**PROPOSAL TUGAS AKHIR IF540 – MACHINE LEARNING**

**PREDIKSI KASUS TERINFEKSI COVID-19 DI**

**INDONESIA MENGGUNAKAN SIMPLE LINEAR**

**REGRESSION MACHINE LEARNING**



**Disusun Oleh Kelompok 8:**

**Felix Sutanto-00000045539**

**Jeffriel Putra Gunawan-00000052616**

**Shyehan Rafael Adlinugroho-00000052738**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS TEKNIK DAN INFORMATIKA**

**UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2022DAFTAR ISI**

**ABSTRAKSI**

# JUDUL: PREDIKSI KASUS TERINFEKSI COVID-19 DI INDONESIA MENGGUNAKAN SIMPLE LINEAR REGRESSION MACHINE LEARNING

## 1. Latar Belakang Masalah

Penyakit virus corona (COVID-19) telah menjadi pandemi di dunia [1]. COVID-19 adalah virus penyebab pneumonia, juga dikenal sebagai *Corona Virus Disease* atau penyakit virus corona. Wabah pertama ditemukan di Wuhan, Cina, di provinsi Hubei pada Desember 2019 [2]. Ketika Organisasi Kesehatan Dunia (WHO) mendeklarasikan pandemi penyakit coronavirus di seluruh dunia (COVID19) dan memperkirakan kesehatan darurat global pada 11 Maret 2020, aktivitas dunia berubah menjadi yang terburuk. *Severe Acute Respiratory Syndrome Coronavirus* 2 (SARS-COV2), satu-satunya penyakit yang berpotensi mematikan, adalah penyebab COVID-19. Penularannya menyebar melalui droplet dari orang ke orang, sehingga pemerintah membuat kebijakan menjaga jarak, memakai masker, dan mencuci tangan sebagai bagian dari protokol kesehatan. Virus ini mempengaruhi sel inang dan dapat menyebabkan reaksi terkait peradangan. Masa inkubasi hingga 14 jam dapat menyebabkan kematian dan mengindikasikan demam, leukopenia, sindrom pernapasan, trombositopenia, dan kondisi multiorgan. [3] Pemahaman tentang ciri-ciri virus ini diperlukan untuk mengendalikan angka kematian.

Menurut data Humanitarian Data Exchange (HDX), epidemi pertama di Indonesia pertama kali tercatat pada 2 Maret 2020, dan pada 26 Juli 2020 telah menyebar ke lebih dari 90.000 orang [4] dan telah mempengaruhi lebih dari 200 negara di seluruh dunia [5]. Menyikapi lonjakan kasus di Indonesia setelah pengumuman tersebut, beberapa provinsi di Indonesia menerapkan pembatasan sosial berskala besar, dan lembaga serta sekolah ditutup akibat pembatasan tersebut. Sementara itu, kualitas udara di lingkungan luar biasa dan patut diperhatikan karena memiliki tingkat kontaminasi ringan akibat pencemaran udara di lingkungan [6]. Lonjakan kasus terus meningkat selama beberapa tahun terakhir, sehingga berbagai metode digunakan untuk mengendalikan penyebaran virus ini.

Pemerintah telah menerapkan langkah-langkah seperti pembatasan aktivitas di luar rumah, ibadah yang dilakukan di rumah, pembelajaran yang dilakukan daring dari rumah, dan langkah-langkah lain untuk menghentikan perkembangan COVID-19 di Indonesia. Pertambahan kasus baru yang relatif masih tinggi mengindikasikan bahwa Indonesia belum melewati puncak pandemi. Kapan kasus positif COVID-19 di Indonesia akan mencapai puncaknya masih menjadi misteri karena penularan virus yang cepat dan peningkatan jumlah kasus positif. Dalam wawancara dengan BBC News Indonesia, Wiku Adisasmito (2020) menyatakan bahwa puncak pandemi masih belum dapat diprediksi karena kasus sangat dinamis dengan perilaku masyarakat Indonesia [7]. Oleh karena itu, diperlukan penelitian untuk menghitung dan menghasilkan data prediksi puncak kasus COVID-19 di Indonesia. Hasil investigasi ini dapat membantu para ilmuwan dan profesional lainnya dalam mengembangkan studi atau desain yang dapat membantu memperlambat atau menghentikan pertumbuhan virus COVID-19 di Indonesia.

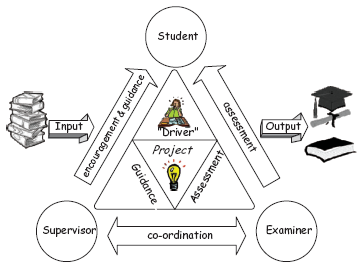
Algoritma *Machine Learning* telah digunakan di banyak aplikasi prediksi, terutama untuk masalah deret waktu. Salah satu algoritma yang banyak digunakan adalah Linear Regression. Berdasarkan data yang disajikan dalam makalah ini, akan dicoba untuk memodelkan prediksi jangkauan virus, jumlah kasus di Indonesia. Sepanjang makalah ini, model *Machine Learning* yang digunakan dalam makalah ini adalah model *Machine Learning Linear Regression*. Berdasarkan model yang diperoleh dalam prediksi, kita dapat memeriksa implikasi yang lebih luas untuk suatu wilayah serta tren baik dalam jangka pendek maupun jangka panjang dalam hal implikasi untuk wilayah tersebut [8]. Untuk mengevaluasi hasil dari pandemi ini, prediksi ini digunakan untuk menghitung intensitas wabah. Penting bagi kebijakan publik dan kesadaran individu untuk bekerja sama dalam mengatasi pandemi ini. Menurut beberapa peneliti, teknik kecerdasan komputasional masih dianggap cukup baik dan banyak digunakan dalam berbagai aplikasi. Akibatnya, dapat dikatakan bahwa kinerja dapat mencapai tingkat penggunaan yang dapat diterima ketika evaluasi dapat diukur. Itu membuat teknik ini dan algoritma *Machine Learning* yang dihasilkan adalah solusi untuk masalah di masa depan.

Penelitian yang dilakukan oleh Ashish U Mandayam, Rakshith A.C, S Siddesha, dan S K Niranjan [9] mengenai *Machine Learning* untuk memprediksi jumlah kasus positif di masa depan untuk tindakan dan kontrol yang lebih baik. Untuk mempelajari kinerja prediksi, dilakukan perbandingan antara *Linear Regression* dan *Support Vector Regression* (SVR) menunjukkan bahwa *Linear Regression* memiliki hasil yang lebih baik dibandingkan dengan SVR. Hal ini dibuktikan dengan nilai , *Mean Absolute Error* (MAE) dan *Mean Squared Error* (MSE) berturut-turut sebesar 0.99632163, 56956.1964 dan 4317204498, sedangkan metode SVR memiliki nilai sebesar 0.3061489, nilai MAE sebesar 1128135.156, dan nilai MSE sebesar 1.3784E+12. Kedua metode diuji pada pembagian persentase data train dan data test yang sama yaitu 85:15.

Penelitian serupa juga sudah dilakukan oleh Ramesh Kumar Mojjada, Arvind Yadav, A.V. Prabhu, dan Yuvaraj Natarajan [10]. Penelitian ini secara khusus berfokus pada prediksi COVID-19 dan penelitian ini bahkan berfokus pada prediksi dan respon awal wabah COVID-19 Untuk memantau situasi saat ini dengan pengambilan keputusan melalui proses prediksi untuk mengarahkan intervensi sebelumnya dalam manajemen yang efisien dari penyakit. Analisis ini didasarkan pada beberapa *Machine Learning* *Regression Model* (LR), *Lowest Absolute and Selective Shrinking Operator* (LASSO), *Vector Supports* (SMS) dan *Exponential Smoking* (ES). Model dilatih dengan data dari statistik pasien COVID-19 Johns Hopkins. Pengumpulan data disiapkan dan dibagi menjadi dua sub-set: *preparation* (85%) dan *assesment* (15%). Kinerja masing-masing model pembelajaran dievaluasi dalam penilaian  , *Modified R-Square* (), MSE, *Mean Absolute Error* (MAE) dan *Root Means Square Error* (RMSE). Dalam analisis ini, kinerja setiap model pembelajaran dievaluasi. Temuan penelitian menunjukkan bahwa mengingat keberadaan dan dimensi kumpulan data, ES berkinerja lebih baik dalam domain perkiraan saat ini. LR dan LASSO keduanya efektif dalam memperkirakan dan memverifikasi tingkat kematian sampai tingkat tertentu. Efek naik turunnya pengumpulan data sangat buruk dicapai oleh SVM.

Berdasarkan data dari penelitian-penelitian sebelumnya, telah dibuktikan bahwa penggunaan metode *Linear Regression* untuk melakukan prediksi memiliki reputasi akurasi yang baik. Model *Linear Regression* relatif sederhana yang dapat menghasilkan prediksi. *Linear Regression* adalah teknik yang sangat fleksibel yang dapat digunakan untuk menjawab berbagai pertanyaan penelitian dan tujuan penelitian, karenanya *Linear Regression* dapat diterapkan pada berbagai bidang dalam studi bisnis dan akademis. Penggunaan *Linear Regression* dapat dijumpai dalam segala hal mulai dari ilmu biologi, perilaku, lingkungan, dan sosial hingga bisnis. Karena *Linear Regression* adalah prosedur statistik yang sudah lama ada, sifat-sifat model *Linear Regression* dipahami dengan baik dan dapat dilatih dengan sangat cepat.

Regresi adalah istilah statistik yang digunakan untuk menggambarkan model yang memperkirakan hubungan antar variabel. Model *Linear Regression* mempelajari hubungan antara satu variabel terikat Y dan satu atau lebih variabel bebas, dilambangkan dengan X. Ada dua jenis *Linear Regression*: Sederhana (Simple) dan Berganda (Multiple) [11]. Jika hanya ada satu variabel bebas X, disebut *Simple Linear Regression*; jika terdapat lebih dari satu variabel bebas disebut *Multiple Linear Regression*. Pada penelitian ini hanya akan berfokus pada *Simple Linear Regression*.



Gambar 1. Contoh gambar skema pembelajaran [12]

## 2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan diatas, maka rumusan

masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana memprediksi puncak COVID-19 di Indonesia dengan metode *Simple Linear Regression*?
2. Bagaimana cara mengukur hasil yang dihasilkan menggunakan perhitungan *Mean Squared Error* (MSE), , dan *Mean Absolute Error* (MAE)?

## 3. Batasan Masalah

Batasan masalah pada penelitian ini dapat dijabarkan menjadi sebagai berikut:

1. Batasan area yang dijadikan objek penelitian adalah wilayah yang berada di Indonesia.
2. Sampel data pada penelitian ini menggunakan data penyakit COVID-19 yang tercatat sejak 22 Januari 2020 sampai dengan 15 November 2022. Dataset diambil situs web resmi github John Hopkins University [sumber: https://github.com/CSSEGISandData/COVID-19/tree/master/csse covid 19 data]

## 4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk:

1. Membangun model yang akan memprediksi jumlah kasus penyakit COVID-19 berdasarkan jumlah penyakit COVID-19 harian.
2. Menilai hasil dari metode *Simple Linear Regression* yang digunakan dengan *Mean Squared Error* (MSE), , dan *Mean Absolute Error* (MAE).

## 

## 5. Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Memberikan ilmu pengetahuan tentang metode *Linear Regression* pada data kasus COVID-19 di Indonesia.
2. Memberikan gambaran kasus COVID-19 di Indonesia dan menunjukkan hasil evaluasi paling baik.

## 6. Telaah Literatur

**6.1. *Machine Learning***

Apa itu *Machine Learning*? Anda mungkin menggunakannya puluhan kali sehari tanpa menyadarinya. Setiap kali Anda menggunakan mesin pencari web seperti Google atau Bing untuk mencari di internet, salah satu alasan yang bekerja dengan baik adalah karena algoritma pembelajaran, yang diterapkan oleh Google atau Microsoft, telah mempelajari cara menentukan *rank* halaman web. Saat aplikasi foto Facebook atau Apple mengenali teman Anda di gambar Anda, itu juga *Machine Learning*. Setiap kali Anda membaca email Anda dan filter spam menlindungi Anda dari banyak spam, itu karena komputer Anda telah belajar membedakan spam dari email non-spam. Jadi, itulah *Machine Learning*. Ada ilmu untuk membuat komputer belajar tanpa diprogram secara eksplisit.

Dua definisi *Machine Learning* ditawarkan. Arthur Samuel menggambarkannya sebagai: ”bidang studi yang memberi komputer kemampuan untuk belajar tanpa diprogram secara eksplisit.” Ini adalah definisi informal yang lebih tua. Tom Mitchell memberikan definisi yang lebih modern: ”Sebuah program komputer dikatakan belajar dari pengalaman E sehubungan dengan beberapa kelas tugas T dan ukuran kinerja P, jika kinerjanya pada tugas-tugas di T, yang diukur dengan P, meningkat dengan pengalaman E.” *Machine Learning* adalah teknologi khusus yang berada di bawah *artificial intelligence* (AI). Cabang AI ini berfokus pada penggunaan data dan algoritma untuk meniru *human learning*, memungkinkan mesin untuk berkembang seiring waktu, menjadi semakin akurat saat membuat prediksi atau klasifikasi. *Machine Learning* bekerja dalam tiga cara dasar, dimulai dengan menggunakan kombinasi data dan algoritma untuk memprediksi pola dan mengklasifikasikan kumpulan data, *error function* yang membantu mengevaluasi akurasi, dan kemudian proses optimasi untuk menyesuaikan titik data ke dalam model yang terbaik.

*Machine Learning* melibatkan menampilkan sejumlah besar data ke mesin sehingga dapat mempelajari dan membuat prediksi, menemukan pola, atau mengklasifikasikan data. Tiga jenis *Machine Learning* adalah *supervised learning*, *unsupervised learning*, dan *reinforcement learning*.

**Definisi 1.** Cara penulisan definisi seperti yang telah diberikan. Sementara untuk penulisan rumus atau persamaan seperti berikut [13].

(1)

Penulisan tabel dapat dilihat pada contoh berikut.

Tabel 1. Contoh tabel dalam proposal [14]

|  |  |
| --- | --- |
| Level () |  |
| 1 | 0 |
| 2 | 1 |
| … | … |

**6.1.1. *Supervised Learning***

Perusahaan konsultan bisnis Gartner memperkirakan bahwa *supervised learning* akan tetap menjadi metode *Machine Learning* paling populer di antara para pemimpin perusahaan teknologi informasi pada tahun 2022. Dengan memasukkan data *input* dan *output* historis ke dalam algoritma *Machine Learning*, jenis *Machine Learning* ini memungkinkan algoritma untuk mengubah model untuk membuat *output* sedekat mungkin dengan hasil yang diinginkan. Algoritma umum yang digunakan selama *supervised learning* meliputi *Neural Networks*, *Decision Trees*, *Linear Regression*, dan *Support Vector Machines*. Jenis *Machine Learning* ini mendapatkan namanya karena mesin ”supervised” saat sedang belajar, yang berarti Anda memasukkan informasi algoritma untuk membantunya belajar. Hasil yang Anda berikan ke mesin diberi label data, dan sisa informasi yang Anda berikan digunakan sebagai fitur *input*.

**6.2. Literatur B**

Perhatikan untuk memilih teori yang benar-benar perlu untuk dijelaskan saja dan bukan merupakan pengetahuan umum [15]. Kriteria utama dalam penerimaan atau penolakan usulan topik suatu Proposal Skripsi sangat tergantung pada kejelasan dan substansi isi dari topik Skripsi yang dituangkan dalam Proposal Skripsi ini. Penulisan literatur wajib disertai dengan sitasi sumber pustakanya [16].

## 7. Metodologi Penelitian

Pada bagian ini dijabarkan langkah-langkah yang hendak dilakukan dalam menyusun dan mengerjakan penelitian. Langkah-langkah penelitian yang dijabarkan dimulai dari awal hingga akhir selesai (dokumentasi). Beberapa contohnya yang dapat diikuti antara lain studi literatur, pengumpulan data, pengambilan sampel, analisis kebutuhan, perancangan, implementasi, pengujian, evaluasi, dan dokumentasi [17].

## 8. Spesifikasi Sistem

Pada bagian ini dijelaskan secara singkat *software* dan *hardware* yang akan digunakan untuk pengerjaan penelitian. [18].

**9. Jadwal Penelitian**

Pada bagian ini dijelaskan rencana waktu jadwal penelitian dari inisialisasi hingga pembuatan laporan menggunakan Gantt Chart seperti pada contoh di Tabel 2 untuk memberikan gambaran waktu tentang kegiatan penelitian.

Tabel 2. Rencana waktu penelitian

| Kegiatan | Minggu ke- | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | … | | | | n |
| Isi sesuai metode penelitian |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |  | | | |  |
|  |  |  |  |  |  |  |  |  |  | | | |  |
|  |  |  |  |  |  |  |  |  |  | | | |  |
|  |  |  |  |  |  |  |  |  |  | | | |  |
|  |  |  |  |  |  |  |  |  |  | | | |  |
|  |  |  |  |  |  |  |  |  |  | | | |  |
|  |  |  |  |  |  |  |  |  |  | | | |  |
|  |  |  |  |  |  |  |  |  |  | | | |  |
|  |  |  |  |  |  |  |  |  |  | | | |  |
|  |  |  |  |  |  |  |  |  |  | | | |  |
|  |  |  |  |  |  |  |  |  |  | | | |  |
|  |  |  |  |  |  |  |  |  |  | | | |  |
|  |  |  |  |  |  |  |  |  |  | | | |  |

Tabel 2. Rencana waktu penelitian (lanjutan)

| Kegiatan | Minggu ke- | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | … | n |
| Isi sesuai metode penelitian |  |  |  |  |  |  |  |  |  |  |  |

## 10. Hasil Penelitian

Pada bagian ini dijelaskan demo dari system yang sudah dibangun, seperti langkah-langkah penggunaan system beserta tangkapan layar, disertai alamat url dari system yang sudah publik. [18].

## 11. Evaluasi dan Kesimpulan

Pada bagian ini dijelaskan evaluasi pada system yang dilakukan terhadap expert atau expert knowledge yang didapatkan dari referensi, dilakukan benchmarking hasil apakah system dapat mengakomodir keperluan kasus yang diambil beserta kesimpulannya. [18].

**12. Daftar Pustaka**

**\*IEEE** *citation style***dan jumlah referensi paling sedikit 10 (kurun waktu tidak lebih dari 10 tahun terakhir dan minimal 70% artikel internasional)** yang dapat bersumber dari: buku, jurnal, conference proceeding, situs web (institusi / perusahaan / organisasi), disertasi, dan wawancara.

[1] X. Yuan, J. Xu, S. Hussain, H. Wang, N. Gao, and L. Zhang, “Trends and prediction in daily incidence and deaths of COVID-19 in the United States: a search-interest based model,” *medRxiv*, Oct. 2020, doi: 10.1101/2020.04.15.20064485.

[2] Z. Yang *et al.*, “Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions,” *J Thorac Dis*, vol. 12, no. 3, pp. 165–174, Oct. 2020, doi: 10.21037/JTD.2020.02.64.

[3] M. Sukmana and F. A. Yuniarti, “The Pathogenesis Characteristics and Symptom of Covid-19 in the Context of Establishing a Nursing Diagnosis,” *Jurnal Kesehatan Pasak Bumi Kalimantan*, vol. 3, no. 1, pp. 21–28, Oct. 2020, doi: 10.30872/J.KES.PASMI.KAL.V3I1.3748.

[4] S. Setiati and M. K. Azwar, “COVID-19 and Indonesia,” *Acta Med Indones*, vol. 52, no. 1, p. 84, Oct. 2020, [Online]. Available: https://www.actamedindones.org/index.php/ijim/article/view/1426

[5] D. Ivanov, “Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case,” *Transp Res E Logist Transp Rev*, vol. 136, p. 101922, Oct. 2020, doi: 10.1016/J.TRE.2020.101922.

[6] J. Wang, Y. Zhou, Y. Sun, and F. W. Wibowo, “Prediction of air quality in Jakarta during the COVID-19 outbreak using long short-term memory machine learning,” *IOP Conf Ser Earth Environ Sci*, vol. 704, no. 1, p. 12046, Oct. 2021, doi: 10.1088/1755-1315/704/1/012046.

[7] C. Wijaya, “Virus corona: ‘Puncak pandemi Covid-19 di Indonesia sulit diprediksi akibat jumlah tes Covid-19 yang tak konsisten’, ujar peneliti - BBC News Indonesia.” Oct. 2020. [Online]. Available: https://www.bbc.com/indonesia/indonesia-52616740

[8] F. Liu *et al.*, “Predicting and analyzing the COVID-19 epidemic in China: Based on SEIRD, LSTM and GWR models,” *PLoS One*, vol. 15, no. 8, Oct. 2020, doi: 10.1371/JOURNAL.PONE.0238280.

[9] A. U. Mandayam, A. C. Rakshith, S. Siddesha, and S. K. Niranjan, “Prediction of Covid-19 pandemic based on Regression,” *Proceedings - 2020 5th International Conference on Research in Computational Intelligence and Communication Networks, ICRCICN 2020*, pp. 1–5, Oct. 2020, doi: 10.1109/ICRCICN50933.2020.9296175.

[10] R. K. Mojjada, A. Yadav, A. v Prabhu, and Y. Natarajan, “Machine learning models for covid-19 future forecasting,” *Mater Today Proc*, 2021, doi: 10.1016/J.MATPR.2020.10.962.

[11] S. I. Bangdiwala, “Regression: simple linear,” *https://doi.org/10.1080/17457300.2018.1426702*, vol. 25, no. 1, pp. 113–115, Oct. 2018, doi: 10.1080/17457300.2018.1426702.

[12] M. a. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*. 2011. doi: 10.1017/CBO9780511976667.

[13] H. G. Munthe, “Sistem Pendukung Keputusan Penentuan Prioritas Usulan Sertifikasi Guru Dengan Metode Simple Additive Weighting,” 2013.

[14] D. J. Moylett, N. Linden, and A. Montanaro, “Quantum speedup of the traveling-salesman problem for bounded-degree graphs,” *Phys Rev A  (Coll Park)*, 2017, doi: 10.1103/PhysRevA.95.032323.

[15] M. Kumar, M. Husain, N. Upreti, and D. Gupta, “Genetic Algorithm: Review and Application,” *SSRN Electronic Journal*, 2020, doi: 10.2139/ssrn.3529843.

[16] E. Rieffel and W. Polak, “Quantum computing: a gentle introduction,” *Choice Reviews Online*, 2011, doi: 10.5860/choice.49-0911.

[17] G. Llorach and J. Blat, “Say hi to eliza: An embodied conversational agent on the web,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017. doi: 10.1007/978-3-319-67401-8\_34.

[18] O. Kramer, *Genetic Algorithm Essentials*. 2017. doi: 10.1007/978-3-319-52156-5.