José Miguel Barrón-Adame¹, Luis Alberto Holgado-Apaza², María Susana Acosta-Navarrete¹, Rafael Guzmán-Cabrera³, Luis Beltran Palma-Ttito⁴, Solinka Suma-Salas⁵, Ralph Miranda-Castillo⁶

> ¹ Universidad Tecnológica del Suroeste de Guanajuato, Comité de Investigación, Mexico

> > ² Universidad Nacional del Altiplano, Peru

³ Universidad de Guanajuato, Mexico

⁴ Universidad Nacional de San Antonio Abad del Cusco, Peru

> ⁵ Universidad Nacional de Huancavelica, Peru

⁶ Universidad Nacional Amazónica de Madre de Dios, Peru

{mbarrona, msacosta}@utsoe.edu.mx, {lholgado, rmiranda}@unamad.edu.pe, guzmanc@ugto.mx, luis.palma@unsaac.edu.pe, solinkasumasalas@gmail.com

Abstract. In the present paper are analyzed the correlations between the meteorological variables such as maximum temperature, minimum temperature, relative humidity, precipitation with the number of infections by SARS-CoV-2. The data set for the meteorological variables was obtained from web site of the National Service of Meteorology and Hydrology of Peru (SENAMHI), while the number of infections by SARS-CoV-2 was obtained from the denominated data positive by COVID 19 reported in web site by the Ministry of Health from Peru. After the preprocessing and the fusion of the data sets, it was obtained a data subset with 365 registers and 6 columns. To detect the correlations between the meteorological variables and the number of infections by SARS-CoV-2, the Spearman's rank correlation coefficient was computed. The results show significant correlations between the variables minimum temperature and number of infections by SARS-CoV-2 for a rho=-0.45, pvalor=0.00<0.05; relative humidity and number of infections by SARS-CoV-2 for a rho=-0.24, p-valor=0.00<0.05; precipitation and number of infections by SARS-CoV-2 for a rho=-0.24, p-valor=0.00<0.05. According to the results obtained, we concluded that minimum temperatures facilitate the SARS-CoV-2 transmissibility.

Keywords. Data mining; COVID-19, environmental variables.

1 Introduction

On December 31st, 2019, China's National Health Commission, reported to the World Health Organization Office (WHO) the identification of a new type of coronavirus in Wuhan, Hubei, China [1]. The coronavirus was isolated on January 7th, 2020, and on January 11th-12th, 2020, WHO

Computación y Sistemas, Vol. 26, No. 1, 2022, pp. 399–409 doi: 10.13053/CyS-26-1-4011



Fig.1. Process for finding relationships between natural environmental factors and SARS-CoV-2 transmissibility

received information that 41 cases of infection with the new coronavirus had been preliminarily diagnosed in Wuhan city, in which 7 were critically ill.

The outbreak is presumed to be associated with exposures at a seafood market in Wuhan, however, at that time the evidence was still highly suggestive. As a precaution measure, the market had been closed on January 1st, 2020.

At this stage of the outbreak, infection among health care workers is not reported yet, being a new disease, there was no clear evidence of person-to-person transmission [2].

To date, SARS-CoV-2 is known to be transmitted from person to person by inhalation or contact with respiratory droplets and aerosols [3].

The reported numbers WHO for this disease worldwide at 10:42 am CET on May 11th, 2021 amounts to 158,551,526 confirmed cases and 3,296,855 deaths [4].

On the other hand, there are several domains where large volumes of information are stored in centralized and distributed databases [5], such as in Multimedia, for the retrieval of information by content through the similarity of objects [6]. Commerce, customer behavior is evaluated by purchasing criteria [7], in Social Sciences, to determine socio-economic patterns [8] or in

Computación y Sistemas, Vol. 26, No. 1, 2022, pp. 399–409 doi: 10.13053/CyS-26-1-4011 Medicine, to identify relationships in drug supply [9].

In this context, many efforts are made to understand the behavior of SARS-CoV-2, which is the reason of the present research it is verified whether there is a correlation between the number of SARS-CoV-2 infection and environmental factors, so that this knowledge can be used in favor of people's health.

2 Related Work

There are investigations that analyze the influence of daily temperatures on the number of covid-19 cases in the provinces of Spain and conclude that no evidence was detected suggesting a decrease in covid-19 cases at warmer mean, minimum and maximum temperatures, but they also warn that these results should be taken with caution [15]. [16] states that there is an optimal climatic zone in which SARS-CoV-2 is concentrated, and this temperature is 5°C to 15° C, and therefore suggest that the air temperature of hospitals and homes should be outside this range.

[17] indicates that there is evidence of a lower number of cases of COVID infections in countries where temperatures are warm despite the low

Date	(°C) Maximum Temperature	(°C) Minimum Temperature	(%)Relative Humidity	(mm/day) Precipitation		
2020-03-01	35	24,4	87,5	0		
2020-03-02	34,8	22,6	89,8	0		
2020-03-03	35	23,8	86,3	0		
2020-03-04	34,8	24	89,3	0		
2020-03-05	34,6	24	89,6	0		
2020-03-06	34,6	24,6	89,7	0		
2020-03-07	34,4	23,6	90,4	0		
2020-03-08	34,6	24	89	0		
2020-03-09	34,2	24,2	90,9	10		
2020-03-10	34,2	24	91,9	0		

Table 1: Structure of the Meteorological data set obtained from the Maldonado MET CP station

socio-economic situation, and they find a correlation between temperature and total COVID-19 cases of 0.59 and 0.42 with total COVID-19 cases and deaths. [18] collected meteorological data from 122 cities between January 2020 and February 2020 and get results showing a positive linear relationship, between mean temperature and confirmed COVID-19 cases, with a threshold of 3° C. They reiterate that the results are robust to sensitivity analyses.

They further clarify that there is no evidence that COVID-19 cases decrease as the climate becomes warmer.

[19] studied the influence of temperature and humidity on the number of COVID-19 infections in 100 Chinese cities and 1005 U.S. countries and found a statistically significant negative correlation between temperature and relative humidity in both China and the U.S. They claim that higher temperature and higher relative humidity potentially suppress the transmission of COVID-19.

In addition, [20] indicate that there is no clear evidence of environmental factors such as temperature, humidity or UV radiation in the spread of COVID-19 or slowing its transmission. Each of the phases is detailed below.

And they state that the scientific evidence of SARS-CoV-2 behavior with environmental factors is limited and scarce and is based more on ecological studies that do not allow measuring causality and suggest doing more in-depth studies with epidemiological designs and with more extensive data series.

3 Methodology

The process to find the relationship between natural environmental factors and the transmissibility of SARS-CoV-2 is shown in Figure 1, where 3 phases are observed: (1) Raw Data Base, (2) Data exploration and pre-processing, (3) Relationship search. The data used in this research belongs to the Tambopata District, Department of Madre de Dios, Peru, from 2020 and the beginning of the year 2021. In order to fulfill the objectives, a series of steps had to be followed, which are summarized below.

3.1 Download Raw Data

First, it was necessary to obtain environmental data from the National Service of Meteorology and Hydrology of Peru [10], and then to get medical data on the number of SARS-CoV-2 infections in Tambopata, Madre de Dios, Peru [11].

The hydrometeorological dataset was downloaded from the SENAMHI portal, which offers data from 370 stations located throughout Peru [10]. For the purposes of this study, the daily data records of the variables minimum and maximum temperature (°C), relative humidity (%) and precipitation (mm/day) of the Puerto

Region	Province	Distrit	MetodX	Age	Gender	Result_Date
San Martin	Mariscal Caceres	Juanjui	PR	44	Male	20201110
Ucayali	Atalaya	Sepahua	PR	42	Male	20201012
Ucayali	Coronel Portillo	Calleria	PR	43	Male	20201019
Amazonas	Condorcanqui	Nieva	PR	49	Male	20201017
Amazonas	Condorcanqui	Nieva	PR	46	Male	20201103
Amazonas	Condorcanqui	Nieva	PR	49	Male	20201108
Amazonas	Bagua	Aramango	PR	49	Male	20200611
Loreto	Maynas	San Juan Bautista	PR	46	Male	20201029
Loreto	Maynas	Belen	PR	46	Male	20201104
San Martin	San Martin	Tarapoto	PR	45	Male	20201105

Table 2. Structure of the data set COVID-19 Data Positive reported by the Ministry of Health (Peru)

Maldonado MET CP station (code 100109) of the conventional-meteorological type, from March 1st, 2020 to February 28th, 2021 with a total of 365 records, were chosen because the first reported case of SARS-CoV-2 infection in the department of Madre de Dios occurred on March 18th, 2020.

Table 1 shows a sample of this dataset corresponding to the first 10 records of the aforementioned dataset. Table 1 shows from left to right the values recorded for the variables: date, maximum temperature, minimum temperature, relative humidity and precipitation.

In relation to the transmissibility of SARS-CoV-2, we chose to download the dataset called Data positive for COVID-19 reported by the Ministry of Health in the National Open Data Platform of the Peruvian government, which contains the total number of reported cases that tested positive for COVID-19 in Peru, containing variables such as; department, province and district, method of diagnosis, age, sex and date of result, as of March 28th, 2021, with a total of 1,048,575 records [21].

Note Table 2, which shows the first 10 records of the aforementioned dataset.

Table 2 shows from left to right the values recorded for the variables: department, province, district, method of diagnosis, age, sex and date of result.

In this sample, the cut-off date columns were omitted, which for all the records corresponded to 2021/03/28, indicating that the downloaded data contain information up to the indicated date, and the column called UUID, that is the universal and unique identifier for each row.

3.2 Raw Data Exploring and Preprocessing

A second step was the preparation of data, using simple statistical techniques, visualization techniques and joins, to ensure that the data were free of errors [12].

In this phase, the quality of the data was verified, starting with the identification of missing values in the two data sets. Figure 2 shows the amount of missing data for each data set, in the case of meteorological data (Figure 2a). It can be observed that the variables minimum temperature, precipitation and relative humidity have 1, 1 and 2 missing values, respectively.

In case of the set of positive data for COVID-19 (Figure 2b), it can be observed that the variables age and date have 24 and 2023 missing values respectively, these values were imputed by their corresponding median, because these variables have extreme values (Figure 3a).

Since COVID-19 positive data, another subset was generated with information from the Tambopata province with the variables date of results and number of infected, this was possible by means of the functions *filter()*, *group_by()* and *summarise()* both from the library "dplyr" which is part of the R language, The next procedure was to



Fig. 2: (a) Amount of missing values in the meteorological data set. (b) Amount of missing values in the positive COVID-19 data set.





join this subset with the meteorological data set by means of the function *full_join()* also belonging to the aforementioned package.

This resulted in a set of 365 rows and 6 columns representing the variables result date, maximum temperature, minimum temperature, relative humidity, precipitation and number of infected, which was used for the correlation analysis. Next, the extreme values were identified and treated. Figure 3a shows that all the variables have extreme values, however, the variables precipitation and number of infected have very high extreme values

This is due to the fact that in the Madre de Dios region there are seasons with torrential rains, with respect to the variable number of infected, these extreme values are due to the maximum number of infections that occurred in the province of Tambopata during the study period.

Computación y Sistemas, Vol. 26, No. 1, 2022, pp. 399–409 doi: 10.13053/CyS-26-1-4011

			,	Vari	able	s						St	atist	ic			F	o-val	or		
maximul	m tei	mper	atura	a								(0,963	}				0,00	0		
minimun	n ten	npera	atura									(0,866	5				0,00	0		
relative l	humi	idity										(0,991					0,03	4		
Precipita	ation											(0,526	5				0,00	0		
number	infec	ted										(),526	5				0,00	0		
																			_		
precipitation-	-0.31	0.04	-0.12	0.29	1		precipitation-	-0.31	0.04	-0.13	0.29	1		precipitation-	-0.31	0.04	-0.18	0.29	1		
relative_ humidity-	-0.75	-0.24	-0.19	1	0.29	Corr	relative_ humidity-	-0.75	-0.24	-0.24	1	0.29	Corr	relative_ humidity-	-0.75	-0.24	-0.22	1	0.29	Corr	
number_ infected-	0.04	-0.42	1	-0.19	-0.12	1.0 0.5 0.0	number_ Infected-	0.1	-0.43	1	-0.24	-0.13	1.0 0.5 0.0 -0.5	number_ infected-	0.06	-0.43	1	-0.22	-0.18	0.5	
minimum_ temperature-	0.5	1	-0.42	-0.24	0.04	-0.5	minimum_ temperature-	0.5	1	-0.43	-0.24	0.04	-1.0	minimum_ temperature-	0.5	1	-0.43	-0.24	0.04	-0.5	
maximum_ temperature-	1	0.5	0.04	-0.75	-0.31		maximum_ temperature-	1	0.5	0.1	-0.75	-0.31		maximum_ temperature-	1	0.5	0.06	-0.75	-0.31		
	maximum_ temperature-	minimum_ temperature-	number_ infected-	relative_ humidity-	precipitation-			maximum_ temperature-	minimum_ temperature-	number_ infected-	relative humidity-	precipitation-			maximum	minimum_ temperature-	number_ infected-	relative_ humidity-	precipitation-		
	maxi	tempe	ē.E	5 Z	precip			temp	temp		6 E	preci			maximur temperatu	tempe	2.5	5 <u>7</u>	precip		
			a)							b)							C))			
precipitation-	-0.31	0.04	-0.13	0.29	1		precipitation-	-0.31	0.04	-0.12	0.29	1		precipitation	-0.31	0.04	-0.12	0.29	1		
relative_ humidity-	-0.75	-0.24	-0.19	1	0.29	Corr 1.0	relative_ humidity-	-0.75	-0.24	-0.19	1	0.29	Corr 1.0	relative humidity	-0.75	-0.24	-0.19	1	0.29	Corr 1.0	
number_ infected-	0.07	-0.43	1	-0.19	-0.13	0.0	number_ infected-	0.08	-0.41	1	-0.19	-0.12	0.5 0.0 -0.5	number infected		-0.41	1	-0.19	-0.12	-0.5	
minimum_ temperature-	0.5	1	-0.43	-0.24	0.04	-1.0	minimum_ temperature-	0.5	1	-0.41	-0.24	0.04	-10	minimum_ temperature	0.5	1	-0.41	-0.24	0.04	-1.0	
maximum_ temperature-	1	0.5	0.07	-0.75	-0.31		maximum_ temperature-	1	0.5	0.08	-0.75	-0.31		maximum_ temperature		0.5	0.08	-0.75	-0.31		
	maximum_ temperature-	minimum_ temperature-	number infected-	relative_ humidity-	precipitation-			maximum_ temperature-	minimum_ temperature-	number_ infected-	relative_ humidity-	precipitation-			maximum_ temperature-	minimum	number_ Infected-	relative_ humidity-	precipitation-		
			d)							e))						f)				
precip	pitation-	-0.31	0.04	-0.15	0.29	1	precip	vitation-	-0.31	0.04	-0.12	0.29	1	pre	cipitation-	-0.31	0.04	-0.17	0.29	1	
	elative_ umidity-	-0.75	-0.24	-0.16	1	0.29		lative_ unidity-	-0.75	-0.24	-0.2	1	0.29	Car	relative_ humidity-	-0.75	-0.24	-0.23	1	0.29	Co
	umber_ nfected-	0.07	-0.4	1	-0.16	-0.15		umber_ nfected-	0.09	-0.37	1	-0.2	-0.12		number_ infected-	0.13	-0.4	1	-0.23	-0.17	
	mum_ erature-	0.5	1	-0.4	-0.24	0.04		mum_ erature-	0.5	1	-0.37	-0.24	0.04		inimum_ iperature-	0.5	1	-0.4	-0.24	0.04	1
	imum_ erature-	1	0.5	0.07	-0.75	-0.31		imum_ erature-	1	0.5	0.09	-0.75	-0.31		aximum_ 1perature-	1	0.5	0.13	-0.75	-0.31	
		maximum_ temperature-	minimum_ temperature-	number_ infected-	relative_ humidity-	precipitation-			maximum_ temperature-	minimum_ temperature-	number_ infected-	relative_ humidity-	precipitation-			maximum_ temperature-	minimum_ temperature-	number_ infected-	relative_	precipitation-	
		·	g)			-			-	h)			4				i)			-	

Table 3: Results obtained from the Shapiro-Wilks normality test

Fig. 4: Correlation among the meteorological variables and the variable of number of infections by SARS-CoV-2, with a delay of (a) 2 days, (b) 3 days, (c) 4 days, (d) 5 days, (e) 6 days, (f) 7 days, (g) 8 days, (h) 9 days, (i) 10 days, (j) 11 days, (k) 12 days, (l) 13 days, (m) 14 days and (n) 15 days

Computación y Sistemas, Vol. 26, No. 1, 2022, pp. 399–409 doi: 10.13053/CyS-26-1-4011



Fig. 4. Continuation

In order to correct this, the imputation by the values of the 5th and 95th percentiles was applied to each of the variables.

It should be clarified that extreme values were considered to be those above Q3 + 1.5 * IQR or below Q3 - 1.5 * IQR, being Q1 and Q3 the quartiles 1 and 3 respectively, IQR = Q3 - Q1.

In addition, it was necessary to apply the Yeo-Johnson transformation using the *yeojohnson()* function found in the "bestNormalize" library of the *R* language for the purpose of centering and scaling the variables. Note the distribution of the data in figure 3b after this procedure.

3.3 Search of Relationships

The third step was to verify whether the data presented correctly, and it could be seen that the data was not uniform among themselves; it was necessary to transform them in order to stabilize

Computación y Sistemas, Vol. 26, No. 1, 2022, pp. 399–409 doi: 10.13053/CyS-26-1-4011

ISSN 2007-9737

406 José Miguel Barrón-Adame, Luis Alberto Holgado-Apaza, et al.

Experiment	Days late	rho, p- valor	(^o C) maximum temperatura	(^o C) minimum temperatura	(%) relative humidity	(mm/day)] Precipitation	number infected
1	2	rho	0,04	-0,42	-0,19	-0,12	1,00
I	2	p-valor	0,41	0,00	0,00	0,03	-
2	3	rho	0,10	-0,43	-0,24	-0,13	1,00
2	3	p-valor	0,06	0,00	0,00	0,01	-
3	4	rho	0,06	-0,43	-0,22	-0,18	1,00
3	4	p-valor	0,23	0,00	0,00	0,00	-
4	5	rho	0,07	-0,43	-0,19	-0,13	1,00
4	5	p-valor	0,21	0,00	0,00	0,01	-
5	C	rho	0,08	-0,41	-0,19	-0,12	1,00
5	6	p-valor	0,12	0,00	0,00	0,02	-
6	7	rho	0,08	-0,41	-0,19	-0,12	1,00
0	/	p-valor	0,13	0,00	0,00	0,02	-
7	0	rho	0,07	-0,40	-0,16	-0,15	1,00
7	8	p-valor	0,20	0,00	0,00	0,01	-
	0	rho	0,09	-0,37	-0,20	-0,12	1,00
8	9	p-valor	0,10	0,00	0,00	0,02	-
	10	rho	0,13	-0,40	-0,23	-0,17	1,00
9	10	p-valor	0,02	0,00	0,00	0,00	-
10		rho	0,06	-0,42	-0,19	-0,24	1,00
10	11	p-valor	0,24	0,00	0,00	0,00	-
	10	rho	0,04	-0,44	-0,15	-0,19	1,00
11	12	p-valor	0,43	0,00	0,00	0,00	-
12	10	rho	0,01	-0,45	-0,14	-0,18	1,00
	13	p-valor	0,84	0,00	0,01	0,00	-
10		rho	0,01	-0,45	-0,13	-0,17	1,00
13	14	p-valor	0,87	0,00	0,02	0,00	-
	4-	rho	0,00	-0,42	-0,12	-0,14	1,00
14	15	p-valor	1,00	0,00	0,02	0,01	-

them and reduce the distance between them; an alternative was sought to make them comparable, using the Johnson and Johnson method [13].

The procedures that were followed in this phase were as follows.

Normality Test

 The Shapiro-Wilks normality test was done. This test was chosen because it is one of the

Computación y Sistemas, Vol. 26, No. 1, 2022, pp. 399–409 doi: 10.13053/CyS-26-1-4011

most powerful tests for all types of distribution and sample size [22]. The procedure followed is presented below:

Statistical Hypothesis Statement

- H0: The variable under a study that presents a normal distribution.
- H1: The variable under a study that does not present a normal distribution.

Significance Level

5% (alpha=0.05).

Decision Rule

- p-value>alpha: H0 is not rejected (The variable under a study that presents a normal distribution).
- p-value<alpha: H0 is rejected (The variable under a study that does not have a normal distribution).

The results of the Shapiro-Wilks normality test are shown in Table 3. It shows that the p-value for all the variables of analysis is less than 0.05, indicating that none of them presents a normal distribution, therefore, to find the correlations the Spearman's nonparametric test was used.

Correlation Analysis

 Finally, Spearman's rank correlation coefficient was applied to detect the correlation between the number of SARS-CoV-2 infections and the environmental variables [14]. They were able to identify that minimum temperatures are related to the increase in the number of SARS-CoV-2 infections.

Spearman's nonparametric correlation test was applied to find possible relationships of each of the meteorological variables with the variable number of SARS-CoV-2 infections. The next stage shows the experiments carried out by taking into account a delay of 2, 3, 4, 4, 5, 6, 7, 7, 8, 9, 10, 11, 12, 13, 14 and 15 days for the variable number of infected by COVID-19.

In agreement with the results of the study carried out by [23], where they state that the median incubation period of SARS-Cov-2 virus was 4 days (interquartile range, 2 to 7) in 1099 patients with Covid-19, as well as [24] estimated a median incubation period of 5-6 days, with a range of 1 to 14 days.

4 Experiments and Results

Figure 4 shows the summary of the Spearman correlation plots. Each image corresponds to the results of the correlation experiments between

meteorological variables and the number of SARS-CoV-2 infections with lags of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15 days.

After applying the statistical test to the preprocessed data, correlation coefficients (rho) and statistical significances (p-value) were obtained between the meteorological variables and the variable number of SARS-CoV-2 infections. Table 4 shows in rows the correlation experiments performed for different values of lag days (from 2 to 15) for the variable number of SARS-CoV-2 infected with their respective rho and p-value values, and in columns the meteorological variables.

Of the 4 studied meteorological variables, 3 showed significant correlation with the number of SARS-CoV-2 infections. The most significant coefficient was obtained with the minimum temperature variable (rho=-0.45, p-value = 0.00<0.05) corresponding to the experiments with 13 and 14 days of delay. In relation to the relative humidity variable the most significant value (rho=-0.24, p-value=0.00<0.05) was obtained for 3 days of delay of the variable in question. Finally, the precipitation variable achieve his most significant value (rho=-0.24, p-value = 0.00<0.05) for the experiment with 11 days of delay.

5 Conclusions

In this research, the existence of a correlation between the level of SARS-CoV-2 infection and environmental factors was verified. For this purpose, data mining was used. The R programming language allowed the collection, exploration and preprocessing of the data, in order to find possible relationships between the meteorological variables maximum temperature, minimum temperature, relative humidity, precipitation and the number of SARS-CoV-2 infections.

In the data exploration and preprocessing stage, the presence of extreme values was observed in all the variables; to correct this, imputation by the 5th percentile was used for those values below Q1 - 1.5 * IQR and by the 95th percentile for those values above Q3 + 1.5 * IQR, in addition to which it was necessary to apply the Yeo-Johnson transformation implemented in the R

library "bestNormalize" to center and scale the variables.

The most significant results show a moderate negative correlation (rho=-0.45, p-value = 0.00 < 0.05) between minimum temperature and the number of SARS-CoV-2 infections. Thus demonstrating that minimum temperatures are related to the increase in the number of SARS-CoV-2 infections.

References

- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet, 395(10223), 497–506. DOI: 10.1016/S0140-6736(20)30183-5.
- WHO (2020). OMS, Nuevo coronavirus China. https://www.who.int/csr/don/12-january-2020-novelcoronavirus-china/en/.
- Molero-García, M. J., Arranz-Izquierdo, J., Pérez-Gutiérrez, M. I. (2020). Aspectos básicos del COVID-19 para el manejo desde Atención Primaria. Atención Primaria, 1–13. DOI: 10.1016/j.aprim.2020. 12.007.
- 4. WHO. (2021). Panel de control de la enfermedad por coronavirus de la OMS (COVID-19).
- Riquelme, J. C. Ruiz, R. Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 10, núm. 29, pp. 11–18.
- Fernández, J., Gil-Costa, V., Ludueña, V., Reyes, N., Roggero, P., Chávez, E. (2011). Indexación y recuperación de información multimedia. La plata Buenos aires, Universidad Michoacana de San Nicolás de Hidalgo, pp. 324–328.
- Mylonakis, J., Diacogannis, G. (2010). Evaluating the likelihood of using linear discriminant analysis as a commercial bank card owners credit scoring model. International Business Research, Vol. 3, No. 2, pp. 9–20.
- Gallager, B., Eliassi-Rad, T. (2010). Leveraging label independent features for classification in sparsely labeled networks: An empirical study. In: Advances in social network mining and analysis. Pennsylvania: Springer Verlag Heidelberg, pp. 1–19.
- Tari, L., Anwar, S., Liang, S., Cai, J., Baral, C. (2010). Discovering drug-drug interactions: a textmining and reasoning approach based on properties

Computación y Sistemas, Vol. 26, No. 1, 2022, pp. 399–409 doi: 10.13053/CyS-26-1-4011 of drug metabolism. Bioinformatics, Vol. 26, pp. 547–553.

- **10. SENAMHI.** (2021). SENAMHI Perú. https://www.senamhi.gob.pe/?&p=estaciones
- MINSA. (2021). Plataforma Nacional de Datos Abiertos. https://www.datosabiertos.gob.pe/dataset/ casos-positivos-por-covid-19-ministerio-de-saludminsa/resource/690e57a6-a465-47d8-86fd.
- 12. Faceli, K., Lorena, A. C., Gama, J., Ponce de León Carvallo, A. C. (2011). Inteligencia Artificial: Uma Abordagem de Aprendizado de Maquina.
- Agudelo-Rodríguez, L. A., Flórez-Londoño, A. D. J., Correa-Mejía, D. A. (2020). Efectos de la aplicación del marco normativo para entidades de gobierno en los indicadores financieros. Entramado, 16(2), 90–110. DOI: 10.18041/1900-3803/ entramado.2.6545.
- 14. Martínez-Ortega, R. M., Tuya-Pendás, L. C., Martínez-Ortega, M., Pérez-Abreu, A., Cánovas, A. M. (2009). El coeficiente de correlacion de los rangos de spearman caracterizacion. Revista Habanera de Ciencias Medicas, 8(2).
- **15. Briz-Redón, Á., Serrano-Aroca, Á. (2020).** A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. Science of the Total Environment, 728, 138811. DOI: 10.1016/j.scitotenv.2020.138811.
- 16. Huang, Z., Huang, J., Gu, Q., Du, P., Liang, H., Dong, Q. (2020). Optimal temperature zone for the dispersal of COVID-19. Science of the Total Environment, 736, 139487. DOI: 10.1016/j. scitotenv.2020.139487.
- 17. Iqbal, M. M., Abid, I., Hussain, S., Shahzad, N., Waqas, M. S., Iqbal, M. J. (2020). The effects of regional climatic condition on the spread of COVID-19 at global scale. Science of the Total Environment, 739. DOI: 10.1016/j.scitotenv.2020.140101.
- Xie, J., Zhu, Y. (2020). Association between ambient temperature and COVID-19 infection in 122 cities from China. Science of the Total Environment, 724, 138201. DOI: 10.1016/j.scitotenv.2020.138201.
- Wang, J., Tang, K., Feng, K., Lin, X., Lv, W., Chen, K. (2020). Impact of temperature and relative humidity on the transmission of COVID-19: A modeling study in China and the United States. SSRN, 11(2).
- 20. Linares Gil, C., Sánchez Martínez, G., Díaz Jiménez, J. (2020). ¿Influyen el clima y la contaminación atmosférica en la transmisión e incidencia de la nueva enfermedad COVID19? Revista Diecisiete: Investigación Interdisciplinar para los Objetivos de Desarrollo Sostenible. Vol. 3, pp. 43–54.

- 21. Plataforma Nacional de Datos Abiertos (2021). Plataforma Nacional de Datos Abiertos. https://www.datosabiertos.gob.pe/dataset/casospositivos-por-covid-19-ministerio-de-saludminsa/resource/690e57a6-a465-47d8-86fd.
- **22. Razali, N. M., Wah, Y. B. (2011).** Power comparisons of Shapiro-Wilks, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. Journal of Statistical Modeling and Analytics, 2, 21–33.
- 23. Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, D. S. C., Du, B., Li, L., Zeng, G., Yuen, K.-Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., Zhong, N. (2020). Clinical Characteristics of Coronavirus Disease 2019

in China. New England Journal of Medicine, 382(18), 1708–1720.

24. López, V., Vázquez, T., Alonso-Titos, J., Cabello, M., Alonso, A., Beneyto, I., Crespo, M., Díaz-Corte, C., Franco, A., González-Roncero, F., Gutiérrez, E., Guirado, L., Jiménez, C., Jironda, C., Lauzurica, R., Llorente, S., Mazuecos, A., Paul, J., Rodríguez-Benot, A., Hernández, D. (2020). Recomendaciones en el manejo de la pandemia por coronavirus SARS-CoV-2 (Covid-19) en pacientes con trasplante renal. Nefrología, 40(3), 265–271. DOI:10.1016/j.nefro.2020.03.002.

Article received on 06/03/2021; accepted on 15/08/2021. Corresponding author is Rafael Guzmán-Cabrera.