**UNIVERSIDAD NACIONAL DE COLOMBIA**

**Técnicas De Aprendizaje Estadístico**

**Cristian Londoño Franco**

**Paula Bermeo**

**Sara Carolina Ruiz Echeverry**

**Sebastian Ruiz Melo**

**Accidentalidad en Medellín**

**Juan David Ospina Arango**

**Sede Medellín**

**2021-2**

**Medellín, Antioquia**

**6 / diciembre / 2021**

**Entendimiento del Desarrollado del Trabajo**

En este trabajo, se abordará el problema de predecir la accidentalidad en la ciudad de Medellín, a partir de la historia reciente de los accidentes reportados. El insumo principal de este trabajo son los datos abiertos de incidentes viales que publica la Alcaldía de Medellín en el portal **MeData.**

Para el trabajo se desarrolló un modelo de predicción **(regresión lineal),** para poder visualizar el comportamiento de los datos y poder **tener un valor aproximado de futuros accidentes** en la ciudad de Medellín, también se agrupo los barrios de Medellín, de acuerdo con su accidentalidad, junto a un mapa, lo cual **nos permite poder ver patrones de los accidentes** en la ciudad.

Para presentar el trabajo, se realizó una **aplicación,** que permita seleccionar una ventana de tiempo, predecir la accidentalidad por tipo de accidente utilizando una ventana y visualizar los grupos de barrios en un mapa; junto a un **video promocional** que muestra sus beneficios y su uso.

**Metodología**

Para este usamos un base de datos que tiene registrado los accidentes en la ciudad de Medellín, esta base de datos se obtuvo del portal **MeDeta.**

Ya con la base de datos la **limpiamos en Barrios, Comuna, etc.…** con el **lenguaje de programación R**

**Modelo Predictivo:**

Para el modelo predictivo, decidimos hacer, un modelo de regresión lineal simple usando Rstudio, donde primero convertimos las variables **AÑO y CLASE ACCIDENTE en factor;** Después usamos **la fórmula de regresión lineal simple** con las variables **AÑO y CLASE DE ACCIDENTE,** para cada tipo de accidente (ya que son pocos las diferentes clases de accidentes nos pudimos permitir hacerlo una por una).

**Agrupamiento:**

Para el agrupamiento, creamos un data frame con todos los **barrios y su frecuencia,** luego creamos un data frame donde están agrupados por **barrio, clase de accidente y gravedad**, después creamos un data frame con la **frecuencia de cada combinación.**

**Mapa:**

Para el mapa se usó la biblioteca **shp** con la base de datos extraída de la página **Geo Medellín,** después con **ggplot** se dibujó la ciudad de Medellín y agregamos una capa con los barrios que se necesitan y se puso los datos de frecuencia con un degradado.

**Aplicación:**

La aplicación fue creada, con el lenguaje de Python, con la librería **flask,** la cual usando **templates** de **html** y hace uso de las librerías **panda** y **numpy** para las operaciones de búsqueda; también se usa bootstrap **5** para dar un formato responsive en diferentes plataformas.

**Video:**

Se hizo un video donde una compañera (Paula) explica los beneficios de la aplicación y como se usa.

**Bibliografía**

* **Data Frame:**

[**http://medata.gov.co/dataset/incidentes-viales**](http://medata.gov.co/dataset/incidentes-viales)

[**https://geomedellin-m-medellin.opendata.arcgis.com/datasets/barrio-vereda-2014/explore?location=6.268888%2C-75.595550%2C12.00&showTable=true**](https://geomedellin-m-medellin.opendata.arcgis.com/datasets/barrio-vereda-2014/explore?location=6.268888%2C-75.595550%2C12.00&showTable=true)

* **Limpiar La Base de Datos:**

[**https://rpubs.com/camilamila/limpieza**](https://rpubs.com/camilamila/limpieza)

* **Github:**

[**https://github.com/CristianLF0903/AccidentalidadEnMedellin/**](https://github.com/CristianLF0903/AccidentalidadEnMedellin/)

* **Regresión lineal y predicción:**

[**https://rpubs.com/Joaquin\_AR/254575**](https://rpubs.com/Joaquin_AR/254575)

* **Gráficas y Mapa:**

[**https://www.r-graph-gallery.com/choropleth-map.html**](https://www.r-graph-gallery.com/choropleth-map.html)