

Predicting the Ending Time of COVID-19 Pandemic in Turkey

Ekin Kara, Önder Ertan*, Furkan Yalçinkaya and Meltem Eryilmaz

Atilim University, Ankara, Turkey

*Corresponding Author: Önder Ertan, Atilim University, Ankara, Turkey.

Received: September 01, 2022

Published: December 07, 2022

© All rights are reserved by Önder Ertan, *et al.*

Abstract

Coronavirus pandemic has been going on since late 2019, millions of people died worldwide, vaccination has recently started in many countries and new strategies are sought by countries since they are still struggling to defeat the virus. So, this research is made to predict the possible ending time of the coronavirus pandemic in Turkey using data mining and statistical studies. Data mining is a computer science study that processes large amounts of data and produces new useful information. It is especially used to support decision making in companies today. So, this project could support the decision making of authorities in developing an effective strategy against the on-going pandemic. During the research we have practiced on Turkey's coronavirus and vaccination data between 13 January 2021 and 28 May 2021. We used Rapidminer and the Random Forest method for data mining. After all the simulations we have applied and observed during our research, it was clearly seen that vaccination parameters were decreasing the new cases. Also, the stringency index did not affect the new cases. As a conclusion of our research and observations, we think that the government should vaccinate as many people as it can in order to relax restrictions for the last time.

Keywords: Coronavirus; COVID-19; Turkey

Introduction

New coronavirus disease 2019 (COVID-19) is a type of virus that can be transmitted by respiratory tract and contact. Firstly, COVID-19 appeared in a group of patients who developed respiratory tract symptoms such as fever, coughing, difficulty in breathing in Wuhan, China in late December 2019. After studies that were carried on this group of patients, the new virus, COVID-19, was identified [4,21].

As it is known, the world has been struggling with the COVID-19 virus for a year. Currently, Turkey has been trying to carefully go through the normalization process. Last time the government went for the partial normalization in late May 2020 [1], unfortunately because of a variety of disregarded reasons, the country went back to where it first started. Number of cases, the positivity rate and deaths soared up. After weeks of healthcare

providers' and authorities' battling with the virus and a-17-days-long full lockdown [7], the country managed to maintain control of the pandemic. At the moment, according to the officially given information, cases and deaths are declining. For the present the country is normalizing again after readjusting the pandemic values, the future of Turkey's fight against coronavirus is not certainly known. Medical experts are afraid of living through the previous vicious cycle again.

We think that due to the lack of processed and analyzed data about the virus, today we are struggling to develop a full scale and efficient strategy against it. Therefore, we have decided to do this research on COVID-19 and vaccination data using data mining and statistical studies. At this point giving a brief information about data mining would be great for the readers. Data mining is the process of sorting, analyzing and generating useful information from large

data sets which are retrieved from databases or companies. Data mining has been increasingly used since the 20th century, due to it being productive and helpful in various types of industrial sectors.

Thus, as is seen, the goal of this research is to make a prediction for the possible ending time of the pandemic especially based on the number of cases and vaccination progress in Turkey.

In the following parts of the paper we are going to explain how we performed our research step by step.

Material and Methodology

Material

As the first step of our study we had to collect data to work on. Daily data have been officially announced by the Ministry of Health in Turkey. But in order to get a more detailed and large amount of data we needed datasets. So we preferred to retrieve data from a repository that belonged to “Our World in Data” (OWID) which is an online publication website that focuses on problems and crises of the world by storing, analyzing and interpreting data. OWID stored its data on GitHub [3]. The GitHub repository includes the COVID-19 data in a spreadsheet which is continuously updated.

Stringency index was calculated and added to the excel file by OWID however it is important to mention that the calculation is made by using values that are known as indicators and related to government pandemic policies [13].

Methodology

In this part of the article you will be given information about the stages we went through.

Conceptual modelling

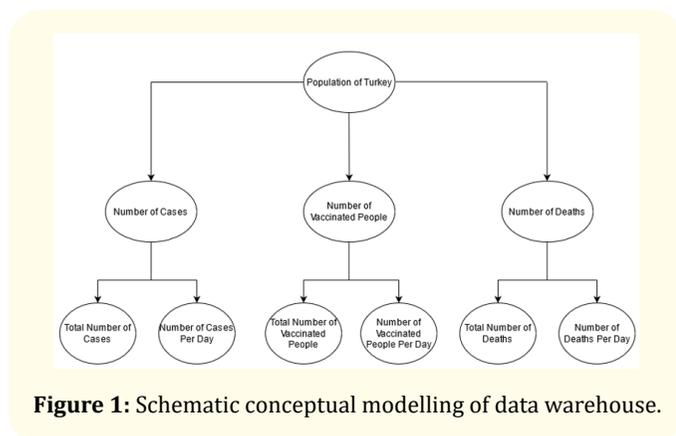


Figure 1: Schematic conceptual modelling of data warehouse.

The process of developing a data model for data that will be stored in a database is known as data modelling. On the other hand conceptual modelling defines what the system includes. The goal of conceptual modelling is to organize, scope and define business concepts and rules.

Here in figure 1, we designed the schematic conceptual modelling of the data warehouse by dividing the population of Turkey, which can be seen as a universal set, into three different branches that are “Number of Cases”, “Number of Vaccinated People”, “Number of Deaths”. Each branch also includes two sub branches of “Total Number” and “Number ... Per Day”. The model covers concepts of our project and focuses on representing data to be seen as “real world” data.

Preprocessing

Data preprocessing is one of the crucial stages of data mining. In preprocessing, data are literally filtered by cleaning the dataset from nonessential and noisy data. The OWID excel file included other countries’ COVID-19 information, since we did not need them. Since our study focuses only on Turkey and the course of the pandemic as part of our research, we applied the “Extract, Transform, Load” (ETL) [16] process by cleaning unnecessary cells and removing all the other countries and unneeded parts in the dataset. Also in order to make sure the data was correct we have compared the OWID dataset with official publications of the Turkish Ministry of Health, as a result of which we have verified the correctness [17].

Hence our final version of excel file had the following attributes left: date, total_cases, new_cases, total_deaths, new_deaths, reproduction_rate, new_tests, total_tests, positive_rate, tests_per_case, total_vaccinations, people_vaccinated, people_fully_vaccinated, new_vaccinations, stringency_index.

Thus in preprocessing we used the following operations to simplify our dataset:

- **Data Cleaning:** In our data set, we cleaned the attributes and data that we did not need as well as duplicate data such as constant number of population. On the other hand, there was not any incomplete, noisy, inconsistent or intentional data.

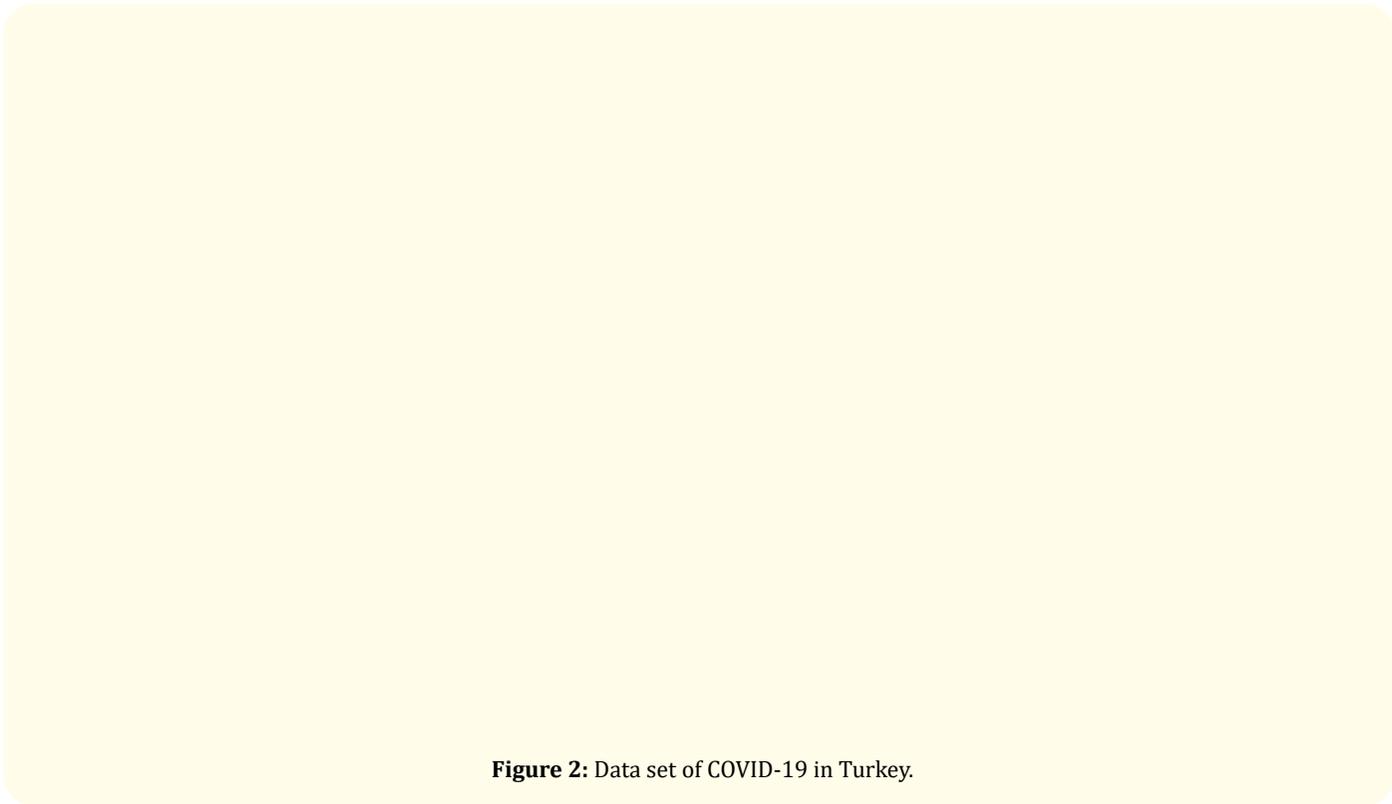


Figure 2: Data set of COVID-19 in Turkey.

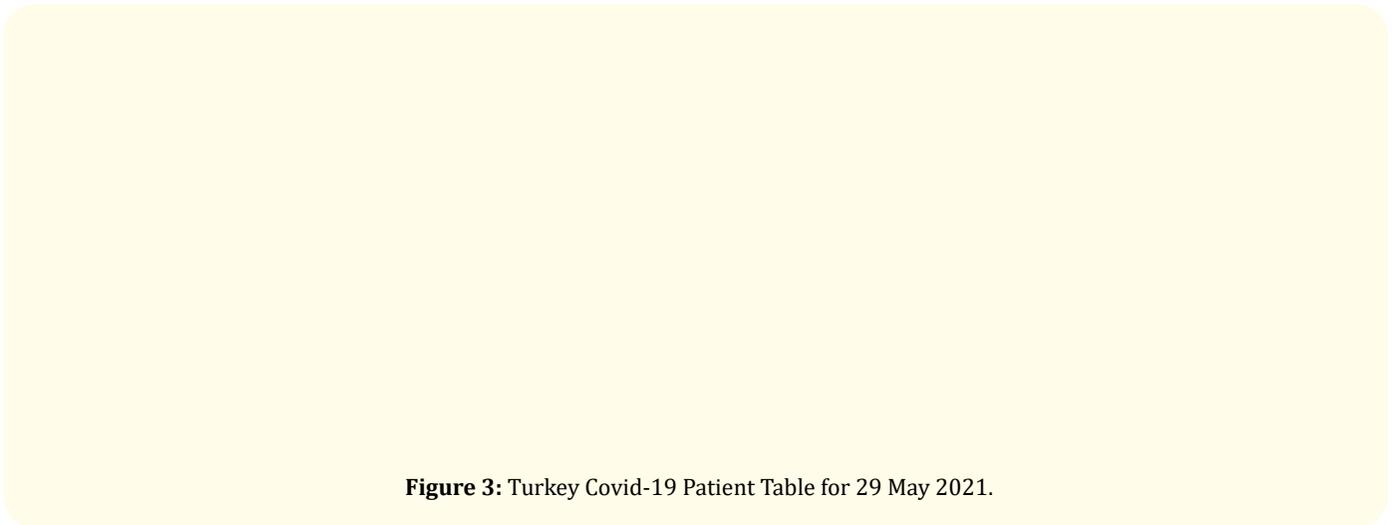


Figure 3: Turkey Covid-19 Patient Table for 29 May 2021.

- **Data Reduction:** In this technique we got rid of other countries and their data since we focus only on Turkey and we limited our data with the data as of the first time vaccination started in Turkey which is 13 January 2021. Then we focused on the total number of the attributes.
- **Data Transformation and Data Discretization:** Since there was not any incomplete, noisy, inconsistent or intentional data as we mentioned before, we have not transformed our data. There was no need for it. But there were missing data in early vaccination.

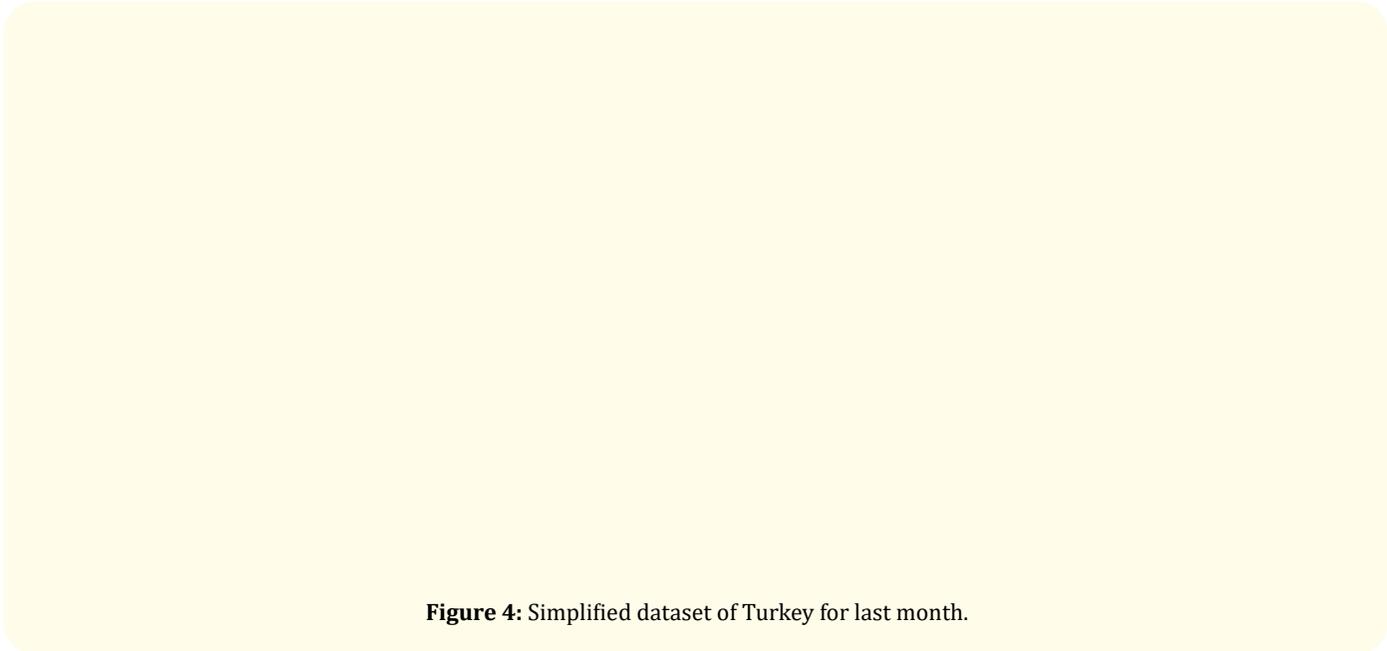


Figure 4: Simplified dataset of Turkey for last month.

Choosing data mining method

After processing and preparing our data in a spreadsheet file, we used the Random Forest method to model and analyze our data in Rapidminer which is an advanced data mining tool that comes with many widgets and tools to make a prediction [20]. Random Forest is a popular machine learning model because it produces

decent results without the need for hyperparameter estimation and can be used to solve both regression and classification issues [15]. We tried to use the traditional Decision Tree method too however, since one of the biggest problems of this method is excessive learning-data overfitting [18]. Also after using Random Forest we realised that our results were more realistic and logical, in addition this method performed the best among the others.

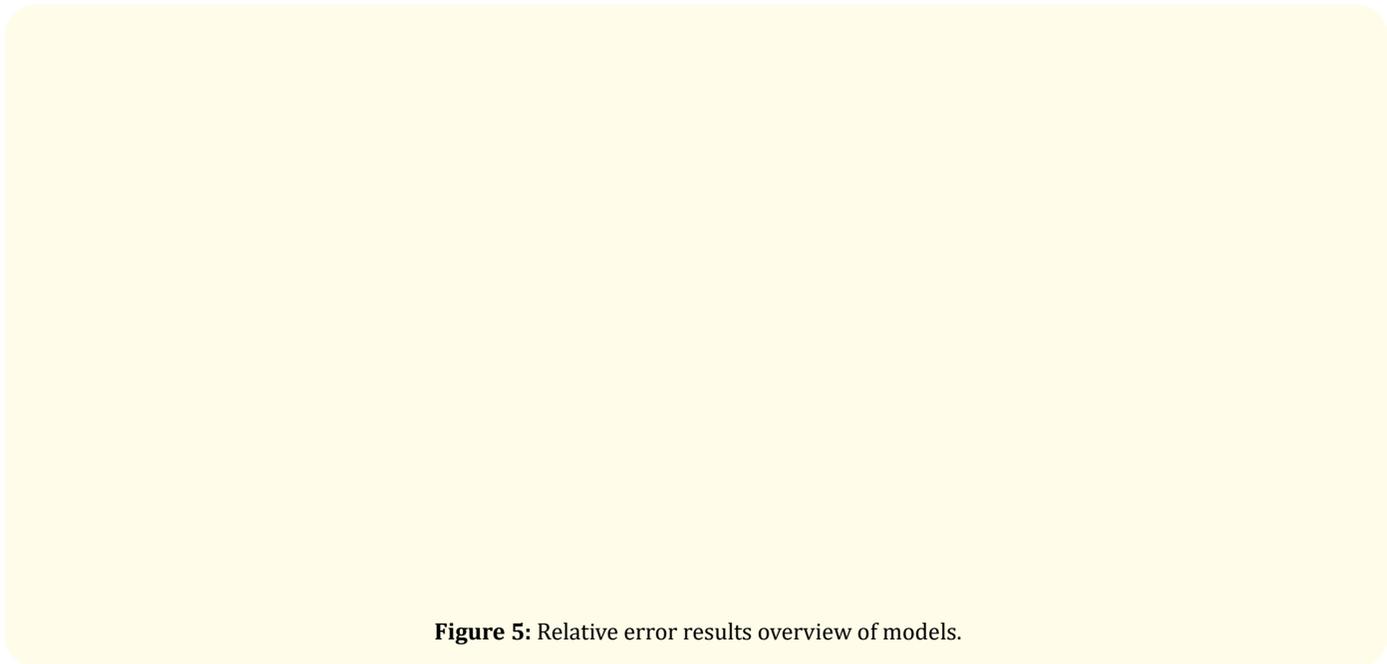


Figure 5: Relative error results overview of models.

Results

In this part of the research we will be talking about the prediction stage that we have done using Rapidminer. We have encountered 8.4% of relative error and 2.1% of standard deviation in Random Forest results. Rapidminer took more time for analyzing the data in the Random Forest method compared to other methods in

modelling. Green and tones of it support the prediction, while red and tones of it contradict the prediction of Rapidminer (as it can be seen in Figure 8). The data mining tool completed empty spaces in the input data table (such as data related to vaccination which was partially missing until 11 February) using its deep learning and prediction methods, but these assumptions of the program did not affect the results.

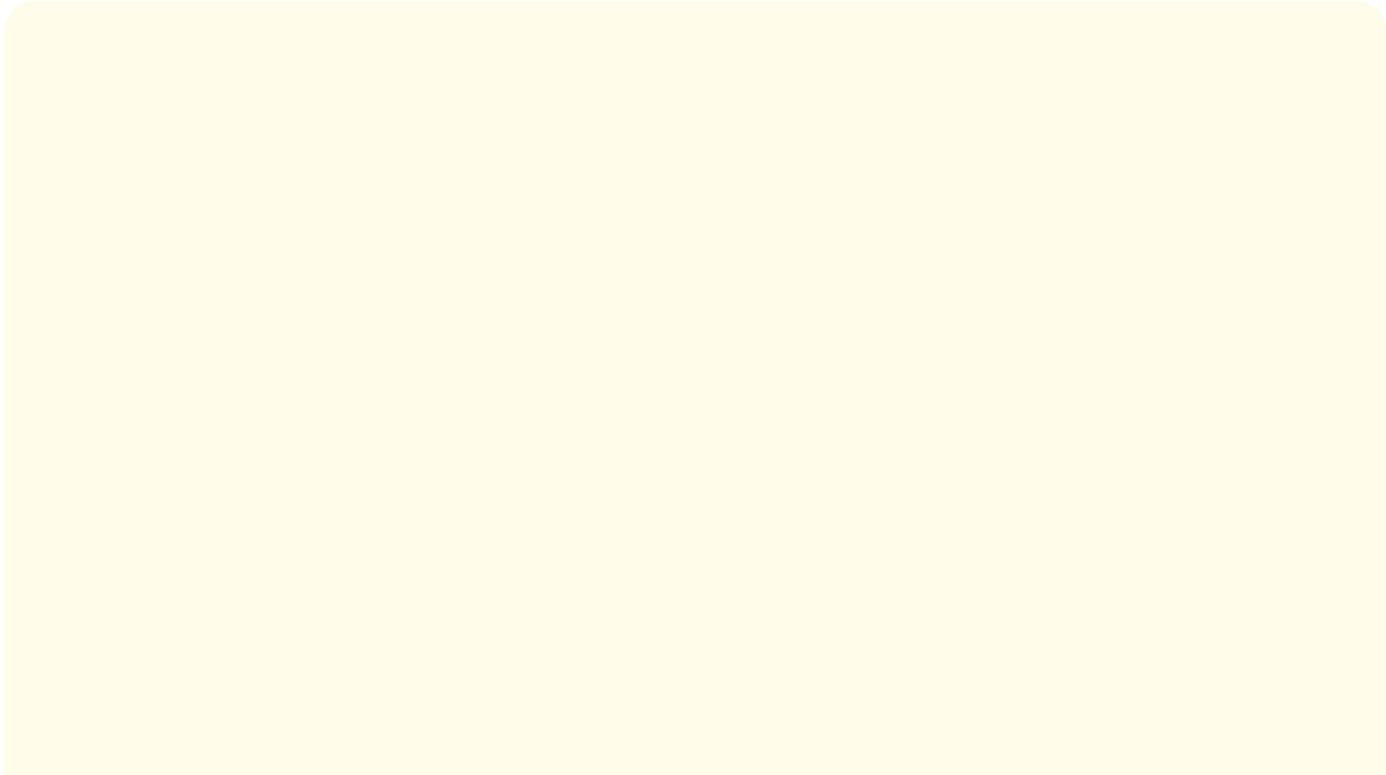


Figure 6: Random Forest Prediction.

Rapidminer used `new_cases`, `total_cases`, `total_deaths`, `reproduction_rate`, `new_tests`, `total_tests`, `positive_rate`, `total_vaccinations`, `people_vaccinated`, `people_fully_vaccinated`, `new_vaccinated`, `stringency_index` and made predictions for `new_cases`.

`people_vaccinated` represents the people who are vaccinated with the first dose, `people_fully_vaccinated` represents people who are vaccinated with the second dose as well. As it can be seen in Figure 6 new cases are predicted to be growing.

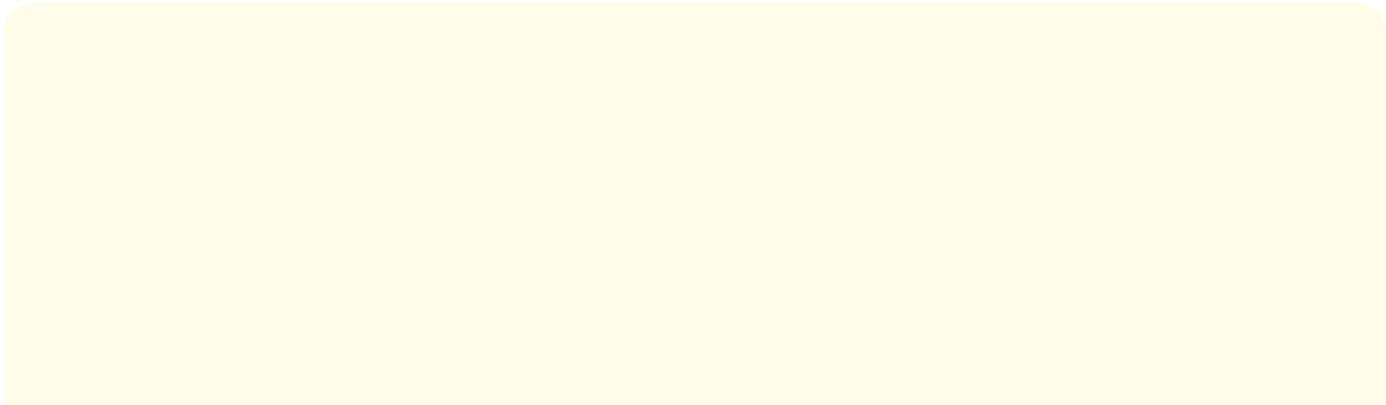


Figure 7: Random Forest Tree.

Optimal Parameters for Random Forest modelling was also calculated by Rapidminer. As maximal depth decreases and number of trees increase, error rate increases too. As it can be seen in figure

8 the tool chose the parameters with least error rate which had 60 trees and maximal depth of 7.

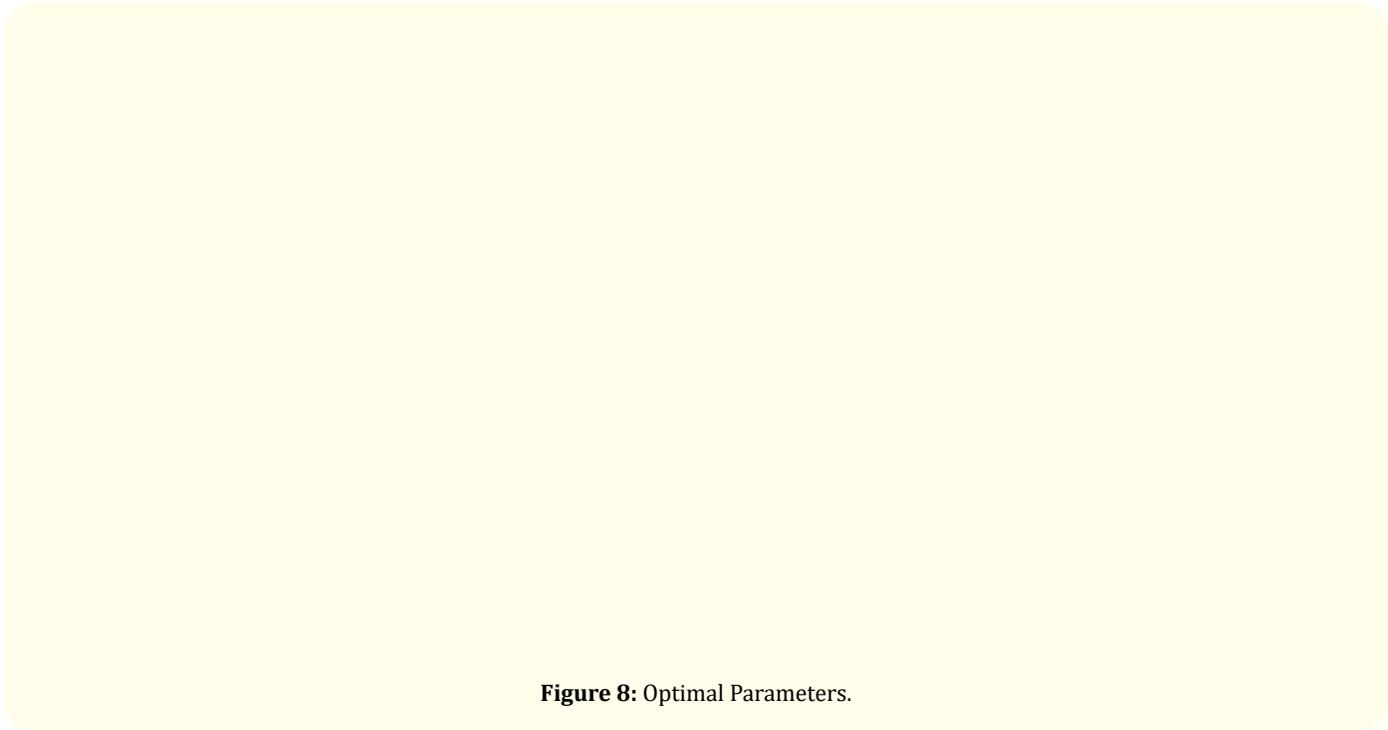


Figure 8: Optimal Parameters.

According to our experiments in the simulator; increasing the stringency index when the number of cases is high does not help the course of Covid-19 pandemic. As we increased the positive_rate in the simulation, the new cases proportionally increased as well. Also if new_tests and total_tests are increased new_cases

are increased in proportion to these. On the other hand when we simulated vaccination data, this decreased the new cases a little bit even though this has a big importance in our observation. But with the combination of vaccination data and some other related data, new cases decreased proportionally.

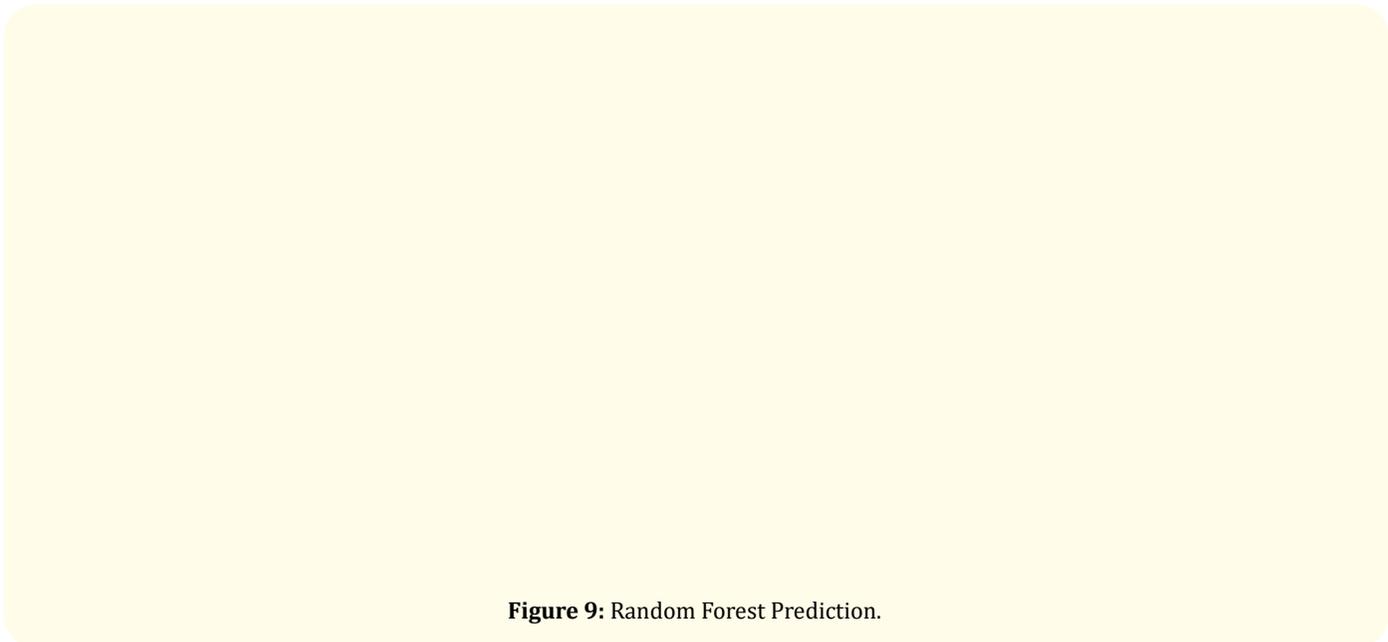


Figure 9: Random Forest Prediction.

Conclusion

To conclude our research we can say that related to the results and state of worldwide pandemic, parallel to the speed of the vaccination process the pandemic can be taken under control. Even if the people get infected when they are vaccinated they can survive it. The vaccination speed should be increased in Turkey in order to take the pandemic under control and slow down the rise of new cases and deaths in the short term. As we have experienced in the simulations, the stringency index did not affect the rate of cases or the numbers vitally. So the stringency index which represents the physical measures, especially lockdowns, is not the strategy we should be applying when coping with the pandemic. Also, while lockdowns are pointless, they affect the economy and social life negatively [11]. In fact, although measures are taken most of the time, the ability of flu-like viruses to spread rapidly, the insufficient amount of vaccines and, as the most important reason, the weakness of the health systems in countries where the disease continues, cause these pandemics to carry on. However, if we take cognizance of countries who vaccinated most of their population such as Israel and USA, the restrictions are reduced and people are returning to their old normal lives [10,19]. Therefore, as mentioned the only weapon we have against the virus is vaccines right now and if we manage to vaccinate most of Turkey, we may have a one way ticket to normalization. If Turkey succeeds in vaccinating most of its population, considering the possibility of vaccinated people being less mortal against variants of the virus, as the number of cases and the existence of infection decrease, we expect that the pandemic will end completely by the end of 2022 in Turkey.

Bibliography

1. "Prof. Serap Şimşek Yavuz: Salgın yönetimi Bilim Kurulu'nda olsaydı bunları yaşamazdık". *Gazete Duvar*.
2. F Budak and Ş Korkmaz. "COVID-19 pandemi sürecine yönelik genel bir değerlendirme: türkiye örneği". *Sosyal Araştırmalar ve Yönetim Dergisi* 1 (2020): 62-79.
3. Owid. "owid/covid-19-data". GitHub.
4. N Rezaei. *Coronavirus Disease - COVID-19*. Cham, Switzerland: Springer International Publishing, (2021).
5. PG Auwaerter. "Coronavirus COVID-19 (SARS-CoV-2): Johns Hopkins ABX Guide". *Coronavirus COVID-19 (SARS-CoV-2) | Johns Hopkins ABX Guide*.
6. H Ritchie., *et al.* "Coronavirus Pandemic (COVID-19) - Statistics and Research". *Our World in Data*, 05-Mar- (2020).
7. TC İçişleri Bakanlığı Bilgi İşlem Dairesi Başkanlığı. "81 İl Valiliğine Tam Kapanma Tedbirleri Genelgesi Gönderildi". T.C. İçişleri Bakanlığı, 26-Apr (2021).
8. Mathieu E., *et al.* "A global database of COVID-19 vaccinations". *Nature Human Behaviour* (2021).
9. Hasell J., *et al.* "A cross-country database of COVID-19 testing". *Scientific Data* 7 (2020): 345.
10. E Leshem and A Wilder-Smith. "COVID-19 vaccine impact in Israel and a way out of the pandemic". *The Lancet* 397.10287 (2021): 1783-1785.
11. Kochhar A S., *et al.* "Lockdown of 1.3 billion people in India during Covid-19 pandemic: A survey of its impact on mental health". *Asian Journal of Psychiatry* 54 (2020): 102213.
12. D Witteveen and E Velthorst. "Economic hardship and mental health complaints during COVID-19". *Proceedings of the National Academy of Sciences* 117.44 (2020): 27277-27284.
13. Oxford University, Calculation and presentation of the Stringency Index 4.0 (2020): 1-3.
14. H N Nawazish and V K Shukla. "Integrating "Random Forest" with Indexing and Query Processing for Personalized Search". 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), (2020): 280-283.
15. Y M C Zhang. "Ensemble Machine Learning". Springer, New York, NY, USA, (2012).
16. P Vassiliadis and A Simitsis. "Extraction, Transformation, and Loading". *Encyclopedia of Database Systems* (2009): 1095-1101.
17. TC Sağlık Bakanlığı.
18. Song YY and Lu Y. "Decision tree methods: applications for classification and prediction". *Shanghai Archives of Psychiatry* 27.2 (2015): 130-135.
19. Bagcchi S. "The world's largest COVID-19 vaccination campaign". *The Lancet Infectious diseases* 21.3 (2021): 323.
20. M Karimi. "Data Analysis using RapidMiner" (2017).
21. TC Sağlık Bakanlığı. "COVID-19 Nedir?". COVID-19 Nedir?.