



Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico Matemáticas



Minería de Datos

Prof. Mayra Cristina Berrones Reyes

Avance I. Proyecto Integrador

Grupo: 03

Equipo: 04

Matricula: Integrante:

1821952	Alemán Alvarado Anahí
1819032	Gauna Rodríguez Karyme Mayela
1795134	Sánchez Olvera Ricardo Zarek

San Nicolás de los Garza a 30 de octubre del 2020

1. Título de la base de datos

- a) Nombre: Novel Corona Virus 2019 Dataset
- b) <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>

2. Descripción de los datos

- a) Tipos de datos: Tablas con los datos organizados y gráficas
- b) Descripción de las columnas: "Resumen", "Ubicación", "País", "Género", "Edad", "Inicio de síntomas", "Ciudad", "Provincia", "Latitud", "Longitud", "Resolución geográfica", "Fecha de observación", "Ultima actualización", "Número acumulado de casos confirmados", "Número acumulado de muertes", "Número acumulado de recuperados"

3. Justificación del uso de datos

- a) Una de las principales razones para trabajar con esta base de datos es que maneja datos acerca de un tema que para el día de hoy sigue siendo algo nuevo día con día, el covid-19 es una enfermedad que muestra diversos comportamientos y la misma base de datos nos da una muestra de características y conceptos que nos dan la oportunidad de pensar en más de una forma para aplicar la minería de datos por lo que la hizo una base muy accesible para nosotros.
- b) El mayor beneficio que podemos obtener de esta base es encontrar datos específicos y verídicos de los primeros casos de covid-19 en el mundo, tanto fechas como características, que nos ayudarán a cumplir nuestro objetivo.

4. Planteamiento del problema

Actualmente no existe un programa que busque de manera eficiente los casos de personas recuperadas y de muertes por covid-19 en el país de China.

5. Objetivo final

Objetivo principal: Realizar un programa que nos otorgue de manera efectiva los casos de recuperación y las muertes por covid-19 en China de acuerdo a su edad y género.

6. Planeación de la herramienta a utilizar

- a) Las técnicas a utilizar son:
 - Clasificación: Para poder mostrar los datos clasificados de acuerdo al género y por edad.
 - Regresión: Al tomar el histórico de datos se puede minimizar el error predictivo según la cantidad de datos tomados, y así hacer una predicción mas exacta