Mühendislik Bilimleri ve Tasarım Dergisi
10(4), 1212 – 1221, 2022
e-ISSN: 1308-6693

Journal of Engineering Sciences and Design
DOI: 10.21923/jesd.976865

Araştırma Makalesi Research Article

PREDICTING DIAGNOSIS OF COVID-19 DISEASE WITH ADABOOST AND NAIVE BAYES MACHINE LEARNING ALGORITHMS

Ahmet ÇELİK*

Kütahya Dumlupınar University, Tavşanlı Vocational School, Department of Computer Technologies, Kütahya, Türkiye

Keywords

Machine Learning, Data Mining, Naive Bayes, AdaBoost, Classification, COVID-19.

Abstract

Coronavirus disease (COVID-19) has infected millions of people all over the world and caused the death of many people. Identifying people with this disease as soon as possible is an important factor to prevent the disease from spreading. For disease detection, PCR (Polymerase Chain Reaction) tests performed. The results of tests always cannot give 100% accurate. In addition, obtaining information about test results sometimes may take a few days. Regarding the persons who applied to health institutions with suspicion of that illness, the diagnosis of COVID-19 disease takes place with the emergence of different disease symptoms. In this study, diagnostic estimates made for patients in the COVID-19 Surveillance dataset implementing Adaboost and Naive Bayes machine learning (ML) algorithm. It is possible to make predictions about new data by gaining experience from pre-existing data by means of using ML algorithms. In dataset determined within international disease codes for COVID-19 disease diagnosis estimates. Symptoms of patients used as attribute data in the dataset and used in binary format to be suitable for machine learning algorithms. According to the results obtained in this study, the classification forecast made with 85% accuracy with the Naive Bayes algorithm and 100% with the Adaboost algorithm.

COVID-19 HASTALIK TEŞHİSİNİN ADABOOST VE NAİVE BAYES ALGORİTMALARIYLA TAHMİN EDİLMESİ

Anahtar Kelimeler

Makine Öğrenmesi, Veri Madenciliği, Naive Bayes, AdaBoost, Sınıflandırma, COVID-19.

Öz

Koronavirüs hastalığı (COVID-19) , tüm dünyada milyonlarca insana bulaşmış ve birçok insanın ölümüne sebep olmuştur. Bu hastalığı taşıyan kişilerin en kısa sürede tespit edilmesi, hastalığın yayılmasına engel olmaktadır. Hastalık tespiti için PCR (Polymerase Chain Reaction) testleri yapılmaktadır. Bu testleri sonuçları %100 doğrulukta olmamaktadır. Ayrıca test sonuçlarının öğrenilmesi bazı durumlarda birkaç gün zaman alabilmektedir. Hastalık şüphesiyle sağlık kuruluşlarına başvuran kişilerin COVID-19 hastalık teşhisi farklı hastalık belirtilerinin varlığı kullanılarak gerçekleşebilmektedir. Bu çalışmada, Adaboost ve Naive Bayes denetimli makine öğrenme algoritması kullanılarak COVID-19 Surveillance veri seti içindeki hastaların COVID-19 teşhis tahminleri gerçekleştirilmiştir. Makine öğrenmesi algoritmaları kullanılarak önceden var olan verilerden tecrübe kazanarak yeni veriler hakkında tahminler yapılabilmektedir. Bu çalışmada, COVID-19 hastalık teşhis tahminlerinde, uluslararası hastalık kodlarıyla belirtilen veriler kullanılmıştır. Veri setindeki hastaların gösterdiği belirtiler öznitelik bilgisi olarak kullanılmıştır. Öznitelik verileri makine öğrenme algoritmalarına uygun olması için ikili formatta kullanılmıştır. Bu çalışmada elde sonuçlara göre Naive Bayes algoritmasıyla %85, Adaboost algoritmasıyla %100 doğrulukta sınıflandırma tahmini gerçekleştirilmiştir.

Alıntı / Cite

Çelik, A., (2022). Predicting Diagnosis of COVID-19 Disease with Adaboost and Naive Bayes Machine Learning Algorithms, Journal of Engineering Sciences and Design, 10(4), 1212-1221.

* İlgili yazar / Corresponding author: ahmet.celik@dpu.edu.tr, +90-274-443-6745

Yazar Kimliği / Author ID (ORCID Number)	Makale Süreci / Article Process	
A. Çelik, 0000-0002-6288-3182	Başvuru Tarihi / Submission Date	31.07.2021
	Revizyon Tarihi / Revision Date	11.07.2022
	Kabul Tarihi / Accepted Date	24.07.2022
	Yayım Tarihi / Published Date	30.12.2022

1. Introduction

There are many types of Coronaviruses around the world with different symptoms. Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) are the most common types. Coronavirus Disease 2019 (COVID-19) with different symptoms as a species was first seen in China (Wiguna and Riana, 2020).

This virus is very fast contagious, therefore, so it is extended to many countries around the world and millions of people lost their lives. By PCR (Polymerase Chain Reaction) test, COVID-19 patients are identified. However, getting information about PCR test results may take hours or days. This situation can prolong the treatment process of the patients and the isolation period of the persons at risk of infection. Since millions of people around the world are likely to get this disease, it will be useful to diagnose the disease by a very rapid decision system.

However, depending on the number of patients, a large data stack has been formed. Still, the number of data is gradually increasing (Çelik, 2020). In addition, large amounts of processed or unprocessed data are created by many organizations around the world over the years. These data need to be analyzed quickly. This can be accomplished by data mining methods (Kumar and Singh, 2019). Data Mining can be defined as finding data that has the potential to be useful. Data mining is to reveal the unknown important and effective information from the database for a specific decision-making task. From metadata stored in data storage areas, there is a possibility to find those with interesting features (Chen et al., 2018).

Social media usage data, weather forecast data, internet usage data, cyber security data, electronic shopping data, mobile operator customer data, industry production data, geographic data, mobile operator data, diagnosis and treatment data in the field of health, text content data, image data is in very large amounts. It is very useful to analyze this huge amount of data on the basis of data mining and gain the ability to predict and make decisions about new data by a system using machine learning algorithms.

Wiguna and Riana (2020), in their study, used the C4.5 decision tree algorithm on COVID-19 Surveillance dataset. C4.5 decision tree algorithm is a data mining method using for machine learning. In this study, a system that automatically decides on COVID-19 symptom data has been developed.

Çelik (2020), in his study used the Apriori algorithm to detect COVID-19 disease. Apriori algorithm was used on symptom data shown with ICD (International Classification of Diseases) codes. COVID-19 Surveillance dataset was used as dataset. This dataset contains information on 14 patients showing seven symptoms. In this study, it was observed that the patients with the A01, A02 and A04 symptoms were 100% COVID-19.

Kumar and Singh (2019), in their study, huge data in health services were analyzed with the tools available in the Hadoop system. In this study, by examining the medical database, electronic health records, image-text and clinical decision support system data, predictions have been made with data learning methods.

Randhawa et al. (2018), in their study, detected credit card fraud by using machine learning algorithms in their study. Credit card fraud is a huge problem in financial services. Customers who own a credit card lose billions of dollars each year due to credit card counterfeits. In this study, AdaBoost and majority voting methods were applied separately and hybrid.

Xiao et al. (2019), in their study, used the Adaboost-based method to quickly and accurately estimate joint motion from the surface electromyogram (sEMG). Although different loads on the joint were applied in the study, accurate prediction performance was obtained.

Wang et al. (2019), in their study, based on feature learning, AdaBoost + SVM machine learning algorithms were used together. First, classifiers were taught with Adaboost. Later, these were accepted as features and used on the SVM algorithm. In this study, tests were carried out on four class, ionosphere, chess, monk1 datasets.

Xu and Yuan (2020), in their study, developed an AdaBoost classifier for sonar images with low resolution and noise in the Cifar-10 dataset. Directed gradient (HOG) Histogram is used to perform feature extraction and Support vector machine (SVM) algorithms have been used. Then Adaboost algorithm was used on these properties.

According to the results, approximately 92% was obtained. This result reveals the higher level compared to general methods.

Wu and Zhu (2008), in their study, classified the sample and shape information on the invoice as an attribute using the Bayesian classification method in their study. In this study, 96% success rate was obtained. In addition, manual classification realized in 36 minutes, has been reduced to 2.7 minutes with the method.

Yılmaz and Öztürk (2019), in their study, created a model based on the Bayes algorithm on the Geographical Information System (GIS) data obtained from reference sources in order to create a forest fire risk map. Self-attributes were created by creating variable definitions and ranges of CBS data. In this study, the risk assessment of forest fires creating negative results in many aspects has been applied on a dynamic model.

Olgun and Özdemir (2013), in their study, compared the classification success of Artificial Neural Networks and Bayes machine learning algorithms of Shewhart control graphic patterns based on pattern recognition in their study. For the control chart patterns, normal, repetitive, up trending, down trending, Up Sudden change, Down Sudden change have been used. As a result, it has been revealed that Bayesian pattern recognition has better classification success than artificial neural networks.

In this study, Adaboost and Naive Bayes machine learning algorithms were applied on the data in the COVID-19 Surveillance dataset published in the UCI (University of California, Irvine) machine learning repository. Decision making success rates were tested using seven disease symptoms of 14 patients in this dataset. Data of patients in the same dataset were used to test their success performance. In the results obtained, it has been observed that Adaboost algorithm gave more accurate results. More reliable results will be obtained as the number of sample patients increases.

2. Material and Method

AdaBoost and Naive Bayes algorithms were used to predict PUS (Patient Under Supervision), PIM (Person in Monitoring) and PWS (Person without Symptoms) classification on the COVID-19 Surveillance dataset. Prediction success was measured using the unclassified data of the patients in the dataset. The flow chart of the study is shown on Figure 1. The patient with the PUS Class should be quarantined at home until the PCR test result is available. The patient with the PIM Class should be kept under observation for COVID-19. The patient with PWS Class does not show any signs of disease.

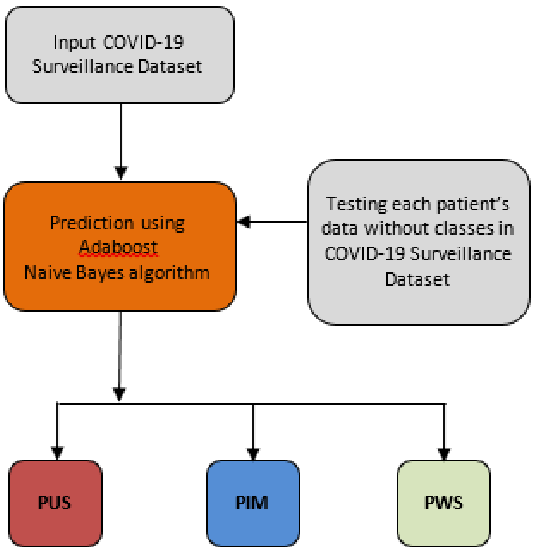

Figure 1. Flow Diagram of The Implemented Model

In the COVID-19 Surveillance dataset, seven disease symptoms are given as attribute information. Disease symptoms are given in ICD (International Classification of Diseases) codes. Symptoms of many diseases have been published with these codes in medical literature and are known all over the world. Latest ICD-11 list was published at 2018 by World Health Organization (WHO) (WHO, 2005). The ICD disease codes within the dataset are shown in the Table 1. The table shows the symptoms and effects of the disease indicated by ICD codes A01-A07.

Table 1. A00-A09 Diseases and Codes (WHO, 2005).

Disease Name	Disease Code
Typhoid and paratyphoid fevers	A01
Other salmonella infections	A02
Shigellosis	A03
Other bacterial intestinal infections	A04
Other bacterial foodborne intoxications	A05
Amoebiasis	A06
Other protozoal intestinal diseases	A07

The COVID-19 Surveillance dataset used in this study was published in the UCI data repository on April 24, 2020. This dataset contains ICD code information for seven disease symptoms that can only be used in the diagnosis of Coronavirus (Dua and Graff, 2019). The content of the COVID-19 Surveillance dataset is shown in the Table 2.

Table 2. COVID-19 Surveillance Dataset Data (Dua and Graff, 2019).

A01,A02,A03,A04,A05,A06,A07	Classes
+,+,+,+,+,-,-	PUS
+,+,-,+,+,-,-	PUS
+,+,+,+,-,+,-	PUS
+,+,-,+,-,+,-	PUS
+,-,-,-,-,-,+	PUS
+,+,+,-,-,-,+	PUS
+,+,-,-,-,-,+	PUS
+,+,+,+,-,-,-	PUS
+,-,-,+,+,-,-	PIM
-,+,-,+,+,-,-	PIM
+,-,-,+,-,+,-	PIM
-,+,-,+,-,+,-	PIM
-,+,-,-,-,-,+	PIM
-,-,-,-,-,-,+	PWS

In the dataset, defined by the ICD codes, if a disease symptom, it is indicated with the "+" symbol, if no disease symptom, it is indicated with the "-" symbol. In the COVID-Surveillance dataset, patients are divided into classes. These classes are PUS (Patient Under Supervision), PIM (Person in Monitoring) and PWS (Person without Symptoms) (Dua and Graff, 2019).

2.1. AdaBoost Algorithm

The AdaBoost (Adaptive Boosting) algorithm was developed by Freund and Schapire in 1996 (Freund and Schapire, 1999). The AdaBoost algorithm is an algorithm that raises the weak classification in a dataset to a strong classification (Wang et al., 2019).

The AdaBoost algorithm was developed as an effective reinforcement algorithm to improve the accuracy of the classification of a "weak" learning algorithm. By doing the learning correctly, this algorithm can classify the examples that are difficult to classify correctly. AdaBoost algorithm can be applied to most classifier learning algorithms. In the AdaBoost algorithm, the sample weight value is realized by using both upward samples and downward samples. It automatically updates the data space for optimum classification. By weighing the samples, adaptation to new samples is achieved with little loss (Sun et al., 2006).

AdaBoost is used with different types of algorithms to increase application performance. The misclassified data and fewer samples are paid more attention to data samples with AdaBoost algorithm. However, it is also sensitive to noise and external sample values. AdaBoost algorithm can improve results from different algorithms as long as the classification is not random (Randhawa, 2018).

In the AdaBoost algorithm, x instances in the dataset using attribute values can be classified into classes. Equation 1 shows these classes.

$$x_i, x_{i+1}, x_{i+2} \ \ and \ y_n \in \{-1,1\} \qquad (1)$$

The class variable y_n takes the values -1 or 1. The variable n indicates the number of data x in the dataset. i is to show metadata. The weight calculation in the dataset is shown in equation 2.

$$w_i = \frac{1}{n} \qquad (2)$$

w_i values, indicates the weight of the x_i.

The error rate of misclassified x_i data is shown in equation 3.

$$\epsilon_i = \frac{\sum_{k=1}^{e} w_k}{\sum_{i=1}^{n} w_i} \qquad (3)$$

ϵ_i is the error rate value, e is the number of incorrectly classified data and w_k is the weights of the error data. The performance value is calculated using the error rate. Calculation of the performance value is shown in equation 4.

$$\propto_i = \frac{1}{2}\ln\left(\frac{1-\epsilon_i}{\epsilon_i}\right) \qquad (4)$$

$\propto_i$ value, shows the performance value of x_i data. Using the performance value, new weight values are calculated. The calculation of the new weight value is shown in equation 5.

$$w_{new_i} = w_i e^{\mp \propto_i} \qquad (5)$$

w_{new_i} is the new calculated weight value of data x_i. After the new weight value found, the weight values should be normalized. The calculation of the new normalized weight values is shown on equation 6.

$$W_{norm_i} = \frac{w_{new_i}}{\sum_{i=1}^{n} w_{new_i}} \qquad (6)$$

W_{norm_i} is the new normalized weight value of data x_i. Classification is carried out using normalized weight values. The classification process is shown in equation 7.

$$C_{x_i} = \arg max_i \left(\sum_{i=1}^{n} \propto_i [y_i]\right) \qquad (7)$$

C_{x_i} shows the class of the x_i.

2.2. Naive Bayes Algorithm

Naive Bayes algorithm was developed by British mathematician Thomas Bayes. This method is based on the data probability (Balaban and Kartal, 2018). Naive Bayes algorithm is one of the statistics based algorithms. In this algorithm, by using previously classified data in the dataset, there is a possibility to find out which of the existing classes belongs to the new data (Silahtaroğlu, 2016).

Naive Bayes algorithm offers a good solution for binary classification problems. It is widely used in computational and real-time operations. But the attribute species should be well understood (Randhawa, 2018). The Naive Bayes algorithm is very efficient when analyzing large datasets. Bayes classifier takes the probability values of the attribute data into account (Wu and Zhu, 2008).

Bayes algorithm is one of the supervised learning algorithms since the classification objectives are known in advance. It is based on the classification effect of Naive Bayes probability factors. Generally, the probability factor in the Bayesian method has a positive effect on datasets (Orhan and Adem, 2012).

First of all the frequencies of class values are found in Bayes algorithm. Equation 8 shows the calculation of class frequency values.

$$p(c_i) = \frac{c_{i\_n}}{n} \quad (8)$$

$p(c_i)$ shows the frequency value of the c_i class, c_{i_n} shows the total number of the c_i class. n, shows the number of x samples in the dataset. After that, the probability of each x_i data is calculated in classes. The probability of belonging x_i to each class shown in equation 9.

$$p(x_i|c_i) = \frac{x_i}{c_{i\_n}} \quad (9)$$

$p(x_i|c_i)$ is the frequency of x_i data in the c_i class. c_{i_n} shows the number of instances of the c_i class. The belonging frequency of x_i to all classes is shown in equation 10.

$$p(x_i) = \sum_{i=1}^{n} p(x_i|c_i) * p(c_i) \quad (10)$$

$p(x_i)$ is the frequency x_i in the dataset. At the last stage, the probability of belonging x_i to classes in the dataset is shown on equation 11.

$$p(c_i|x_i) = \frac{p(x_i|c_i) * p(c_i)}{p(x_i)} \quad (11)$$

$p(c_i|x_i)$, probability of belonging x_i to classes in the dataset. According to obtained result, the x_i data belongs to the class which high probability value.

3. Experimental Results

In the study, firstly, the process of converting the attribute data represented by ICD codes to binary digital type was performed in the COVID-19 Surveillance dataset. AdaBoost and Naive Bayes algorithms need to be implemented using this format. The data converted into binary digital format is shown in the Table 3. The ICD code's disease symptom of each patient is available in the table. A value of "1" indicates that there is disease symptom represented by the ICD code. A value of "0" indicates that there is no disease symptom represented by the ICD code.

Table 3. Binary Format Of Surveillance COVID-19 Data.

Patients	A01	A02	A03	A04	A05	A06	A07
p-1	1	1	1	1	1	0	0
p-2	1	1	0	1	1	0	0
p-3	1	1	1	1	0	1	0
p-4	1	1	0	1	0	1	0
p-5	1	0	0	0	0	0	1
p-6	1	1	1	0	0	0	1
p-7	1	1	0	0	0	0	1
p-8	1	1	1	1	0	0	0
p-9	1	0	0	1	1	0	0
p-10	0	1	0	1	1	0	0
p-11	1	0	0	1	0	1	0
p-12	0	1	0	1	0	1	0
p-13	0	1	0	0	0	0	1
p-14	0	0	0	0	0	0	1

As the first method, the data in the COVID-19 Surveillance dataset was analyzed with the AdaBoost algorithm and PUS, PIM classification forecast (decision) was made for each patient. The classification forecast was made by the AdaBoost algorithm of the patients is shown in Table 4. The classification forecasts made by the AdaBoost algorithm are shown in the AdaBoost (Result) field. Classifications are PUS and PIM. When the patient attributes are analyzed with the AdaBoost algorithm, the probability of belonging to the PIM class is shown in the AdaBoost (PIM) Prediction field. When the patient attributes are analyzed with the AdaBoost algorithm, the probability of belonging to the PUS class is shown in the AdaBoost (PUS) Prediction field. According to the results %92% prediction with accuracy using the AdaBoost algorithm were established on the Surveillance COVID-19 dataset.

Table 4. Prediction And Probability Values Made for Each Patient By AdaBoost Algorithm.

Patients	AdaBoost (Result)	AdaBoost (PIM) Prediction	AdaBoost (PUS) Prediction
p-1	PUS	0.002	0.998
p-2	PUS	0.002	0.998
p-3	PUS	0.002	0.998
p-4	PUS	0.002	0.998
p-5	PIM	0.998	0.002
p-6	PUS	0.002	0.998
p-7	PUS	0.002	0.998
p-8	PUS	0.002	0.998
p-9	PIM	0.998	0.002
p-10	PIM	0.998	0.002
p-11	PIM	0.998	0.002
p-12	PIM	0.998	0.002
p-13	PIM	0.998	0.002

For each patient's class prediction was found probability of the highest 99.8% and the lowest 0.2%. The class whose high probability value is calculated with the AdaBoost algorithm shows the class of the new instance belongs. Figure 2 shows the graph of the classification forecast results in color using AdaBoost. The red column shows the probability of AdaBoost (PIM) and the blue column shows the probability of AdaBoost (PUS). The 5th patient's class was estimated as PIM while he was expected to be PUS. With the AdaBoost algorithm, only one patient's disease diagnosis prediction was made incorrectly.

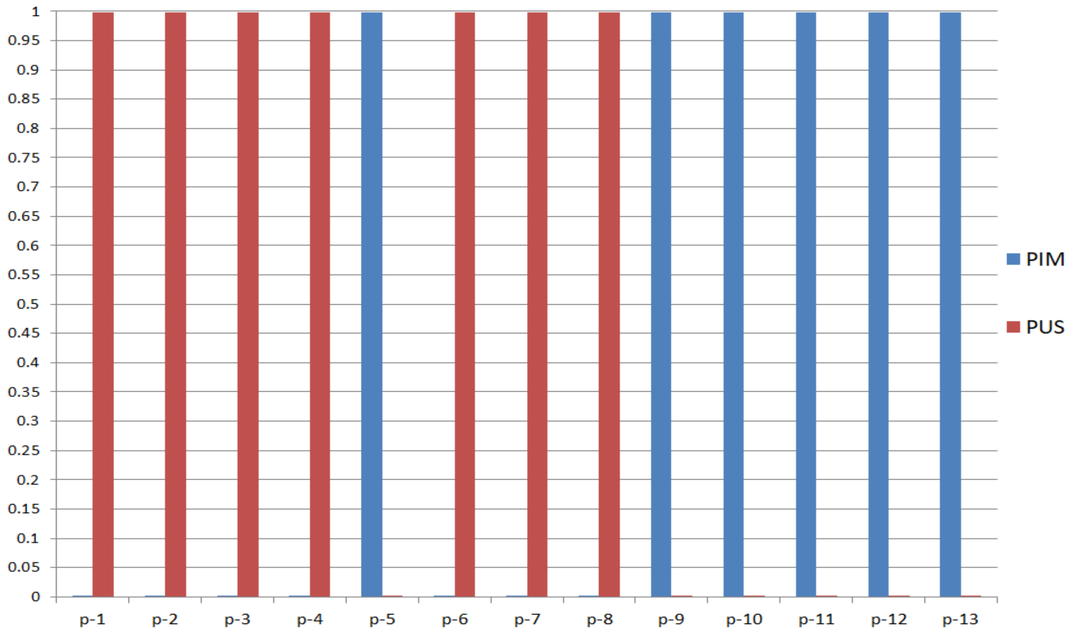

Figure 2. Classification Forecast Performed by AdaBoost Algorithm

As the second method, the data in the COVID-19 Surveillance dataset was analyzed with the Naive Bayes algorithm and PUS, PIM classification forecast (decision) was made for each patient. The classification forecast of the patients was made by Naive Bayes algorithm of is shown in Table 5.

Table 5. Prediction And Probability Values Made for Each Patient By Naive Bayes Algorithm.

Patients	Naive Bayes (Result)	Naive Bayes (PIM) Prediction	Naive Bayes (PUS) Prediction
p-1	PUS	0.232	0.768
p-2	PIM	0.643	0.357
p-3	PUS	0.231	0.769
p-4	PIM	0.643	0.356
p-5	PIM	0.651	0.349
p-6	PUS	0.056	0.944
p-7	PUS	0.262	0.738
p-8	PUS	0.117	0.823
p-9	PUS	0.489	0.511
p-10	PIM	0.829	0.171
p-11	PUS	0.489	0.511
p-12	PIM	0.829	0.171
p-13	PUS	0.305	0.695

The classification forecasts made by the Naive Bayes algorithm are shown in the Naive Bayes (Result) field. Classifications are PUS and PIM When the patient attributes are analyzed with the Naive Bayes algorithm, the probability of belonging to the PIM class is shown in the Naive Bayes (PIM) Prediction field. When the patient attributes are analyzed with the Naive Bayes algorithm, the probability of belonging to the PUS class is shown in the Naive Bayes (PUS) Prediction field.

According to the results obtained, the prediction was made with 54% accuracy with the Naive Bayes algorithm on the Surveillance COVID-19 dataset. Incorrect classification forecast was made for one patient in the PUS class and one in the PIM class. The classification of 2th, 4th, 5th, 9th 11th and 13th patient was realized incorrect prediction. The 2th, 4th and 5th patient's class was estimated as PIM while expected to be PUS. The 9th, 11th and 13th patient's class was estimated as PUS while he was expected to be PIM.

Figure 3 shows the graph of the classification forecast results in color using Naive Bayes. The red column shows the probability of Naive Bayes (PIM) and the blue column shows the probability of Naive Bayes (PUS).

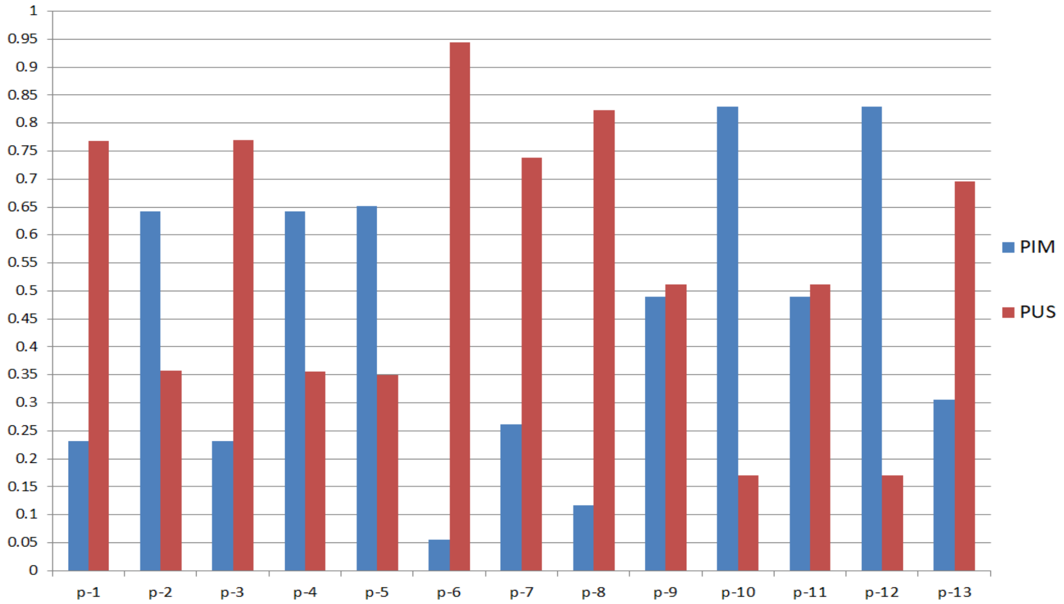

Figure 3. Classification Forecast Performed by Naive Bayes Algorithm

4. Result and Discussion

In this study, classification forecasts of the patients in the COVID-19 Surveillance dataset were made using AdaBoost and Naive Bayes machine learning algorithms on the basis of data mining. In the dataset, fourteen patients had symptoms specified by seven ICD codes. Using these signs, patients can be classified as PUS, PIM, and PWS. With machine learning methods, predictions can be made on new data by gaining experience from the data in the dataset. There are large amounts of digital data in many areas around the world and the numbers continue to increase. Machine learning is carried out with and without supervision. The target points in supervised learning are predetermined, but the target points in unsupervised learning are not pre-determined. AdaBoost and Naive Bayes machine learning algorithms are supervised learning methods. COVID-19 has spread all over the world and millions of people have been infected and died due to this pandemic disease. Detection of this disease is carried out with PCR tests and obtaining information about the test results may take a few days in some cases. Unfortunately, PCR test results of the disease caused by COVID-19 mutation viruses may take longer. In the study, classification forecasts were made separately with AdaBoost and Naive Bayes algorithms using the unclassified data of the patients in the dataset. As a result of the study, the prediction was made with 92% accuracy with the AdaBoost algorithm and with 54% accuracy with the Naive Bayes algorithm. According to the results, AdaBoost machine learning algorithm is more successful than Naive Bayes machine learning algorithm in the COVID-19 Surveillance dataset. This study proved that using machine learning methods in the diagnosis of COVID-19 can be useful and fast.

Conflict of Interest

No conflict of interest was declared by the author.

References

Balaban, M. E., Kartal, E. (2018). Veri madenciliği ve makine öğrenmesi temel algoritmaları ve R Dili ile Uygulamaları, İkinci Baskı. İstanbul, Türkiye, Çağlayan Kitap & Yayıncılık & Eğitim.

Chen, Y., Chang, Y., Kan, Y., Chen, R. S., Wu, S. F. (2018). Using Data Mining Technique to Improve Billing System Performance in Semiconductor Industry, 2018 International Conference on Information and Computer Technologies (ICICT), DeKalb, IL, 23-25 March, 2018.

Çelik, A. (2020). Using Apriori Data Mining Method in COVID-19 Diagnosis. Journal of Engineering Technology and Applied Sciences, 5(3), 121-131.

Dua, D., Graf, F. C. (2019). UCI Machine Learning Repository, University of California, School of Information and Computer Science, Irvine, USA.

Freund, Y, Schapire, R. E. (1999). A Decision-Theoretic Generalization of On-line Learning and An Application to Boosting, Proc. Eur. Conf. Comput. Learn. Theory, 119-139.

Kumari, S., Singh, M. (2019). Big Data Analytics for Healthcare Industry: Impact, Applications, and Tools, in Big Data Mining and Analytics, 2(1), 48-57.

Olgun, M., Özdemir, G. (2013). İstatiksel Özellik Temelli Bayes Sınıflandırıcı Kullanarak Kontrol Grafiklerinde Örüntü Tanıma, Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 27(2), 303-311.

Orhan, U., Adem, K. (2012). The Effects of Probability Factors in Naive Bayes Method, 2012 Elektrik-Elektronik ve Bilgisayar Mühendisliği Sempozyumu, Bursa, Turkey, 29 Nov-01 Dec. 2012.

Randhawa, K., Loo, C. K., Seera, M., Lim, C. P., Nandi, A. K. (2018). Credit Card Fraud Detection Using AdaBoost and Majority Voting, IEEE Access, 6,14277-14284.

Silahtaroğlu, G. (2016). Veri madenciliği (Kavram ve algoritmaları), Üçüncü Baskı, İstanbul, Türkiye, Papatya Yayıncılık Eğitim.

Sun, Y., Kamel, M. S., Wang, Y. (2006). Boosting for Learning Multiple Classes with Imbalanced Class Distribution, Sixth International Conference on Data Mining (ICDM'06), Hong Kong, China, 18-22 Dec 2006.

Wang, F., Li, F., He, F., Wang, R., Yu, W., Nie, F. (2019). Feature Learning View Point of Adaboost and a New Algorithm, IEEE Access, 7, 149890-149899.

Wiguna, W., Riana, D. (2020). Diagnosis of Coronavirus Disease 2019 (Covid-19) Surveillance Using C4.5 Algorithm, Journal PILAR Nusa Mandiri, 16, 71-80.

World Health Organization (WHO). (2005). International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10 manual), Centers for Disease Control and Prevention, 1: 1-1268.

Wu, X., Zhu, X. (2008). Mining With Noise Knowledge: Error-Aware Data Mining, IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 38(4), 917-932.

Xiao, F., Wang, Y., He, L., Wang, H., Li, W., Liu, Z. (2019). Motion Estimation From Surface Electromyogram Using Adaboost Regression and Average Feature Values, IEEE Access, 7,13121-13134.

Xu, H., Yuan, H. (2020). An SVM-Based AdaBoost Cascade Classifier for Sonar Image, IEEE Access, 8: 115857-115864.

Yılmaz, İ., Öztürk, D. (2019). Mekânsal Planlama ve Karar Problemlerinde CBS ve Bayes Ağlarının Entegrasyonu: Orman Yangın Tehlikesi Örneği, TMMOB 6. Coğrafi Bilgi Sistemleri Kongresi, Ankara, Türkiye, 23-25 Ekim 2019