

Predección de la accidentalidad en Medellín



Jaime Andrés Molina Correa
Valentina García Velasquez
Felipe Villarreal Piedrahita
Daniel Chanci Restrepo
Ricardo Penaloza Velasquez

Profesor

Juan David Ospina Arango

Curso

Técnicas de Aprendizaje Estadístico

Universidad Nacional de Colombia
Sede Medellín
2020

Introducción

Objetivos

Descripción de los datos

Los datos utilizados se encuentran en las bases de datos de Geomedellín (Portal Geográfico del Municipio de Medellín) y contienen información referente a múltiples accidentes de tránsito, en los que se detalla el tipo de accidente, dónde y cuándo ocurrió. La base de datos utilizada en este trabajo es una unión de las bases de incidentes en 2014, 2015, 2016, 2017 y 2018, se usará este último año para validar los modelos planteados.

La base de datos original cuenta con las siguientes variables:

- X: Componente de coordenada.
- Y: Componente de coordenada.
- OBJECTID: Id de cada incidente.
- RADICADO: Código emitido por la secretaría de movilidad de Medellín.
- FECHA: Fecha del incidente.
- HORA: Hora del incidente.
- DIA: Día del mes en el que ocurre el incidente
- PERIODO: Año del incidente.
- CLASE: Tipo de accidente
- DIRECCION.
- DIRECCION_ENC.
- CBML: Código de ubicación del predio en la ciudad.
- TIPO_GEOCOD: Tipo de
- GRAVEDAD: repercusiones del accidente
- BARRIO.
- COMUNA.
- DISEÑO: clasificación del lugar del accidente.
- DIA_NOMBRE: Día de la semana en el que ocurre el incidente
- MES: Mes del incidente en número (del 1 al 12)
- MES_NOMBRE: Mes del incidente
- X_MAGNAMED: Componente de coordenada
- Y_MAGNAMED: Componente de coordenada
- LONGITUD: Componente de coordenada
- LATITUD: Componente de coordenada

Se deciden eliminar las variables “X”, “Y”, “X_MAGNAMED” y “Y_MAGNAMED” ya que cumplen la misma función que las variables “LONGITUD” y “LATITUD”. También se elimina la variable “RADICADO” ya que sirve para identificar a un respectivo incidente, al igual que la variable “OBJECTID”. Se elimina la variable “MES_NOMBRE” ya que es redundante en la base de datos, debido a que se encuentra la misma información en la variable “MES”. Además ponerle los nombres a partir de la variable “MES” traería un aumento al coste computacional innecesario. También se decide eliminar las variables “DIRECCION”, “DIRECCION_ENC”, “CBML”, “TIPO_GEOCOD” y “HORA” ya que no serán de utilidad en el modelo.

Además se crean otras dos variables “SEMANA” y “FECHA_ESPECIAL”. Para poder analizar el número de accidentes por semana y en días festivos y demás fechas especiales. Por lo tanto, las variables de interés que se usan en este proyecto son las siguientes:

- OBJECTID: Id de cada incidente.
- CLASE: Tipo de accidente

- GRAVEDAD: repercusiones del accidente
- COMUNA.
- BARRIO.
- DISEÑO: clasificación del lugar del accidente.
- LATITUD: Componente de coordenada
- LONGITUD: Componente de coordenada
- FECHA: Fecha del incidente.
- DIA: Día del mes en el que ocurre el incidente
- MES: Mes del incidente en número (del 1 al 12)
- PERIODO: Año del incidente.
- DIA_NOMBRE: Día de la semana en el que ocurre el incidente
- SEMANA.
- FECHA_ESPECIAL.

Definido lo anterior, se realiza una limpieza en la base de datos, ya que hay registros u observaciones que están mal escritos o simplemente no deberían estar. A continuación se explican los cambios realizados.

Depuración

Para COMUNA

La variable COMUNA debe tener las comunas urbanas de Medellín y sus corregimientos.

Se sabe que en Medellín hay 16 comunas urbanas y 5 corregimientos, ambos están compuestos por barrios, por lo que la variable debería tener $16 + 5 = 21$ niveles. Pero al observar los niveles de la variable “COMUNA” en el conjunto de entrenamiento, se evidencian 84 comunas, por lo cual se decide buscar las razones de esto y tratar de corregirlo.

Comunas: Popular, Santa Cruz, Manrique, Aranjuez, Castilla, Doce de Octubre, Robledo, Villa Hermosa, Buenos Aires, La Candelaria, Laureles-Estadio, La América, San Javier, El Poblado, Guayabal, Belén.

Corregimientos: Corregimiento de San Cristóbal, Corregimiento de San Antonio de Prado, Corregimiento de Santa Elena, Corregimiento de Altavista, Corregimiento de San Sebastián de Palmitas.

Análisis descriptivo