

BAB I

PENDAHULUAN

1.1. Latar Belakang

Coronavirus adalah salah satu patogen utama yang terutama menargetkan sistem pernapasan manusia. Wabah sebelumnya dari coronavirus (CoV) termasuk sindrom pernafasan akut yang parah (SARS) -CoV dan sindrom pernapasan Timur Tengah (MERS) -CoV yang sebelumnya telah ditandai sebagai agen yang merupakan ancaman kesehatan masyarakat yang besar. Pada akhir Desember 2019, pasien rawat inap dirawat di rumah sakit dengan diagnosis awal pneumonia dari etiologi yang tidak diketahui. Pasien-pasien ini secara epidemiologis terkait dengan pasar grosir makanan laut dan hewan basah di Wuhan, Provinsi Hubei, Cina (Bogoch et al., 2020). Perkiraan timbulnya wabah Coronavirus pada laporan awal coronavirus, memiliki potensial menjadi pandemi mengingat pertumbuhan jumlah kasus untuk novel 2019 coronavirus (COVID-19, dinamai oleh WHO pada 11 Februari 2020) yang dianggap secara signifikan bertambah setiap waktu secara cepat (Zhao et al., 2020).

Gejala infeksi COVID-19 muncul setelah masa inkubasi sekitar 5,2 hari (Q. Li et al., 2020). Periode dari timbulnya gejala COVID-19 hingga kematian berkisar antara 6 hingga 41 hari dengan median 14 hari (P. Li et al., 2020). Periode ini tergantung pada usia pasien dan status sistem kekebalan pasien. Gejala infeksi yang ditimbulkan lebih pendek di antara pasien diatas 70 tahun dibandingkan dengan mereka yang berusia di bawah 70 (Wang et al., 2020). Gejala yang paling umum pada awal penyakit COVID-19 adalah demam, batuk, dan kelelahan, sementara gejala lainnya termasuk produksi dahak, sakit kepala, hemoptisis, diare, dyspnoea, dan limfopenia (Ren et al., 2020).

Menurut gugusan dari kesimpulan paragraf di atas disimpulkan bahwa, semakin hari semakin bertambahnya kasus Covid-19 yang banyak diantaranya diatas umur 70 tahun di seluruh dunia. Covid-19 di Indonesia telah menjadi sebuah virus mematikan yang tingkat pertumbuhannya di setiap provinsi sangat cepat dan terbanyak di Asia Tenggara dan berada di posisi ke 11 di dunia. Peneliti memerlukan sebuah data untuk melakukan evaluasi penanganan dan mengetahui statistik Covid-19 di seluruh provinsi Indonesia. Namun ketersediaan data Covid-19 masih terbilang tersebar (terpisah) dan sulit untuk pengaksesan datanya. Maka diperlukan sebuah sistem untuk melakukan

pengumpulan data secara terpusat dalam melakukan pencarian dan pengumpulan data yang diinginkan. Penelitian ini menggunakan bantuan mesin pencari untuk mencapai harapan data terpusat dan dapat diakses dengan mudah. Tidak dapat dipungkiri bahwa mesin pencari tidak terlepas dari *web crawler*. Mesin pencari (*web crawler*) berfungsi sebagai penjaring *web*, serta membantu mengunduh laman *web* dengan cepat dan menjadi salah satu komponen penting dari mesin pencari *web* yang menemukan halaman *web* dengan skala besar (Wardekar & Gupta, 2018).

Pada penelitian terdahulu (Shi et al., 2016) dijelaskan bagaimana implementasi menggunakan arsitektur *web crawler* yang bernama *scrapy*, lalu *scrapy* melakukan *crawling* informasi berita dari laman *web* dengan menggunakan metode *web crawler*. Pada akhirnya, *web crawler* dapat memantau laman *web* dan menyimpan informasi (Shi et al., 2016). Pada penelitian kali ini implementasi *web crawler* dengan arsitektur *beautifulsoup*, lalu *beautifulsoup* melakukan *crawling* informasi pada laman *web* yang dibutuhkan, dan menggunakan empat tahap metode sekaligus yaitu metode *web crawling*, *converting*, *preprocessing* dan *integrate*. Penelitian ini diharapkan dapat digunakan untuk mengumpulkan dan mengintegrasikan data-data tersebut secara otomatis dalam masalah data sebaran Covid-19 yang terpisah.

1.2. Identifikasi Masalah

Berdasarkan latar belakang di atas, beberapa masalah dapat diidentifikasi sebagai berikut:

1. Data tidak bersifat terpusat atau tergolong terpisah pada setiap provinsi.
2. Data masih sulit untuk dilakukan pengaksesannya.

1.3. Rumusan Masalah

Berdasarkan identifikasi masalah di atas, beberapa rumusan masalah sebagai berikut:

1. Bagaimana metode *crawling* untuk mendapatkan data berformat *pdf* dilaman *web* yang dirujuk?
2. Bagaimana mekanisme *preprocessing* untuk menghasilkan data yang baik, melakukan *converting* data pada setiap format berbeda, dan menyatukan seluruh data menjadi satu format hingga berbentuk *time series*?

1.4. Tujuan Penelitian

Adapun tujuan yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Dapat membangun aplikasi *web crawling* untuk mengunduh data yang dibutuhkan pada laman *web* yang dituju.
2. Dapat membangun aplikasi *web crawling* untuk menghasilkan data yang baik, melakukan *convert* pada setiap format *file* berbeda, dan menyatukan seluruh data menjadi satu format hingga terbentuk *time series*.

1.5. Manfaat Penelitian

Adapun Manfaat yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Memahami metode *web crawling*, metode *converting*, dan *preprocessing* data sehingga menghasilkan data yang terbaik untuk memudahkan dilakukan pengolahan lebih lanjut.
2. Mendapatkan data informasi pada *website* untuk mencari tahu perkembangan data sebaran Covid-19 di Indonesia.

1.6. Batasan Masalah

Adapun batasan masalah pada penelitian ini sebagai berikut:

1. Data *file* pada laman web yang *dicrawling* memiliki data dengan format yang berbeda.
2. *File* dari laman *web* yang *dicrawling* pada situs loker.bnpb.go.id adalah *file* yang berada dalam satu halaman *web* yang sama.
3. *Crawling* hanya sampai kedalaman level 1.
4. *File* yang hanya dapat di-*crawling* dan diunduh dengan bantuan *chrome driver*.