

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi menyebabkan banyaknya data yang beredar di internet yang menyebabkan data-data tersebut menjadi tidak berguna apabila tidak bisa dimanfaatkan dengan baik. Tidak hanya itu, perkembangan dari komputer juga memungkinkan komputer untuk dapat memproses data-data menjadi bentuk yang lebih kompleks atau dapat melakukan berbagai macam komputasi. Kedua hal tersebut menyebabkan berkembangnya satu cabang ilmu pengetahuan yang bernama *data analysis* di mana ilmu tersebut memanfaatkan serta mengolah data agar data tersebut dapat berguna dan memiliki manfaat. Hasil-hasil dari *data analysis* ini dapat menjadi faktor dalam pengambilan keputusan berdasarkan data yang sudah diolah (*data driven decision making*). Hal ini menjadi sangat penting untuk dapat membantu manusia dalam mengambil keputusan-keputusan penting yang sulit untuk diambil hanya dengan mengandalkan manusia itu sendiri.

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) merupakan virus yang menjadi penyebab pandemi yang tengah dihadapi oleh seluruh negara di seluruh penjuru dunia. Setidaknya sudah 223 negara yang terkonfirmasi kasus virus COVID-19. Pada Tanggal 5 Januari 2020 Organisasi Kesehatan Dunia atau biasa dikenal dengan WHO (*World Health Organization*) menyatakan bahwa terdapat wabah penyakit yang disebabkan oleh varian virus

baru. Kemudian, 25 hari setelahnya atau tepat pada 30 Januari 2020 WHO menyatakan bahwa COVID-19 merupakan wabah penyakit dalam skala besar atau pandemi dengan total kasus yang tercatat pada saat itu sebanyak 7818 kasus di seluruh penjuru dunia [1].

Berbagai macam upaya dilakukan baik dari pemerintah, organisasi, aktivis sosial maupun individu untuk dapat mengurangi penyebaran dan penularan virus COVID-19. Vaksinasi merupakan salah satu cara atau metode yang digunakan untuk menstimulasi infeksi dan mempersiapkan sistem imun terhadap penyakit yang sebenarnya di dalam tubuh [2]. Vaksinasi ini dilakukan melalui proses penyuntikan vaksin pada manusia. Tujuan dari vaksinasi adalah memberikan tubuh paparan terhadap mikroba yang telah dilemahkan atau sebagiannya saja agar tubuh dapat membentuk respon imun sehingga ketika tubuh terpapar dengan penyakit sebenarnya maka tubuh sudah siap [3]. Masalah yang dihadapi dengan terus berjalannya vaksinasi adalah efektivitas vaksinasi tersebut terhadap kasus baru.

Dengan adanya data vaksinasi diharapkan dapat dilihat hasil korelasi antara vaksinasi yang sudah dilakukan dengan kasus per hari. Kemudian, dari data yang ada juga akan dilakukan prediksi untuk melihat perkembangan kasus baru setiap harinya.

Penelitian ini diharapkan dapat menjawab permasalahan dari banyaknya data COVID-19 yang ada lalu mengolah data tersebut menjadi hal yang berguna dan melihat efektivitas vaksinasi terhadap kasus baru. Algoritma yang

digunakan dalam melakukan korelasi dan prediksi dalam penelitian ini adalah *Correlation Heatmap* serta *Linear Regression* dan *Random Forest Regression*. Selanjutnya, dari data yang sudah diolah tersebut dapat didapatkan hasil yang dapat dimanfaatkan sebagai dasar pengambilan keputusan atas peraturan atau kebijakan yang akan diambil dari suatu negara. Dengan demikian diharapkan juga pandemi yang tengah terjadi akan segera reda dengan memanfaatkan sumber daya ilmu pengetahuan yang ada yang dijadikan landasan sebagai bahan pertimbangan pembuatan keputusan.

1.2. Maksud dan Tujuan

Skripsi ini memiliki tujuan untuk menganalisis setiap data yang ada dari *dataset* yang sudah didapatkan. Kemudian, dari data tersebut akan dilihat korelasi antara parameter satu dengan yang lain dan juga model prediksinya. Setelah itu, analisis deskriptif akan dilakukan yang bertujuan untuk memperlihatkan bagaimana hubungan dari korelasi dan juga prediksi yang dapat digunakan sebagai bahan pertimbangan pengambilan kebijakan pemerintah untuk dapat menangani pandemi yang tengah terjadi.

1.3. Batasan Masalah

Pada skripsi ini terdapat batasan masalah sebagai berikut:

1. *Dataset* yang digunakan adalah *dataset* yang berasal dari <https://ourworldindata.org>
2. Metode korelasi yang digunakan adalah *Correlation Heatmap*.

3. Metode prediksi yang digunakan adalah prediksi dengan menggunakan algoritma *machine learning Linear Regression* dan *Random Forest Regression*.
4. Adapun bahasa pemrograman yang digunakan adalah *Python 3* serta *library machine learning* yang digunakan adalah *Scikit-learn*.
5. Negara yang akan diambil datanya adalah negara Inggris, India, Indonesia, dan Amerika.

1.4. Metode Penelitian

Penelitian ini terdiri dari beberapa langkah antara lain studi literatur, mengintegrasikan *dataset* dan memperbarui *dataset*, *data cleansing*, *Exploratory Data Analysis* (EDA), pembuatan model dan pengujian model yang telah dibuat.

1.5. Sistematika Penelitian

Skripsi ini terdiri dari enam bab dengan sistematika penulisan sebagai berikut:

1. BAB I: PENDAHULUAN

Bab ini berisikan tentang latar belakang penelitian, tujuan penelitian, batasan masalah, metode penelitian dan sistematika penulisan.

2. BAB II: LANDASAN TEORI

Bab ini berisi teori dasar tentang *data science*, *Python 3*, *Jupyter Notebook*, *Anaconda*, *machine learning*, *predictive analytics*, *correlation*, *Correlation Heatmap*, dan *Linear Regression*.

3. BAB III: METODE PENELITIAN

Bab ini membahas mengenai langkah pengerjaan penelitian meliputi detail dari sistem yang digunakan.

4. BAB IV: PERANCANGAN

Bab ini memuat tentang proses yang dilakukan penulis dalam melakukan penelitian.

5. BAB V: ANALISIS DAN PEMBAHASAN

Bab ini menjelaskan tentang hasil dari analisis serta penjelasannya.

6. BAB VI: PENUTUP

Bab ini memuat kesimpulan dari penelitian serta saran untuk penelitian lebih lanjut.

