial Instagram dilaksanakan selama 6 bulan, dimulai pada bulan Maret 2021 hingga bulan September 2021. Pengerjaan proyek dilakukan dengan melalui beberapa tahapan. Tahapan pertama adalah mempelajari mengenai bahasa pemrograman Python, *scraping*, dan cara melakukan *scraping* menggunakan bahasa pemrograman Python selama dua minggu. Setelah melakukan pencarian mengenai hal tersebut, kemudian dilanjutkan dengan membuat kode pemrograman yang dapat memberikan akses untuk masuk ke dalam media sosial Instagram. Proses tersebut diawali dengan melakukan import library. Beberapa library yang digunakan yaitu json, csv, dan juga pandas. Selain itu juga digunakan Instagram *Private* API. Kemudian program memasukkan *username* dan *password* yang disimpan pada masing-masing variabel, dapat dilihat pada Gambar 3.21.

```

17 device_id = None
18 try:
19     settings_file = "credentials.json"
20     if not os.path.isfile(settings_file):
21         # settings file does not exist
22         print('[W] Unable to find auth cookie file: {0:s} (creating a new one...').format(settings_file)
23
24         # login new
25         api = Client(
26             username, password,
27             on_login=lambda x: onlogin_callback(x, settings_file))
28     else:
29         with open(settings_file) as file_data:
30             cached_settings = json.load(file_data, object_hook=from_json)
31             print(cached_settings)
32
33             device_id = cached_settings.get('device_id')
34             # reuse auth settings
35             api = Client(
36                 username, password,
37                 settings=cached_settings)
38
39             print('[I] Using cached login cookie for "' + str(api.authenticated_user_id) + '".')
40
41 except (ClientCookieExpiredError, ClientLoginRequiredError) as e:
42     print('[E] ClientCookieExpiredError/ClientLoginRequiredError: {0:s}'.format(e))
43
44     # Login expired
45     # Do relogin but use default ua, keys and such
46     if username and password:
47         api = Client(
48             username, password,
49             device_id=device_id,
50             on_login=lambda x: onlogin_callback(x, settings_file))
51     else:
52         print("[E] The login cookie has expired, but no login arguments were given.")
53         print("[E] Please supply --username and --password arguments.")
54         print('-' * 70)
55         sys.exit(0)
56
57 except ClientLoginError as e:
58     print('[E] Could not login: {s}.\n[E] {s}\n\n{s}'.format(
59         json.loads(e.error_response).get("error_title", "Error title not available."),
60         json.loads(e.error_response).get("message", "Not available"), e.error_response))
61     print('-' * 70)
62     sys.exit(9)
63 except ClientError as e:
64     print('[E] Client Error: {s}'.format(e.error_response))
65     print('-' * 70)
66     sys.exit(9)
67 except Exception as e:
68     if str(e).startswith("unsupported pickle protocol"):
69         print("[W] This cookie file is not compatible with Python {}.format(sys.version.split(' ')[0][0])
70         print("[W] Please delete your cookie file 'credentials.json' and try again.")
71     else:
72         print('[E] Unexpected Exception: {0:s}'.format(e))

```

Gambar 3.21 Penyimpanan *username* dan *password* pada *variable*

Untuk menghindari terjadinya proses *login* secara berulang maka dilakukan pembuatan *credential* yang dapat menyimpan *cookies* sehingga proses *login* dapat dilewati. Kemudian program akan mengakses halaman dari akun yang dituju menggunakan *user id* dari akun tersebut. Unggahan foto dan video yang berada di beranda dari akun tersebut dikumpulkan dan disimpan ke dalam *file* dengan format JSON dengan menggunakan potongan kode yang ditampilkan pada Gambar 3.22.

```
results = api.user_feed('3535868226')
next_max_id = results.get('next_max_id')
hasil = json.dumps(results, indent = 4)
# print(hasil)

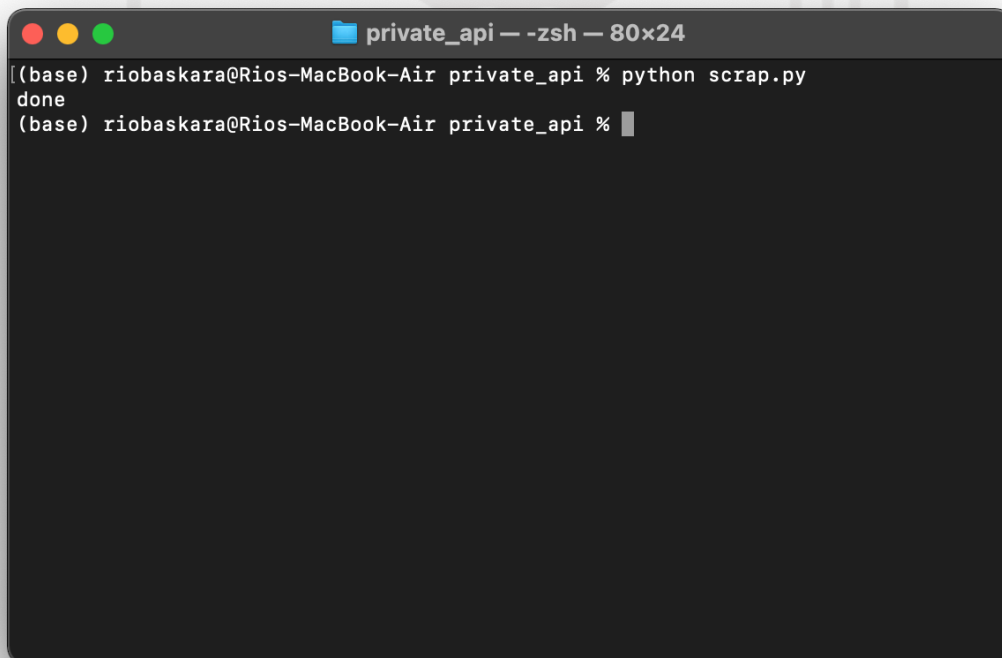
while next_max_id:
    results = api.user_feed('3535868226', max_id = next_max_id)
    next_max_id = results.get('next_max_id')

with open('data_new_update.json', 'w') as outfile:
    json.dump(results, outfile, indent = 4)

print("done")
```

Gambar 3.22 Potongan kode scraping dan penyimpanan format json

Setelah tahap penulisan kode program telah selesai dilakukan, kode program dijalankan menggunakan *terminal* atau *command prompt* seperti yang ditampilkan pada Gambar 3.23. *Scraping* dilakukan terhadap satu akun yaitu kominfodiy. Proses *running* program dapat selesai dalam waktu yang beragam tergantung dengan kecepatan dan kestabilan jaringan internet yang digunakan ketika melakukan *running* program.



```
private_api -- zsh -- 80x24
(base) riobaskara@Rios-MacBook-Air private_api % python scrap.py
done
(base) riobaskara@Rios-MacBook-Air private_api %
```

Gambar 3.23 Proses running menggunakan terminal

Setelah data-data dikumpulkan, data tersimpan dalam bentuk *file* JSON. Format *file* tersebut diubah dari format JSON menjadi format CSV. Perubahan dilakukan menggunakan kode yaitu dengan membuka *file* pertama yang masih berformat file JSON dan dilakukan penyimpanan ulang menjadi format *file* CSV seperti potongan kode pada Gambar 3.24. Data yang telah disimpan menjadi format *file* CSV kemudian dilakukan *parsing* atau memisahkan data-data yang akan digunakan dan data-data yang tidak digunakan dari keseluruhan data.

```
with open('data_new.json', 'w') as outfile:  
    json.dump(result, outfile, indent = 4)  
  
df = pd.read_csv("Parsing.csv", sep=";")  
# mediaid = df["media_id"]
```

Gambar 3.24 Potongan kode mengubah JSON menjadi CSV

Berdasarkan hasil dari *scraping* dari akun tersebut, data-data yang berhasil dikumpulkan yaitu tanggal, bulan, tahun, dan jam ketika foto atau video tersebut diunggah, id dari foto atau video, *caption* yang telah dituliskan pada foto atau video, dan *link* dari unggahan foto atau video tersebut seperti yang ditampilkan pada Gambar 3.25 yang kemudian akan digunakan untuk keperluan situs Jogja Center.

Tanggal	Timestamp	Link gambar	Caption	Link post	Media_id		
05/08/21	1628129892	https://insta	Hai Lur!\nKa	https://www	2633266855686853562_3535868226		
04/08/21	1628086393	https://insta	Hasil laporar	https://www	2632901954711610127_3535868226		
04/08/21	1628063733	https://insta	ÄÚSaya suda	https://www	2632711871105217551_3535868226		
04/08/21	1628051653	https://insta	Berkaitan de	https://www	2632610541166590612_3535868226		
03/08/21	1627998671	https://insta	Rekapitulasi	https://www	2632166088460491304_3535868226		
03/08/21	1627996397	https://insta	Hasil laporar	https://www	2632147015559296065_3535868226		
02/08/21	1627916220	https://insta	Hai Lur! Men	https://www	2631474440667690661_3535868226		
02/08/21	1627911833	https://insta	Hasil laporar	https://www	2631437642981745432_3535868226		
02/08/21	1627887588	https://insta	Hai Lur! \nM	https://www	2631234017189875923_3535868226		
01/08/21	1627824884	https://insta	Hasil laporar	https://www	2630708262777240449_3535868226		
01/08/21	1627809711	https://insta	Pemda DIY la	https://www	2630580978334312085_3535868226		
01/08/21	1627784612	https://insta	Hai Lur!\n\nl	https://www	2630369108461722592_3535868226		
31/07/21	1627740685	https://insta	Hasil laporar	https://www	2630001952091809450_3535868226		
31/07/21	1627723789	https://insta	Berkaitan de	https://www	2629860217718142644_3535868226		
30/07/21	1627645103	https://insta	Hasil laporar	https://www	2629200148932455629_3535868226		
30/07/21	1627635341	https://insta	Gubernur DI	https://www	2629118259894690767_3535868226		
30/07/21	1627610659	https://insta	Hai Lur!\n\nl	https://www	2628910891072300586_3535868226		
29/07/21	1627559657	https://insta	Hasil laporar	https://www	2628483375625587372_3535868226		
28/07/21	1627484781	https://insta	Halo Sahaba	https://www	2627855267964750978_3535868226		
28/07/21	1627481534	https://insta	Hasil laporar	https://www	2627828028292338295_3535868226		
27/07/21	1627385106	https://insta	Hasil laporar	https://www	2627019133655444669_3535868226		
27/07/21	1627367347	https://insta	Laporan Ban	https://www	2626870159585794892_3535868226		
27/07/21	1627355254	https://insta	Hai Lur!\n\nl	https://www	2626768715670936300_3535868226		
26/07/21	1627302610	https://insta	Hasil laporar	https://www	2626327111562721245_3535868226		
26/07/21	1627277186	https://insta	Dalam konfe	https://www	2626113840608000623_3535868226		
26/07/21	1627263267	https://insta	Hai Lur Infor	https://www	2625997075899090516_3535868226		
25/07/21	1627219178	https://insta	Hasil laporar	https://www	2625627229730431950_3535868226		
25/07/21	1627208453	https://insta	Hai Lur! \n\n	https://www	2625537266204725954_3535868226		
25/07/21	1627190768	https://insta	Salam Hang	https://www	2625388425579332516_3535868226		
24/07/21	1627131455	https://insta	Hasil laporar	https://www	2624891360291806769_3535868226		

Gambar 3.25 Hasil scraping

3.2.3 Pemantauan dan Pengendalian Proyek

Kegiatan magang di Dinas Komunikasi dan Informatika Daerah Istimewa Yogyakarta dilaksanakan secara luring. Untuk memastikan proyek yang dilakukan sesuai dengan arahan, pemantauan dan pengendalian proyek dilakukan secara rutin bersama mentor dan peserta magang lainnya. Pemantauan dilaksanakan paling sedikit sebanyak satu kali dalam satu minggu, setiap hari Jum'at menggunakan aplikasi pihak ketiga seperti *google meet* atau *zoom* untuk mengurangi adanya kontak secara langsung di masa pandemi. Pada rapat pemantauan scraping yang dilangsungkan secara daring pada tanggal 4 Juni 2021 terdapat kendala dimana Ketika menjalankan kode yang telah dibuat diperlukan *login* secara berulang kali yang

mengakibatkan akses terblokir. Mentor memberikan solusi dengan memberikan Langkah untuk membuat *credentials* yang dapat menyimpan *cookies* untuk dapat melewati proses *login* tersebut. Peserta magang akan melakukan *share screen* dan memaparkan perkembangan pekerjaan dalam proyek yang telah diberikan. Baik mentor maupun peserta magang lainnya dapat memberikan masukan jika terdapat hal-hal yang kurang sesuai.

3.2.4 Penutupan Proyek

Proyek *scraping* media sosial Instagram akan terus dilakukan karena kebutuhan data perlu untuk selalu diperbarui untuk memberikan hasil analisis yang tepat. Dalam akhir periode magang, kode pemrograman yang telah dibuat diberikan kepada mentor selaku pembimbing saat magang agar dapat diteruskan sehingga kode program yang telah dibuat dapat dikembangkan.



BAB IV

REFLEKSI PELAKSANAAN MAGANG

4.1 Teknis

Selama kegiatan magang, penulis terlibat dalam proyek Jogja Center. Dalam pengerjaan proyek tahap awal diperlukan penyesuaian dan adaptasi terhadap aktivitas dan juga proyek yang diberikan. Terdapat banyak hal baru yang dipelajari untuk menjadi penunjang dalam pengerjaan proyek. Selama menjalani kegiatan magang, penulis terlibat dalam proyek Jogja Center untuk melakukan *web scraping* pada media sosial Instagram.

Pada saat tahapan awal pengerjaan proyek, terdapat beberapa hal yang menjadi kendala terkait langkah awal yang harus dipersiapkan sebelum melakukan *web scraping*. Setelah proses mempelajari mengenai *web scraping* melalui artikel-artikel dan juga beberapa dokumentasi yang ditemukan, penulis dapat memahami tahapan yang diperlukan, mulai dari persiapan hingga pelaksanaan *web scraping*.

Dalam melakukan *scraping*, terdapat beberapa kendala yang dialami oleh penulis seperti adanya kesalahan dalam penulisan *script* yang digunakan untuk *scraping*. Kesalahan-kesalahan dalam penulisan *script* terjadi karena diperlukan pembelajaran lebih lanjut mengenai bahasa pemrograman python pada tingkatan yang lebih tinggi. Kesalahan tersebut dievaluasi oleh mentor dan juga peserta magang lainnya, sehingga dapat diperbaiki dan penulis mengantisipasi adanya kesalahan serupa di kemudian hari. Program yang digunakan dibuat oleh penulis dengan bantuan mentor karena program untuk melakukan *scraping* yang dapat ditemukan di internet belum memberikan output sesuai dengan keinginan. *Scraping* yang dilakukan baru dapat dilakukan terhadap 1 akun karena adanya kendala dalam pembuatan kode pemrograman dimana belum menemukan cara untuk melakukan *scraping* terhadap lebih dari 1 akun.

Selain melakukan *scraping*, penulis juga terlibat dalam beberapa proyek kolaborasi yang dikerjakan bersama peserta magang lainnya. Beberapa proyek kolaborasi yang dikerjakan yaitu yang pertama, penulis diminta untuk melakukan pendataan terhadap tempat-tempat wisata yang ada di Daerah Istimewa Yogyakarta berdasarkan kategori yang telah ditentukan. Proyek kolaborasi yang kedua yaitu penulis diminta untuk melakukan pendataan dan rekapitulasi direktori putusan Pengadilan Negeri dan Pengadilan Agama dari 4 kabupaten dan 1 kota di provinsi Daerah Istimewa Yogyakarta. Untuk proyek kolaborasi yang ketiga, penulis diminta

untuk melakukan visualisasi perbandingan data Covid-19 dan data vaksinasi dengan data mobilitas di Daerah Istimewa Yogyakarta. Proyek kolaborasi yang terakhir yaitu penulis diminta untuk menganalisis sentimen publik berdasarkan kata kunci yang telah ditentukan serta mencari mengenai *update* berita hoaks yang beredar di Daerah Istimewa Yogyakarta.

Dalam melaksanakan magang di Dinas Komunikasi dan Informatika Daerah Istimewa Yogyakarta, sering kali kode pemrograman dibuat menggunakan bahasa pemrograman Python. Dengan adanya mata kuliah di Jurusan Informatika UII yang menggunakan bahasa pemrograman Python seperti *deep learning*, dapat menjadi bekal yang bermanfaat ketika membuat kode pemrograman karena sudah diberikan dasar dari bahasa pemrograman tersebut.

4.2 Nonteknis

Selama pelaksanaan program magang di Dinas Komunikasi dan Informatika Daerah Istimewa Yogyakarta, banyak hal dapat dijadikan pelajaran. Dalam setiap pengerjaan proyek dibutuhkan komunikasi, baik kepada sesama peserta magang maupun kepada mentor. Hal ini menuntut penulis untuk dapat menyesuaikan cara berkomunikasi sesuai dengan situasi dan kondisi. Pada awal kegiatan magang dilaksanakan, masih terdapat keraguan ketika diharuskan untuk berpendapat maupun bertanya. Seiring berjalannya waktu penulis dapat menyesuaikan diri dan mengurangi rasa ragu hingga akhirnya penulis tidak segan untuk berpendapat dan bertanya ketika terdapat kendala.

Lingkungan kerja yang sangat berbeda dengan lingkungan perkuliahan juga menjadi kendala pada tahap awal pelaksanaan magang. Setiap pekerjaan harus dikerjakan dengan sungguh-sungguh karena pekerjaan yang diberikan bukan hanya untuk kepentingan pribadi namun juga kepentingan bersama. Pada lingkungan kerja, penulis juga dituntut untuk dapat mempertanggungjawabkan setiap pekerjaan yang telah dikerjakan.

Salah satu manfaat terbesar dalam melakukan magang adalah dapat memperkaya relasi. Dengan relasi yang luas, memperbesar memungkinkan adanya kemudahan dalam mendapatkan pekerjaan nantinya, karena dengan begitu memungkinkan kita mendapatkan rekomendasi pekerjaan dari rekan kerja. Selain itu kegiatan magang juga dapat mengisi *curriculum vitae* (CV) sebagai pengalaman kerja, sehingga dapat menjadi pertimbangan oleh perusahaan saat melakukan pendaftaran kerja.

BAB V

KESIMPULAN

5.1. Kesimpulan

Selama melaksanakan magang di Dinas Komunikasi dan Informatika Daerah Istimewa Yogyakarta, penulis berkontribusi pada proyek *scraping* media sosial instagram, pendataan tempat wisata di daerah istimewa yogyakarta, pendataan dan rekapitulasi direktori putusan pengadilan, visualisasi perbandingan data covid-19 dan vaksinasi dengan data mobilitas, dan analisis sentimen publik berdasarkan kata kunci.

Seiring dengan berjalannya waktu, dengan pelaksanaan magang, baik *soft* maupun *hard skills* yang penulis miliki dirasakan telah meningkat. Hal tersebut juga dapat terjadi karena adanya bimbingan dari mentor yang senantiasa membantu serta memberikan arahan. Selain itu, dengan menjalani magang, penulis juga mendapatkan relasi baru serta menambah pengalaman yang dapat memberikan gambaran mengenai dunia kerja yang sesungguhnya.

5.2. Saran

Terdapat beberapa saran yang dapat disampaikan kepada beberapa pihak yang terkait mengenai pelaksanaan magang di Dinas Komunikasi dan Informatika Daerah Istimewa Yogyakarta, yaitu:

- Dinas Komunikasi dan Informatika Daerah Istimewa Yogyakarta

Setelah menyelesaikan kegiatan magang selama 6 bulan di Dinas Komunikasi dan Informatika, diharapkan waktu penerimaan mahasiswa magang dapat dimulai lebih cepat karena terdapat beberapa persiapan yang perlu dilakukan mahasiswa mengenai administrasi penjaluran.

- Prodi Informatika – Program Sarjana

Dengan banyaknya perusahaan yang telah menjalin kerja sama dengan Informatika UII, diharapkan informasi yang diberika kepada mahasiswa dapat disampaikan dengan jangka waktu yang cukup jauh dari batas pendaftaran penjaluran, sehingga mahasiswa dapat melakukan persiapan seperti memilih mata kuliah yang dapat mendukung berjalannya magang di perusahaan yang akan

dipilih. Pemantauan kepada mahasiswa yang sedang menjalani program magang agar lebih ditingkatkan lagi.



DAFTAR PUSTAKA

- Perez, F., Granger, B. E., & Hunter, J. D. (2011). Python: An Ecosystem for Scientific Computing. *Computing in Science & Engineering*, 13(2), 13–21
- Jarmul K, Lawson R. (2017). *Phyton Web Scraping* 2nd edition. Packt Publishing.
- Sahria Y. (2022) implementasi teknik web scraping pada Jurnal SINTA untuk analisis topik penelitian kesehatan Indonesia. *University Research Colloquium*. (pp 297–306). Yogyakarta
- Josi A., Abdillah L.A., Suryayusra. (2014) Penerapan teknik web scraping pada mesin pencari artikel ilmiah. (pp 159-164). Palembang
- Bansal, S. K., & Kagemann, S. (2015). *Integrating Big Data: A Semantic Extract-Transform-Load Framework*. *Computer*, 48(3), 42–50.
- Annur, C. (2021, November 15). Ada 91 Juta Pengguna Instagram di Indonesia, Mayoritas Usia Berapa. Retrieved from <https://databoks.katadata.co.id/datapublish/2021/11/15/ada-91-juta-pengguna-instagram-di-indonesia-mayoritas-usia-berapa>
- NapoleonCat. (2021, October) Instagram users in Indonesia. Retrieved from <https://napoleoncat.com/stats/instagram-users-in-indonesia/2021/10/>

LAMPIRAN

