



Tecnológico de Monterrey

Reporte preparación de la base de datos (Delivery)

Escuela: Instituto Tecnológico de Estudios Superiores de Monterrey

Materia: Desarrollo de proyectos y análisis de datos

Profesor: Alfredo García Suárez

Nivel Académico: Profesional

Ciudad: Puebla

Autores

Omar Eduardo Pelcastre Reyes

Saúl Jesús Cuervo Méndez

Juan José Lara García

Cristian Marino Gutiérrez Jiménez

Kevin Vergara Lara

Marco Ivan Olalde Gonzalez

A01735985@tec.mx

A01735937@tec.mx

A01736667@tec.mx

A01736337@tec.mx

A01735970@tec.mx

A01733378@tec.mx

A la hora de realizar los análisis de regresión de este archivo, nos encontramos con una problemática bastante grande, la escasez de valores numéricos (sean entero o flotantes), exactamente solo 4 valores numéricos, siendo dos de estos latitud y longitud, los cuales consideramos que sirven realmente poco a la hora de hacer análisis de regresión por como funcionan esos valores, por otro lado, al solo tener 2 valores numéricos realmente usables y que consideramos no tienen ninguna relación entre sí, decidimos que sea el único archivo sobre el cual no haremos un análisis de regresión lineal, pero decidimos hacer análisis de regresión logísticos, esto usando estas dos variables como las independientes, además de uno usando latitud y longitud, siendo estos problemas algo a tener en cuenta a la hora de entender los resultados que se verán posteriormente.

```
Vars_Indep= df[['_latitude', '_longitude']]
Var_Dep= df['garage_blocking']
```

```
Vars_Indep= df[["number_of_available_trolleys"]]
Var_Dep= df['traffic_congestion']
```

```
Vars_Indep= df[["number_of_operators", "number_of_available_trolleys"]]
Var_Dep= df['type_of_vehicle']
```

```
Vars_Indep= df[['number_of_operators', 'number_of_available_trolleys']]
Var_Dep= df['used_traffic_cone']
```

```
Vars_Indep= df[["number_of_operators", "number_of_available_trolleys"]]
Var_Dep= df['accident']
```

Aquí se puede ver las variables que utilizamos para realizar los análisis de regresión, en estos casos en la mayoría las variables independientes son las mismas por lo dicho anteriormente, solo en uno usamos las variables de latitud y longitud, pero como se podrá ver a continuación los resultados de casi ninguno son realmente satisfactorios.

```
Precisión del modelo:
0.0
```

```
Precisión del modelo:
1.0
```

```
Exactitud del modelo:
0.9572649572649573
```

```
Exactitud del modelo:
1.0
```

```
Sensibilidad del modelo:
0.0
```

```
Sensibilidad del modelo:
1.0
```

```
Puntaje F1 del modelo:
0.0
```

```
Puntaje F1 del modelo:
1.0
```

Para ejemplificar lo mal que salieron la mayoría de resultados tenemos a la izquierda el uso de la variable dependiente de si se pusieron conos de tráfico y a la derecha si se produjeron accidentes, pero como se puede observar, en el de los conos de tráfico nos da un resultado de no tener nada de relación, y en el otro nos dice tener una relación perfecta, lo cual es algo que es realmente poco útil ya que una correlación perfecta suele ser algo más idílico que algo realmente real.

[[110 0] [7 0]]	Precisión del modelo: 0.9401709401709402	Exactitud del modelo: 0.9401709401709402
Sensibilidad del modelo: 0.9401709401709402	Puntaje F1 del modelo: 0.9401709401709402	

Por último, tenemos este modelo que pertenece a la variable de congestión de tráfico y a pesar de que los resultados del modelo nos demuestran lo que pareciera ser un modelo realmente bueno y realista, si vemos la matriz de confusión, podemos observar que en realidad la calidad del modelo se debe a la gran cantidad de verdaderos positivos que hay.

En conclusión, considero que para nuestro archivo los resultados realmente son deficientes ya que no nos permiten tener una afirmación concreta, siendo por lo tanto inconclusos, probablemente esto se deba a la gran cantidad de repeticiones en muchas de las variables y a nuestra incapacidad de realizar análisis de regresión que no contengan valores numéricos