

Analisis Kinerja Support Vector Machine (SVM) Pada Klasifikasi Jumlah Kasus Bulanan Covid-19

Rifaldo Sitepu¹, Aniq A. Rohmawati, S.Si., M.Si.², Sri Suryani Prasetyowati, S.Si., M.Si.³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹rifaldositepu@students.telkomuniversity.ac.id, ²aniqatiqi@telkomuniversity.ac.id,

³srisuryani@telkomuniversity.ac.id

Abstrak

Covid 19 merupakan sebuah penyakit yang disebabkan oleh virus corona baru yang dinamai *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2). Pada dasarnya penyakit yang disebabkan oleh virus tersebut mengakibatkan infeksi akut pada sistem pernafasan. Penyebaran wabah Covid 19 sangat cepat menyebar keseluruh dunia. Para ahli menilai bahwa faktor penyebaran yang cepat dan semakin luas disebabkan oleh pola mobilitas manusia. Sehingga beberapa negara melakukan beberapa kebijakan untuk mencegah dan menekan pola mobilitas seperti melarang kerumunan massal, menjaga jarak, memakai masker, dan melakukan pembatasan perjalanan baik antar wilayah maupun antar negara. Penelitian ini mengusulkan untuk menganalisa kinerja *Support Vector Machine* (SVM) pada klasifikasi terhadap jumlah kasus bulanan covid-19. Dataset yang digunakan adalah data covid-19 perkelurahan di kota Bandung bulan November 2020 hingga bulan Desember 2021 yang berasal dari Kantor Dinas Kesehatan Kota Bandung. Data yang dikumpulkan juga termasuk data vaksinasi covid-19, kepatuhan memakai masker dan menjaga jarak. Selain itu, peneliti juga menggunakan data iklim yang berasal dari data BMKG stasiun Bandung, seperti suhu, curah hujan, dan penyinaran matahari. Dari hasil evaluasi yang dilakukan didapat hasil akurasi terbaik pada bulan Desember 2021 dengan akurasi 100%, dilanjutkan dengan bulan Juli dan Agustus 2021 dengan akurasi 97% dan Oktober dengan akurasi 90%. Dari hasil tersebut juga di dapat bahwa rata rata yang menghasilkan akurasi, precision, f1-score dan recall terbaik ada pada kernel RBF, Sehingga dapat disimpulkan bahwa support vector machine (svm) cukup baik dalam mengklasifikasi jumlah kasus bulanan covid-19.

Kata kunci : klasifikasi, covid-19, support vector machine

Abstract

Covid-19 is disease caused by the new corona virus called *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2). The effect of this virus usually causes infection on respiratory system. Covid-19 was rapidly spread globally. Experts said that the factor that caused this to spread rapidly is human mobility. Therefore, several countries create new rules so that it can suppress the spreading of this disease, by prohibiting a large scale gathering, keeping away distance with each other, mandatory rule of using mask, and the prohibition for the entry of their country. This research proposes a performance analysis of *Support Vector Machine* (SVM) to classify the monthly data of covid-19. The dataset used was the covid-19 data of towns in Bandung from November 2020 to December 2021 which was acquired from The Public Health Office of Bandung. The data that was collected includes the vaccination data of covid-19, the obligatory to use mask in public and keeping up distance with each other. This research also uses the weather data from BMKG Bandung, which includes temperature, rainfall pattern, and sunshine pattern. From the evaluation it is found that the best accuracy comes from December 2021 with 100% accuracy followed by August 2021 with 97% accuracy and October 2021 with 90%. From the result it is also found that the average that produces the best accuracy, precision, f1-score, and recall came from RBF kernel. Which can be concluded that support vector machine (SVM) is good to classify the amount of active cases of covid-19.

Keywords: classification, covid-19, support vector machine

1. Pendahuluan

1.1 Latar Belakang

Wabah pneumonia antipikal akut pertama kali dilaporkan oleh pejabat di kota Wuhan, provinsi Hubei, China, tepatnya pada bulan Desember 2019 [1]. Penyakit tersebut diketahui disebabkan oleh virus corona baru yang dinamai *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2)[1],[4],[6]. Penyakit yang disebabkan oleh virus SARS-COV-2 tersebut dinamai Covid-19 [2],[3],[4],[5],[6]. Penyebaran wabah Covid-19 dengan cepat terus menyebar ke seluruh negara china dan menyebar ke seluruh dunia, hingga pada 11 Maret 2020 Organisasi Kesehatan Dunia (WHO) menyatakan dan mengklasifikasikan penyakit yang di sebabkan oleh SARS-CoV-2 atau Covid 19 sebagai pandemi [2],[4],[6].

Indonesia sebagai negara dengan total penduduk 270,20 juta jiwa berdasarkan data tahun 2020 juga ikut merasakan dampak dari penyebaran virus SARS-COV-2[17]. Dimana pada Januari 2021 Indonesia masih terus menunjukkan tren kasus yang terus meningkat, dimana hingga per 7 Januari 2021 kasus positif bertambah 9.321 kasus, dimana sudah sebanyak 23.520 kasus orang meninggal dunia, 659.437 kasus sembuh, dan kasus aktif berjumlah 114.766 kasus, yang dalam presentase 14,4% dari rata-rata dunia[22]. Di tahun yang sama berdasarkan hasil pantauan Satgas penanganan covid-19 terdapat beberapa provinsi di Indonesia yang mengalami tren peningkatan kasus selama 7 hari terakhir dari 7 hari sebelumnya atau disebut *seven day moving average* pertanggal 31 Oktober 2021, salah satunya provinsi Jawa Barat. Dimana kasus yang dilaporkan ada sebanyak 741 kasus baru dengan kasus aktif 1.424 kasus[23]

Faktor penyebaran yang semakin luas dan cepat disebabkan karena pola mobilitas manusia. Beberapa negara langsung menerapkan beberapa kebijakan yang bertujuan mencegah dan menekan mobilitas manusia, seperti menghindari kerumunan massal, menjaga jarak, memakai masker, dan melakukan pembatasan perjalanan baik antar wilayah maupun antar negara [2],[5],[6],[7]. Hingga saat ini varian virus baru SARS-CoV-2 terus bermunculan, varian baru dinilai lebih menular dan terkadang lebih mematikan sehingga saat ini sebagian negara di seluruh dunia masih mengalami gelombang baru penularan covid-19 [6].

Pada tahun 2020 terdapat penelitian yang membahas mengenai estimasi kinerja algoritma *machine learning* dalam analisis faktor dataset covid-19[11]. Tujuan dari penelitian tersebut memprediksi dan mengklasifikasi terkait covid-19. Dimana penelitian ini menggunakan algoritma *machine learning* seperti *linear* dan *logistic regression*, *Decision Tree* (DT), *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), dan SVM dengan *grid search*. Data yang digunakan pada penelitian ini juga menggunakan dataset covid-19 yang diperoleh dari *Jhons Hopkins University* (JHU) *Center for Systems Science and Engineering* (CSSE). Dari hasil penelitian tersebut menunjukkan bahwa penggunaan SVM dengan metode pencarian *grid* menunjukkan akurasi tertinggi sekitar 95%, lalu diikuti oleh algoritma *decision tree* dengan akurasi sekitar 94%. Namun meskipun mendapat akurasi yang baik, pada studi dan penelitian tersebut juga memiliki beberapa keterbatasan. Salah satunya yaitu terkait keterbatasan ketersediaan data pasien covid-19[11].

Pada studi lainnya terdapat penelitian yang membahas mengenai prediksi pasien Covid 19 menggunakan algoritma *machine learning* terbimbing. Dimana pada penelitian ini menggunakan lima metode klasifikasi diantaranya *Nave Bayes*, *Support Vector Machine* (SVM), Regresi Logistik, *K-Nearest Neighbor* (KNN), dan *Decision Tree* [19]. Dari hasil penelitian tersebut dapat bahwa akurasi *Decision Tree* memperoleh akurasi tertinggi yaitu 94,5% diikuti oleh KNN, SVM, LR dan GNB. Dimana untuk metode SVM sendiri memperoleh hasil akurasi sebesar 85% dan untuk nilai presisinya sekitar 81,1%. Namun pada penelitian ini belum menggunakan data *time series*.

Berdasarkan penelitian yang telah dilakukan dan penjabaran di atas, dapat disimpulkan bahwa *Support Vector Machine* cukup baik dalam memprediksi dan juga mengklasifikasi data *time series* terkait Covid 19. Hal itu bisa dilihat dari kelebihan metode svm dari metode *machine learning* yang lain, SVM memiliki dasar teoritis yang dapat dengan mudah dipahami dan diimplementasikan. Dalam batasan sample yang tersedia, metode svm juga salah satu pendekatan yang cocok dalam menangani masalah masalah berdimensi tinggi dan sample yang terbatas [24],[26],[28]. Dari penelitian tersebut, belum ada ditemukan studi dan penelitian yang dimaksud, sehingga pada penelitian ini mengusulkan untuk dilakukan Analisis Kinerja *Support Vector Machine* (SVM) pada klasifikasi Jumlah Kasus Bulanan Covid-19. Tujuan dari penelitian ini diharapkan dapat memberikan gambaran secara luas mengenai kinerja model SVM dalam melakukan klasifikasi yang kedepannya bisa dimanfaatkan untuk membantu para peneliti dalam memilih algoritma *machine learning* dalam memprediksi penyakit[20] yang diharapkan dapat bermanfaat bagi masyarakat ataupun pemerintah.

1.2 Batasan Masalah

Pada penelitian ini, terdapat batasan masalah yang diantaranya berfokus pada kasus Covid 19 per kelurahan di kota Bandung periode November 2020 hingga Desember 2021. Dimana data yang digunakan yaitu data konfirmasi bulanan setiap kelurahan di kota Bandung. Selain data tersebut terdapat beberapa data yang digunakan yaitu data curah hujan, penyinaran matahari, temperature rata-rata, temperatur minimal, temperatur maksimal, data pendudukan berdasarkan pendidikan, data penduduk berdasarkan kelamin, data vaksinasi covid 19, data ketaatan memakai masker dan menjaga jarak. Adapun metode yang digunakan pada penelitian ini yaitu *Support Vector Machine*, dengan bahasa pemrograman yang digunakan yaitu Python.

1.3 Tujuan

Tujuan dari penelitian ini adalah menganalisis kinerja *support vector machine* (svm) pada klasifikasi jumlah kasus bulanan covid-19. Dimana nantinya diharapkan dengan penelitian ini dapat mengetahui efektifitas sistem dan metode yang digunakan dalam algoritma yang dibangun serta diharapkan dari analisis tersebut

dapat diketahui besaran performansi/akurasi menggunakan metode *support vector machine* pada kasus bulanan covid19.

1.4 Organisasi Tulisan

Susunan penelitian dilakukan berdasarkan ketentuan dan mekanisme institusi yang berlaku, dimana tahapannya sebagai berikut yaitu, pendahuluan, literatur review, perancangan sistem, evaluasi, dan kesimpulan. Bagian pendahuluan berisi latar belakang penelitian, tujuan dan harapan yang diinginkan pada penelitian tersebut, literatur review berisi ringkasan dan ulasan Jurnal atau paper terkait penelitian yang berkaitan erat dengan topik atau judul penelitian, perancangan sistem merupakan sebuah perencanaan terhadap sistem yang akan dibangun. Evaluasi merupakan suatu tahap pengujian dari suatu sistem yang sudah dibangun. Kesimpulan merupakan sebuah hasil rangkuman dari suatu penelitian yang sudah dilakukan.

2. Studi Terkait

Covid 19 merupakan sebuah penyakit yang disebabkan oleh virus baru yang disebut *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2). Virus SARS-CoV-2 ini pada dasarnya mengakibatkan infeksi pada sistem pernafasan. Gejala yang ditimbulkan cukup beragam, diantaranya : demam, batuk kering, sakit kepala, pusing, lemas, muntah, dan diare [4]. Pada 11 Maret 2020 WHO menetapkan covid 19 sebagai pandemi, hal ini menjadi perhatian dan berdampak besar bagi banyak orang di seluruh dunia. Pencarian informasi mengenai covid 19 pun menjadi meningkat dan hampir mendominasi percakapan di seluruh *media social*. Hal ini membuat penyebaran berita palsu pun semakin marak di berbagai *media platform*. Diantaranya, pemberitaan pengobatan dini yang dinilai tidak efektif dan berbahaya. Penggunaan obat-obatan yang diluar konteks, dan test sederhana yang tidak mendasar [6]. Misalnya saja pada beberapa kasus, pasien menolak penggunaan ibuprofen atau obat anti-inflamasi lainnya dengan alasan dapat meningkatkan terinfeksi virus corona, padahal gagasan tersebut tidak mendasar dan menyesatkan. Ada juga beberapa kasus lainnya yaitu penyalahgunaan vitamin D dan terjadinya keracunan massal akibat penggunaan/asupan methanol [8]. Selain dari pada itu pernyataan para ahli yang menggiring kekhawatiran berlebihan pun menyebabkan terjadinya panic buying sehingga dalam suatu waktu terjadi kelangkaan atau kurangnya alat pelindungi diri (APD) di bidang kesehatan, dan terjadinya kenaikan terhadap beberapa harga produk secara tidak proporsional [8].

Berdasarkan laporan yang didapat dari Pusat Sains dan Rekayasa Sistem (CSSE) di Universitas Jhon Hopkins, terdapat kurang lebih 1.400.000 kasus yang sudah dilaporkan dari seluruh dunia (per tanggal 7 April 2020) [4]. Dan hingga per tanggal 23 Oktober 2021, wabah covid 19 telah menginfeksi lebih dari 243,547,503 kasus di seluruh dunia. Dengan total kematian 4,949,697 kasus, dan pemulihan yang dilaporkan sebanyak 220,661,965 kasus [9]. Dari kasus pertama yang dilaporkan secara global, Seorang peneliti yang bernama Jiang et al mengemukakan bahwa presentasi tingkat kematian Covid-19 seluruh dunia adalah 4,5%, yang dimana 8,0% merupakan pasien dengan rentan umur 70-79 tahun, dan 14,8% merupakan pasien dengan rentan umur 80 keatas [7]. Selain dari pada itu Jiang et al juga mengungkapkan bahwa pasien yang berumur diatas 50 tahun dengan memiliki penyakit kronis juga memiliki resiko yang tinggi sehingga penting melakukan tindakan pencegahan [7]. Hingga saat ini varian virus corona terus bermunculan, dan membuat gelombang baru penularan terus terjadi di seluruh dunia.

Pada tahun 2020 telah dilakukan penelitian untuk mengetahui prediksi penyebaran virus corona (Covid-19) menggunakan model SVM. penelitian ini dilakukan berdasarkan data *time series* yang dikumpulkan dari 22 Januari 2020 hingga 25 April 2020 di seluruh dunia, dimana data yang digunakan termasuk atribut lokasi, kasus meninggal dan kasus sembuh. Data tersebut diperoleh dari *Jhons Hopkins University* (JHU) *Public Repository Center for System Science and Engineering* (CSSE). Pada penelitian ini metode SVM digunakan untuk mengeksplorasi dampak pada identifikasi kematian dan pemulihan. Hasil penelitian tersebut didapat bahwa prediksi dari pasien covid-19 bergantung pada skor atribut saat menggunakan model perhitungan SVM. Selain itu, peneliti juga menemukan bahwa *hyperplane* yang ideal menggunakan RBF dan C, parameter tersebut cukup membantu dalam membandingkan parameter *hyperplane* dalam memeriksa *vector* pendukung dengan lebih baik. Pada saat yang bersamaan peneliti juga melakukan analisis statistik melalui diagram batang untuk membedakan kelompok subjek. Dengan menggunakan fungsi kernel SVM menghasilkan nilai kinerja yang optimal dalam memprediksi kasus covid 19. Peneliti juga mengatakan bahwa model svm merupakan salah satu teknik *machine learning* yang baik dalam memprediksi karena kesederhanaannya [10].

Pada tahun 2020 terdapat penelitian yang membahas mengenai estimasi kinerja algoritma *machine learning* dalam analisis faktor dataset covid-19. Tujuan dari penelitian tersebut memprediksi dan mengklasifikasi terkait covid-19. Dimana penelitian ini menggunakan algoritma *mechine learning* seperti *linear* dan *logistic regression*, *Decision Tree* (DT), *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), dan SVM dengan *grid search*. Data yang digunakan pada penelitian ini juga menggunakan dataset covid-19 yang diperoleh dari *Jhons Hopkins University* (JHU) *Center for Systems Science and Engineering* (CSSE). Dari hasil penelitian