

# Predicción de la Efectividad de las Pruebas Rápidas Realizadas a Pacientes con COVID-19 mediante Regresión Lineal y Random Forest



## Predicting the Effectiveness of Rapid Tests Performed to Patients with COVID-19 through Linear Regression and Random Forest

Patiño Pérez, Darwin; Munive Mora, Celia; Cevallos-Torres, Lorenzo; Botto-Tobar, Miguel

**Darwin Patiño Pérez**  
darwin.patinop@ug.edu.ec  
Universidad de Guayaquil, Ecuador  
**Celia Munive Mora**  
cm3877@desales.edu  
DeSales University, Estados Unidos  
**Lorenzo Cevallos-Torres**  
lorenzo.cevallost@ug.edu.ec  
Universidad de Guayaquil, Ecuador  
**Miguel Botto-Tobar**  
miguel.bottot@ug.edu.ec  
Universidad de Guayaquil, Ecuador

**Ecuadorian Science Journal**  
GDEON, Ecuador  
ISSN-e: 2602-8077  
Periodicidad: Semestral  
vol. 5, núm. 2, 2021  
esj@gdeon.org

Recepción: 18 Junio 2021  
Aprobación: 30 Agosto 2021

URL: <http://portal.amelica.org/ameli/jatsRepo/606/6062590003/index.html>

DOI: <https://doi.org/10.46480/esj.5.2.108>



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivar 4.0 Internacional.

Como citar: Patiño Pérez, D., Munive Mora, C., Cevallos-Torres, L., & Botto-Tobar, M. (2021). Predicción de la Efectividad de las Pruebas Rápidas Realizadas a Pacientes con COVID-19 mediante Regresión Lineal y Random Forest. Ecuadorian Science Journal, 5(2), 31-43. DOI: <https://doi.org/10.46480/esj.5.2.108>

**Resumen:** La rapidez de propagación del SARS-CoV2 (COVID-19), ha ocasionado un colapso de los sistemas de salud a nivel mundial, por lo que una estrategia para controlar la propagación es la detección oportuna del virus por medio de pruebas rápidas, que permitan tomar acciones y así dar un tratamiento oportuno que reduzca su propagación. Con la ayuda de las técnicas de inteligencia artificial, dentro del subcampo del aprendizaje automático o machine learning, se han dado avances significativos que permiten agilizar el análisis de grandes volúmenes de datos. Este estudio tiene como objetivo, determinar la efectividad de las pruebas rápidas en la detección del covid-19, mediante el uso de machine learning aplicando una metodología que implica la creación de los modelos de regresión lineal y Random Forest con el lenguaje de programación Python. En la metodología usada se crearon los modelos, que luego fueron definidos y entrenados y después de realizar las pruebas y predicciones, las métricas de validación determinaron la precisión y efectividad de estos modelos. De los resultados obtenidos se concluye que el modelo random forest es bueno dado que proporcionó una precisión del 61% sin embargo con el modelo de regresión lineal se determinó que este tiene un nivel de precisión de aproximadamente el 90%, por lo que finalmente con estos modelos los profesionales de la salud podrán realizar predicciones confiables en cuanto a la efectividad de las pruebas rápidas como un mecanismo que ayudarán a detectar rápidamente la presencia del virus y así reducir la propagación del virus.

**Palabras clave:** COVID-19, Machine Learning, Regresión Lineal, Random Forest, Predicción.

**Abstract:** The rapid spread of SARS-CoV2 (COVID-19) has caused a collapse of health systems worldwide, so a strategy to control the spread is the timely detection of the virus through rapid tests, which allows acting and thus giving a timely treatment that reduces its spread. With the help of artificial intelligence techniques, within the subfield of machine learning or machine learning, there have been significant advances that allow speeding up the analysis of large volumes of data. This study

aims to determine the effectiveness of rapid tests in detecting covid-19, using machine learning, applying a methodology that involves the creation of linear regression and Random Forest models with the Python programming language. In the methodology used, the models were created, which were then defined and trained, and after performing the tests and predictions, the validation metrics determined the precision and effectiveness of these models. From the results obtained, it is concluded that the random forest model is good since it provided a precision of 61%, but with the linear regression model, it was determined that it has a precision level of approximately 90%, so finally, with these results, health professionals will be able to make reliable predictions regarding the effectiveness of rapid tests as a mechanism that will help to quickly detect the presence of the virus and thus reduce the spread of the virus.

**Keywords:** COVID-19, Machine Learning, Linear Regression, Random Forest, Prediction.

## INTRODUCCIÓN

En el mes de diciembre del 2019, en la ciudad de Wuhan, China, se detectó un nuevo brote de neumonía de origen desconocido el cual se ha expandido rápidamente por todo el mundo.(Wu, Chen, & Chan, 2020). Esta nueva neumonía logró ser aislada y su genoma fue secuenciado recibiendo el nombre de SARS-CoV-2 o COVID-19, presentando una similitud genética de un 79% con el SARSCoV y un 50% con MERS-CoV, todos de la familia de los  $\beta$ -coronavirus.(Latini et al., 2020)

El COVID-19 se caracteriza por una inmediata propagación, existe transmisión directa persona a persona por inhalación de gotitas producidas al hablar en voz alta o reír ya que la saliva contiene virus en pacientes infectados(W. Li et al., 2020). Los datos epidemiológicos indican que todas las edades son susceptibles, pero presentan una mayor prevalencia los hombres con edad media de 56 años, individuos con contacto cercano con personas infectadas sintomáticas o asintomáticas, incluyendo otros pacientes que se encuentren en los centros de salud, al igual que el equipo médico quienes conforman un porcentaje importante de las personas infectadas(Rothan & Byrareddy, 2020). Los síntomas típicos son fiebre, tos seca, mialgia, fatiga y disnea con una tomografía torácica (TC) anormal observando neumonía bilateral, opacidad como vidrio esmerilado y/o sombras irregulares bilaterales. Los síntomas atípicos son producción de esputo, cefaleas, hemoptisis, diarrea, mareo, dolor abdominal, náuseas, vomito, confusión y dolor de garganta(Guñez-Coelho, 2020)

Se cuenta con dos tipos de pruebas para el diagnóstico de SARS-CoV-2: las pruebas moleculares que son el Gold estándar y que se basan en la detección del virus (qRT-PCR y RT-LAMP) y las pruebas rápidas que detectan anticuerpos generados por la respuesta inmunológica del paciente(Z. Li et al., 2020). Las pruebas moleculares pueden detectar la presencia del virus en media, una semana antes y hasta una semana después de la aparición de los síntomas(Ahmad, Ali, Kausar, Misbah, & Wahid, 2020).

Sin embargo, qué ocurre con los individuos asintomáticos que no fueron sometidos a ningún tipo de prueba diagnóstica y que ignoran si fueron infectados o no, y en qué momento fueron infectados. Para este grupo existen las pruebas rápidas, pruebas que detectan la respuesta inmunológica del individuo(Rajendra Santosh, Krishnamurthy, & Reddy Baddam, 2020).

A diferencia de las pruebas moleculares que utilizan secreciones mucosas que contienen virus, las pruebas rápidas y en general las actuales pruebas serológicas son de primera generación y aún están siendo perfeccionadas, estas pruebas serológicas incluyen Ensayos por inmunoadsorción ligado a enzimas (ELISA), Inmunoensayos de quimioluminiscencia (CLIA), e Inmunoensayos de electro quimioluminiscencia (EIA),

analizan principalmente sangre, suero o plasma. Se sabe que las pruebas serológicas que usan suero y plasma son consideradas como fuentes de baja contaminación viral y tienen mejor sensibilidad que las realizadas con sangre total o capilar (Valencia Portillo et al., 2020)

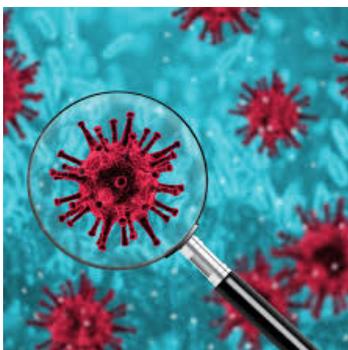


FIGURA 1.  
Virus SARS-CoV-2.  
Autores.

Durante ese periodo de tiempo se han recogido diversos datos relacionados a causas y efectos provocados por el virus denominado “Coronavirus” o COVID-19, datos los cuales han sido procesados y estudiados por distintas entidades. Entre estas investigaciones, se encuentran las realizadas en el área de informática, mediante el uso de datasets para la predicción de covid-19 usando técnicas de machine learning (Pérez, Bustillos, Mora, & Botto-Tobar, 2020). Existen diferentes tipos de técnicas dentro del campo de la inteligencia artificial para la elaboración de predicciones, unas tomando características de entrada y salida extraídas de los dataset y que se ajustan al tipo de aprendizaje profundo (Perez, Bustillos, Botto-Tobar, & Mora, 2021) y otras que solo toman las entradas de los datasets y generan sus propias salidas.

## MATERIALES Y MÉTODOS

### Algoritmos para Predicción

Los algoritmos de aprendizaje supervisado que se usaran como modelo de predicción y clasificación son Regresión Lineal (Rath, Tripathy, & Tripathy, 2020) para la predicción y Random Forest (Muhammad, Islam, Usman, & Ayon, 2020) para la clasificación; el segundo maneja eficientemente las distribuciones para cálculos de probabilidad; estos algoritmos en relación al COVID-19 han sido usados en diversas investigaciones para realizar predicciones en diferentes casos, como los efectos del distanciamiento social en la población, predecir el resultado de salud de los pacientes y como se encontrara en términos de salud y los factores climáticos influyentes.

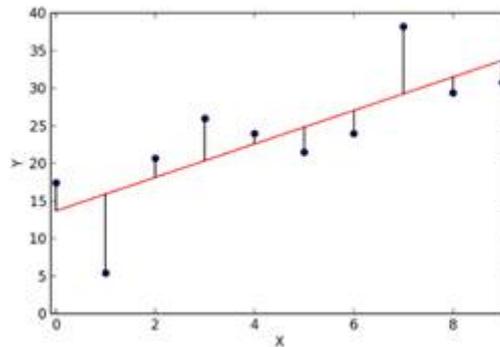


FIGURA 2.  
Regresión Lineal  
(Propia)

Cuando se tratan de muchas características o variables independientes, tomadas del conjunto de datos o dataset se usa regresión lineal múltiple, donde el modelo matemático que se plantea tiene la forma de la ecuación(1).

$$y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_nX_{ni} \tag{1}$$

Es importante tener en cuenta que, con la regresión lineal, se está tratando de predecir una variable continua resultados (Chanchí Golondrino, Campo Muñoz, & Sierra Martinez, 2020). Además, en un modelo de regresión, se trata de minimizar los errores al encontrar la "línea de mejor ajuste"; la línea de regresión de los errores sería mínima (Lee et al., 2020). Al tratar de minimizar la longitud de las líneas negras (o más exactamente, la distancia desde los puntos azules hacia la línea roja) desde la línea roja, lo más cerca posible de cero. Está relacionado con (o equivalente a) minimizar el error cuadrático medio (MSE) o la suma de cuadrados del error (SSE), también llamada "suma de cuadrados residual".

Por otra parte, Random Forest o bosque aleatorio una técnica de aprendizaje automático que se utiliza para resolver problemas de regresión y clasificación (Yoo et al., 2020). Utiliza el aprendizaje por conjuntos, que es una técnica que combina muchos clasificadores para proporcionar soluciones a problemas complejos. Un algoritmo de bosque aleatorio consta de muchos árboles de decisión. El "bosque" generado por el algoritmo de bosque aleatorio se entrena mediante agrupamiento o agregación de arranque. El agrupamiento es un meta-algoritmo de conjunto que mejora la precisión de los algoritmos de aprendizaje automático. El algoritmo (bosque aleatorio) establece el resultado en función de las predicciones de los árboles de decisión. Predice tomando el promedio o la media de la salida de varios árboles(Sharma, Lilhore, Simaiya, & Trivedi, 2021). El aumento del número de árboles aumenta la precisión del resultado. Un bosque aleatorio erradica las limitaciones de un algoritmo de árbol de decisión(Srivatsan, Indi, Agrahari, Menon, & Ashok, 2020). Reduce el sobreajuste de conjuntos de datos y aumenta la precisión. Genera predicciones sin requerir muchas configuraciones en paquetes de Python (como scikit-learn).

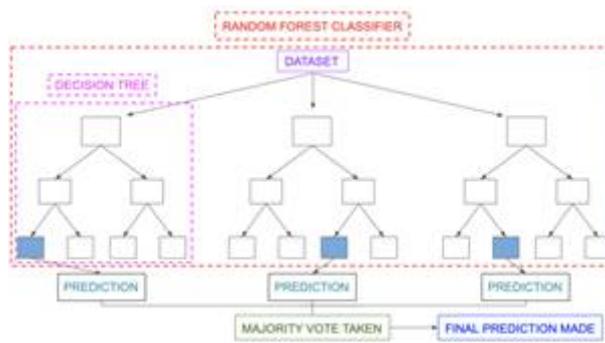


FIGURA 3.  
Random Forest  
(Propia)

Este algoritmo posee ventajas ante otros algoritmos predictivos, como el manejo efectivo de grandes cantidades de datos por su esquema de descomposición según la Figura 3. Los árboles de decisión son los componentes básicos de un algoritmo de bosque aleatorio. Un árbol de decisión es una técnica de apoyo a las decisiones que forma una estructura en forma de árbol, consta de tres componentes: nodos de decisión, nodos hoja y un nodo raíz. Un algoritmo de árbol de decisión divide un conjunto de datos de entrenamiento en ramas, que se segregan aún más en otras ramas (Elshazli et al., 2020). Esta secuencia continúa hasta que se alcanza un nodo hoja. El nodo hoja no se puede segregar más. Los nodos del árbol de decisiones representan atributos que se utilizan para predecir el resultado. Los nodos de decisión proporcionan un enlace a las hojas según la Figura 4 en la que se reflejan los tres tipos de nodos en un árbol de decisión (Vinod & Prabakaran, 2020).

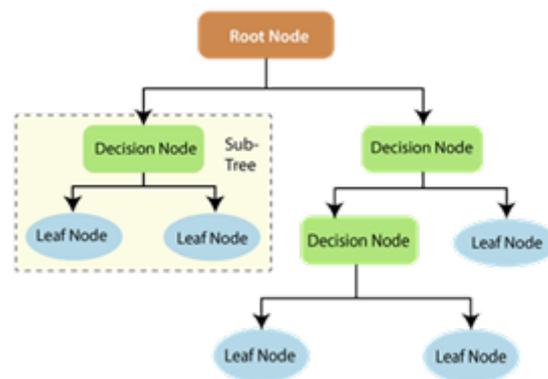


FIGURA 4.  
Árbol de Decisión  
(Propia)

## Métricas

Error absoluto medio (MAE), es la métrica de error más simple utilizada en problemas de regresión. Básicamente es la suma del promedio de la diferencia absoluta entre los valores predichos y reales. En palabras simples, con MAE podemos tener una idea de cuán equivocadas fueron las predicciones. MAE no indica la dirección del modelo, es decir, no hay indicios de rendimiento insuficiente o excesivo del modelo (Ballı, 2021). La siguiente es la fórmula para calcular MAE.

$$MAE = \frac{1}{n} \sum |y - y^{\wedge}|$$

(2)

La matriz de confusión es una métrica que se usa en problemas de clasificación, La matriz de confusión es una métrica que ayuda a describir el desempeño de un modelo de clasificación. Para construir una matriz de confusión, todo lo que se necesita hacer es crear una tabla de valores reales y valores predichos(Pourhomayoun & Shakibi, 2021).

		Predicted Values	
		Good	Bad
Actual Values	Good	True Positive (TP)	False Negative (FN)
	Bad	False Positive (FP)	True Negative (TN)

FIGURA 5.  
Matriz de Confusión  
(Propia)

True Positive. El verdadero positivo no es más que el caso en el que el valor real y el valor predicho son verdaderos. El paciente ha sido diagnosticado con covid y el modelo también predijo que el paciente tenía covid.

False Negative. En falso negativo, el valor real es verdadero, pero el valor predicho es falso, lo que significa que el paciente tiene covid, pero el modelo predijo que el paciente no tenía covid.

False Positive. Este es el caso en el que el valor predicho es verdadero, pero el valor real es falso. Aquí, el modelo predijo que el paciente tenía covid, pero en realidad, el paciente no tiene covid. Esto también se conoce como error tipo 1.

True Negative. Este es el caso en el que el valor real es falso y el valor predicho también es falso. En otras palabras, al paciente no se le diagnostica covid y nuestro modelo predijo que el paciente no tenía covid.

De los valores de esta matriz TP, FN, FP, TN son utilizados para calcular: El accuracy, que, en los problemas de clasificación, se refiere al número de predicciones correctas realizadas por el modelo predictivo sobre el resto de predicciones(Wiguna & Riana, 2020). Se lo usa cuando las clases de variables de destino en los datos están casi equilibradas y no se lo usa cuando las variables de destino en los datos son la mayoría de una clase(Amin, Uddin, Al-Baity, Zeb, & Khan, 2021). La Precision, que significa en qué proporción de todas las predicciones que se hacen con nuestro modelo predictivo son realmente verdaderas. El recall or sensitivity, es la medida que indica qué proporción de pacientes que realmente tenían covid también se predijo que lo tendrían; responde a la pregunta: "¿Qué tan sensible es el clasificador para detectar instancias positivas?". La Specificity, responde a la pregunta: "¿Qué tan específico o selectivo es el clasificador para predecir instancias positivas?". Una especificidad de 0,61 significa que el 61 por ciento de todos los pacientes que no tenían covid se predice correctamente. F1 Score, esta métrica no es más que la media armónica de precision y recall, cuando F1 score es alta, es decir, tanto la precision como el recall del clasificador indican buenos resultados.

## Metodología

Se usará Python como lenguaje de programación, que es el que más soporte tiene en el área de la inteligencia artificial, ya que se distingue por que existe una comunidad que proporciona el soporte necesario para las bibliotecas usadas dentro de la computación científica y la inteligencia artificial. Además, se usará Google COLAB por ser un entorno de desarrollo Online muy potente en la nube de Google. ("A Study of Real World Data V isualization of COVID-19 Dataset Using Python," 2020)

El conjunto de datos de pruebas rápidas de pacientes con COVID tiene 20724 instancias con 8 atributos: provincial\_iso, fecha, num\_casos, num\_casos\_prueba\_pcr, num\_casos\_prueba\_test\_ac, num\_casos\_prueba\_ag, num\_casos\_prueba\_elisa, num\_casos\_prueba\_desconocida.

Se tomó como referencia la fase metodológica dentro de aprendizaje supervisado que implica: obtención de los datos, separación de datos para entrenamiento y para pruebas, definición del modelo, entrenamiento del modelo, ajuste y pruebas.

## Preparación del Conjunto de Datos

Se extrajo un conjunto de datos con solo 1000 instancias y 6 atributos que incluyen: num\_casos, num\_casos\_prueba\_pcr,

num\_casos\_prueba\_test\_ac, num\_casos\_prueba\_ag, num\_casos\_prueba\_elisa, num\_casos\_prueba\_desconocida.

La Tabla 1 refleja la muestra de algunas instancias del conjunto de datos.

TABLA 1.  
Muestra de las instancias del conjunto de datos

num_casos	num_casos_prueba_pcr	num_casos_prueba_test_ac	num_casos_prueba_ag	num_casos_prueba_elisa	num_casos_prueba_desconocida
1	1	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
2	2	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

Propia

## Creación del Modelo de Regresión

1) Incluir paquetes o librerías necesarias.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn import linear_model
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import mean_squared_error, r2_score
```

Se han declarado los tres alias, np para numpy que sirve para el procesamiento numérico y de matrices, pd de pandas para cargar el dataset y poder crear el dataframe, sns para seaborn que se lo emplea para manejar un mejor entorno gráfico. Luego se incluyen algunos métodos de sklearn como linear\_model para crear el modelo de regresión lineal, train\_test\_split que sirve para poder dividir la data en entrenamiento y prueba, además de las funciones que sirven para calcular las métricas adecuadas del modelo.

2) Leer el dataset y convertirlo en un dataframe por medio de la función read\_csv().

dataframe = pd.read\_csv("casos\_diagnostico\_provincia.csv")

3) Mostrar la información de la cabecera del dataframe con el método head().

```
dataframe.head()
print('Información en el dataset:')
print(dataframe.keys())
```

4) Mostrar el cuadro estadístico y el diagrama de correlación por medio de las funciones describe() y corr() de pandas, además del mapa de calor por medio de plt de pyplot y sns de seaborn.

```
dataframe.describe()
```

**TABLA 2.**  
Muestra de las instancias del conjunto de datos

num_casos	num_casos_grueba_pcr	num_casos_grueba_test_ac	num_casos_p_rueba_ag	num_casos_p_rueba_elisa	num_casos_grueb_a_desconocida	
count	20723.000000	20723.000000	20723.000000	20723.000000	20723.000000	20723.000000
mean	125.481735	103.724219	0.229938	21.239685	0.178160	0.109733
std	351.665446	296.064538	1.395562	108.869959	2.047763	1.269657
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	2.000000	1.000000	0.000000	0.000000	0.000000	0.000000
50%	26.000000	22.000000	0.000000	0.000000	0.000000	0.000000
75%	112.000000	95.000000	0.000000	4.000000	0.000000	0.000000
max	6743.000000	6726.000000	32.000000	3411.000000	86.000000	65.000000

Propia

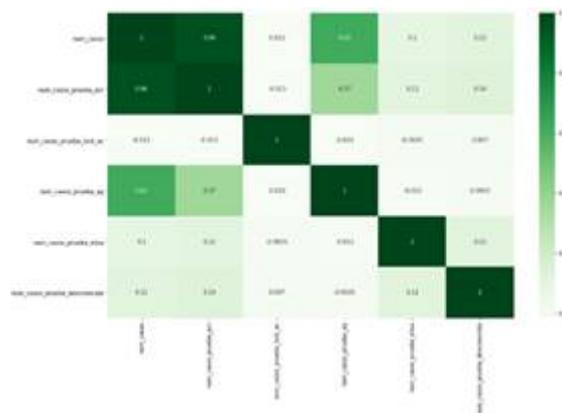
```
dataframe.corr()
```

**TABLA 3.**  
Muestra de las instancias del conjunto de datos

num_casos	num_casos_grueba_pcr	num_casos_grueba_test_ac	num_casos_grueba_ag	num_casos_p_rueba_elisa	num_casos_grueb_a_desconocida	
num_casos	1.000000	0.917927	-0.012092	0.621968	0.100367	0.119690
num_casos_grueba_pcr	0.917927	1.000000	-0.013171	0.371216	0.116053	0.137927
num_casos_grueba_test_ac	-0.012092	-0.013171	1.000000	-0.016049	-0.002524	0.006373
num_casos_grueba_ag	0.621968	0.371216	-0.016049	1.000000	-0.011609	0.002528
num_casos_p_rueba_elisa	0.100367	0.116053	-0.002524	-0.011609	1.000000	0.122816
num_casos_grueb_a_desconocida	0.119690	0.137927	0.006373	-0.002528	0.122816	1.000000

Propia

```
plt.figure(figsize=(16,10))
sns.heatmap(dataframe.corr(),annot=True,cmap = 'Greens')
```



**FIGURA 5.**  
Diagrama de Correlación  
(Propia)

6) Preparación de la Data

Puesto que es un modelo de regresión lineal, del dataframe se extraen, en X una característica de entrada o variable independiente, en y se guarda la variable dependiente, salida o destino (número de casos) tal como se aprecia en las tablas 4, 5.

```
X = dataframe[['num_casos_prueba_pcr']]
y = dataframe[['num_casos']]
y.head()
```

TABLA 4.  
Número de Casos

num_casos	
0	1
1	0
2	0
3	0
4	2

Propia

```
X.head()
```

TABLA 5.  
Características de Entrada

num_casos_prueba_pcr	
	107
	26
	29
	5
	1

Propia

7) Dividir la data en datos para entrenamiento (train) y datos para pruebas (test).

TABLA 6.  
Casos y pruebas para entrenamiento

17965	144
11387	26
20563	47
11574	5
7484	1

Propia

```
X_train.head()
```

TABLA 7.  
Casos y pruebas para entrenamiento

num_casos_prueba_pcr	
17965	107
11387	26
20563	29
11574	5
7484	1

Propia

8) Creación, entrenamiento y predicción con el modelo de regresión lineal.

```
modelo = linear_model.LinearRegression()
```

```
modelo.fit(X_train,y_train)
```

```
y_test_pred = modelo.predict(X_test)
```

9) Aplicación de las métricas.

```
mse=mean_squared_error(y_test, y_test_pred)
```

```
r2 = r2_score(y_test,y_test_pred)
```

Luego se genera un diagrama de densidad con los números de casos según Figura 6.

```
fig, ax = plt.subplots(figsize=(16,7))
```

```
sns.distplot(df.iloc[:,2])
```

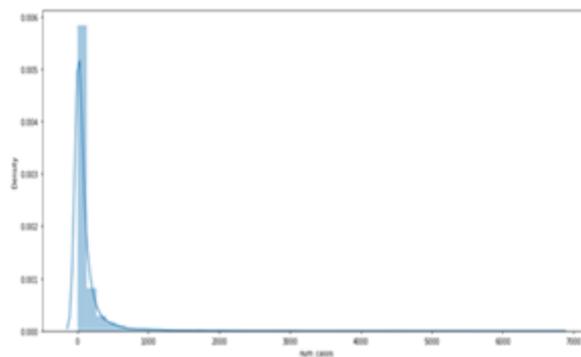


FIGURA 6.  
Diagrama de Densidad  
(Propia)

### Creación del Modelo Random Forest

Para este algoritmo, se toman los datos del dataframe según el código en referencia.

```
X
=
df[['num_casos','num_casos_prueba_test_ac','num_casos_prueba_ag','num_casos_prueba_elisa','num_casos_prueba_d
y = df[['num_casos_prueba_pcr']].values
y = y.reshape(-1,1)
modelo = RandomForestClassifier(n_estimators=40)
modelo.fit(X_train, y_train)
print(modelo.score(X_test,y_test))
y_predicted =modelo.predict(X_train)
```

```
accuracy_score(y_train, y_predicted)
0.6087585957292798
y_predicted =model.predict(X_test)
accuracy_score(y_test, y_predicted)
0.5765983112183354
```

## RESULTADOS

-Como resultado de la predicción por el modelo de regresión lineal usando la respectiva data se obtuvo  $m = \text{modelo.coef\_}$ ,  $b = \text{modelo.intercept\_}$  cuyos valores fueron para el coeficiente es  $m = 1.139$  y para el término independiente es  $b = 7.524$ . Por lo que la función objetivo hallada  $f(x)$  o  $y = mX + b$  quedaría representada por  $y = 1.14 X + 7.52$ .

-De las métricas relacionadas con el modelo de regresión lineal usadas para la predicción se obtuvo el error cuadrático medio de 8948.91 y una varianza de  $r^2 = 0.923014$  con las respectivas funciones.

-La precisión (accuracy) del modelo Random Forest, con los datos de entrenamiento (train) es del 60.9% y con los datos de prueba (test) es del 57.5%.

## CONCLUSIONES

-La confiabilidad de los modelos creados es muy importante, con el modelo random forest se calculó la precisión tanto con los datos de entrenamiento y datos de prueba en el caso del accuracy con train y test tiene una diferencia del 3% aproximadamente por lo que se puede concluir que el modelo está generalizando muy bien y se lo puede dar por bueno.

-La predicción de la efectividad de las pruebas rápidas con el modelo de regresión lineal ha dado un error aceptable con una varianza del 0.92 muy cercano a 1 lo que indica que la predicción es muy buena. Por otra parte, el modelo de Random Forest ha determinado que aproximadamente entre el 58% y 61% de los pacientes con pruebas rápidas han sido clasificados adecuadamente.

## REFERENCIAS BIBLIOGRÁFICAS

- A Study of Real World Data Visualization of COVID-19 dataset using Python. (2020). *International Journal of Management and Humanities*, 4(8). <https://doi.org/10.35940/ijmh.h0834.044820>
- Ahmad, S., Ali, N., Kausar, M., Misbah, H., & Wahid, A. (2020). Road toward rapid-molecular point of care test to detect novel SARS-coronavirus 2019 (COVID-19): Review from updated literature. *Allergologia et Immunopathologia*, Vol. 48. <https://doi.org/10.1016/j.aller.2020.06.001>
- Amin, S., Uddin, M. I., Al-Baity, H. H., Zeb, M. A., & Khan, M. A. (2021). Machine learning approach for COVID-19 detection on twitter. *Computers, Materials and Continua*, 68(2). <https://doi.org/10.32604/cmc.2021.016896>
- Balli, S. (2021). Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. *Chaos, Solitons and Fractals*, 142. <https://doi.org/10.1016/j.chaos.2020.110512>
- Chanchí Golondrino, G. E., Campo Muñoz, W. Y., & Sierra Martínez, L. M. (2020). Aplicación de la regresión polinomial para la caracterización de la curva del COVID-19, mediante técnicas de machine learning. *Investigación e Innovación En Ingenierías*, 8(2). <https://doi.org/10.17081/invinno.8.2.4103>
- Elshazli, R. M., Toraih, E. A., Elgaml, A., El-Mowafy, M., El-Mesery, M., Amin, M. N., ... Kandil, E. (2020). Diagnostic and prognostic value of hematological and immunological markers in COVID-19 infection: A meta-analysis of 6320 patients. *PLoS ONE*, 15(8 August). <https://doi.org/10.1371/journal.pone.0238160>

- Guíñez-Coelho, M. (2020). Impacto del COVID-19 (SARS-CoV-2) a Nivel Mundial, Implicancias y Medidas Preventivas en la Práctica Dental y sus Consecuencias Psicológicas en los Pacientes. *International Journal of Odontostomatology*, 14(3). <https://doi.org/10.4067/s0718-381x2020000300271>
- Latini, A., Agolini, E., Novelli, A., Borgiani, P., Giannini, R., Gravina, P., ... Novelli, G. (2020). COVID - 19 and genetic variants of protein involved in the SARS - CoV - 2 entry into the host cells. *Genes*, 11(9). <https://doi.org/10.3390/genes11091010>
- Lee, H., Park, S. J., Lee, G. R., Kim, J. E., Lee, J. H., Jung, Y., & Nam, E. W. (2020). The relationship between trends in COVID-19 prevalence and traffic levels in South Korea. *International Journal of Infectious Diseases*, 96. <https://doi.org/10.1016/j.ijid.2020.05.031>
- Li, W., Zhang, B., Lu, J., Liu, S., Chang, Z., Peng, C., ... Chen, J. (2020). Characteristics of Household Transmission of COVID-19. *Clinical Infectious Diseases*, 71(8). <https://doi.org/10.1093/cid/ciaa450>
- Li, Z., Yi, Y., Luo, X., Xiong, N., Liu, Y., Li, S., ... Ye, F. (2020). Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. *Journal of Medical Virology*, 92(9). <https://doi.org/10.1002/jmv.25727>
- Muhammad, L. J., Islam, M. M., Usman, S. S., & Ayon, S. I. (2020). Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients' Recovery. *SN Computer Science*, 1(4). <https://doi.org/10.1007/s42979-020-00216-w>
- Perez, D. P., Bustillos, R. S., Botto-Tobar, M., & Mora, C. M. (2021). X-Ray Images Analysis by Medium Artificial Neural Network. *Ecuadorian Science Journal*, 5(1), 55–60. <https://doi.org/10.46480/esj.5.1.50>
- Pérez, D. P., Bustillos, R. S., Mora, C. M., & Botto-Tobar, M. (2020). Prediction of Covid19 with the use of Random Forests Algorithm and Artificial Neural Networks. *Ecuadorian Science Journal*, 4(2), 101–110. <https://doi.org/10.46480/esj.4.2.41>
- Pourhomayoun, M., & Shakibi, M. (2021). Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health*, 20. <https://doi.org/10.1016/j.smhl.2020.100178>
- Rajendra Santosh, A. B., Krishnamurthy, K., & Reddy Baddam, V. R. (2020). Proposal of research model for the detection of COVID-19 among asymptomatic carriers. *International Archives of Otorhinolaryngology*, 24(3). <https://doi.org/10.1055/s-0040-1712936>
- Rath, S., Tripathy, A., & Tripathy, A. R. (2020). Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, 14(5). <https://doi.org/10.1016/j.dsx.2020.07.045>
- Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of Autoimmunity*, Vol. 109. <https://doi.org/10.1016/j.jaut.2020.102433>
- Sharma, S. K., Lilhore, U. K., Simaiya, S., & Trivedi, N. K. (2021). An improved random forest algorithm for predicting the COVID-19 pandemic patient health. *Annals of the Romanian Society for Cell Biology*, 25(1).
- Srivatsan, R., Indi, P. N., Agrahari, S., Menon, S., & Ashok, S. D. (2020). Machine learning based prognostic model and mobile application software platform for predicting infection susceptibility of COVID-19 using healthcare data. *Research on Biomedical Engineering*. <https://doi.org/10.1007/s42600-020-00103-6>
- Valencia Portillo, R. T., Amorín Uscata, B., Gonzales-Zubiate, F. A., Juscamaíta Medina, K., Sevillano, O. R., & Ramos-Sanchez, E. M. (2020). Pruebas rápidas para COVID-19, la mejor alternativa para Ecuador. *Bionatura*, 5(3). <https://doi.org/10.21931/rb/2020.05.03.21>
- Vinod, D. N., & Prabakaran, S. R. S. (2020). Data science and the role of Artificial Intelligence in achieving the fast diagnosis of Covid-19. *Chaos, Solitons and Fractals*, 140. <https://doi.org/10.1016/j.chaos.2020.110182>
- Wiguna, W., & Riana, D. (2020). DIAGNOSIS OF CORONAVIRUS DISEASE 2019 (COVID-19) SURVEILLANCE USING C4.5 ALGORITHM. *Jurnal Pilar Nusa Mandiri*, 16(1). <https://doi.org/10.33480/pilar.v16i1.1293>
- Wu, Y. C., Chen, C. S., & Chan, Y. J. (2020). The outbreak of COVID-19: An overview. *Journal of the Chinese Medical Association*, 83(3), 217–220. <https://doi.org/10.1097/JCMA.0000000000000270>

Yoo, S. H., Geng, H., Chiu, T. L., Yu, S. K., Cho, D. C., Heo, J., ... Lee, H. (2020). Deep Learning-Based Decision-Tree Classifier for COVID-19 Diagnosis From Chest X-ray Imaging. *Frontiers in Medicine*, 7. <https://doi.org/10.3389/fmed.2020.00427>