Prediction System for the Spread of Corona Virus in Central Java with K-Nearest Neighbor (KNN) Method

1stFarid Fitriyadi, 2ndMuqorobin

¹Informatika Universitas Sahid Surakarta, ²Informatika ITB AAS Indonesia ¹Universitas Sahid Surakarta, ²Institut Teknologi Bisnis AAS Indonesia ¹faridfitriyadi@gmail.com, ²robbyaullah@gmail.com

Abstract—Corona Virus is currently spreading very rapidly in many parts of Indonesia, including Central Java Province. According to the current data of corona database in Central Java, today on 17th of August 2021, the number of confirmed cases is; Confirmed in Treatment (Active Cases): 16.344, Confirmed Recovered: 408.697, and Confirmed Dead: 29.148. Therefore, the total number of cases is 454.189, obtained from the sum of the number of being treated, recovered, and dead. Corona Virus is a collection of viruses that can infect the respiratory system, generally mild, such as common cold, although, some forms of diseases like; SARS, MERS, and COVID-19 are more deadly. In anticipating this case, the government has created some policies which include; limiting activities outside the house, having school activities done from home, working from home, and even having religious activities done from home too. The purpose of this study was to predict the possible rate of new cases in one of Central Java areas with confirmed cases of corona virus. Thus, it can be used as information material for the public to anticipate early. The research method applied in this research is problem analysis and literature study, data collection and implementation. The application of the K-Nearest Neighbor (KNN) method is expected to be able to predict the level of spread of COVID-19 in Central Java. The results of the research on testing the prediction system for the new cases level were tested in the Sragen area. Testing is carried out by taking samples for new cases, namely Kudu Regency/City, Confirmed: 17,599, Treated: 89, Recovered: 18,303, Died: 1,721, Suspected: 87 and Discarded Suspected: 1,711. After doing the prediction with K-NN algorithm, it showed the Condition: High.

Keywords:K-Nearest Neighbor (KNN) Method, Corona Virus

I. INTRODUCTION

The world is currently cautious of the spread of a virus known as the corona virus. Corona viruses (CoV) are part of a family of viruses which cause illnesses ranging from the flu to more severe diseases such as Middle East Respiratory Syndrome (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS-CoV). The disease caused by the corona virus known as COVID 19 is a new type that was discovered in 2019 and has never been identified to attack humans before (World Health Organization, 2019). The case of the corona virus appeared and attacked humans for the first time in the province of Wuhan, China. Initially, it was suspected that it was pneumonia, with flu-like symptoms in general. These symptoms include cough, fever, fatigue, shortness of breath and no appetite. However, unlike influenza, the corona virus can develop quickly to cause more severe infections and organ failure (Alvina Felicia Watratan, 2020).

This emergency condition especially occurs in patients with previous health problems. Due to the very fast transmission of the corona virus, the World Health Organization (WHO) established the corona virus as a pandemic on March 11, 2020. The status of a pandemic or global epidemic indicates that the spread of COVID-19 is occurring very quickly that almost no country in the world can ensure that they are safe from the corona virus (Widiyani, 2020).

COVID-19 has spread to 196 countries. There have been 414,179 positive confirmed cases and 18,440 dead from positive confirmed cases. Meanwhile, COVID-19 was first confirmed in Indonesia on March 2, 2020 as of March 25, there were cumulatively 790 confirmed positive cases (cumulatively), including 58 died, 31 recovered, and 701 were in treatment. Based on these conditions, Indonesia is in a state of alert to the threat of the corona virus and so far no vaccine for COVID-19 has been found, so a form of prevention from the spread of the virus can be done by breaking the chain of transmission. One method to break the chain of transmission is by doing social distancing. With the social restrictions, it is hoped that every community will not become infectious or infected because they do not make contact with anyone so that the rate of spread can decrease. In this study, we will apply the Naive Bayes algorithm to predict the level of spread of COVID-19 in Indonesia by implementing the Naive Bayes Algorithm, which is useful as a measure to anticipate the COVID-19 pandemic (Alvina Felicia Watratan, 2020).

Data mining is a process of extracting or mining large data and information, which have not been known previously, but can be understood and useful from large databases and are used to make very important business decisions. Data mining describes a collection of techniques with the aim of finding unknown patterns in the data that has been collected. Data mining allows users to find knowledge in database data that is impossible for the user to know the existence of. Data mining is a process of automatically searching for useful information in large data storage places. As a series of processes, data mining can be divided into several stages of the process. These stages are interactive; the user is directly involved or through a knowledge base.

The stages of Data Mining are:

a. Data Cleaning

Data cleaning is the process of removing noise and inconsistent or irrelevant data.

b. Data Integration

Data integration is the merging of data from various databases into one new database.

c. Data Selection

The data in the database is often not used entirely, therefore only data that is suitable for analysis will be retrieved from the database.

d. Data Transformation

Data is converted or merged into a format which is suitable for processing in Data Mining.

e. Mining Process

It is a major process when methods are applied to find valuable and hidden knowledge from data. Several methods can be used based on data grouping. Pattern Evaluation is to identify interesting patterns in the knowledge based that are found.

f. Knowledge Presentation

It is a visualization and presentation of knowledge about the methods used to obtain the knowledge obtained by users.

K-Nearest Neighbor is one of the methods in making decisions using supervised learning where the results of the new input data are classified based on the closest in the value data [3].

The K-Nearest Neighbor (KNN) algorithm is a method for classifying objects based on the learning data that is closest to the object. KNN is a supervised learning algorithm where the results of the new query instance are classified based on the majority of the categories in the KNN algorithm. Where the class appears the most will be the class resulting from the classification [4].

Closeness is defined in terms of metric ranges, such as Euclidean range. Euclidean range [5] can be found using equation 1 below.

$$Dxy = \sqrt{\sum_{l=1}^{n} (x_l - y_l)^2}$$
 (1)

D : range of closeness

- x : data training
- y : data testing
- n : the number of individual attributes between 1 to n
- f : similarity function between case X and case Y
- i : individual attributes between 1 and n

The steps to calculate the K-Nearest Neighbor method [6] include;

- 1. Determine the parameter K (number of closest neighbors).
- 2. Calculate the square of the Euclidean range (query instance) of each object to the given sample data using equation 1.
- 3. Then sort the objects into groups that have the smallest Euclidean range.
- 4. Collect category Y (Nearest Neighbor Classification)
- 5. By using the category of Nearest Neighbor which is the majority, the value of the calculated query instance can be predicted.

PHP according to Anhar (2020: 3) is a server-side web programming language that is open source, PHP is also a script that is integrated with HTML and is located on the server (server side HTML embedded script). PHP is also a script used to create very dynamic web pages, dynamic means that the display page that will be displayed is created when the page is requested by the client. PHP was first created by Rasmus Lerdorf, a reliable C programmer from Greenland Denmark in 1995, PHP was named FI (Form Interpreted) which was used to manage from the web. In its development, the codes used were released to the public so that many programmers around the world began to develop them. In 1997 PHP was released with version 2.0, this version was already integrated with the C programming language and was equipped with modules so that the quality of PHP's work was significantly improved. In the same year, a programming company called Zend re-released this version of PHP with a better, cleaner and faster version. Along with the development of the era in 1994, PHP version 4.0 began to be released and this version was most widely used in the early 21st century because this version of PHP was able to build complex webs with high speed stability. In 2004, the programming company Zend released PHP again with its latest version 5.0 which the core of the PHPH interpreter underwent major changes. This version also incorporated an object-oriented programming model into PHP to answer the development of programming languages towards an object-oriented paradigm (Anhar, 2020).

Previous research by Rizki Tri Prasetio and Sari Susanti, from BSI University, 2019 with the entitled "Life Expectancy Prediction of Lung Cancer Patients Post Thoracic Surgery Using Boosted k-Nearest Neighbor" concluded that lung cancer ranks sixth out of the ten leading causes of death in Indonesia. Factors causing lung cancer are dominated by cigarette smoke. Thoracic surgery is one of the main solutions for lung cancer. However, there are many risks and complications after thoracic surgery that can lead to death. In this study, the life expectancy of lung cancer patients will be predicted after living one year of life after thoracic surgery using computer aided diagnosis (CAD). This prediction is done by analyzing the patient's condition before and after surgery. The data used in the study was secondary data which contained 470 data with a distribution of 400 patient data that were alive (survival) and 70 data of patients who died (die). Adaptive Boost was used as a level optimization algorithm on the k-nearest neighbor algorithm. The results of the study showed that the proposed method produces an accuracy of prediction of life expectancy of 85.11% using 10-fold cross validation with the k parameter in the k-nearest neighbor algorithm with a value of 5.

Previous research by RahandanuRachmat from Jendral Ahmad Yani University, 2020 with the title "Water Quality Prediction System in TirtaRaharja Regional Drinking Water Company Using K – Nearest Neighbors (K – NN)" concluded that PDAM (Regional Drinking Water Company) TirtaRaharja is one the only Regional Owned Enterprise having task of providing clean water services for the people of Cimahi City. Clean water is a basic need that is always consumed by the community and affects community activities. The development of the city of Cimahi is currently quite rapid, with the plan to build a smart city, causing the domestic demand for clean water to increase. K – Nearest Neighbor (KNN) is a classification

algorithm that considers several supporting parameters to carry out the classification process which results in ease of calculation time and power. KNN can be considered as one of the most well-known non-parametric models. In the research and implementation process of data mining on determining the feasibility of water quality at PDAM TirtaRaharja using K – Nearest Neighbors, it can be concluded that K – Nearest Neighbors which is implemented in the process of testing the feasibility of drinking water at PDAM TirtaRaharja, produces an accuracy of 93% for testing with the Eligible label. Drinkable, and 98% for accuracy testing with the label Not Worth Drinking with a K value of 14 where the K value is the most ideal amount that has been processed through K – Fold Cross Validation from a total of 1,818 data.

II. RESEARCH METHODS

The research method is a scientific process or method to obtain data that will be used for research purposes. The research was carried out in stages starting from planning, determining the research focus, research time, data collection, analysis, and presentation of research results. The research methods applied in this study are as follows:

1. Problem Analysis and Literature Study

This stage is the first step to determine the formulation 3. of the research problems, which in this case is observing problems related to the level of spread of COVID-19 that occurs especially in Indonesia. The existing 4. problems are then analyzed to find out how to solve them and determine the scope of the problems to be studied. The theoretical basis from various literatures regarding the application of the Naive Bayes method, concepts and theories of data mining and prediction of the rate of spread of COVID-19 in Indonesia, from journals are studied in order to obtain a knowledge base to conduct further research.

2. Collecting Data

The systematic procedure used to collect data is quantitative research methods. The implementation of quantitative research methods focuses on the use of numbers, tables, graphics and diagrams to display the results of the data obtained and those will be used for data analysis of the K-Nearest Neighbor method. As for data collection, researchers took data directly on the official web address: https://corona.jatengprov.go.id/data. After the data is collected, data analysis is carried out to adjust the data process to be processed in the K-Nearest Neighbor method.

3. Implementation

In accordance with data processing, the implementation stage is about how the data processing is applied in a system. As for testing this system, the researcher built a prototype using the PHP programming language to implement the Prediction of the Spread of Corona Virus Cases in Central Java. Further testing was done to determine whether the research conducted was in accordance with the expected objective which is to predict the level of spread of COVID-19 in Central Java Province.

III. RESULT AND DISCUSSION

In this discussion, the researcher wrote down the stages of the data processing process to produce a prediction system for the spread of the corona virus in Central Java.

1. Data on the spread of the corona virus in Indonesia increased in July. Meanwhile, in August it has started to experience a decline



Figure 1. Condition of the Spread of the Corona Virus in Indonesia

- 2. At this initial stage, researcher obtained the data on the spread of corona cases in Central Java by from official website: https://corona.jatengprov.go.id/data
- 3. After that, the data was taken and data processing was carried out as needed in the K-Nearest Neighbor Algorithm.
- 4. The data that researcher obtained from the official website is still in the form of original data as shown in Figure 2.

and middle comments	× +					0	
C a coronajateng	prov.go.id/data					Q 🕁	*
as 🚊 Kurausi Manajornen	🕲 Tokobagus - Situs j M Ginal	🖬 YouTube 🐮 Maps 🏟 Settings -	Reservorda				
Tanggap COVID- Provinci Jawa Tangah	19	Boranda Data Vaksin	asi Borita R	IS Rujukan Donasi	FAQ Downi	load Link Deteksi M	Indiri
	1	Tabel Sebaran COVID-19	Jawa Teng	gah			
Rumah Sakit	Kabupaten/Kota						
O Data dibawah meru	pakan data pasien berdasarka	n Kabupaten/Kota Domisili					
0 Data dibawah meru	pakan data pasien berdasark:	an Kabupaten/Kota Domisili					
O Data dibawah meru Kabupaten/Kata	pakan data pasien berdasark: Tekorfirmat	an Kabupatèn/Kota Domisili Drawi Isramit-singkt-singer namini	Sembuh	Meninggal	Surges	Suspek Discorder	
Data dibawah meru Kabupaten/Kata Kota Senangang	pakan data pasien berdasərk Tettorfirməl 33.710	an Kabupaten/Kota Domistil Drevet Novel 1 State Temperatur 1 State Temperatur	Semboli 29.635	Meninggal 3.724	Surgesk 52%	Suspek Discercie 11:34	
Data dibawah meru Kabupaten/Rata Kora Sebagaten Kuta	pakan data pasien berdasarkt TenanTimat 81.776 25.776	an Kabupatèn/Kota Domisili Drewet Inner * sigit * sener Bit 1579	Sembuti 29.655 24.113	Meningai 3.724 1.624	Sunpek 578	Suspek Discercie 11.54 3.78	
Date dibawah meru Katupeter/Tota KOTA SENARJANG KLATEN BANYUMAS	pakan data pasien berdasarka Tekan meta Kutha Safte 26.475	an Kabupaten/Kota Domisili Demos Isrone*sjupt+sonemet Str 1379 723	Sembuti 29.635 24.113 24.919	Meninggal 1.724 1.024 828	Surpek 526 992 1.186	Surgek Discertie 11.34 3.78 6.22	
Data dibawah meru Kaaupeter/Kata Kota sekuatang KLATEN BANYUMAS JEFRAA	pakan data pasien berdasarka Tehartheat State 28.476 28.475 17,482	an Kabupaten / Kota Domisili Israe "sight-sam nami 1377 288 183	Sembuch 29.835 24.113 24.919 18.571	Meningen 1.724 1.024 628 809	Surpek 52h 592 5.186 395	Sunpek Discerde 11.34 3.78 4.22 3.55	
Data dibawah meru Kasajaren/Tata Kota senakajag KLATEN BANYUMAS JEFABA MAGELANS	pakan data pasien berdasarka Tarbat finad State	an Kabupaten Kota Domisili Israe Tanjat Saar nadi 1375 1381 1383 1383 1383 1383	Semburti 29.855 24.919 36.871 15.870	Meningai 1.724 1.024 8.08 8.08 8.08 732	Surpek 53% 592 5.186 395 106	Surpek Discertin 11.34 3.75 6.22 3.15 2.12	
Data dibawah menu Kebuperentara Kota sekarang KLATEN Boahrunas Jenaka Magitakag KESUMEN	Image: select berdsaukt Todustmass 2 3 4 3	n Kabupatén Késa Demisili Demi Pasat ngat nga nga 1994 1995 1995 1995 1995 1995 1995 1995	Semburh 29.635 24.113 24.919 16.571 15.570 15.275	Meningan 1.724 1.024 828 808 732 940	Surgek 528 992 1.186 195 108 206	Surpek Discertin 11.34 3.76 4.22 3.15 2.12 2.12 2.12	

Figure 2. Data Source of the Confirmation of the Spread of the Corona Virus in Central JavaProvince

- 5. After that, the researcher took and processed the data for the calculation process using the K-Nearest Neighbor method.
- 6. In the data, there was no classification label, so the researcher added a column to determine the classification of labeling at the case level.
- 7. As for giving classification labels, the researchers made four case categories, namely Very High, High, Medium and Low.
- 8. The way to determine these categories is to carry out a data normalization process based on the highest, average and lowest number of cases, then the rank of range between cases is obtained based on case data

dated 17 August 2021. Then researchers can retrieve data as shown in table 1.

Description	Case	Total number
Average	12.734	12.734
MIN	3.802	3.802
MAX	33.710	33.710
Range - High	33.710 - 12.734 = 20.976	20.976
Range- Average	12.734 - 3.802 = 8.932	8.932

9. Based on the table above, a category of case conditions can be made, as shown in table 2

Table 2. Number of Level Category of Corona Virus Cases in Central Java

Number of Cases			Category
20.976	-	33.710	Very high
12.734	-	20.976	High
8.932	-	12.734	Average
3.357	-	8.932	Low

10. After that, the data of corona cases in Central Java will change with the addition of column of the condition for the level of virus spread, so that the data can be interpreted as training data, this can be seen in table 3. Table 3. Training Data of Corona Virus

with Training Category

	D (61)				CONTRACTO
No	Regency/City	Confirmed	Suspects	Discarded	CONDITION
				Suspects	RESULT
1	Semarang Ci	33.710	526	11.348	Very high
2	Klaten	26.716	992	3.788	Very high
3	Banyumas	26.475	1.186	6.224	Very high
4	Jepara	17.562	395	3.155	High
5	Magelang	17.336	108	2.125	High
6	Kebumen	16.911	206	2.232	High
7	Kudus	16.685	49	2.711	High
8	Cilacap	16.161	713	1922	High
9	Kendal	15.815	215	3676	High
10	Purworejo	15665	403	2001	High
11	Surakarta	15323	244	2079	High
12	Sragen	14776	180	1430	High
13	Karanganyar	14.465	614	2236	High
14	Tegal	14.390	671	4613	High
15	Semarang	14288	170	1850	High
16	Wonosobo	12816	114	2875	High
17	Demak	11682	129	3996	Average
18	Pemalang	11434	144	3929	Average
19	Blora	11181	326	2113	Average
20	Temannggung	10939	130	1452	Average
21	Bovolali	10579	251	1368	Average
22	Sukoharjo	10401	621	2772	Average
23	Wonogiri	10089	94	1980	Average
24	Purbalingga	9456	300	1468	Average
25	Banjarnegara	9034	483	3626	Average
26	Brebes	8245	313	1466	Rendah
27	Batang	7560	31	1018	Rendah
No	City	Confirmed	Suspects	Discarded	CONDITION
			~	Suspects	RESULT
28	Pati	7359	31	2236	Low
29	Pekalongan	7210	127	1520	Low
30	Rembang	6600	11	889	Low
31	Grobogan	6138	74	2188	Low
32	Magelang	5591	22	682	Low
33	Pekalongan	5086	12	340	Low
34	Salatiga	4219	65	609	Low
35	Tegal	3802	124	512	Low
55	10500	5002	147	512	L0 W

11. After that, the researcher conducted a calculation on the K-Nearest Neighbor method. The formula is as follows:

$$Dxy = \sqrt{\sum_{l=1}^{n} (x_l - y_l)^2}$$
 (1)

- 12. As for system testing, researcher will use data testing, which is providing new training data by making case examples as shown in table 4.
- 13. Based on the testing data, the researchers performed calculations by calculating using the K-Nearest Neighbor (KNN) method to predict classification based on past data.

Table 5. K-NNThe Result of K-NNMethodCalculation.

1	KOTA SEMARANG	259.564.321	68.644	128.414.224	4.012.009	192.721	92.871.769	485.123.688
2	KLATEN	83.119.689	2.220.100	33.756.100	485.809	819.025	4.313.929	124.714.652
3	BANYUMAS	78.783.376	408.321	43.771.456	797.449	1.207.801	20.367.169	145.335.572
4	JEPARA	1.369	8.836	2.999.824	833.569	94.864	2.085.136	6.023.598
5	MAGELANG	69.169	297.025	5.442.889	978.121	441	171.396	6.959.041
6	KEBUMEN	473.344	352.836	9.144.576	595.984	14.161	271.441	10.852.342
7	KUDUS	835.396	784	9.000.000	160.000	1.444	1.000.000	10.997.624
8	CILACAP	2.067.844	1.142.761	15.832.441	1.085.764	391.876	44.521	20.565.207
9	KENDAL	3.182.656	47.524	13.191.424	781.456	16.384	3.861.225	21.080.669
10	PURWOREJO	3.740.356	528.529	16.337.764	1.283.689	99.856	84.100	22.074.294
11	KOTA SURAKARTA	5.180.176	746.496	23.213.124	698.896	24.649	135.424	29.998.765
12	SRAGEN	7.969.329	442.225	30.118.144	264.196	8.649	78.961	38.881.504
13	KARANGANYAR	9.821.956	819.025	32.947.600	660.969	277.729	275.625	44.802.904
14	TEGAL	10.297.681	177.241	28.217.344	692.224	341.056	8.421.604	48.147.150
15	SEMARANG	10.962.721	165.649	28.976.689	720.801	6.889	19.321	40.852.070
16	WONOSOBO	22.877.089	199.809	42.497.361	1.500.625	729	1.354.896	68.430.509
17	DEMAK	35.010.889	7.396	63.920.025	272.484	1.764	5.221.225	104.433.783
18	PEMALANG	38.007.225	29.584	63.266.116	804.609	3.249	4.919.524	107.030.307
19	BLORA	41.190.724	21.025	62.631.396	1.352.569	57.121	161.604	105.414.439
20	TEMANGGUNG	44.355.600	65.536	65.804.544	1.737.124	1.849	67.081	112.031.734
21	BOYOLALI	49.280.400	182.329	73.925.604	1.857.769	26.896	117.649	125.390.647
22	SUKOHARJO	51.811.204	399.424	92.659.876	515.524	285.156	1.125.721	146.796.905
23	WONOGIRI	56.400.100	202.500	96.079.204	451.584	49	72.361	153.205.798
24	PURBALINGGA	66.308.449	361.201	97.752.769	1.879.641	45.369	59.049	166.406.478
25	BANJARNEGARA	73.359.225	201.601	104.060.401	1.760.929	164.836	3.667.225	183.214.217
26	BREBES	87.497.316	78.400	128.754.409	641.601	51.076	60.025	217.082.827
27	BATANG	100.781.521	2.809	128.936.025	1.565.001	3.136	480.249	231.768.741
28	PATI	104.857.600	1.024	150.234.049	279.841	3.136	275.625	255.651.275
29	PEKALONGAN	107.931.321	16.900	139.145.616	1.530.169	1.600	36.481	248.662.087
30	REMBANG	120.978.001	1.089	154.380.625	1.113.025	5.776	675.684	277.154.200
31	GROBOGAN	131.354.521	2.304	168.143.089	1.115.136	169	227.529	300.842.748
32	KOTA MAGELANG	144.192.064	63.001	176.570.944	2.205.225	4.225	1.058.841	324.094.300
33	KOTA	156.575.169	1	184.796.836	2.050.624	5.625	1.879.641	345.307.896
34	KOTA SALATIGA	179.024.400	12.321	207.907.561	2.515.396	484	1.214.404	390.674.566
35	KOTA TEGAL	190.357.209	13.924	223.382.916	2.199.289	1.369	1.437.601	417.392.308

- 14) After that, the researcher took the closest range from the calculation results and categorized it according to the condition category.
- 15) Based on the calculation of the K-NN Algorithm, the researchers took the value of K = 3. There, it can be seen that there are more number of "Highest" condition categories, so the prediction results state that the testing data produces the following condition: Highest.

Tabel 7. Hasil KlasifikasiMetode K-NN The Classification Results of K-NN Method

JEPARA	1369	8836	2999824	833569	94864	2085136	6023598	Tinggi
MAGELANG	69169	297025	5442889	978121	441	171396	6959041	Tinggi
KEBUMEN	473344	352836	9144576	595984	14161	271441	10852342	Tinggi
KUDUS	835396	784	9000000	160000	1444	1000000	10997624	Tinggi
CILACAP	2067844	1142761	15832441	1085764	391876	44521	20565207	Tinggi
KENDAL	3182656	47524	13191424	781456	16384	3861225	21080669	Tinggi
PURWOREJO	3740356	528529	16337764	1283689	99856	84100	22074294	Tinggi
KOTA SURAKARTA	5180176	746496	23213124	698896	24649	135424	29998765	Tinggi
SRAGEN	7969329	442225	30118144	264196	8649	78961	38881504	Tinggi
SEMARANG	10962721	165649	28976689	720801	6889	19321	40852070	Tinggi
KARANGANYAR	9821956	819025	32947600	660969	277729	275625	44802904	Tinggi
TEGAL	10297681	177241	28217344	692224	341056	8421604	48147150	Tinggi
WONOSOBO	22877089	199809	42497361	1500625	729	1354896	68430509	Tinggi
DEMAK	35010889	7396	63920025	272484	1764	5221225	104433783	Sedang

16. Based on the results of the classification above, the system testing is stated for the following condition: Highest.

Table 8. The Prediction Result of the Condition of Corona

Virus Spread							
Regency	Confirmed	Dead	Suspects	Discarded Suspects	Result		
Kudus	17.599	1.721	87	1711	The highest		

- 17. To make the implementation of the system easie, the researchers made a prototype system to run the system flow using the K-NN Algorithm method
- 18. The interface is still made simple using the PHP programming language, the following is a display of the Input Data Testing Interface:



Figure 3. The Implementation of Prediction System

19. After inputting the system, it will produce the following



20. The test is stated valid because the results of the appointment are the same as the tests carried out on manual calculations and the prototype program obtains the same results, the system is stated valid.

VI. CONCLUSION

Based on study on the Prediction of the Spread of the Corona Virus in Central Java Province with the K-NN Method, the followings are the conclusions that can be drawn:

- a. The K-NN method utilizes training data to generate the probability of each criteria for a different class, so that the probability values of these criteria can be optimized to predict the level of Covid-19 spread in Central Java based on the classification process which is carried out by the Naive Bayes method itself.
- b. Based on the data on COVID-19 Cases Per Regency/City which was used as training data, the K-NN method successfully classified those which were tested. Hence, the K-NN method is successful in predicting the number of COVID-19 cases per Regency/City.
- c. The implementation of the system by making a prototype of the PHP programming language has been able to test the system by producing test results which are in accordance with the algorithm calculation.

REFERENCES

- Mona, Nailul. 2020. Konsep Isolasi Dalam Jaringan Sosial Untuk Meminimalisasi Efek Contagious (Kasus Penyebaran Virus Corona Di Indonesia). Jurnal Sosial HumanioraTerapan (JSHT), 2(2), pp. 117. doi: https://doi.org/10.7454/jsht.v2i2.86.
- Saleh, Alfa. 2015. Implementasi Metode Klasifikasi [2] Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. Creative Information Technology Journal (Citec Journal), 2(3), 208-216. doi: https://doi.org/10.24076/ pp. citec.2015v2i3. 49.
- [3] Bustami. 2013. PenerapanAlgoritma Naive Bayes UntukMengklasifikasi Data NasabahAsuransi. TECHSI (JurnalPenelitian Teknik Informatika), 3(2),pp. 127-146. doi: https://doi.org/10.29103/techsi.v5i2.154.
- [4] Ridwan, M., Suyono, H., Sarosa, M. 2013. Penerapan Data Mining untukEvaluasiKinerja AkademikMahasiswaMenggunakanAlgoritma Naive Bayes Classifier. Jurnal EECCIS, 1 (7), pp. 59-64.
- [5] Mujiasih, S. 2011. Pemanfaatan Data Mining UntukPrakiraanCuaca. JurnalMeteorologidan Geofisika, 12(2), pp. 189-195. doi: http://dx.doi.org/10.31172/jmg.v12i2.100
- [6] Yunus, N. R., &Rezki, Annisa. 2020. KebijakanPemberlakuan Lockdown SebagaiAntisipasiPenyebaran Corona Virus Covid-19. SALAM; JurnalSosial&BudayaSyar-i, 7(3), pp. 227-238. doi: <u>https://doi.org/10.15408/sjsbs.v7i3.15083</u>.
- [7] Muqorobin, M., &Rais, N. A. R. (2020). Analysis of the Role of Information Systems Technology in Lecture Learning during the Corona Virus Pandemic. International Journal of Computer and Information System (IJCIS), 1(2), 47-51.
- [8] Hikmah, I. N., &Muqorobin, M. (2020). Employee Payroll Information System On Company Web-Based Consultant Engineering Services. International Journal of Computer and Information System (IJCIS), 1(2), 27-30.

- [9] Muslihah, I., & Muqorobin, M. (2020). Texture Characteristic of Local Binary Pattern on Face Recognition with PROBABILISTIC LINEAR DISCRIMINANT ANALYSIS. International Journal of Computer and Information System (IJCIS), 1(1), 22-26.
- [10] Muryani, A. S., & Muqorobin, M. (2020). Decision Support System Using Cloud-Based Moka Pos Application To Easy In Input In Orange Carwash Blulukan Flash NO. 110 Colomadu. International Journal of Computer and Information System (IJCIS), 1(3), 66-69.
- [11] Jannah, N. F., & Muqorobin, M. (2021). Analysis Of Kasir Applications In Sales Management Information Systems at ASRI Store. International Journal of Computer and Information System (IJCIS), 2(2), 40-44.
- [12] Santoso, L. P., Muqorobin, M., & Fatkhurrochman, F. (2020). Online Analysis System of Application of Partners for Land Asrocument Officers of Sukoharjo District. International Journal of Computer and Information System (IJCIS), 1(3), 59-61.
- [13] Muqorobin, M., Apriliyani, A., & Kusrini, K. (2019). Sistem Pendukung Keputusan Penerimaan Beasiswa dengan Metode SAW. Respati, 14(1).
- [14] Muqorobin, M., & Rais, N. A. R. (2020, November). ANALISIS PERAN TEKNOLOGI SISTEM INFORMASI DALAM PEMBELAJARAN KULIAH DIMASA PANDEMI VIRUS CORONA. In Prosiding Seminar Nasional & Call for Paper STIE AAS (pp. 157-168).
- [15] Muqorobin, M., & Rais, N. A. R. (2020). Analysis of the Role of Information Systems Technology in Lecture Learning during the Corona Virus Pandemic. International Journal of Computer and Information System (IJCIS), 1(2), 47-51.