



Visualisasi dan Analisa Data Penyebaran Covid-19 dengan Metode Klasifikasi *Naïve Bayes*

Muhammad Ikbāl¹, Septi Andryana², Ratih Titi Komala Sari³

^{1,2,3} Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional

article info

Article history:

Received 18 November 2020

Received in revised form

17 December 2020

Accepted 2 January 2021

Available online October 2021

DOI:

<https://doi.org/10.35870/jtik.v5i4.233>

Keywords:

System, Covid, Naïve Bayes Classifier.

Kata Kunci:

Sistem, Klasifikasi, Covid, Naïve Bayes.

abstract

The covid-19 virus became a pandemic in 2020. The spread of covid cases has hit the whole world, reaching 63 million cases in 190 countries as of November 2020. Information regarding the spread of covid is necessary for the general public. This research will produce a system that can provide information on the geographic distribution of covid cases. The data on the distribution of covid cases in this study were also used to analyze the classification using the Naive Bayes Classifier method. The Naive Bayes Classifier method works by using probability calculations so that this research can be used to classify the covid status in an area. The results of this study have succeeded in providing information on the status of the covid pandemic based on data on covid cases that have occurred around the world. Covid case data becomes training data for the analysis of the Naive Bayes classifier method so that it can determine the status of the Covid pandemic based on test data provided by system users. This research has succeeded in helping users to know the status of the Covid pandemic in an area well because it has reliable training data.

abstrak

Virus covid-19 menjadi pandemi pada tahun 2020. Persebaran kasus covid telah melanda seluruh dunia, hingga mencapai 63 juta kasus di 190 Negara per november 2020. Informasi terakut persebaran covid perlu bagi masyarakat secara umum. Penelitian ini akan menghasilkan sistem yang dapat memberikan informasi persebaran kasus covid secara geografis. Data Persebaran kasus covid pada penelitian ini juga digunakan untuk melakukan analisa klasifikasi dengan metode Naive Bayes Classifier. Metode Naive Bayes Classifier bekerja dengan menggunakan perhitungan probabilitas, sehingga pada penelitian dapat digunakan untuk klasifikasi status covid pada suatu wilayah. Hasil penelitian ini berhasil memberikan informasi status pandemi covid berdasarkan data kasus covid yang terjadi diseluruh dunia. Data kasus covid menjadi data latih untuk analisa metode naive bayes classifier, sehingga dapat menentukan status pandemi covid berdasarkan data uji yang diberikan oleh pengguna sistem. Penelitian ini berhasil membantu pengguna, mengetahui status pandemi covid di suatu wilayah dengan baik, karena memiliki data latih yang dapat dipertanggung jawabkan.

*Corresponding author. Email: ikbalpro1995@gmail.com¹.

© E-ISSN: 2580-1643.

Copyright © 2021. Published by Lembaga Informasi dan Riset (KITA INFO dan Riset), Lembaga KITA (<http://creativecommons.org/licenses/by/4.0/>).

1. Latar Belakang

Covid-19 merupakan salah satu jenis virus biologis yang saat ini tengah menjadi pandemi di seluruh dunia. Virus ini dapat melakukan penularan dari manusia ke manusia secara mudah melalui *droplet* atau titik air yang berasal dari tubuh manusia yang sudah terinfeksi [1]. *Droplet* mengandung virus dapat menempel pada benda-benda yang sering dijamah, sehingga virus berpindah ke tubuh manusia lain yang sehat. Karena kemudahan penularan, maka per tanggal 30 November 2020 jumlah kasus positif di seluruh dunia mencapai 63 juta kasus. Kasus positif covid telah menyebar ke 190 Negara [2].

Informasi terkait sebaran kasus *covid* di seluruh dunia dapat diketahui pada media *online* atau berita surat kabar. Namun penyampaian informasi saat ini cenderung bersifat tekstual, sehingga kurang informatif. Penyampaian informasi tekstual terbatas pada persentase angka dan nama daerah. Dengan jumlah penyebaran virus *covid 19* di 190 negara, penyampaian informasi secara tekstual tidak relevan dan tidak menarik minat pembaca.

Penelitian ini dilakukan dengan tujuan untuk menampilkan informasi penyebaran virus *covid-19* di seluruh dunia secara informatif dan menarik minat pembaca. Tampilan informasi geografis menjadi solusi untuk menampilkan kasus penyebaran virus *covid 19* di seluruh Dunia. Informasi geografis menampilkan jumlah kasus penyebaran dalam bentuk peta dunia. Dengan tampilan informasi geografis, pengguna dapat mengetahui jumlah kasus pada letak daerah secara pasti. Selain itu tampilan informasi geografis mempermudah pembaca untuk mendapatkan informasi secara lengkap dalam satu tampilan [3].

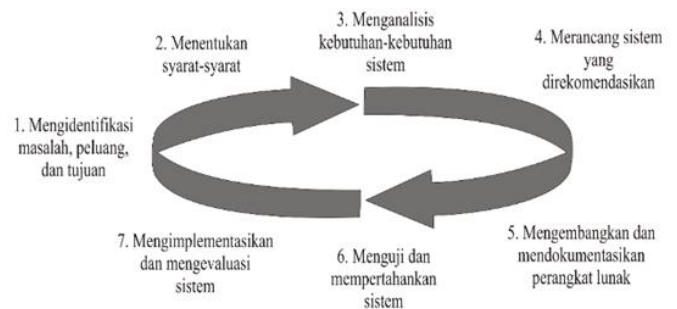
Penelitian kasus klasifikasi data *covid* telah dilakukan dengan tujuan klasifikasi data penyebaran kasus *covid* di Indonesia [4]. Metode yang digunakan adalah *K-Nearest Neighbor* (KNN). Pada prosesnya analisa metode KNN membutuhkan nilai K sebagai acuan standarisasi dan perhitungan jarak antar data. Penentuan nilai K bersifat random atau *subyektifitas*, sehingga tidak memiliki dasar data yang *obyektif*.

Hal ini berbeda dengan penelitian ini yang menggunakan metode *Naive Bayes Classifier* (NBC).

Metode NBC membutuhkan data latih untuk proses analisa. Data latih yang digunakan merupakan data kasus yang terjadi di seluruh negara di Dunia. Data latih pada penelitian ini bersifat obyektif, sehingga hasil dari analisa dapat dipertanggungjawabkan. Penggunaan Data latih mempengaruhi hasil analisa [5].

2. Metodologi Penelitian

Pada penelitian kali ini penulis menggunakan metode Algoritma Naïve Bayes untuk analisa hasil datanya dan tahapan *System Development Life Cycle* (SDLC) untuk pengembangan sistemnya. SDLC atau *Systems Life Cycle*, dalam rekayasa sistem dan rekayasa perangkat lunak, adalah proses pembuatan dan pengubahan sistem serta model dan metodologi yang digunakan untuk mengembangkan sistem-sistem tersebut. Konsep ini umumnya merujuk pada sistem komputer atau informasi [6]. Konsep ini umumnya merujuk pada sistem komputer atau informasi [7].



Gambar 1. Tahap Kegiatan Penelitian

Berikut adalah tujuh tahap pengembangan sistem dengan metode *SDLC* pada penelitian ini :

- Tahap awal adalah mengetahui pokok permasalahan dan kebutuhan masyarakat terhadap informasi penyebaran covid 19 secara geografis.
- Data penyebaran *covid 19* diperoleh dari *worldometers* per bulan September 2020. Begitu juga data *refrensi* penerapan metode *Naive Bayes Classifier* diperoleh dari studi pustaka
- Tahap menentukan fungsi-fungsi sistem berdasarkan tujuan penelitian dan data yang berhasil diperoleh.
- Kegiatan membuat desain sistem dengan menggunakan diagram perancangan *Unified Modelling Language* (UML).
- Tahap pembuatan sistem dengan menggunakan bahasa pemrograman *PHP Framework Codeigniter*.

- f) Sistem yang telah selesai dibuat wajib untuk diuji terlebih dahulu, agar pengguna terhindar dari *error sistem*. Pengujian sistem pada penelitian ini menggunakan metode *blackbox testing*. Metode *Black Box* berfokus pada ketepatan fungsi-fungsi sistem.
- g) Tahap akhir dari pengembangan sistem di penelitian ini adalah melakukan implementasi sistem, agar dapat digunakan oleh pengguna secara umum. Hasil dari implementasi sistem menjadi dasar evaluasi *accepted test* pengguna terhadap sistem yang telah selesai dibuat.

Metode Naive Bayes Classifier

Algoritma *Naive Bayes* adalah salah satu algoritma dalam teknik klasifikasi yang mudah diimplementasikan dan cepat prosesnya [8]. Algoritma *Naive Bayes* didasarkan pada probabilitas bersyarat. Ini menggunakan Teorema *Bayes*, rumus itu menghitung probabilitas dengan menghitung frekuensi nilai dan kombinasi nilai dalam sejarah data. Teorema *Bayes* menemukan probabilitas suatu peristiwa terjadi mengingat probabilitas peristiwa lain yang telah terjadi [9]. Metode *Naive Bayes Classifier* (NBC) menggunakan model statistik untuk melakukan proses klasifikasi data. Metode NBC menghitung nilai *probabilitas* data uji berdasarkan data kasus yang sudah pernah terjadi atau data latih [10]. Pada metode ini, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain. Dasar formula *Naive Bayes*, yang digunakan adalah :

$$P(C|X) = \frac{P(x|c)P(c)}{P(x)}$$

Keterangan :

- X : Data dengan kelas yang belum diketahui
 H : Label Kelas
 P(H) : Probabilitas dari Hipotesa H
 P(X) : Probabilitas X
 P(C|X) : Probabilitas Hipotesis H berdasarkan kondisi X.
 P(X|C) : Probabilitas X, berdasarkan kondisi hipotesis H.

3. Hasil dan Pembahasan

Tahap awal analisa dilakukan dengan menentukan kelompok klasifikasi pandemi yaitu normal atau buruk. Kelompok klasifikasi ditentukan dari 3 kriteria yaitu, jumlah kasus mati, jumlah kasus sembuh dan jumlah kasus aktif. Masing-masing kriteria memiliki sub kriteria positif dan negatif. Penentuan status sub kriteria dilakukan dengan cara menghitung perbandingan nilai kriteria dibandingkan dengan nilai rata-rata kriteria dari seluruh negara. Berikut ini contoh penentuan status kriteria dari Negara Albania.

- Nilai *Death* 10.704 dari nilai rata-rata seluruh data adalah 146.686,78
- Nilai *Recovered* 322 dari nilai rata-rata seluruh data adalah 4.756,66
- Nilai *Active* 6.284 dari nilai rata-rata seluruh data adalah 98.867,25

Dari perbandingan nilai dari masing-masing kriteria di Negara Albania dengan nilai rata-rata masing-masing kriteria, maka status status masing-masing kriteria adalah sebagai berikut.

- Status *Death* Negatif
- Status *Recovered* Negatif
- Status *Active* Negatif

Selanjutnya adalah menentukan status pandemi dari Negara Albania. Penentuan status pandemi dilakukan dengan cara membandingkan nilai total kasus di Negara Albania dengan nilai rata-rata kasus yang terjadi di 190 negara. Dengan nilai total kasus 17.310 dan nilai rata-rata total kasus adalah sebanyak 250.310,70. Maka negara albania memiliki status pandemi normal.

Berdasarkan hasil perbandingan untuk seluruh negara, maka hasil klasifikasi pandemi dari 190 negara adalah sebagai berikut.

- Nilai klasifikasi pandemi buruk ada di 25 negara.
- Nilai klasifikasi pandemi normal ada di 165 negara.

Setelah mendapatkan status kriteria masing-masing data latih, maka selanjutnya dilakukan pengujian dengan data uji. Data uji pada contoh kasus ini adalah sebagai berikut.

- Nilai *Death* 1000
- Nilai *Recovered* 1000
- Nilai *Active* 20

Data uji tersebut terlebih dahulu diberikan status kriteria dengan cara yang sama pada data latih yaitu membandingkan nilai kriteria dengan nilai rata-rata dari seluruh data latih. Hasil perbandingan masing-masing kriteria data uji menunjukkan status kriteria sebagai berikut ini.

- Status *Death* Negatif
- Status *Recovered* Negatif
- Status *Active* Negatif

Status kriteria data uji menjadi acuan analisa perhitungan probabilitas dengan metode *Naive Bayes Classifier*. Berikut ini analisa perbandingan status kriteria data latih berdasarkan referensi data uji dengan nilai total kasus status pandemi data latih.

- Status *Death* Negatif / Total Kasus Pandemi Buruk = $2/25 = 0,080$
- Status *Death* Negatif / Total Kasus Pandemi Normal = $165/165 = 1$
- Status *Recovered* Negatif / Total Kasus Pandemi Buruk = $4/25 = 0,160$
- Status *Recovered* Negatif / Total Kasus Pandemi Normal = $159/165 = 0,964$
- Status *Active* Negatif / Total Kasus Pandemi Buruk = $2/25 = 0,080$
- Status *Active* Negatif / Total Kasus Pandemi Normal = $163/165 = 0,988$

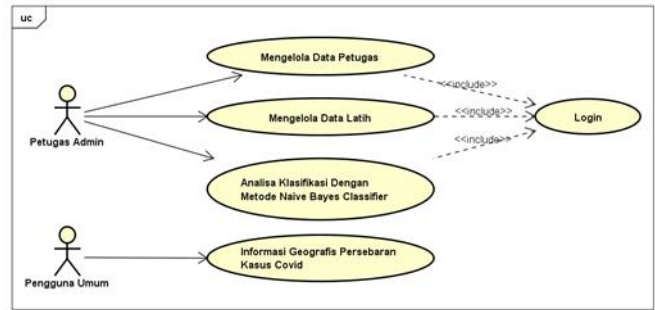
Selanjutnya adalah melakukan perhitungan perkalian nilai perbandingan dari masing-masing total kasus status pandemi. Berikut ini perhitungan yang dilakukan.

- Status Pandemi Buruk = $0,080 * 0,160 * 0,080 = 0,001$
- Status Pandemi Normal = $1 * 0,964 * 0,988 = 0,952$

Dari hasil perhitungan diatas, nilai terbesar adalah status pandemi normal. Oleh karena itu data uji masuk dalam klasifikasi status pandemi normal.

Perancangan Sistem

Perancangan sistem dibuat dengan mengacu pada analisa kebutuhan sistem. Berikut ini desain perancangan sistem usulan menggunakan *Use Case Diagram*.



Gambar 2. Diagram Use Case Sistem Usulan

Implementasi Sistem

Berikut ini penjelasan masing-masing hasil implementasi tampilan sistem. Halaman home ditampilkan saat pertama kali sistem dibuka oleh pengguna. Pada halaman ini ditampilkan persebaran kasus covid dalam bentuk geografis. Pengguna umum dapat mengetahui jumlah kasus di masing-masing negara dengan mengarahkan kursor pada peta negara. Berikut ini tampilan halaman home pada sistem usulan.

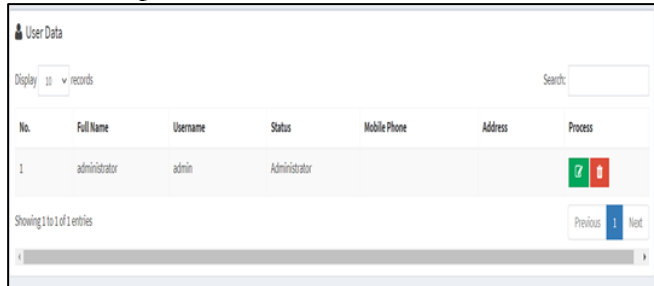


Gambar 3. Halaman Home Sistem Usulan

Halaman login ditampilkan saat pengguna tekan tombol login. Pada halaman ini petugas yang terdaftar dapat menginputkan *username* dan *password* yang dimiliki. Berikut ini tampilan halaman *login* pada sistem usulan.

Gambar 4. Halaman Login Sistem Usulan

Halaman menu *user* akan menampilkan data petugas yang dapat melakukan pengelolaan data latih dan melakukan analisa klasifikasi. Pada halaman ini, petugas admin dapat menginputkan data user baru, rubah data user dan hapus data user. Berikut ini adalah tampilan halaman menu *user*.



Gambar 5. Halaman Menu User

Pada halaman menu data latih akan menampilkan data latih yang sudah diinput pada sistem. Halaman ini petugas dapat melakukan penambahan data latih, merubah data latih dan menghapus data latih. Data latih menjadi data acuan untuk analisa dan data yang ditampilkan pada informasi geografis.

Gambar 6. Halaman Menu Data Latih

Halaman menu analisa menampilkan kolom inputan nilai kasus untuk masing-masing kriteria. Berikut ini tampilan halaman menu analisa.

Gambar 7. Halaman Menu Analisa Sistem Usulan

Data nilai kasus ini akan dianalisa klasifikasi dengan menggunakan metode *Naive Bayes Classifier*. Hasil analisa ditampilkan dengan detail nilai hasil perhitungan dan kesimpulan dari analisa klasifikasi. Berikut ini tampilan halaman hasil analisa.

Gambar 8. Halaman Hasil Analisa Sistem Usulan

4. Kesimpulan

Penelitian ini berhasil menghasilkan sistem informasi geografis persebaran kasus covid di seluruh Dunia. Penyampaian informasi ditampilkan dalam bentuk peta dunia yang dapat menampilkan informasi jumlah kasus pada wilayah yang dipilih oleh pengguna. Sistem usulan juga berhasil memberikan hasil klasifikasi kasus pandemi covid berdasarkan nilai angka kasus pada masing-masing kategori yaitu kasus kematian, kasus sembuh dan kasus aktif yang diinput oleh pengguna. Hasil analisa memberikan informasi status pandemi di wilayah tersebut buruk atau normal. Penelitian ini dapat dikembangkan sebagai sistem pelaporan kasus covid, sehingga data latih dapat selalu terbaru.

5. Daftar Pustaka

- [1] Susilo, A., Rumende, C.M., Pitoyo, C.W., Santoso, W.D., Yulianti, M., Herikurniawan, H., Sinto, R., Singh, G., Nainggolan, L., Nelwan, E.J. and Chen, L.K., 2020. Coronavirus Disease 2019: Tinjauan Literatur Terkini. *Jurnal Penyakit Dalam Indonesia*, 7(1), pp.45-67.
- [2] Azimah, R.N., Khasanah, I.N., Pratama, R., Azizah, Z., Febriantoro, W. and Purnomo, S.R.S., 2020. Analisis Dampak Covid-19 Terhadap Sosial Ekonomi Pedagang Di Pasar Klaten Dan Wonogiri. *EMPATI: Jurnal Ilmu Kesejahteraan Sosial*, 9(1), pp.59-68.

- [3] Prasetyo, Y. and Haryanto, H., 2017. Visualisasi Berbasis Naïve Bayes untuk Pemetaan Penyebaran Penyakit Infeksi Saluran Pernafasan Akut. *Sisfotenika*, 7(1), pp.74-84.
- [4] Wibawa, A.P., Kurniawan, A.C., Adiperkasa, R.P., Putra, S.M., Kurniawan, S.A. and Nugraha, Y.R., 2019. Naïve Bayes Classifier for Journal Quartile Classification. *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*, 7(2), pp.91-99.
- [5] Fitriani, A.S., Fajrillah, F. and Novarika, W., 2019. Implementation of Data Mining Using Naïve Bayes Classification Method To Predict Participation of Governor And Vocational Governor Selection In Jemirahan Village, Jabon District. *The IJICS (International Journal of Informatics and Computer Science)*, 3(2), pp.66-79.
- [6] Erdiansyah, M.Z., Taufik, T. and Raharjana, I.K., 2016. Visualisasi Data Menggunakan Sistem Informasi Geografis untuk Potensi Bank Sampah di Surabaya. *Journal of Information Systems Engineering and Business Intelligence*, 2(1), pp.40-49.
- [7] Setyoningrum, N.R., 2016. Perbandingan Antara Tiga Sdlc Methodology, Parallel, Iterative Dan Agile Development. *Jurnal Bangkit Indonesia*, 5(1), pp.32-32.
- [8] Saputra, M.F.A., Widiyaningtyas, T. and Wibawa, A.P., 2018. Illiteracy Classification Using K Means-Naïve Bayes Algorithm. *JOIV: International Journal on Informatics Visualization*, 2(3), pp.153-158.
- [9] R. Anusha and A. Prof., 2020. Predicting the Student's Preference Between Conventional Learning and E-Learning. *Int. J. Adv. Sci. Technol.*, vol. 29, no. 4, pp. 5917–5922.
- [10] Erkan, U. and Gökrem, L., The Classification of The Students Success Via The Informations Existing In E-School System. *Group*, 10(80), p.1.