

Analisis Dataset COVID-19 menggunakan Algoritma KNN dan Random Forest

Wilfandi Aser Hizkia Samosir^{#1}, Tiur Gantini^{*2},

[#] Program Studi Sistem Informasi, Universitas Kristen Maranatha
Jl. Surya Sumantri No. 65 Bandung, Indonesia.

¹1573901@maranatha.ac.id

²tiur.gantini@it.maranatha.edu

Abstract — In the current pandemic era, all levels of society are struggling to get out of adversity, as well as in the health sector vying to solve this pandemic. starting from reducing the number of cases that occur every day. The COVID-19 dataset itself is available on the website of the Ministry of Research and Technology - National Research and Technology Agency. The purpose of this research itself is to find out how data mining works and processes on the COVID-19 dataset using the KNN and Random Forest algorithms.

Keywords— COVID-19, Data Mining, KNN, Random Forest.

I. PENDAHULUAN

Coronavirus disease 2019 (COVID-19) adalah penyakit yang disebabkan oleh coronavirus jenis baru yaitu Sars-CoV-2, yang dilaporkan pertama kali muncul di Wuhan, menurut World Health Organization(WHO) sampai pada tanggal 20 agustus 2021 sudah ada dua ratus juta lebih kasus yang sudah terkonfirmasi, empat juta diantaranya meninggal dunia.

Disisi lain pada era teknologi dan informasi yang telah berkembang sejauh seperti sekarang ini, kecerdasan buatan dan ilmu data memiliki peran penting dalam industri kesehatan. Teknologi seperti Data Mining sangat terhubung dengan baik sehingga para profesional medis dapat dengan mudah mengelola peran dan perawatan pasien mereka. Kemampuan mendasar dari Data mining itu sendiri adalah untuk mengubah data mentah menjadi informasi yang berguna bagi organisasi untuk membantu organisasi tumbuh dengan membuat keputusan yang berlandaskan pada data yang telah dihasilkan. Hasil analisis data mining dapat membantu pengambil keputusan dalam membuat keputusan serta didukung oleh pengetahuan untuk meningkatkan layanan dari institusi kesehatan, rumah sakit dan sebagainya. Dengan cara yang sama, Data Mining dapat digunakan untuk membantu tugas memprediksi kasus aktif SARS-CoV-2 dan juga membantu memperkirakan penghitungan kasus aktif SARS-CoV-2 di masa mendatang.

II. KAJIAN TEORI

A. Data Mining

Menurut Suntoro, *data mining* adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan[1].

Proses yang umumnya dilakukan oleh *data mining* antara lain:

1. Asosiasi
2. Clustering
3. Deskripsi
4. Estimasi
5. Klasifikasi
6. Prediksi

B. Klasifikasi

Klasifikasi merupakan salah satu dari metode yang paling sering digunakan di data mining. Klasifikasi termasuk kedalam kelompok supervised learning karena didalam proses klasifikasi terdapat proses pembelajaran dengan data yang sudah atau telah lampau[2].

C. K-Nearest Neighbor

Algoritma K-Nearest Neighbor (KNN) merupakan sebuah dan salah satu metode untuk melakukan klasifikasi terhadap objek yang didasarkan pada data pembelajaran yang jaraknya paling dekat dengan objek tersebut. KNN merupakan algoritma *supervised learning* dimana hasil dari *query instance* yang baru akan diklasifikasikan berdasarkan mayoritas dari kategori pada algoritma KNN. Dimana kelas yang paling banyak muncul yang akan menjadi kelas hasil dari klasifikasi [3].

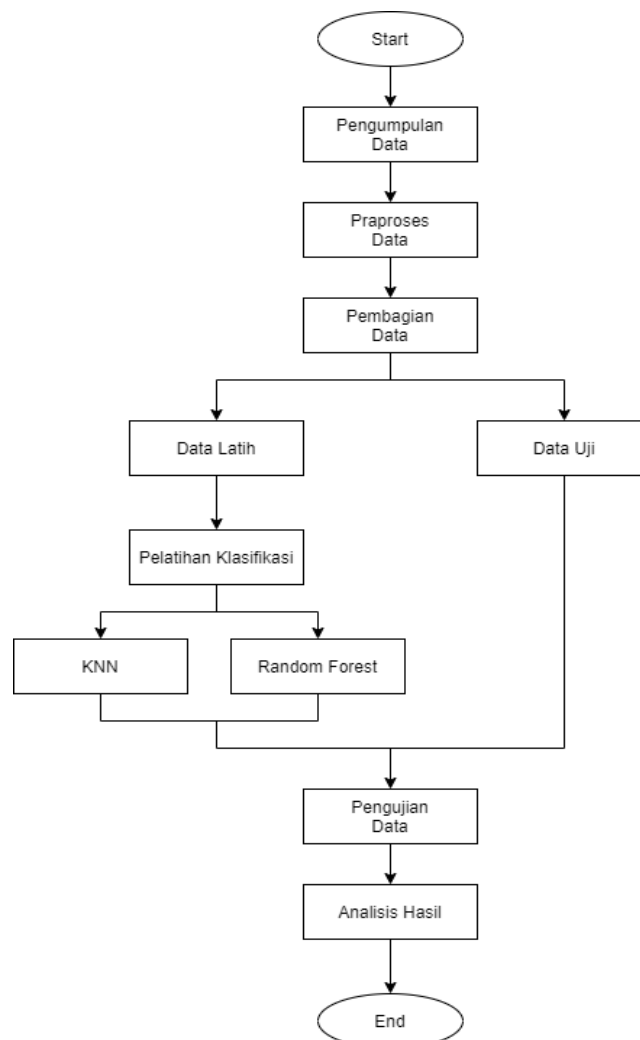
D. Random Forest

Menurut Adnyana, random forest merupakan hasil pengembangan dari decision tree yang menggunakan beberapa decision tree dimana setiap dari decision tree yang sudah dilakukan training menggunakan sampel individu dan juga untuk setiap atribut telah dipecah pada tree yang dipilih dengan menggunakan atribut subset yang bersifat acak dan juga didalam proses klasifikasi, lalu setiap individu akan didasarkan pada vote yang berasal dari suara terbanyak dari kumpulan populasi tree [4].

III. METODOLOGI PENELITIAN

A. Tahapan Penelitian

Pada Gambar 1 memperlihatkan bagaimana tahapan pada penelitian analisis dataset COVID-19 menggunakan algoritma KNN dan Random Forest dilakukan.



Gambar 1. Tahapan Penelitian

B. Pengumpulan Data

Pengumpulan Data yang dipakai pada penelitian ini berasal dari website Kementerian Riset dan Teknologi - Badan Riset dan Inovasi Nasional Republik Indonesia, data yang diambil berupa dataset yang berjudul COVID-19 di Indonesia pada tanggal 20 agustus 2021.

C. Praproses Data

Pada tahapan ini file Excel yang telah diunduh akan dibersihkan dari data yang tidak penting menggunakan aplikasi Microsoft Excel. Tujuannya untuk membuang data yang tidak diperlukan seperti data yang tidak valid atau data ganda.

D. Pembagian Data

Pada bagian ini data yang telah dibersihkan dibagi menjadi 2 set yaitu data latih dan data uji. Tujuan proses ini adalah untuk memvalidasi model dari algoritma KNN dan Random Forest.

E. Pelatihan Klasifikasi

Pada tahap ini data akan dimasukkan kedalam aplikasi WEKA dan akan diklasifikasikan dengan 2 algoritma yaitu KNN dan Random Forest pada fitur klasifikasi. Pelatihan Klasifikasi bertujuan untuk membuat model klasifikasi dari algoritma yang akan digunakan pada penelitian ini.

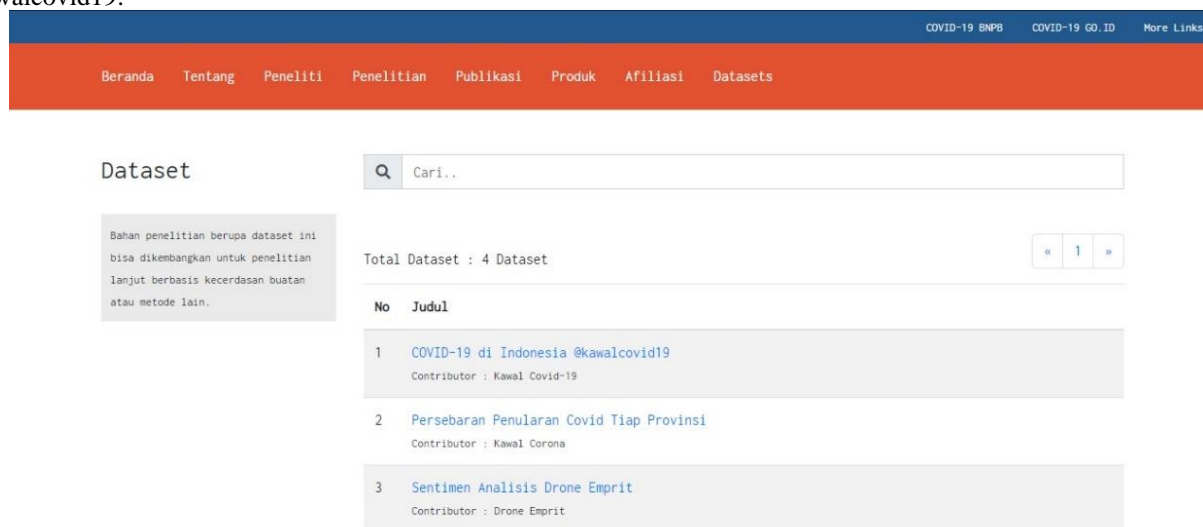
F. Pengujian Data

Pada tahap pengujian data ini, akan dimuat data uji dan model yang telah disimpan pada tahapan pelatihan klasifikasi. Kemudian, data uji akan dievaluasi menurut model klasifikasi KNN dan *Random Forest*. Proses ini bertujuan untuk mengetahui kinerja dari algoritma KNN dan *Random Forest* dalam memprediksi data pada kelas yang telah ditentukan.

IV. HASIL PENELITIAN

A. Pengambilan Data

Pada Gambar 2 memperlihatkan tampilan pilihan dataset COVID-19 di website Kementerian Riset dan Teknologi - Badan Riset dan Inovasi Nasional Republik Indonesia. Dataset yang akan digunakan berjudul COVID-19 di Indonesia @kawalcovid19.



Gambar 2. Pengumpulan Dataset

B. Praproses Data

Pada tahapan praproses data, hal yang akan dilakukan adalah membersihkan dan mengelola data pada *dataset*. Data yang diambil berasal dari *sheet Timeline* berupa data mentah untuk kelas Tanggal, Provinsi dan Kasus_Harian dengan rentang waktu dari tanggal 01-Januari-2021 sampai 30-Juni-2021 dan juga dilakukan penambahan kelas berupa kelas Hari dan Negara. Pada Kelas Kasus_Harian dilakukan pengelompokan jumlah kasus dengan jumlah 0-99,100-999 hingga 1000-9999, tujuannya agar mempermudah algoritma untuk melakukan klasifikasi data, lalu setelah dibersihkan data berjumlah sebanyak 30770. Gambar 3 menunjukkan kelas-kelas yang ada didalam *dataset* sebelum dilakukan pembersihan data.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Total Kasus	Aceh	Bali	Banten	Babel	Bengkulu	DIY	Jakarta	Jambi	Jabar	Jateng	Jatim
555	22 Sep	37,239	111,804	131,069	50,740	22,972	154,214	856,585	29,380	701,263	480,069	393,960
556	23 Sep	37,309	111,942	131,142	50,830	22,980	154,314	856,750	29,407	701,506	480,402	394,180
557	24 Sep	37,394	112,059	131,210	50,881	22,990	154,399	856,931	29,435	701,723	480,688	394,399
558	25 Sep	37,482	112,164	131,254	50,940	23,004	154,495	857,064	29,445	701,911	480,997	394,614
559	26 Sep	37,526	112,249	131,301	50,991	23,010	154,556	857,232	29,468	702,068	481,173	394,784
560	27 Sep	37,589	112,316	131,325	51,027	23,011	154,616	857,335	29,479	702,217	481,358	394,876
561	28 Sep	37,631	112,408	131,369	51,088	23,018	154,687	857,439	29,510	702,422	481,666	395,055
562	29 Sep	37,693	112,496	131,416	51,147	23,019	154,762	857,616	29,529	702,561	481,865	395,280
563	30 Sep	37,733	112,582	131,475	51,209	23,021	154,829	857,765	29,547	702,722	482,009	395,475
564												
565	Kasus Harian	Aceh	Bali	Banten	Babel	Bengkulu	DIY	Jakarta	Jambi	Jabar	Jateng	Jatim
566	15-Mar							19			2	
567	16-Mar			1				14		1	1	
568	17-Mar											
569	18-Mar			4			1	30		12	2	
570	19-Mar			10			2	52		2	4	1
571	20-Mar		3	10			-1	5		15		6
572	21-Mar			4			1	44		14	2	11
573	22-Mar							40		4	1	15
574	23-Mar		3	9				44	1			

Gambar 3. Sebelum pembersihan data

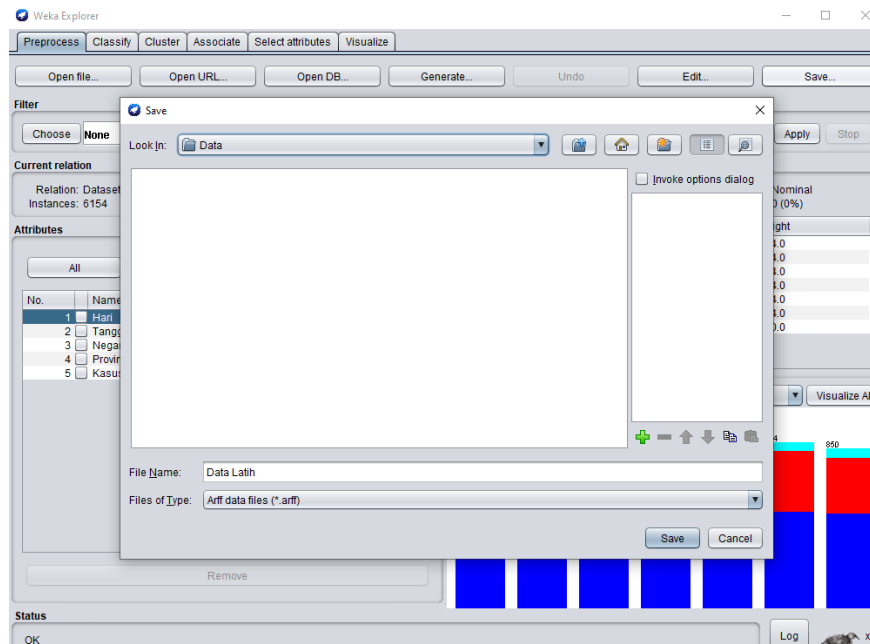
Setelah data dibersihkan maka dataset terlihat seperti gambar 4.

A1	A	B	C	D	E	F
1	Hari	Tanggal	Negara	Provinsi	Kasus_Harian	
2	Jumat	01/01/2021	Indonesia	Maluku Utara	0-99	
3	Jumat	01/01/2021	Indonesia	Aceh	0-99	
4	Jumat	01/01/2021	Indonesia	Papua	0-99	
5	Jumat	01/01/2021	Indonesia	Sulawesi Barat	0-99	
6	Jumat	01/01/2021	Indonesia	Nusa Tenggara Timur	0-99	
7	Jumat	01/01/2021	Indonesia	Papua Barat	0-99	
8	Jumat	01/01/2021	Indonesia	Kalimantan Barat	0-99	
9	Jumat	01/01/2021	Indonesia	Kepulauan Riau	0-99	
10	Jumat	01/01/2021	Indonesia	Gorontalo	0-99	
11	Jumat	01/01/2021	Indonesia	Maluku	0-99	
12	Jumat	01/01/2021	Indonesia	Kalimantan Utara	0-99	
13	Jumat	01/01/2021	Indonesia	Jambi	0-99	
14	Jumat	01/01/2021	Indonesia	Sulawesi Utara	0-99	
15	Jumat	01/01/2021	Indonesia	Nusa Tenggara Barat	0-99	
16	Jumat	01/01/2021	Indonesia	Sulawesi Tengah	0-99	
17	Jumat	01/01/2021	Indonesia	Bengkulu	0-99	
18	Jumat	01/01/2021	Indonesia	Sumatera Selatan	0-99	
19	Jumat	01/01/2021	Indonesia	Sulawesi Tenggara	0-99	
20	Jumat	01/01/2021	Indonesia	Bangka Belitung	0-99	
21	Jumat	01/01/2021	Indonesia	Lampung	0-99	
22	Jumat	01/01/2021	Indonesia	Sumatera Utara	0-99	

Gambar 4. Setelah pembersihan data.

C. Pembagian Data

Pada tahapan pembagian data, *dataset* kemudian dibagi menjadi data latih dan data uji, dengan data latih sebanyak 30770 data dan data uji sebanyak 1020 data yang diambil dari tanggal terakhir pada setiap bulan dalam kurun waktu dari bulan januari hingga juni tahun 2021, lalu setelah itu data latih akan dikonversi menjadi format ARFF dari format CSV dengan menggunakan aplikasi WEKA seperti yang terlihat di Gambar 5.



Gambar 5. Konversi Data Latih menggunakan WEKA

D. Pelatihan Klasifikasi

Pelatihan Klasifikasi dilakukan dengan menggunakan aplikasi WEKA, tahapan pertama data latih yang berformat .arff dimuat kedalam aplikasi kemudian dengan menggunakan fitur *Classify* akan dipilih algoritma yang telah ditentukan yang kemudian akan digunakan untuk melakukan proses klasifikasi. Setelah melakukan *10-folds-cross-validation* maka model klasifikasi telah berhasil dibuat, model klasifikasi itu kemudian akan disimpan untuk kedepannya digunakan untuk melakukan pengujian data. Berikut adalah model klasifikasi KNN dan *Random Forest* yang telah berhasil dibuat:

Time taken to build model: 0.47 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	4860	78.973 %
Incorrectly Classified Instances	1294	21.027 %
Kappa statistic	0.5703	
Mean absolute error	0.2294	
Root mean squared error	0.3181	
Relative absolute error	67.2778 %	
Root relative squared error	77.0311 %	
Total Number of Instances	6154	

Gambar 6. Hasil Summary Random Forest

Pada Gambar 6, gambar menunjukkan hasil validasi klasifikasi Random Forest akurasi data yang menghasilkan data yang diklasifikasi dengan benar berjumlah 78.9% dan yang salah berjumlah 21%.

```

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      4585           74.5044 %
Incorrectly Classified Instances    1569           25.4956 %
Kappa statistic                     0.4343
Mean absolute error                 0.2773
Root mean squared error            0.3495
Relative absolute error            81.3142 %
Root relative squared error        84.6385 %
Total Number of Instances         6154

```

Gambar 7. Hasil Summary KNN

Kemudian pada Gambar 7, gambar menunjukkan hasil validasi klasifikasi algoritma KNN dengan akurasi data yang menghasilkan data yang diklasifikasikan secara benar berjumlah 74.5% dan yang salah berjumlah 25.4%.

```

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0.926    0.342    0.808     0.926   0.863     0.621    0.898     0.928     0-99
      0.550    0.087    0.764     0.550   0.640     0.510    0.849     0.752     100-999
      0.755    0.020    0.675     0.755   0.713     0.698    0.983     0.735     1000-9999
Weighted Avg.  0.790    0.239    0.786     0.790   0.780     0.587    0.886     0.859

```

Gambar 8. Hasil Detailed Accuracy Random Forest

Pada Gambar 8 menunjukkan hasil detailed accuracy berdasarkan kelas kasus_harian pada dataset dengan menggunakan algoritma Random Forest.

```

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0.961    0.524    0.740     0.961   0.836     0.523    0.883     0.914     0-99
      0.446    0.073    0.759     0.446   0.562     0.443    0.825     0.708     100-999
      0.166    0.002    0.828     0.166   0.277     0.359    0.982     0.753     1000-9999
Weighted Avg.  0.745    0.344    0.751     0.745   0.714     0.487    0.868     0.835

```

Gambar 9. Hasil Detailed Accuracy KNN

Sedangkan pada Gambar 9 menunjukkan hasil detailed accuracy berdasarkan kelas kasus_harian pada dataset dengan menggunakan algoritma KNN.

```

=== Confusion Matrix ===

      a      b      c  <-- classified as
3470  277    0 |   a = 0-99
 823 1149  116 |   b = 100-999
 0    78   241 |   c = 1000-9999

```

Gambar 10. Hasil Confusion Matrix Random Forest

Hasil confusion matrix yang ditunjukkan pada Gambar 10 berasal dari klasifikasi kelas Kasus_Harian dengan algoritma Random Forest. Dalam mengklasifikasi data Kasus_Harian algoritma Random Forest sering mengalami kesalahan dalam mengklasifikasi matriks b sebagai matriks a sebanyak 823 data.

=== Confusion Matrix ===

```

a    b    c  <-- classified as
3600 147   0 |    a = 0-99
1145 932  11 |    b = 100-999
 117 149  53 |    c = 1000-9999
  
```

Gambar 11. Hasil Confusion Matrix KNN

Sedangkan Gambar 11 menunjukkan hasil confusion matrix dari klasifikasi kelas Kasus_Harian dengan algoritma KNN. Dalam mengklasifikasi data Kasus_Harian algoritma KNN juga sering mengalami kesalahan dalam mengklasifikasi matriks b sebagai matriks a namun lebih banyak dari algoritma Random Forest dan juga salah mengklasifikasikan matriks c sebagai matriks b.

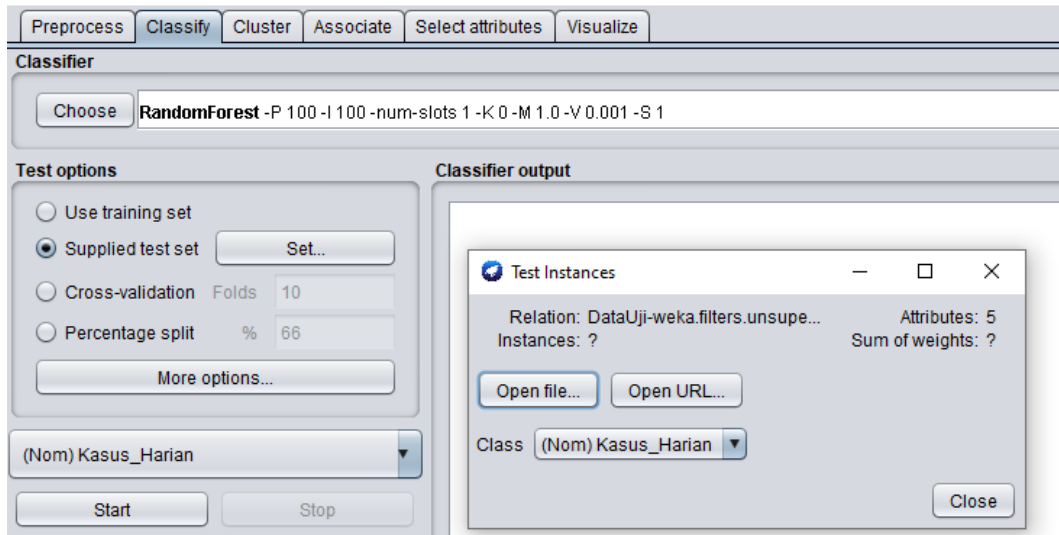
E. Pengujian Data

Pada tahapan pengujian data, data yang sudah ditentukan pada tahapan pembagian data sebagai data uji yang akan diprediksi akan dimuat kedalam excel untuk mengganti kelas yang akan diprediksi menjadi tanda tanya seperti yang terlihat di Gambar 12.

	A	B	C	D	E
	Hari	Tanggal	Negara	Provinsi	Kasus_Harian
1	Minggu	31/01/2021	Indonesia	Aceh	0-99
2	Minggu	31/01/2021	Indonesia	Bali	100-999
3	Minggu	31/01/2021	Indonesia	Bangka Belitung	0-99
4	Minggu	31/01/2021	Indonesia	Banten	100-999
5	Minggu	31/01/2021	Indonesia	Bengkulu	0-99
6	Minggu	31/01/2021	Indonesia	DI Yogyakarta	100-999
7	Minggu	31/01/2021	Indonesia	DKI Jakarta	1000-9999
8	Minggu	31/01/2021	Indonesia	Gorontalo	0-99
9	Minggu	31/01/2021	Indonesia	Jambi	0-99
10	Minggu	31/01/2021	Indonesia	Jawa Barat	1000-9999
11	Minggu	31/01/2021	Indonesia	Jawa Tengah	1000-9999
12	Minggu	31/01/2021	Indonesia	Jawa Timur	100-999
13	Minggu	31/01/2021	Indonesia	Kalimantan Barat	0-99
14	Minggu	31/01/2021	Indonesia	Kalimantan Selatan	100-999
15	Minggu	31/01/2021	Indonesia	Kalimantan Tengah	0-99
16	Minggu	31/01/2021	Indonesia	Kalimantan Timur	100-999
17	Minggu	31/01/2021	Indonesia	Kalimantan Utara	0-99
18	Minggu	31/01/2021	Indonesia	Kepulauan Riau	100-999
19	Minggu	31/01/2021	Indonesia	Lampung	100-999
20	Minggu	31/01/2021	Indonesia	Maluku	0-99
21	Minggu	31/01/2021	Indonesia	Maluku	0-99

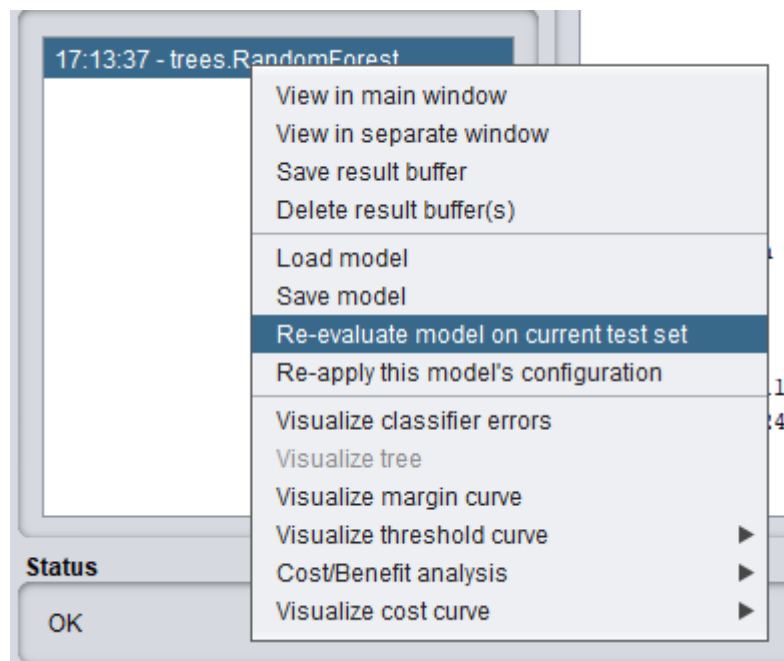
Gambar 12. Tampilan Data Uji yang akan diprediksi

Kemudian file tersebut akan dimuat kedalam aplikasi WEKA untuk mengkonversi format file menjadi .arff dan menyamakan atribut data uji dengan data latih. Langkah ini bertujuan agar WEKA bisa memvalidasi data uji pada data latih yang telah disiapkan sebelumnya. Setelah atribut disamakan maka data uji dimuat kembali kedalam aplikasi WEKA untuk proses pengujian data. Pada tab Classify data uji akan dimasukkan dengan pilihan Supplied test set dan memilih kelas yang akan diuji, pada penelitian ini akan dipilih kelas Kasus_Harian seperti yang terlihat pada Gambar 13.



Gambar 13. Memuat data Uji pada WEKA

Kemudian proses selanjutnya adalah memuat model klasifikasi dari algoritma KNN dan Random Forest yang telah dibuat pada tahapan sebelumnya. Setelah model klasifikasi dimuat kedalam aplikasi, data uji sudah bisa diprediksi menggunakan pilihan Re-evaluate model on current test set pada model yang telah dibuat seperti yang terlihat pada Gambar 14.



Gambar 14. Tampilan saat memprediksi Data Uji pada WEKA

Setelah diprediksi menggunakan pilihan Re-evaluate model on current test set maka hasilnya bisa disimpan ke dalam format .arff.

hasilRandomForest.arff

Relation: DataUji-weka.filters.unsupervised.attribute.StringToNominal-Rlast_predicted

No.	1: Hari Nominal	2: Tanggal Nominal	3: Negara Nominal	4: Provinsi Nominal	5: prediction margin Numeric	6: predicted Kasus_Harian Nominal
1	Minggu	31/01/2021	Indonesia	Aceh	0.813893	0-99
2	Minggu	31/01/2021	Indonesia	Bali	-0.771126	100-999
3	Minggu	31/01/2021	Indonesia	Bangka Belitung	0.747777	0-99
4	Minggu	31/01/2021	Indonesia	Banten	-0.694513	100-999
5	Minggu	31/01/2021	Indonesia	Bengkulu	0.876638	0-99
6	Minggu	31/01/2021	Indonesia	DI Yogyakarta	-0.915996	100-999
7	Minggu	31/01/2021	Indonesia	DKI Jakarta	-0.731403	1000-9999
8	Minggu	31/01/2021	Indonesia	Gorontalo	0.955436	0-99
9	Minggu	31/01/2021	Indonesia	Jambi	0.864602	0-99
10	Minggu	31/01/2021	Indonesia	Jawa Barat	-0.739672	1000-9999
11	Minggu	31/01/2021	Indonesia	Jawa Tengah	-0.677142	1000-9999
12	Minggu	31/01/2021	Indonesia	Jawa Timur	-0.855232	100-999
13	Minggu	31/01/2021	Indonesia	Kalimantan Barat	0.831852	0-99
14	Minggu	31/01/2021	Indonesia	Kalimantan Selatan	-0.621725	100-999
15	Minggu	31/01/2021	Indonesia	Kalimantan Tengah	0.784411	0-99
16	Minggu	31/01/2021	Indonesia	Kalimantan Timur	-0.838205	100-999
17	Minggu	31/01/2021	Indonesia	Kalimantan Utara	0.891852	0-99
18	Minggu	31/01/2021	Indonesia	Kepulauan Riau	-0.616929	100-999
19	Minggu	31/01/2021	Indonesia	Lampung	-0.473057	100-999
20	Minggu	31/01/2021	Indonesia	Maluku	0.928296	0-99
21	Minggu	31/01/2021	Indonesia	Maluku Utara	0.939527	0-99
22	Minggu	31/01/2021	Indonesia	Nusa Tenggara Barat	0.902068	0-99
23	Minggu	31/01/2021	Indonesia	Nusa Tenggara Timur	0.650295	0-99
24	Minggu	31/01/2021	Indonesia	Papua	0.872555	0-99
25	Minggu	31/01/2021	Indonesia	Papua Barat	0.90715	0-99
26	Minggu	31/01/2021	Indonesia	Riau	-0.710102	100-999
27	Minggu	31/01/2021	Indonesia	Sulawesi Barat	0.90545	0-99
28	Minggu	31/01/2021	Indonesia	Sulawesi Selatan	-0.603896	100-999
29	Minggu	31/01/2021	Indonesia	Sulawesi Tengah	-0.300162	100-999
30	Minggu	31/01/2021	Indonesia	Sulawesi Tenggara	0.941906	0-99
31	Minggu	31/01/2021	Indonesia	Sulawesi Utara	0.847875	0-99
32	Minggu	31/01/2021	Indonesia	Sumatera Barat	0.51496	0-99
33	Minggu	31/01/2021	Indonesia	Sumatera Selatan	0.63851	0-99
34	Minggu	31/01/2021	Indonesia	Sumatera Utara	0.846526	0-99
35	Minggu	28/02/2021	Indonesia	Aceh	0.873959	0-99
36	Minggu	28/02/2021	Indonesia	Bali	-0.715932	100-999
37	Minggu	28/02/2021	Indonesia	Bangka Belitung	0.706914	0-99
38	Minggu	28/02/2021	Indonesia	Banten	0.61309	0-99

Gambar 15. Hasil Prediksi Algoritma Random Forest

Hasil prediksi kelas Kasus_Harian yang tertera pada Gambar 15 dihasilkan dengan menggunakan model klasifikasi algoritma Random Forest.

hasilKNN.arff						
Relation: Dataset-1_predicted						
No.	1: Hari Nominal	2: Tanggal Nominal	3: Negara Nominal	4: Provinsi Nominal	5: prediction margin Numeric	6: predicted Kasus_Harian Nominal
1	Minggu	31/01/2021	Indonesia	Aceh	0.999513	0-99
2	Minggu	31/01/2021	Indonesia	Bali	-0.999513	100-999
3	Minggu	31/01/2021	Indonesia	Bangka...	0.999513	0-99
4	Minggu	31/01/2021	Indonesia	Banten	-0.999513	100-999
5	Minggu	31/01/2021	Indonesia	Bengkulu	0.999513	0-99
6	Minggu	31/01/2021	Indonesia	DI Yogy...	-0.999513	100-999
7	Minggu	31/01/2021	Indonesia	DKI Jak...	-0.999513	1000-9999
8	Minggu	31/01/2021	Indonesia	Goront...	0.999513	0-99
9	Minggu	31/01/2021	Indonesia	Jambi	0.999513	0-99
10	Minggu	31/01/2021	Indonesia	Jawa B...	-0.999513	1000-9999
11	Minggu	31/01/2021	Indonesia	Jawa T...	-0.999513	1000-9999
12	Minggu	31/01/2021	Indonesia	Jawa Ti...	-0.999513	100-999
13	Minggu	31/01/2021	Indonesia	Kalima...	0.999513	0-99
14	Minggu	31/01/2021	Indonesia	Kalima...	-0.999513	100-999
15	Minggu	31/01/2021	Indonesia	Kalima...	0.999513	0-99
16	Minggu	31/01/2021	Indonesia	Kalima...	-0.999513	100-999
17	Minggu	31/01/2021	Indonesia	Kalima...	0.999513	0-99
18	Minggu	31/01/2021	Indonesia	Kepula...	-0.999513	100-999
19	Minggu	31/01/2021	Indonesia	Lampu...	-0.999513	100-999
20	Minggu	31/01/2021	Indonesia	Maluku	0.999513	0-99
21	Minggu	31/01/2021	Indonesia	Maluku ...	0.999513	0-99
22	Minggu	31/01/2021	Indonesia	Nusa T...	0.999513	0-99
23	Minggu	31/01/2021	Indonesia	Nusa T...	0.999513	0-99
24	Minggu	31/01/2021	Indonesia	Papua	0.999513	0-99
25	Minggu	31/01/2021	Indonesia	Papua ...	0.999513	0-99
26	Minggu	31/01/2021	Indonesia	Riau	-0.999513	100-999
27	Minggu	31/01/2021	Indonesia	Sulawe...	0.999513	0-99
28	Minggu	31/01/2021	Indonesia	Sulawe...	-0.999513	100-999
29	Minggu	31/01/2021	Indonesia	Sulawe...	-0.999513	100-999
30	Minggu	31/01/2021	Indonesia	Sulawe...	0.999513	0-99
31	Minggu	31/01/2021	Indonesia	Sulawe...	0.999513	0-99
32	Minggu	31/01/2021	Indonesia	Sumate...	0.999513	0-99
33	Minggu	31/01/2021	Indonesia	Sumate...	0.999513	0-99
34	Minggu	31/01/2021	Indonesia	Sumate...	0.999513	0-99
35	Minggu	28/02/2021	Indonesia	Aceh	0.999513	0-99
36	Minggu	28/02/2021	Indonesia	Bali	-0.999513	100-999
37	Minggu	28/02/2021	Indonesia	Bangka...	0.999513	0-99
38	Minggu	28/02/2021	Indonesia	Banten	0.999513	0-99

Gambar 16. Hasil Prediksi Algoritma KNN

Kemudian hasil yang ditunjukkan oleh Gambar 4.16 adalah hasil prediksi menggunakan model klasifikasi dengan algoritma KNN.

F. Analisis Hasil Data

Pada tahapan analisis hasil data akan dibandingkan hasil kinerja dari model klasifikasi algoritma KNN dan Random Forest. Tujuannya untuk mengetahui model klasifikasi mana yang menghasilkan hasil yang paling baik dalam hal akurasi pembuatan model dan memprediksi data uji.

TABEL I
HASIL PERBANDINGAN AKURASI

	Random Forest	KNN
TP Rate	0.790	0.745
FP Rate	0.239	0.344

Precision	0.786	0.751
Recall	0.790	0.745
F-Measure	0.780	0.714
MCC	0.587	0.487
ROC Area	0.866	0.868
PRC Area	0.859	0.835
Time(Model)	0.32s	0s
Time(Testing)	0.31s	0.13s

Keterangan Warna: Merah = Buruk, Hijau = Baik

Menurut hasil dari perbandingan akurasi seperti yang terangkum pada Tabel 4.1, algoritma Random Forest menunjukkan hasil kinerja yang lebih baik dari algoritma KNN namun algoritma KNN mempunyai keunggulan didalam waktu pembuatan model dan testing.

Hari	Tanggal	Negara	Provinsi	Kasus_Harian	Prediksi Random Forest	Prediksi KNN
Minggu	31/01/2021	Indonesia	Aceh	0-99	0-99	0-99
Minggu	31/01/2021	Indonesia	Bali	100-999	100-999	100-999
Minggu	31/01/2021	Indonesia	Bangka Belitung	0-99	0-99	0-99
Minggu	31/01/2021	Indonesia	Banten	100-999	100-999	100-999
Minggu	31/01/2021	Indonesia	Bengkulu	0-99	0-99	0-99
Minggu	31/01/2021	Indonesia	DI Yogyakarta	100-999	100-999	100-999
Minggu	31/01/2021	Indonesia	DKI Jakarta	1000-9999	1000-9999	1000-9999
Minggu	31/01/2021	Indonesia	Gorontalo	0-99	0-99	0-99
Minggu	31/01/2021	Indonesia	Jambi	0-99	0-99	0-99
Minggu	31/01/2021	Indonesia	Jawa Barat	1000-9999	1000-9999	1000-9999
Minggu	31/01/2021	Indonesia	Jawa Tengah	1000-9999	1000-9999	1000-9999
Minggu	31/01/2021	Indonesia	Jawa Timur	100-999	100-999	100-999
Minggu	31/01/2021	Indonesia	Kalimantan Barat	0-99	0-99	0-99
Minggu	31/01/2021	Indonesia	Kalimantan Selatan	100-999	100-999	100-999
Minggu	31/01/2021	Indonesia	Kalimantan Tengah	0-99	0-99	0-99
Minggu	31/01/2021	Indonesia	Kalimantan Timur	100-999	100-999	100-999
Minggu	31/01/2021	Indonesia	Kalimantan Utara	0-99	0-99	0-99
Minggu	31/01/2021	Indonesia	Kepulauan Riau	100-999	100-999	100-999
Minggu	31/01/2021	Indonesia	Lampung	100-999	100-999	100-999

	Random Forest	KNN
TRUE	204	204
FALSE	0	0

Gambar 17. Perbandingan Hasil Prediksi

Gambar 17 memperlihatkan perbandingan hasil prediksi dari pengujian data yang menggunakan model klasifikasi algoritma KNN dan Random Forest, dari hasil yang didapat keduanya memiliki hasil yang sangat baik dengan tidak memiliki satupun data yang bernilai false dalam hal memprediksi Kasus_Harian pada setiap tanggal terakhir dari Bulan Januari hingga Juni di tahun 2021.

V. KESIMPULAN

Berdasarkan hasil penelitian yang berjudul Analisis Dataset COVID-19 menggunakan Algoritma KNN dan Random Forest, dapat disimpulkan dengan menggunakan model klasifikasi KNN dan Random Forest menunjukkan bahwa kedua model memberikan hasil yang sangat baik dalam memprediksi kasus aktif harian dengan masing-masing model

memberikan hasil prediksi yang bernilai True berjumlah sebanyak 204 data dan hasil prediksi kasus aktif harian dengan nilai False dengan jumlah nol, lalu mengenai kinerja antara model klasifikasi KNN dan Random Forest hasilnya menunjukkan kinerja algoritma KNN memiliki nilai akurasi yang lebih rendah tetapi dalam pembuatan model klasifikasinya sedikit lebih cepat dibandingkan dengan algoritma Random Forest yang memiliki akurasi lebih tinggi tetapi membutuhkan waktu yang sedikit lebih lama dalam proses pembuatan model.

DAFTAR PUSTAKA

- [1] J. Suntoro, *Data Mining: Algoritma Dan Implementasi Dengan Pemrograman PHP*, 2019.
- [2] I. H. Witten, E. Frank dan M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 2011.
- [3] B. Avelita, *A Klasifikasi K-Nearest Neighbor.*, 2013.
- [4] B. Adnyana, *Prediksi Lama Studi Mahasiswa Dengan Metode Random Forest*, 2016.