**Predicción del número de vehículos registrados en el sistema de tránsito nacional**

El RUNT se define como un sistema de gestión que permite registrar y mantener actualizada, centralizada, autorizada y validada la información sobre los registros de automotores, conductores, licencias de tránsito, empresas de transporte público, infractores, seguros, entre otros. Se trata de un sistema que está siendo actualizado constantemente, girando alrededor de un sistema centralizado de información de tránsito y transporte el cual opera bajo un esquema de colaboración que depende de la información provista, en línea y en tiempo real.

El gran objetivo de este trabajo es la implementación de un modelo que permita predecir el número de vehículos registrados diariamente en el Registro Único Nacional de Tránsito (RUNT), utilizando como insumo principal la base de datos proporcionada para su elaboración, con el nombre de registros\_autos\_entrenamiento, la cual incluye la fecha y el número de registros de vehículos en determinada fecha.

#### **Introducción**

El Registro Único Nacional de Tránsito, es un sistema que permite consolidar y actualizar la información del registro de tránsito, la cual garantiza una alta confiabilidad y transparencia en los procesos. Su idea está concebida como una solución en la que se integran grandes bases de datos con información, permitiendo registrar, validar y autorizar de una forma unificada cada uno de los trámites.

La creación de un modelo que permita predecir el número de vehículos registrados diariamente, puede ser de gran utilidad para el organismo de tránsito de la ciudad, ya que con dicha información se puede replantear futuros mecanismos en la metodología de registros, también obtener datos de gran utilidad, como la cantidad aproximada de vehículos registrados en un determinado año a futuro, lo que podría conducir a la implementación de nuevas medidas de tránsito, también se podría analizar qué días puntuales se ven reflejados cambios drásticos en números de registros en el RUNT.

#### **Justificación de variables**

La base de datos que tenemos como insumo para analizar e implementar modelos predictivos, contiene 2 columnas con las variables **Fecha** dato numérico de fecha y **Unidades** dato entero real.

La estructura de las variables que contiene los datos es la siguiente:

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Se observa que cada mes es variable en su incremento de inscripción de vehículos, sin embargo, en los últimos años la cantidad de registros ha ido disminuyendo parcialmente. Se aprecia que, durante los meses de noviembre, diciembre, enero, febrero y marzo la cantidad de registros es creciente; una hipótesis que explica este hecho podría ser que se solapamiento de las las vacaciones académicas con ingresos extras en los hogares colombianos, pues se espera que la mayoría de las inscripciones corresponden a personas que cumplen la edad requerida para sacar la licencia de conducción. Un acontecimiento para resaltar es que para los años 2012, 2013, 2014 y 2015 existe un crecimiento en el número de registros para el mes de julio que se explicaría a través de la hipótesis anteriormente planteada, sin embargo, este patrón desapareció para los años 2016 y 2017.

#### **Metodología**

Se creo un modelo de predicción de los registros diarios RUNT. Para esto se crean las variables:

* Día de la semana: Al ser una plataforma que depende del horario normal laboral, la cantidad de registros puede variar a lo largo de la semana.
* Época de vacaciones: Se espera que los mayores registros RUNT corresponden a aquellas personas que sacan su licencia de conducción por primera vez, y existe la tendencia de expedirlo recién se cumple la edad requerida. Los estudiantes sacan provecho del tiempo de vacaciones para realizar este procedimiento.
* Festivo: Corresponde si el día es festivo en Colombia o no.

Se evalúa el comportamiento de la predicción para varios modelos, principalmente:

Regresión lineal, regresión de bosques aleatorios, K vecinos más cercanos, Regresión Lasso.

El rendimiento del modelo se evalúo con el R2. Obteniéndose el mas alto para el modelo de regresión de bosques aleatorios. Sin embargo, se sintonizaron dos modelos: Bosque aleatorio y Regresión lineal. El primero porque fue el de mejor rendimiento y el segundo para evaluar su comportamiento con una sintonización de parámetros más detallada por su característica inferencial.

A pesar de que los resultados de sintonización mejoraron el R2 de la regresión lineal, los bosques aleatorios exhiben un mejor comportamiento en casi un 20% mejor que el primero, por lo tanto, el modelo a implementar para la predicción es, Bosques aleatorios.

**Comportamiento del modelo**

Para sintonizar el modelo se crea una malla de posibles valores dada la primera aproximación otorgada de forma automática. Esto se puede observar en notebook, Runt.ipny

La importancia de las variables seleccionadas en el modelo se presenta a continuación:

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Se aprecia que no existe una gran diferencia en el aporte de información de las variables seleccionados, a excepción de la variable vacaciones esto se explica a que este período no fue detallado estrictamente.

La base de datos consta de registros desde el año 2012 al 2017, se realiza un particionamiento para el data set de prueba para los años 2012 hasta 2016 mientras que para el dataset de prueba corresponde al 2017. Al sintonizar los parámetros se obtiene un R2 para la data set de entrenamiento de R2 = 77% y para el data set de prueba se obtiene R2= 69%. Las predicciones para valores futuros se pueden observar en el aplicativo web que se adjunta al final del informe.

**Conclusión**

La existencia de una base de datos con los registros del RUNT es una buena práctica para tomar medidas de control y optimizar el proceso, sin embargo, se cuenta con muy poco detalle en la inscripción de cada registro lo cual limita la creación de variables explicativas que mejoren el rendimiento del modelo. Para trabajos futuros se espera indagar más sobre el fenómeno a predecir, pues se sospecha la existencia de acontecimientos o funcionamiento interno que puede ayudar a predecir mucho la cantidad de registros y que solo se puede tener en cuenta con la ayuda de un experto.

**Referencias**