

## ABSTRAK

Amry Maulana (01032180019)

### **ANALISIS KORELASI DAN ANALISIS PREDIKTIF PERKEMBANGAN DATA VAKSINASI COVID-19 DENGAN MENGGUNAKAN *MACHINE LEARNING***

Skripsi, Fakultas Sains dan Teknologi 2021

(xiv + 80 halaman; 54 gambar; 4 tabel; 93 halaman lampiran)

Perkembangan teknologi dapat memudahkan kita untuk dapat mengakses dan memanfaatkan data dengan baik. *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. Untuk dapat mengatasi virus SARS-CoV-2 maka salah satu upaya yang dilakukan adalah vaksinasi. Masalah yang dihadapi dengan terus berjalannya vaksinasi adalah efektivitas vaksinasi. Selain itu, hal ini juga dapat menyebabkan banyaknya data yang beredar di internet akan menjadi tidak berguna apabila tidak bisa dimanfaatkan dengan baik. Oleh karena itu, penelitian ini dilakukan untuk menjawab persoalan efektivitas vaksinasi dan pemanfaatan data COVID-19. Pada penelitian ini terdapat empat negara yang dianalisis, negara tersebut adalah negara Inggris, India, Indonesia, dan Amerika. Pada penelitian ini dibuat model *correlation* dan juga *prediction* dari *dataset* COVID-19 yang sudah didapatkan dengan menggunakan *machine learning*. Kemudian, dari data tersebut dilihat bagaimana hubungan dari *correlation* dan *prediction* yang dapat digunakan sebagai bahan pertimbangan pengambilan kebijakan. Model *correlation* yang dibuat adalah model *Correlation Heatmap* yang dibagi menjadi tiga bagian yaitu dengan menggunakan variabel *new\_vaccinations*, data dari *first wave*, dan data dari *second wave* serta model *Correlation Heatmap* untuk penurunan kasus pada *first wave* dan *second wave*. Hasil *correlation* dari negara Indonesia sendiri cukup mengkhawatirkan di mana dengan besar nilai koefisien *correlation* pada variabel *new\_cases* dan *new\_deaths* sebesar 0.95 mengindikasikan bahwa meskipun negara ini sudah melakukan upaya sebaik mungkin untuk dapat mengatasi pandemi yang tengah terjadi ini. Nampaknya Indonesia masih kurang bisa menurunkan angka kematian akibat COVID-19. Model *prediction* yang digunakan pada penelitian ini menggunakan *Linear Regression* dan *Random Forest Regression* yang digunakan untuk dapat memprediksi kasus harian dari pandemi COVID-19 dan dilakukan evaluasi dengan menggunakan MSE (*Mean Square Error*), dan RMSE (*Root Mean Square Error*). Kemudian, hasil dari evaluasi tersebut digunakan pada setiap negara pada penelitian ini.

Kata kunci : COVID-19, *Correlation Heatmap*, *Linear Regression*, *Random Forest Regression*, *Correlation*, *Prediction*

Referensi : [1] (2010-2021)

## ABSTRACT

Amry Maulana (01032180019)

### **CORRELATION ANALYSIS AND PREDICTION ANALYSIS OF COVID-19 VACCINATION DATA USING MACHINE LEARNING**

Thesis, Faculty of Science and Technology 2021

(xiv + 80 pages; 54 figures; 4 table; 93 appendices)

The development in technology can make it easier for us to be able to access and make good use of data. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is a virus that is causing a pandemic that is being faced by many countries all around the world. To be able to overcome the SARS-CoV-2 virus, one of the efforts made is vaccination. The problem with continuing vaccination is the effectiveness of vaccination itself. In addition, this can also cause the amount of data circulating on the internet to be useless if it cannot be used properly. Therefore, this study was conducted to answer the issue of the effectiveness of vaccination and the utilization of COVID-19 data. In this study, there are four countries to be analyzed, these countries include the United Kingdom, India, Indonesia, and United States. In this study, correlation and prediction models were made from the COVID-19 dataset that had been obtained using machine learning. Then, from the data seen how the relationship between correlation and prediction can be used as consideration for policymaking. The correlation model made is the Correlation Heatmap model, which is divided into three parts: the new\_vaccinations variable, data from the first wave and data from the second wave, and the Correlation Heatmap model for case reductions in the first wave and second wave. The correlation results from Indonesia itself are quite worrying, with the large correlation coefficient on the new\_cases and new\_deaths variables of 0.95, even though this country has made its best efforts to overcome this ongoing pandemic. It seems that Indonesia is still unable to reduce the number of deaths due to COVID-19. The prediction model used in this study uses Linear Regression and Random Forest Regression which is used to predict new cases of the pandemic COVID-19 and evaluated using MSE (Mean Square Error) and RMSE (Root Mean Square Error). Then, the evaluation results were used for each country in this study.

Keyword : COVID-19, Correlation Heatmap, Linear Regression, Random Forest Regression, Correlation, Prediction

Reference : [1] (2010-2021)