|  |  |
| --- | --- |
| A picture of a winding road and trees  Proyecto final  Minería de datos | Sobre el trabajo  Este documento muestra la investigación, recolección y análisis de datos sobre COVID-19.  Luis Vásquez Benjamyn Recinos  Wilson Nájera |

Índice

Introducción

**Descripción del problema**

El proyecto consta del análisis de una base de datos sobre contagios de COVID-19 a nivel mundial.

El análisis está enfocado al incremento de muertes por covid-19 con relación al tiempo transcurrido, esto quiere decir que, con la base de datos y con algoritmos como k-vecinos se podrá comparar el número de muertes con las fechas que se programen en el algoritmo, para luego ser comparadas con una fecha distinta, poder hacer un análisis de incremento o disminución de muertes. Así como la creación de gráficas sobre las muertes sobre diferentes fechas establecidas.

Creación e interpretación de un árbol de decisiones que permita de manera clara y gráfica la información sobre muertes en algún determinado país o región. Para esto es necesario utilizar el algoritmo de árbol de decisiones.

Desarrollo de un modelo de predicción que permita obtener un dato aproximado de muerte en un día en específico. El algoritmo para utilizar es k-vecinos cercanos.

El archivo necesario para la construcción de este proyecto es el de la base de datos de covid19 a nivel mundial del 2020. Esta base de datos cuenta con los datos necesarios para trabajar la comparación de las muertes correspondientes a fechas.

La entrega consta de lo siguiente:

Documentación: Se entregará un documento con las especificaciones del proyecto

Capturas: Capturas de la implementación de los algoritmos de análisis de datos sobre la base de datos anteriormente mencionada.

**Objetivos**

General

Analizar la estadística de tasa de mortalidad por covid-19 con relación a diferentes fechas por cada país durante con datos recabados en el año 2020.

Específicos

* Crear una gráfica con relación a número de muertes en diferentes fechas.
* Utilizar el algoritmo k-vecinos cercanos para comprobar la relación entre el número de muertes de una fecha establecida por países.
* Utilizar herramientas de análisis de datos para formular conclusiones sobre los datos recabados.

Antecedentes

Herramientas a utilizar

* RapidMiner: RapidMiner es un software dirigido hacia el análisis de datos, con el cual se podrá implementar cada uno de los algoritmos necesarios para la construcción del proyecto.
* Python: Python es un lenguaje con el cual se puede hacer Data Minning sobre una base de datos. Esta herramienta servirá para la creación y ploteo de datos sobre muertes a causa del covid-19.

Conclusiones

Recomendaciones

Anexos