COVID-19 Next Day Trend Forecast

Marcelo Costa[1], Margarida Rodrigues[1], Pedro Baptista[1], João Henriques[1,3,4](✉), Ivan Miguel Pires[1,2,4], Cristina Wanzeller[1,4], and Filipe Caldeira[1,4]

[1] Department of Computer Science, Polytechnic of Viseu, 3504-510 Viseu, Portugal
{estgv16800,estgv16061,estgv17058}@alunos.estgv.ipv.pt,
{joaohenriques,impires,cwanzeller,caldeira}@estgv.ipv.pt

[2] Instituto de Telecomunicações, Universidade da Beira Interior, 6200-001 Covilhã, Portugal

[3] Department of Informatics Engineering, University of Coimbra, 3030-290 Coimbra, Portugal

[4] CISeD—Research Centre in Digital Services, Polytechnic of Viseu, 3504-510 Viseu, Portugal

Abstract. Historically, weather conditions are depicted as an essential factor to be considered in predicting variation infections due to respiratory diseases, including influenza and Severe Acute Respiratory Syndrome SARS-CoV-2, best known as COVID-19. Predicting the number of cases will contribute to plan human and non-human resources in hospital facilities, including beds, ventilators, and support policy decisions on sanitary population warnings, and help to provision the demand for COVID-19 tests. In this work, an integrated framework predicts the number of cases for the upcoming days by considering the COVID-19 cases and temperature records supported by a kNN algorithm.

Keywords: KNN · COVID-19 cases · Temperature

1 Introduction

In 2003, in Guangdong, the outbreak of Severe Acute Respiratory Syndrome (SARS) gradually disappeared with the arrival of warm weather. Temperature and its variations may have affected the SARS outbreak [14]. Few studies report that COVID-19 was related to meteorological factors, which decreased with increasing temperature. On 28 November, Portugal recorded more than 280.000 COVID-19 cases [4]. After the peak in the first wave on 17 April, the curve began to decrease. According to experts, one of the main factors was lockdown [3,5]. By this time of the year, the season in Portugal is Spring, which means the average temperature is around 11 °C and 20 °C. An analysis of the number of COVID-19 cases and the places manifested more intensively was carried out. The conclusion was that a very close relationship between the new cases and the climate could be established. The data analysis revealed that SARS-Cov2

I. M. Pires et al. (Eds.): GOODTECHS 2021, LNICST 401, pp. 44–50, 2021.
https://doi.org/10.1007/978-3-030-91421-9_4

increases virulence with temperatures between 6 °C and 11 °C and the presence of lower humidity levels in the air [2]. According to DGS, Portugal is in the second wave of COVID-19. Portugal is also facing the flu season, with the aggravating factor that the amount of flu vaccines available was manifestly insufficient [8,12]. National Health System includes 21,000 beds as the total capacity, while 17,700 are allocated to COVID-19 assistance [11]. The number of COVID-19 cases growth put pressure on hospitals. On 4 November, Portugal realized 444 hospitalized patients by COVID-19 in the Intensive Care Unit (ICU), which is 90% of the capacity of the ICU [6]. In 2016 the ICU occupation average was above 75% [10], considering this behavior is stable across the years. Another question raises how prevent ICU rupture if these numbers get back to previous years standard value. Given the lack of evidence, it is essential to understand the weather conditions impact on COVID-19 transmission. In this case, the relationship between weather conditions and several COVID-19 cases exists. It will be possible help to manage human resources and keep up with the demand for beds material resources (beds, ventilators, etc.) in hospital facilities and help sanitary populations' warnings decisions.

The remainder of this work is organized as follows. Section 2 presents the related work. Section 3 presents the architecture. Section 4 supporting our approach to assess data quality. In Sect. 5 we present the final product preview. Finally, in Sect. 6 conclusions are drawn, followed by future work guidelines.

2 Related Work

As in a common sense, weather conditions are an essential factor that impacts the number of infections due to respiratory diseases, such as Severe Acute Respiratory Syndrome (SARS) and influenza, also known as "the flu". Studies presented after the outbreak of SARS concluded the disease gradually disappeared with the arrival of warm weather. Temperature and its variations may have affected the SARS outbreak [14]. An exciting study relates the number of COVID-19 deaths in Wuhan with the temperature and humidity. The conclusions reached by this study were that the daily mortality of COVID-19 is positively correlated with DTR (diurnal temperature range) but negatively with absolute humidity. This study suggests the temperature variation and humidity may also be important factors affecting the COVID-19 mortality [9]. Another study pursues the relationship between ambient temperature and daily COVID-19 confirmed cases in 122 Chinese cities. The results indicate that mean temperature linearly relates with the number of COVID-19 cases when the temperature is below 3 °C. Notwithstanding, no evidence supporting such and relationship, especially a decreasing number of COVID-19 cases when the weather gets warmer [15].

3 Architecture

The following elements were identified to obtain prediction results: the sensors (DH11) and the board (Raspberry Pi 3), the API, the database, and the trained prediction model. Each of these is represented on Image 1.

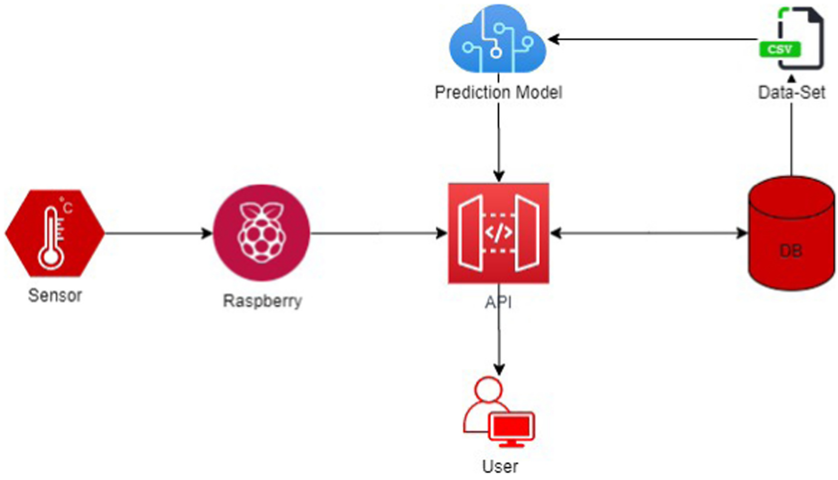

Fig. 1. Solution architecture

Each raspberry pi and the respective DH11 sensor are installed in the area where the user wants to obtain the prediction. It's a requisite to ensure the COVID-19 number of new cases are available for the selected geographic area. Temperature data is stored through API calls on the database server. This persistent data is used for temperature mean calculation and provided when needed to the predictive model. The following variables were selected to train the model:

- Health sub-region;
- The municipalities average maximum temperature in the health sub-region for the respective date;
- The municipalities average maximum temperature in the health sub-region between day n and $n - 7$ days;
- Comparing the average of the number of new cases of COVID-19 in the health sub-region counties between days n and $n - 7$ and the number of new cases of n.

The target consists of predicting whether the number of new cases of $n + 1$ is higher or lower than the number of new cases of n ($n = \text{current day}$).

4 Experiment

The experiment involved a Raspberry PI (version 3) with a DH11 attached. The central server, deployed as a Virtual Machine, consists of 2 cores operating at 3 GHz with 4 GB of RAM, running the API and the predictive Model. The MongoDB Atlas service provided the database support. APIs providing the municipality temperature data [7] and COVID-19 cases [13] were collected with a Python application. This application aggregates municipalities by health sub-region and averages their maximum temperature. After collecting temperature data, we accessed the COVID-19 API and extracted the number of cases corresponding to each health sub-region. The application produces a CSV file and includes the following features: date, health sub-region of the North, number of cases in the North, health sub-region of the "Centro", number of cases in the "Centro", health sub-region of "Lisboa and Vale do Tejo", number of cases in "Lisboa and Vale do Tejo", health sub-region of "Alentejo", number of cases in "Alentejo", health sub-region of "Algarve", number of cases in the "Algarve". Subsequently, another Python program was developed that creates our dataset CSV file including the following features:

- Health sub-region;
- Daily municipalities average maximum temperature for health sub-region;
- Municipalities average maximum temperature for sub-region between n and n − 7 (days);
- Average number of new cases of COVID-19 in the municipalities of the health sub-region between n and n − 7 (days);
- Number of new cases from n.

Our approach is to predict whether the tomorrow's number of new cases (day n + 1) by considering the number of new cases occurred today (day n). A training dataset will be considered as input to the predictive model. With all mounting and setup done, we collected the previous temperature and COVID-19 data to train our model. It's possible to keep "feeding" our dataset through this API and test different predictive models' inputs. A visual explanation is available on Image 2.

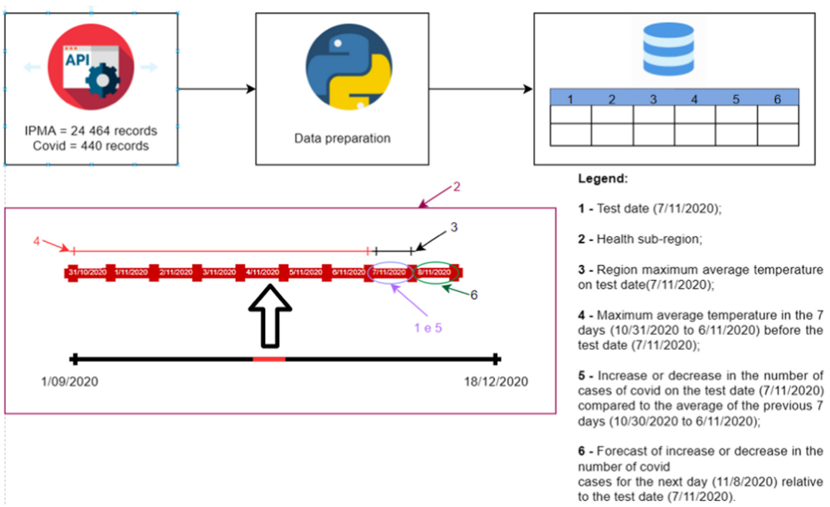

Fig. 2. Dataset diagram

5 Results

Our previous analysis supported selecting the kNN as the first algorithm for our predictive model, which makes predictions using the training dataset directly. Predictions were made for a new instance (x) by searching through the entire training set for the K most similar instances (the neighbors) and summarizing the output variable for those K instances. For regression, this might be the mean output variable. In classification, this might be the mode (or most common) class value [1]. The distance was used to determine the most similar K instances in the training dataset to a new input. For real-valued input variables, the most popular distance measure is Euclidean distance. After tuning the K value, we conclude that 3 corresponds to the value providing the higher accuracy while the chosen metric was Minkowski. The Table 1 shows the accuracy values obtained with kNN for different day intervals and 15% and 30% test size.

Table 1. kNN accuracy scores (k = 3)

Days	15%	30%
6	0.49	0.55
7	0.66	0.58
8	0.54	0.44
9	0.45	0.47
10	0.59	0.50

Further research and testing allowed us to explore the Random Forest algorithm and achieve 72.4% accuracy with 41 n_estimators. This result will define the response of our predictive model setup as presented in Table 2.

Table 2. Random Forest accuracy scores (n_estimators = 41)

Days	15%	30%
6	0.45	0.54
7	0.72	0.61
8	0.56	0.52
9	0.45	0.52
10	0.53	0.47

6 Conclusion

Predictive analytic tools give users deep, real-time insights into an almost endless array of business activities. These tools can be used to predict different behaviors and patterns, including the demand for COVID-19 tests, basing predictions on an analysis of data collected over some time. According to the achieved result, the presented model is achieved 72% accuracy in predicting the number of COVID-19 cases for the next day. For this result, we considered as main variables the temperature and evolution of the last 7 days of the number of COVID-19 cases. This model may be implemented under any circumstances considering temperature collection is ensured, and the number of new cases of COVID-19 in the area concerned is known. Following the implementation of the proposed architecture, it is possible to set up a complete system without substantial investment. It was also essential to state the importance of "days interval" in the model. As the first symptoms appear after the 4 days, knowing that, between the COVID-19 test schedule and the result's registration. Next, the patients wait 2 or 3 days. The results obtained are consistent with the sources consulted. On the seventh day, we established the relationship capable of generating the most satisfactory results regarding the prediction.

It will become essential to consider a more extended period of days as the forecasting weather conditions are pretty precise. Moreover, given the low achieved model accuracy suggests exploring new features, such as the humidity, day of the week, or wind speed.

Acknowledgements. This work is funded by FCT/MEC through national funds and, when applicable, co-funded by the FEDER-PT2020 partnership agreement under the project **UIDB/50008/2020**. (*Este trabalho é financiado pela FCT/MEC através de fundos nacionais e cofinanciado pelo FEDER, no âmbito do Acordo de Parceria PT2020 no âmbito do projeto UIDB/50008/2020*). This work is also funded by National Funds through the FCT - Foundation for Science and Technology, I.P., within the scope of the projects **UIDB/00742/2020** and **UIDB/05583/2020**. This article is based upon work from COST Action IC1303-AAPELE—Architectures, Algorithms, and Protocols for Enhanced Living Environments and COST Action CA16226-SHELD-ON-Indoor living space improvement: Smart Habitat for the Elderly, supported by COST (European Cooperation in Science and Technology). COST is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. It boosts their research, career, and innovation. More information in www.cost.eu. Furthermore, we would like to thank the Research Center in Digital Services (CISeD) and the Polytechnic of Viseu for their support.

References

1. Brownlee, J.: K-nearest neighbors for machine learning, August 2020. https://books.google.pt/books?hl=pt-PT&lr=&id=n-oDwAAQBAJ&oi=fnd&pg=PP1&dq=J.+Brownlee.+K-nearest+neighbors+for+machine+learning&ots=3kfEYbkCsc&sig=eyhGIdPfBSIrFXEo_avVbkfbywk&redir_esc=y#v=onepage&q&f=false

2. Carvalho, P.: Coronavírus: relação "estreita" com o clima aponta para verão mais calmo em portugal, March 2020. https://www.publico.pt/2020/03/13/ciencia/noticia/coronavirus-relacao-estreita-clima-aponta-verao-calmo-portugal-1907601
3. Claudino, H.M.: Sete gráficos para compreender as diferenças entre a primeira e a segunda vaga, November 2020. https://tvi24.iol.pt/geral/29-11-2020/sete-graficos-para-compreender-as-diferencas-entre-a-primeira-e-a-segunda-vaga
4. DGS: Covid-19 - página inicial, November 2020. https://covid19.min-saude.pt/
5. e Pinto, J.V.: Podemos estar a sair do "pico" da segunda vaga, mas "não é altura para descansar", November 2020. https://www.publico.pt/2020/11/27/sociedade/noticia/covid19-incidencia-abranda-portugal-nao-altura-descansar-1940973
6. Inácio, A.M.: Internamentos não atingem estimativas, mas uci só têm vagas quando alguém morre ou tem alta, November 2020. https://www.dn.pt/edicao-do-dia/04-nov-2020/internamentos-nao-atingem-estimativas-mas-uci-so-tem-vagas-quando-alguem-morre-ou-tem-alta-12994838.html
7. IPMA: Temperature API, December 2020. https://api.ipma.pt/
8. Lopes, M.J.: A procura deve aumentar, mas "não vai haver falta de vacinas" da gripe para quem as deve tomar, October 2020. https://www.publico.pt/2020/10/19/sociedade/noticia/procura-aumentar-nao-vai-haver-falta-vacinas-gripe-tomar-1935801
9. Ma, Y.: Effects of temperature variation and humidity on the death of Covid-19 in Wuhan, China, July 2020. https://www.sciencedirect.com/science/article/abs/pii/S0048969720317393
10. SNS: Avaliação da situação nacional das unidades de cuidados intensivos (2016). https://www.sns.gov.pt/wp-content/uploads/2016/05/Avalia%C3%A7%C3%A3o-nacional-da-situa%C3%A7%C3%A3o-das-unidades-de-cuidados-intensivos.pdf
11. SNS: Sns conta com um total de 17.700 camas para assistência à pandemia, October 2020. https://www.sns.gov.pt/noticias/2020/10/23/covid-19-capacidade-instalada/
12. Sol: Falta de vacinas da gripe preocupa farmacêuticos, October 2020. https://sol.sapo.pt/artigo/712179/falta-de-vacinas-da-gripe-preocupa-farmac-uticos
13. VOST: Covid-19 API, December 2020. https://covid19-api.vost.pt/
14. Wallis, P.: Disease metaphors in new epidemics: the UK media framing of the 2003 SARS epidemic, June 2005. https://www.sciencedirect.com/science/article/abs/pii/S0277953604005891?via%3Dihub
15. Xie, J.: Association between ambient temperature and Covid-19 infection in 122 cities from China, July 2020. https://reader.elsevier.com/reader/sd/pii/S0048969720317149