



ANÁLISIS PREDICTIVO DE COVID-19

EQUIPO DE INVESTIGACIÓN

DE LA CRUZ CRUZ J. MURILLO VALDÉS C.A. VILLEGAS RINCÓN L.R.



Introducción

EL COVID-19 es una enfermedad por coronavirus causada por el virus SARS-CoV-2. Actualmente la transmisión del virus ha sido tan extensa que se ha declarado como una pandemia provocando una situación crítica de salud.

El presente trabajo de investigación pretende analizar la tendencia actual de los casos de COVID-19 de manera que, basándonos en los datos analizados obtengamos resultados a nivel mundial y regional para predecir la tendencia futura.

Determinar las probabilidades de los estados de contagio, muerte y recuperación de COVID-19 a nivel mundial y regional. Basándonos en los resultados obtenidos de cada estado de COVID-19 generaremos una predicción de la tendencia futura que puede presentar cada estado.

Recursos

Los datos COVID-19 proceden de bases de datos proporcionados por diversos investigadores de la plataforma de Kaggle. Los datos han sido manipulados por medio del lenguaje de programación Python en conjunto con librerías internas y externas para poder representar los resultados gráficamente.



Metodología

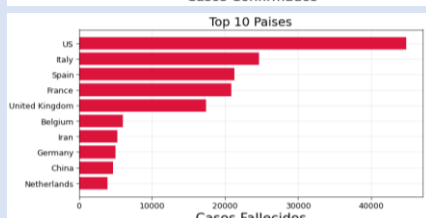
Como proceso en nuestro de plan de trabajo se realizó una limpieza de datos en nuestros dataframes (bases de datos), es decir, la información que se utilizó se sometió a un proceso de eliminación de variables no necesarias, así como el ordenamiento del mismo para una mejor manipulación. Para obtener los resultados como las sumas totales de casos confirmados, muertes y recuperados se crearon funciones para obtener los valores de interés. Por otro lado también se realizó la creación de algoritmos que permitieron la creación de nuevos dataframes para una mejor manipulación de la información mundial y regional.

Datos preliminares mundiales

Los siguientes datos muestran los casos acumulados hasta la fecha del 21 de Abril de 2020.

- Total de casos confirmados: 2561043
- Total de casos de muertes: 176983
- Total de casos recuperados: 679819

Datos preliminares regionales



Como se puede observar en las gráficas los resultados son distintos en cada una de ellas. Estos resultados nos proporcionan la información necesaria para realizar nuestras predicciones ya que representan los casos acumulados de COVID-19. Tomando como referencia la información obtenida en las graficas a nivel regional se determinó analizar al país de US debido a que actualmente es el país con más casos confirmados por lo que la predicción tendrá una mayor precisión.

El **método** empleado para la predicción es la **inferencia bayesiana**, es un tipo de inferencia estadística en la que las evidencias u observaciones se emplean para actualizar o inferir la probabilidad de que una hipótesis pueda ser cierta. En este caso se utilizó para obtener las predicciones de los casos de COVID-19, ya que permite alcanzar conclusiones más sólidas de los datos.

La predicción se obtendrá por inferencia bayesiana a través de un remuestreo aleatorio de 1000 iteraciones con la cual se obtendrá las medias de las frecuencias y tendencias históricas, con una incertidumbre de tendencia de .05% y una flexibilidad de tendencia del 10%.

Agradecimientos

Sudalai Raj Kumar, Data Scientist at H2O.ai and Vansh Jatana, Student at SMR University, Data Scientist Aspirant. Chennai, Tamil Nadu, India.

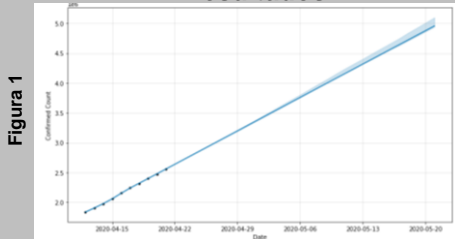
Dataset

Novel Corona Virus 2019 Dataset

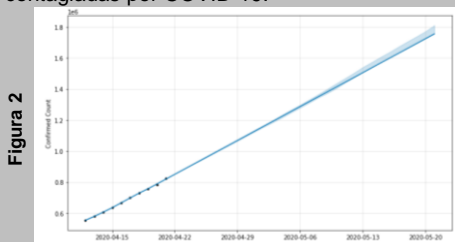
Más información



Resultados



La figura 1 nos muestra una predicción de los casos confirmados mundialmente de 30 días a partir de (21/04/2020) última fecha de registro de nuestra base de datos al (21/05/2020). Sabemos que el total de casos confirmados a nivel mundial al 21/04/2020 es de 2,561,043 de personas contagiadas por COVID-19 y nuestra predicción indica que al 21/05/2020 ascenderán aproximadamente a los 4,952,471 de personas contagiadas por COVID-19.



La figura 2 nos muestra una predicción de los casos confirmados en los Estados Unidos de 30 días a partir de (21/04/2020) última fecha de registro de nuestra base de datos al (21/05/2020) fecha actual. Sabemos que el total de casos confirmados en los Estados Unidos al 21/04/2020 es de 823,786 de personas contagiadas por COVID-19 y nuestra predicción indica que al 21/05/2020 ascenderán aproximadamente a los 1,753,748 de personas contagiadas por COVID-19 en los Estados Unidos.

Conclusión

De acuerdo con los recursos de COVID-19 de Google al 21/05/2020 hay 1.58 millones de casos confirmados en los Estados Unidos y nuestra predicción indica que al 16/05/2020 debieron llegar a 1.59 millones de casos recuperados. Por lo que podemos concluir que al menos en el análisis predictivo (regional) de Estados Unidos el índice de contagio al 21/04/2020 último día de nuestra base de datos y fecha inicial de nuestra predicción ha disminuido. Los casos restantes a nivel mundial y regional son mostrados en la investigación completa debido a que la comparación de los resultados de nuestra predicción con los datos reales es similar a comparación de la conclusión de Estados Unidos antes mencionado. Esta investigación cuenta con las bases suficientes para poder hacer predicciones de cualquier país e inclusive región del mundo. También cuenta con la normalización de los datos adecuada, para poder realizar estudios basados en cálculos de tendencia central. Como recomendación para poder realizar predicciones con un menor grado de incertidumbre es necesario crear un algoritmo que calcule el índice de los estados de contagio diario para posteriormente ser usado en la predicción.