



Prediksi Kenaikan Covid-19 di Tangerang Selatan Menggunakan Metode Klasifikasi Naive Bayes

Raditia Vindua¹, Nurhayati², Wiwin Winarti³

Universitas Pamulang

dosen02380@unpam.ac.id¹, dosen02378@unpam.ac.id², dosen02374@unpam.ac.id³

Kata kunci:

Covid-19; Data Mining; Klasifikasi Nave Bayes

Abstrak

Penyebaran infeksi Covid-19 yang sangat cepat membuat hingga hampir tak ada negara atau wilayah di dunia yang absen dari virus Covid-19. Peningkatan jumlah kasus Covid-19 terjadi dalam waktu yang cepat hingga butuh penanganan secepatnya. Berbagai cara sudah dilakukan oleh Pemerintah dalam mengantisipasi dan menanggulangi pandemi ini. Data Covid-19 Di Tangerang Selatan belum dimanfaatkan untuk mengenali pola (pattern) yang dapat dimanfaatkan untuk memprediksi kenaikan Covid-19 di Tangerang Selatan yang baru. Oleh karena itu penelitian ini dilakukan sebagai langkah antisipasi terhadap pandemi Covid-19 dengan memprediksi kenaikan Covid-19 terutama di kota Tangerang Selatan. Metode penelitian yang di terapkan pada penelitian ini adalah analisis masalah dan studi literatur, mengumpulkan data, implementasi ke dalam bahasa data mining menggunakan metode Klasifikasi Naïve Bayes dan evaluasi. Tujuan dari penelitian ini adalah membuat model data mining untuk prediksi kenaikan Covid-19 di Tangerang Selatan yang baru dengan metode Klasifikasi Naïve Bayes. Sehingga dari pola tersebut di harapkan dapat memprediksi kenaikan Covid-19 di Tangerang Selatan sebagai langkah untuk antisipasi masyarakat Tangerang Selatan terhadap Covid-19. Pada penelitian ini Klasifikasi Naïve Bayes dapat digunakan untuk memprediksi tingkat kenaikan Covid-19 di Kota Tangerang Selatan dengan tingkat akurasi 69.3878% dan nilai error 30.6122%. Hasil prediksi metode klasifikasi Naïve Bayes menggunakan WEKA, ada 136 data Covid-19 yang di prediksi benar dan ada 60 data Covid-19 yang di prediksi salah di Kota Tangerang Selatan.

Pendahuluan

Coronavirus Disease 2019 (Covid-19) adalah penyakit jenis baru yang belum pernah diidentifikasi sebelumnya pada manusia. Virus Covid-19 menyerang bagian pernapasan, sehingga menyebabkan penderitanya kesulitan bernafas (World Health Organization, 2021). Tidak hanya menyerang bagian pernapasan, tetapi penyakit ini juga dapat menular dari manusia ke manusia melalui percikan batuk/bersin. Tanda dan gejala umum infeksi Covid-19 termasuk gejala gangguan pernapasan akut. Seseorang yang terjangkit virus ini akan menunjukkan gejala ringan seperti gejala flu, demam, batuk, dan sesak napas, sehingga

menyebabkan virus ini kesulitan untuk dideteksi. Masa inkubasi rata-rata adalah 5 – 6 hari dengan masa inkubasi demam, batuk, dan sesak napas. Seseorang yang memiliki penyakit bawaan seperti penyakit diabetes, jantung, asma dan kanker akan lebih tinggi resiko mengalami penyakit yang lebih serius akibat virus ini. Pada kasus yang parah, Covid-19 dapat menyebabkan pneumonia, sindrom pernapasan akut, gagal ginjal, dan bahkan kematian (Tosepu et al., 2020).

Corona Virus atau Covid-19 pertama kali mulai muncul pada akhir Desember 2019. Dimana kasusnya dimulai dengan pneumonia atau radang paru-paru misterius pada Desember 2019 di kota Wuhan, provinsi Hubei Cina. Tragedi pada akhir Desember 2019 tersebut terus menular dari manusia ke manusia melalui percikan batuk/bersin (*droplet*) dengan cepat sehingga penyebaran virus Corona mewabah ke seluruh dunia. Presiden Joko Widodo melaporkan pertama kali menemukan dua kasus infeksi Covid-19 di Indonesia pada tanggal 2 Maret 2020 (Djalante et al., 2020).

Pada 11 Maret 2020 Organisasi Kesehatan Dunia (WHO) mendeklarasikan virus Covid-19 sebagai global pandemi. Pandemi Corona virus 2019 (Covid-19) juga menjadi perhatian dunia karena mengancam kesehatan masyarakat secara umum (Emy et al., n.d., 2020). Status pandemi menandakan bahwa penularan Covid-19 berlangsung sangat cepat hingga hampir tak ada negara di dunia dapat terhindar dari virus (Widiyani, 2020). Orang yang paling berisiko tertular penyakit ini adalah orang yang kontak erat dengan pasien Covid-19 termasuk yang merawat pasien Covid-19 (Kementerian Kesehatan Republik Indonesia, 2021). Peningkatan jumlah kasus Covid-19 terjadi dalam waktu yang cepat hingga butuh penanganan secepatnya. Sayangnya, hingga kini belum ada obat spesifik untuk menangani kasus infeksi virus Corona atau Covid-19. Berdasarkan masalah tersebut, maka Indonesia dalam status waspada terhadap ancaman virus Covid-19, maka bentuk pencegahan dari meluasnya penyebaran virus dapat dilakukan dengan cara memutuskan rantai penularannya. Protokol kesehatan serta peraturan pembatasan sosial pun sudah diberlakukan. Namun pembatasan sosial secara keseluruhan dapat menyebabkan penurunan ekonomi. Oleh karena itu di perlukan prediksi daerah mana saja yang akan mengalami peningkatan kasus Covid-19, sehingga diharapkan Covid-19 dapat di cegah dan tidak menginfeksi lebih banyak penduduk Indonesia di kemudian hari.

Masalah yang dihadapi terkait penyebaran Covid-19 terdapat di Jawa Timur yang merupakan peringkat 1 dengan jumlah kasus positif Covid-19 di Indonesia. Pada tanggal 1 Juli 2020 dengan total kasus 12321 dan jumlah kasus aktif mencapai 7196. Dengan menggunakan metode *Vector Autoregressive Moving Average* (VARMA) didapatkan perbandingan data 95:5 yang memiliki hasil peramalan terbaik (Rohman & Rohmawati, 2021).

Dilakukan juga penelitian tentang prediksi Covid-19 di Jombang menggunakan model *Integer-valued Autoregressive* (INAR). Dari proses tersebut, diperoleh model INAR(4) yang terbaik dengan nilai *Akaike's Information Criterion* (AIC) sebesar 1407,224 dan diperoleh hasil prediksi jumlah kasus positif Covid-19 di Jombang pada tanggal 21 Februari 2021 sebanyak 27 kasus dengan estimasi eror prediksi menggunakan *Mean Absolute Percentage Error* sebesar 28,5% (Subiantoro, 2021).

Penelitian berikutnya yang membahas tentang Peramalan Kasus Covid-19 di DKI Jakarta dengan model Arima memiliki hasil akhir bahwa metode ini mampu meramalkan jumlah warga DKI Jakarta yang Positif Covid-19 pada 27 hari kedepan dengan sangat baik (Nurul Qomariasih, 2021).

Sedangkan dengan menggunakan metode lain yaitu Metode Fuzzy Mamdani memiliki tingkat akurasi sebesar 50,16% yang dianggap cukup akurat untuk memprediksi jumlah

pasien yang positif Covid-19 sehingga dapat dilakukan percepatan penanganan agar dapat memberikan kebijakan yang tepat untuk penyedia layanan, fasilitas dan tenaga medis (Prakarsa & Nasution, 2021).

Salah satu pendekatan untuk menjelaskan solusi dari permasalahan yang terjadi dalam dunia nyata adalah memodelkan atau merumuskan permasalahan nyata ke dalam bahasa data mining. Data mining mencari pola yang diinginkan dalam database yang besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti (Hermawati, n.d., 2013). Sehingga diharapkan dapat digunakan untuk membantu mencari solusi terkait dengan penyebaran penyakit tersebut.

Beberapa teknik telah dikembangkan dan diimplementasikan untuk mengekstrak pengetahuan dan informasi untuk menemukan pola pengetahuan yang mungkin berguna untuk pengambilan keputusan. Teknik-teknik yang digunakan untuk pengekstrakan pengetahuan dalam data mining adalah pengenalan pola, klusterisasi, asosiasi, prediksi dan klasifikasi. Klasifikasi merupakan salah satu topik utama dalam data mining atau machine learning. Klasifikasi adalah suatu pengelompokan data dimana data yang digunakan tersebut mempunyai kelas label atau target (Wibowo, 2021). Salah satu algoritma klasifikasi yang sering digunakan dalam memprediksi adalah *Naive Bayes Classifier*. Penerapan metode naive bayes diharapkan mampu memprediksi tingkat penyebaran Covid-19 di Indonesia

Metode Naïve Bayes memiliki nilai probabilitas yang dapat di optimalkan untuk memprediksi tingkat penyebaran Covid-19 berdasarkan proses klasifikasi. Metode ini berhasil mengklasifikasikan 16 data dari 33 data yang diuji dalam kasus Covid-19 per provinsi dengan keakuratan sebesar 48,4848%, di mana dari 33 data yang di uji dalam kasus Covid-19 per provinsi yang diuji terdapat 16 data yang berhasil diklasifikasikan dengan benar (Alvina Felicia Watratan et al., 2020).

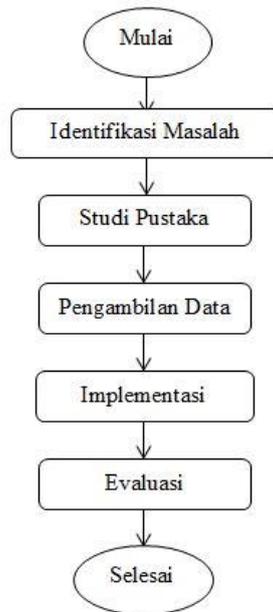
Penelitian selanjutnya yang mengenai Analisa Data Penyebaran Covid-19 dengan metode Naïve Bayes memiliki hasil akhir bahwa penelitian ini berhasil menghasilkan sistem informasi geografis persebaran kasus covid di seluruh Dunia (Ikbal et al., 2021).

Dari kasus tersebut, maka peneliti melakukan penelitian dengan mengangkat judul tentang “Prediksi Kenaikan Covid-19 di Tangerang Selatan Menggunakan Metode Klasifikasi Naive Bayes”. Penelitian dilakukan dengan tujuan untuk memperoleh model dalam memprediksi penambahan kasus Covid-19 di Tangerang Selatan.

Metode

2.1 Tahapan Penelitian

Langkah-langkah metodologi penelitian yang digunakan dalam penelitian ini adalah identifikasi masalah, studi pustaka, pengumpulan data, implementasi dan evaluasi seperti tergambar dalam gambar dibawah ini :



Gambar 1. Tahapan Penelitian

a. Identifikasi Masalah

Identifikasi masalah dilakukan untuk menemukan permasalahan yang ada terutama terhadap kenaikan kasus Covid-19 di Tangerang Selatan. Pada saat ini beberapa wilayah di Tangerang Selatan mengalami kenaikan jumlah kasus Covid-19, untuk mencegah semakin meningkatnya kasus, maka diperlukan kelompok tingkat kenaikan kasus Covid-19, yaitu rendah, sedang, dan tinggi. Dengan mengetahui kelompok kasus Covid-19 ini diharapkan dapat meningkatkan kewaspadaan masyarakat terhadap virus Covid-19 sehingga tertib dalam menjalankan protokol kesehatan dan mengurangi laju penyebaran virus di lingkungan masyarakat.

b. Studi Pustaka

Studi pustaka dilakukan oleh peneliti dengan tujuan untuk mempelajari penelitian sebelumnya yang berkaitan dengan penelitian yang dibuat oleh peneliti saat ini. Dalam hal ini, peneliti mempelajari mengenai data tingkat kenaikan kasus Covid-19 di Tangerang Selatan, metode dari algoritma Naïve Bayes, serta informasi-informasi lainnya yang berkaitan dengan penelitian yang dilakukan oleh peneliti.

c. Pengumpulan Data

Pengumpulan data pada penelitian ini menggunakan metode pengumpulan data *study literature* dan mengunduh data, berupa data sekunder yang didapatkan dari web Tangsel Tanggap Covid. Data yang digunakan dalam penelitian ini adalah dataset tingkat kenaikan kasus Covid-19 di Tangerang Selatan per tanggal 01 Desember 2021 Sampai 31 Januari 2022(Pemerintah Kota Tangerang Selatan, 2021). Setelah data dikumpulkan maka dilakukan analisa data untuk menyesuaikan proses data yang akan diolah pada metode Naive Bayes.

d. Implementasi

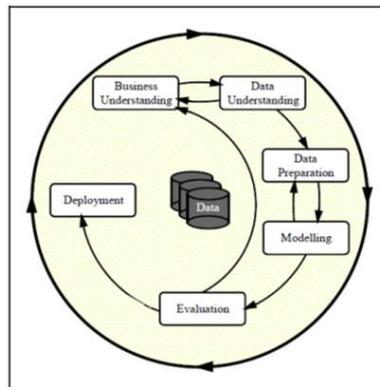
Sesuai dengan pengolahan data maka pada tahap implementasi adalah tentang bagaimana pengolahan datanya diterapkan dalam sebuah *tools*. *Tools* yang akan digunakan dalam implementasi penelitian ini adalah dengan menggunakan Software Weka. Selanjutnya pengujian dilakukan untuk mengetahui apakah penelitian yang dilakukan telah sesuai dengan tujuan yang diharapkan yaitu untuk memprediksi tingkat penyebaran Covid-19 di Tangerang Selatan.

e. Evaluasi

Evaluasi model dilakukan dalam penelitian ini dengan menerapkan perhitungan akurasi untuk mengetahui ketepatan dalam mengelompokkan data dari model yang dibuat terhadap data. Akurasi merupakan rasio perkiraan seberapa akurat model dalam mengelompokkan keseluruhan data dengan tepat.

2.2 Metode Analisis

Metode analisis yang digunakan dalam penelitian ini menggunakan analisis data mining sebagai alat mengolah data menggunakan WEKA. Terdapat enam proses CRISP-DM (*Cross Industry Standard Process for Data Mining*) seperti tergambar dalam Gambar dibawah ini :



Gambar 2. Tahap CRISP-DM

2.2.1 *Bussiness Understanding*

Pada fase ini berfokus pada pemahaman dan perspektif bisnis proses dari suatu sistem. Yaitu penentuan tujuan untuk prediksi Kenaikan Covid19 di Kota Tangerang selatan dan menyiapkan strategi untuk penyampaian tujuan.

2.2.2 *Data Understanding*

Pada fase ini berfokus pada pembelajaran data yang sudah ada, Pengumpulan dan penyeleksian data. dilakukan dengan cara mencari bahan materi yang berhubungan dengan permasalahan, perancangan, guna mempermudah proses implementasi sistem. Pencarian materi dilakukan melalui pencarian di buku panduan dan internet.

2.2.3 *Data Preparation*

Fase persiapan data adalah fase yang terdiri dari pemilihan data, pembersihan data, mengintegrasikan data, dan transformasi data agar dapat dilanjutkan kedalam tahap pemodelan.

2.2.4 *Modeling*

Pada fase ini proses yang terjadi adalah pemilihan model yang sesuai. Pemodelan disini menggunakan klasifikasi Naïve Bayes. Algoritma Naïve Bayes adalah salah satu algoritma dalam teknik klasifikasi yang mudah diimplementasikan dan cepat prosesnya. Karena Algoritma Naïve Bayes tidak memerlukan skema estimasi parameter perulangan yang rumit, artinya dapat diterapkan untuk dataset yang memiliki ukuran yang besar (Vercellis, 2008). Algoritma Naïve Bayes didasarkan pada probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Mochammad Haldi Widiyanto, 2021). 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 Naïve Bayes yang digunakan adalah :

$$P(c|x) = \frac{p(x|c)P(c)}{p(x)} \quad (1)$$

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 Hipotesa H berdasarkan kondisi X.

P(X|C) : Probabilitas X, berdasarkan kondisi hipotesis H.

Agar dapat mengoptimalkan hasil, dataset dilakukan pemodelan dengan algoritma klasifikasi Naïve Bayes terlebih dahulu, setelah modelnya keluar maka akan muncul parameter evaluasi beserta akurasi.

2.2.5 Evaluasi

Pada fase ini akan dilakukan proses evaluasi dari fase sebelumnya. Fase evaluasi ini akan melakukan perbandingan kuantitatif dengan mempertimbangkan nilai akurasi Hasil Klasifikasi Naïve Bayes. Akurasi merupakan rasio perkiraan seberapa akurat model dalam mengelompokkan keseluruhan data dengan tepat.

$$\text{Akurasi}(\%) = \frac{|\text{Nilai Benar} - \text{Hasil Analisis}|}{\text{Nilai Benar}} \times 100\% \quad (2)$$

2.2.6 Deployment

Pada fase ini proses yang terjadi adalah penyusunan laporan atau presentasi dari pengetahuan yang didapat dari evaluasi pada proses data mining. Sehingga dari fase ini menghasilkan prediksi tingkat kenaikan kasus Covid-19 di Tangerang Selatan yang baru.

Hasil dan Pembahasan

3.1 Persiapan Data Mining

Berdasarkan data Covid-19 di kota Tangerang Selatan pada bulan Desember 2021 dan Januari 2022, dalam penelitian ini ada sebanyak 435 data. Data positif Covid-19 di kota Tangerang Selatan didapatkan dari web Tangsel Tanggap Covid dan di jabarkan pada tabel berikut:

Tabel 1. Data Positif Covid-19 di kota Tangerang Selatan

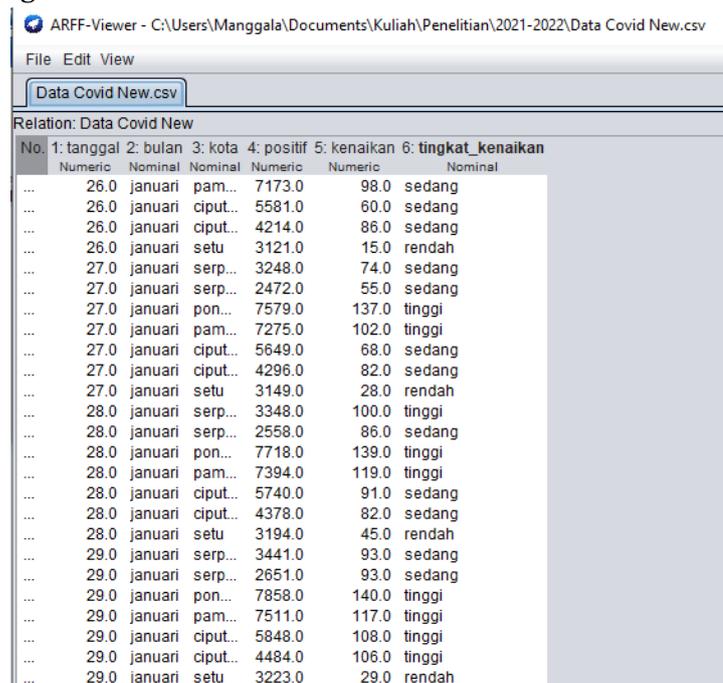
Tanggal	Serpong		Pondok		Ciputat		Setu
	Serpong	Utara	Aren	Pamulang	Timur	Ciputat	
1-Des-21	2913	2167	7048	6844	5330	3906	3038
2-Des-21	2914	2167	7050	6845	5330	3907	3038
3-Des-21	2915	2168	7051	6845	5330	3907	3038
4-Des-21	2915	2169	7052	6845	5330	3907	3038
5-Des-21	2915	2173	7052	6845	5330	3907	3038
6-Des-21	2915	2173	7052	6846	5330	3907	3038
7-Des-21	2915	2175	7052	6846	5330	3907	3038
8-Des-21	2915	2175	7052	6847	5330	3907	3038
9-Des-21	2915	2175	7052	6847	5330	3907	3038
10-Des-21	2915	2175	7053	6847	5330	3907	3039

11-Des-21	2915	2175	7054	6848	5330	3907	3039
12-Des-21	2916	2175	7054	6848	5331	3907	3039
13-Des-21	2916	2175	7054	6848	5331	3907	3039
14-Des-21	2916	2175	7054	6849	5332	3908	3038
15-Des-21	2916	2175	7054	6849	5332	3908	3038
16-Des-21	2916	2175	7054	6849	5332	3908	3038
17-Des-21	2916	2175	7053	6851	5332	3908	3038
18-Des-21	2916	2175	7053	6851	5332	3908	3038
19-Des-21	2916	2175	7053	6851	5332	3908	3038
20-Des-21	2916	2175	7053	6851	5332	3908	3038
21-Des-21	2916	2175	7053	6852	5332	3911	3038
22-Des-21	2916	2175	7053	6852	5332	3915	3038
23-Des-21	2916	2175	7053	6852	5332	3915	3038
24-Des-21	2916	2175	7053	6852	5332	3917	3038
25-Des-21	2916	2175	7053	6852	5332	3917	3038
26-Des-21	2916	2175	7053	6852	5332	3917	3038
27-Des-21	2916	2175	7053	6852	5332	3917	3038
28-Des-21	2916	2177	7053	6852	5332	3918	3038
29-Des-21	2916	2177	7054	6852	5332	3918	3038
30-Des-21	2916	2177	7055	6852	5334	3917	3038
31-Des-21	2917	2177	7056	6852	5334	3917	3038
1-Jan-22	2917	2178	7057	6852	5334	3917	3038
2-Jan-22	2919	2178	7059	6853	5334	3917	3038
3-Jan-22	2919	2181	7060	6854	5334	3917	3038
4-Jan-22	2921	2181	7063	6855	5335	3917	3038
5-Jan-22	2921	2182	7063	6855	5335	3917	3038
6-Jan-22	2923	2183	7064	6858	5335	3917	3038
7-Jan-22	2925	2184	7066	6860	5339	3920	3038
8-Jan-22	2930	2186	7069	6860	5341	3922	3040
9-Jan-22	2930	2190	7071	6860	5342	3924	3042
10-Jan-22	2931	2191	7074	6861	5342	3924	3044
11-Jan-22	2932	2192	7075	6863	5344	3927	3045
12-Jan-22	2935	2194	7075	6865	5344	3934	3046
13-Jan-22	2937	2195	7078	6867	5350	3938	3047
14-Jan-22	2940	2195	7088	6869	5356	3939	3049
15-Jan-22	2943	2200	7096	6870	5361	3943	3049
16-Jan-22	2950	2205	7110	6879	5366	3951	3050
17-Jan-22	2950	2205	7110	6879	5366	3951	3050
18-Jan-22	2966	2215	7127	6895	5377	3955	3056
19-Jan-22	2978	2224	7144	6906	5390	3967	3059
20-Jan-22	2991	2232	7164	6923	5408	3985	3062
21-Jan-22	3010	2249	7184	6940	5423	4010	3072
22-Jan-22	3049	2275	7218	6985	5453	4029	3080
23-Jan-22	3072	2294	7255	7006	5460	4055	3087
24-Jan-22	3088	2311	7284	7025	5477	4077	3098
25-Jan-22	3130	2365	7353	7075	5521	4128	3106

26-Jan-22	3174	2417	7442	7173	5581	4214	3121
27-Jan-22	3248	2472	7579	7275	5649	4296	3149
28-Jan-22	3348	2558	7718	7394	5740	4378	3194
29-Jan-22	3441	2651	7858	7511	5848	4484	3223
30-Jan-22	3516	2732	7996	7622	5915	4555	3240
31-Jan-22	3601	2797	8135	7763	6032	4623	3274

Kemudian data tersebut diolah di *Microsoft Excel* dengan pemilihan atribut yaitu: Tanggal, Bulan, Kota, Banyaknya yang positif Covid-19 dan Tingkat Kenaikan Covid-19 di setiap kota di Tangerang Selatan dengan tiga kelompok yakni ‘rendah’, ‘sedang’, dan ‘tinggi’. Dimana tingkat rendah dengan nilai kenaikan kurang dari 50, tingkat sedang dengan nilai kenaikan lebih dari 50, dan tingkat tinggi dengan nilai lebih dari 100.

Setelah itu data di save dalam bentuk *.csv agar selanjutnya dapat diproses di aplikasi *WEKA* seperti gambar di bawah ini:



No.	1: tanggal	2: bulan	3: kota	4: positif	5: kenaikan	6: tingkat_kenaikan
	Numeric	Nominal	Nominal	Numeric	Numeric	Nominal
...	26.0	januari	pam...	7173.0	98.0	sedang
...	26.0	januari	ciput...	5581.0	60.0	sedang
...	26.0	januari	ciput...	4214.0	86.0	sedang
...	26.0	januari	setu	3121.0	15.0	rendah
...	27.0	januari	serp...	3248.0	74.0	sedang
...	27.0	januari	serp...	2472.0	55.0	sedang
...	27.0	januari	pon...	7579.0	137.0	tinggi
...	27.0	januari	pam...	7275.0	102.0	tinggi
...	27.0	januari	ciput...	5649.0	68.0	sedang
...	27.0	januari	ciput...	4296.0	82.0	sedang
...	27.0	januari	setu	3149.0	28.0	rendah
...	28.0	januari	serp...	3348.0	100.0	tinggi
...	28.0	januari	serp...	2558.0	86.0	sedang
...	28.0	januari	pon...	7718.0	139.0	tinggi
...	28.0	januari	pam...	7394.0	119.0	tinggi
...	28.0	januari	ciput...	5740.0	91.0	sedang
...	28.0	januari	ciput...	4378.0	82.0	sedang
...	28.0	januari	setu	3194.0	45.0	rendah
...	29.0	januari	serp...	3441.0	93.0	sedang
...	29.0	januari	serp...	2651.0	93.0	sedang
...	29.0	januari	pon...	7858.0	140.0	tinggi
...	29.0	januari	pam...	7511.0	117.0	tinggi
...	29.0	januari	ciput...	5848.0	108.0	tinggi
...	29.0	januari	ciput...	4484.0	106.0	tinggi
...	29.0	januari	setu	3223.0	29.0	rendah

Gambar 3. Data .csv yang akan diproses di WEKA

Kemudian data di simpan dalam bentuk *.arff agar dapat diolah menggunakan *WEKA* dan dilanjutkan pemodelan data mining. Pemodelan data mining yang digunakan merupakan klasifikasi *Naïve Bayes*.

3.2 Modeling Klasifikasi Naïve Bayes

Setelah data melewati tahap *preprocessing*, data tersebut di klasifikasikan menggunakan metode klasifikasi *Naïve Bayes*. Hasil dataset menggunakan *Naïve Bayes* dapat dilihat pada Gambar berikut:

```

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0.538    0.268    0.125     0.538   0.203     0.149    0.865    0.526    rendah
      1.000    0.065    0.711     1.000   0.831     0.815    0.976    0.818    sedang
      0.654    0.000    1.000     0.654   0.791     0.527    0.963    0.989    tinggi
Weighted Avg.  0.694    0.027    0.902     0.694   0.757     0.542    0.958    0.935

=== Confusion Matrix ===

 a  b  c  <-- classified as
 7  6  0 | a = rendah
 0 27  0 | b = sedang
49  5 102| c = tinggi
    
```

Gambar 4. Hasil Klasifikasi Naïve Bayes menggunakan *Supplied test set WEKA*

Berdasarkan hasil gambar diatas, dapat di simpulkan :

- a. Baris pertama “7 6 0” menunjukkan bahwa ada (7+6+0) tingkat kenaikan rendah dan 6 salah diklasifikasikan sebagai tingkat kenaikan yang sedang
- b. Baris kedua “0 27 0” menunjukkan bahwa ada 27 tingkat kenaikan sedang dan semua benar diklasifikasikan sebagai tingkat kenaikan yang sedang
- c. Baris ketiga “49 5 102” menunjukkan bahwa ada (49+5+102) tingkat kenaikan tinggi dan 49 salah diklasifikasikan sebagai tingkat kenaikan yang rendah dan 5 salah diklasifikasikan sebagai tingkat kenaikan yang sedang

3.3 Evaluasi

Fase evaluasi ini akan melakukan perbandingan quantitaf dengan mempertimbangkan nilai akurasi dan *error* hasil Klasifikasi *Naïve Bayes* dengan menggunakan data Covid-19 di kota Tangerang Selatan pada bulan Februari 2022. Hasil dapat dilihat pada Gambar berikut:

```

=== Summary ===

Correctly Classified Instances      136          69.3878 %
Incorrectly Classified Instances    60           30.6122 %
Kappa statistic                    0.4333
Mean absolute error                 0.1762
Root mean squared error             0.3691
Relative absolute error             29.2301 %
Root relative squared error         50.0053 %
Total Number of Instances          196
    
```

Gambar 5. Hasil Evaluasi Klasifikasi Naïve Bayes menggunakan *WEKA*

Berdasarkan dari hasil tabel diatas, dapat di simpulkan bahwa tingkat akurasi hasil Klasifikasi Naïve Bayes memiliki nilai akurasi 69,3878% dan nilai error 30.6122%.

3.4 Deployment

Kemudian data tersebut di prediksi menggunakan pola dari metode klasifikasi Naïve Bayes.menggunakan WEKA. Hasil prediksi klasifikasi Naïve Bayes menggunakan WEKA dapat dilihat pada Gambar dan di jadikan Tabel menggunakan excel sebagai berikut:

No	tanggal	bulan	kota	positif	kenaikan	prediction margin	predicted tingkat_kenaikan	tingkat_kenaikan	
1	1			serp...	3703.0	102.0	-0.807884	sedang	rendah
2	1			serp...	2896.0	99.0	0.991898	sedang	sedang
3	1			pon...	8353.0	218.0	1.0	rendah	rendah
4	1			pam...	7916.0	153.0	1.0	rendah	rendah
5	1			ciput...	6141.0	109.0	0.848149	rendah	rendah
6	1			ciput...	4739.0	116.0	0.7724	rendah	rendah
7	1			setu	3320.0	46.0	-0.995954	sedang	tinggi
8	2			serp...	3839.0	136.0	0.992233	rendah	rendah
9	2			serp...	3015.0	119.0	-0.315243	sedang	rendah
10	2			pon...	8577.0	224.0	1.0	rendah	rendah
11	2			pam...	8127.0	211.0	1.0	rendah	rendah
12	2			ciput...	6284.0	143.0	0.999935	rendah	rendah
13	2			ciput...	4894.0	155.0	0.999969	rendah	rendah
14	2			setu	3354.0	34.0	0.864727	tinggi	tinggi
15	3			serp...	4019.0	180.0	1.0	rendah	rendah
16	3			serp...	3179.0	164.0	0.999946	rendah	rendah
17	3			pon...	8973.0	396.0	-0.354538	tinggi	rendah
18	3			pam...	8426.0	299.0	-0.436152	tinggi	rendah
19	3			ciput...	6534.0	250.0	1.0	rendah	rendah
20	3			ciput...	5125.0	231.0	1.0	rendah	rendah
21	3			setu	3461.0	107.0	-0.364649	sedang	rendah

Gambar 6. Hasil Prediksi Klasifikasi Naïve Bayes menggunakan WEKA

Selanjutnya dari hasil prediksi menggunakan WEKA dijadikan Tabel menggunakan excel sebagai berikut:

Tabel 2. Hasil Prediksi Covid19 menggunakan Microsoft Excel

Tanggal	Serpon g	Serpong Utara	Pondok Aren	Pamulan g	Ciputat Timur	Ciputa t	Set u
1-Feb-22	3703	2896	8353	7916	6141	4739	3320
2-Feb-22	3839	3015	8577	8127	6284	4894	3354
3-Feb-22	4019	3179	8973	8426	6534	5125	3461
4-Feb-22	4162	3335	9313	8751	6709	5319	3560
5-Feb-22	4344	3540	9663	9095	6894	5586	3642
6-Feb-22	4558	3728	9948	9450	7117	5791	3700
7-Feb-22	4739	3870	10249	9696	7288	5987	3746
8-Feb-22	4936	4085	10578	9970	7538	6215	3830
9-Feb-22	5147	4286	10982	10383	7781	6489	3921
10-Feb-22	5308	4429	11283	10700	8003	6687	3985
11-Feb-22	5449	4615	11607	10996	8164	6939	4037
12-Feb-22	5711	4854	12161	11483	8469	7221	4153
13-Feb-22	5900	5010	12572	11833	8746	7476	4228
14-Feb-22	6131	5179	12881	12179	8949	7676	4301
15-Feb-22	6361	5249	13242	12656	9232	7965	4412
16-Feb-22	6607	5673	13693	13061	9526	8225	4510

22							
17-Feb-22	6829	5859	14101	13511	9731	8462	4588
18-Feb-22	7013	6014	14463	13890	9930	8684	4650
19-Feb-22	7113	6017	14510	14009	10004	8771	4678
20-Feb-22	7315	6272	15074	14318	10220	9021	4780
21-Feb-22	7422	6405	15235	14475	10348	9140	4816
22-Feb-22	7560	6576	15511	14750	10519	9293	4889
23-Feb-22	7684	6717	15745	14997	10649	9470	4953
24-Feb-22	7799	6843	15934	15198	10783	9615	5000
25-Feb-22	7920	6943	16121	15403	10898	9733	5038
26-Feb-22	8019	7048	16269	15613	10998	9833	5063
27-Feb-22	8095	7128	16417	15720	11065	9912	5091
28-Feb-22	8155	7169	16534	15793	11126	9967	5107

Berdasarkan dari hasil tabel diatas, dapat di simpulkan bahwa ada 136 data Covid19 yang di prediksi benar dan ada 60 data Covid19 yang di prediksi salah (yang berwarna kuning).

Kesimpulan

Berdasarkan analisis yang telah dilakukan , maka dapat diambil kesimpulan bahwa Klasifikasi Naïve Bayes dapat digunakan untuk memprediksi tingkat Covid19 di Kota Tangerang Selatan dengan tingkat akurasi 69.3878% dan nilai error 30.6122%. Dari hasil prediksi metode klasifikasi Naïve Bayes menggunakan WEKA, terdapat 136 data Covid19 yang di prediksi benar dan ada 60 data Covid19 yang di prediksi salah di Kota Tangerang Selatan.

Daftar Pustaka

- Alvina Felicia Watratan, Arwini Puspita. B, & Dikwan Moeis. (2020). Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia. *Journal of Applied Computer Science and Technology*, 1(1), 7–14.
<https://doi.org/10.52158/jacost.v1i1.9>
- Djalante, R., Lassa, J., Setiamarga, D., Sudjatma, A., Indrawan, M., Haryanto, B., Mahfud, C., Sinapoy, M. S., Djalante, S., Rafliana, I., Gunawan, L. A., Surtiari, G. A. K., & Warsilah, H. (2020). Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020. *Progress in Disaster Science*, 6.
<https://doi.org/10.1016/j.pdisas.2020.100091>
- Emy, N. P., Yanti, D., Made, I., Dharma, A., Nugraha, P., Wisnawa, G. A., Putu, N., Agustina,

- D., & Diantari, A. (n.d.). *GAMBARAN PENGETAHUAN MASYARAKAT TENTANG COVID-19 DAN PERILAKU MASYARAKAT DI MASA PANDEMI COVID-19*.
- Hermawati, F. A. (n.d.). *Data Mining*.
- Ikkal, M., Andryana, S., & Komala Sari, R. T. (2021). Visualisasi dan Analisa Data Penyebaran Covid-19 dengan Metode Klasifikasi Naïve Bayes. *Jurnal JTIC (Jurnal Teknologi Informasi Dan Komunikasi)*, 5(4), 389.
<https://doi.org/10.35870/jtik.v5i4.233>
- Kementerian Kesehatan Republik Indonesia. (2021). *Info Infeksi Emerging Kementerian Kesehatan RI*.
- Mochammad Haldi Widiyanto. (2021). *Algoritma Naive Bayes*. Binus University.
- Nurul Qomariasih. (2021). PERAMALAN KASUS COVID-19 DI DKI JAKARTA DENGAN MODEL ARIMA. *Jurnal Syntax Transformation*. <https://doi.org/10.46799/jurnal>
- Pemerintah Kota Tangerang Selatan. (2021). *Tangsel Tanggap Covid-19*.
- Prakarsa, G., & Nasution, V. M. (2021). Penerapan Logika Fuzzy Menggunakan Metode Mamdani Pada Prediksi Jumlah Kasus Positif Covid-19. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(4), 1660. <https://doi.org/10.30865/mib.v5i4.3282>
- Rohman, Z. A., & Rohmawati, A. A. (2021). *Prediksi Penyebaran COVID-19 Harian di Jawa Timur Menggunakan Model Vector Autoregressive Moving Average (VARMA)*. 8(5), 11150–11163.
- Subiantoro, F. S. (2021). PREDIKSI KASUS KONFIRMASI POSITIF COVID-19 DI JOMBANG MENGGUNAKAN MODEL INTEGER-VALUED AUTOREGRESSIVE(p). *MATHunesa Jurnal*.
- Tosepu, R., Gunawan, J., Effendy, D. S., Ahmad, L. O. A. I., Lestari, H., Bahar, H., & Asfian, P. (2020). Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Science of the Total Environment*, 725.
<https://doi.org/10.1016/j.scitotenv.2020.138436>
- Vercellis, C. (2008). *Business intelligence : data mining and optimization for decision making*. Wiley.
- Wibowo, A. (2021). *Klasifikasi*. Binus University.
- Widiyani, R. (2020). *Latar Belakang Virus Corona, Perkembangan hingga Isu Terkini*. Detik.Com.
- World Health Organization. (2021). *Critical preparedness, readiness and response actions for COVID-19*.