

**Autor:**

**ANTONIO BRAVO MUÑOZ**

**TRABAJO DE FIN DE MÁSTER**

**Máster en Data Science**

PREDICCIÓN DE LA CARGA DE VEHÍCULOS EN VÍAS URBANAS DE MADRID

**ÍNDICE DE CONTENIDOS**

1. INTRODUCCIÓN AL TRABAJO REALIZADO……………………………………………………………………………………………….
2. DESCRIPCIÓN DE LOS DATOS DE ENTRADA………………………………………………..
3. METODOLOGÍA………………………………………………………………………………………….
4. RESULTADOS…………………………………………………………………………………………….
5. CONCLUSIONES……………………………………………………………………………………….

ANEXO I. GUÍA DE USUARIO DASHBOARD………………………………………………………..

1. INTRODUCCIÓN AL TRABAJO REALIZADO

En este trabajo de fin de Máster se aborda el problema del tráfico en grandes ciudades como Madrid, concretamente el estudio y la predicción de la carga de tráfico en vías urbanas de esta cuidad.

El crecimiento progresivo de población en las grandes urbes desemboca en un mayor número de vehículos en sus calles, algo que conlleva de forma inevitable a un aumento en los niveles de tráfico. Este trabajo surge a partir de esta idea y de los problemas que conlleva a las personas que se ven forzadas a usar sus vehículos para ir a trabajar en lugar de hacer lo en transporte público.

La idea principal del trabajo es partir de un itinerario ya predefinido y fijo, por ejemplo la ruta que hacemos desde casa al trabajo cada día, y conseguir predecir la carga de vehículos que habrá en un momento dado. Esta idea se basa en la intención de optimizar el tiempo que tenemos. En muchos casos, si por la mañana salimos de casa 15 min más tarde, podemos experimentar un retraso en la llegada considerablemente mayor, de hasta 30 o 45 minutos debido al aumento del tráfico.

Para la realización de este proyecto se usan los datos de intensidad de tráfico que el Ayto. de Madrid (Dirección General de Gestión y Vigilancia de la Circulación) pone a disposición de la ciudadanía a través del siguiente [enlace](https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=33cb30c367e78410VgnVCM1000000b205a0aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD).

Las muestras de tráfico son recogidas por una red de Puntos de Medida que el Ayuntamiento ha desplegado por toda la cuidad. Es necesario usar los datos referentes a estos puntos, los cuales pueden encontrarse a través de este otro [enlace](https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=ee941ce6ba6d3410VgnVCM1000000b205a0aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD).

Se han encontrado trabajos relacionados con esta temática en Internet, los cuales han servido para sentar las bases del procedimiento a aplicar. Los más relevantes se citan a continuación.

1. *[Application of data mining techniques for traffic density estimation and prediction. Jithin Raja, Hareesh Bahuleyana, Lelitha Devi Vanajakshia. 11th Transportation Planning and Implementation Methodologies for Developing Countries, 10-12 December 2014, Mumbai, India.](https://cyberleninka.org/article/n/1462742/viewer)*
2. *[Deep Learning-Based Caution Area Traffic Prediction with Automatic Identification System Sensor Data. Kwang-Il Kim and Keon Myung Lee. Department of Computer Science, Chungbuk National University, Cheongju 28644, Korea.](https://www.mdpi.com/1424-8220/18/9/3172/htm)*

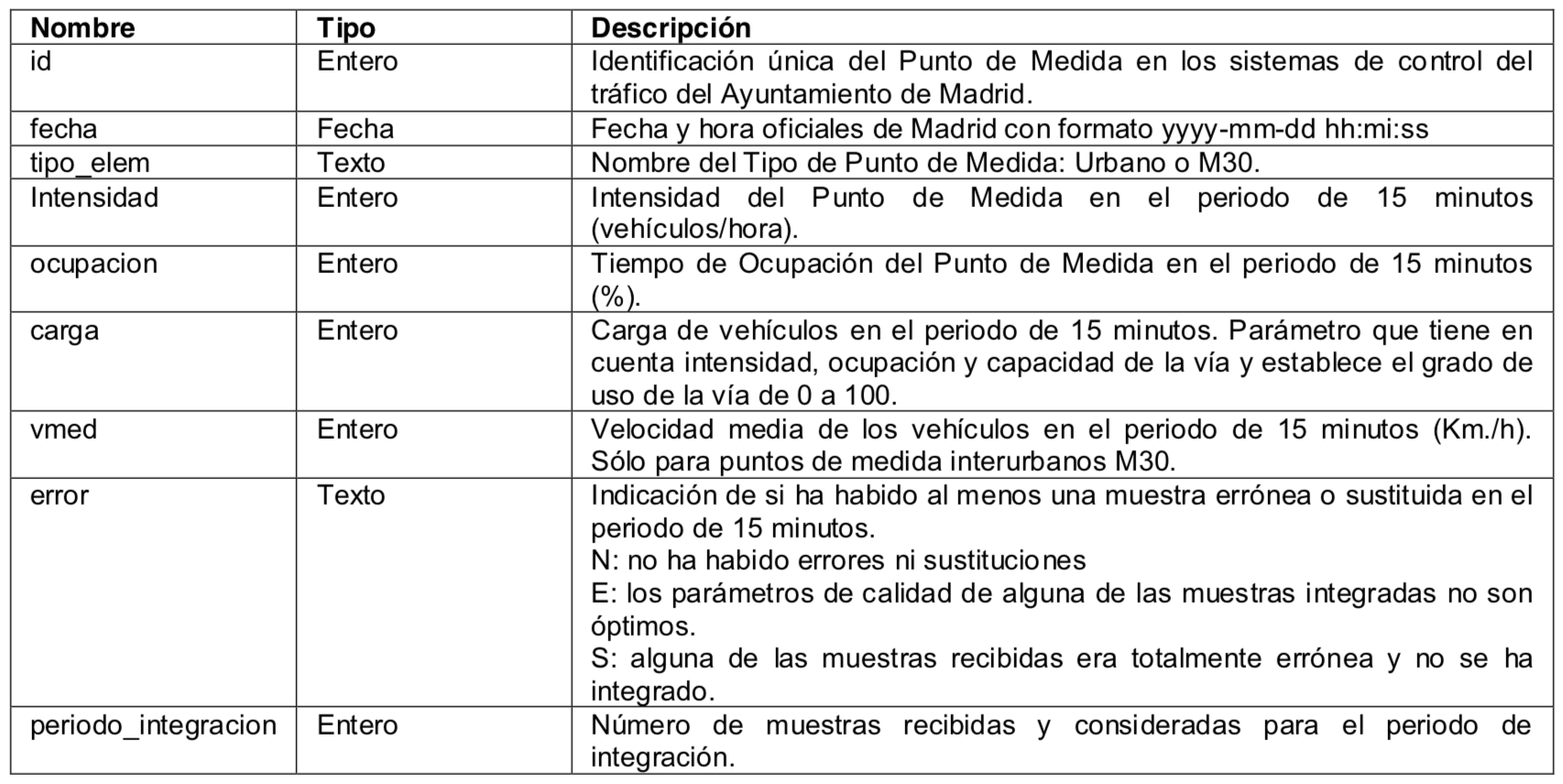
2. DESCRIPCIÓN DE LOS DATOS DE ENTRADA

Los datos usados en este trabajo se clasifican en 2 tipos:

* [Datos Históricos de Intensidad de tráfico desde Julio de 2013 (datos de los puntos de medida).](https://datos.madrid.es/FWProjects/egob/Catalogo/Transporte/Trafico/ficheros/Estructura_DS_Contenido_Trafico_Historico.pdf)
* [Ubicación de los puntos de medida de tráfico.](https://datos.madrid.es/FWProjects/egob/Catalogo/Transporte/Trafico/ficheros/InformacionUbicacionPuntosMedida.pdf)

Para el histórico de datos de la intensidad de tráfico, los datos vienen agrupados por meses desde 2013. Para cada mes se registran lecturas en periodos de 15 minutos de todos los sensores existentes. Los datos se presentan en formato CSV y están disponibles para descargar como archivos comprimidos ZIP.

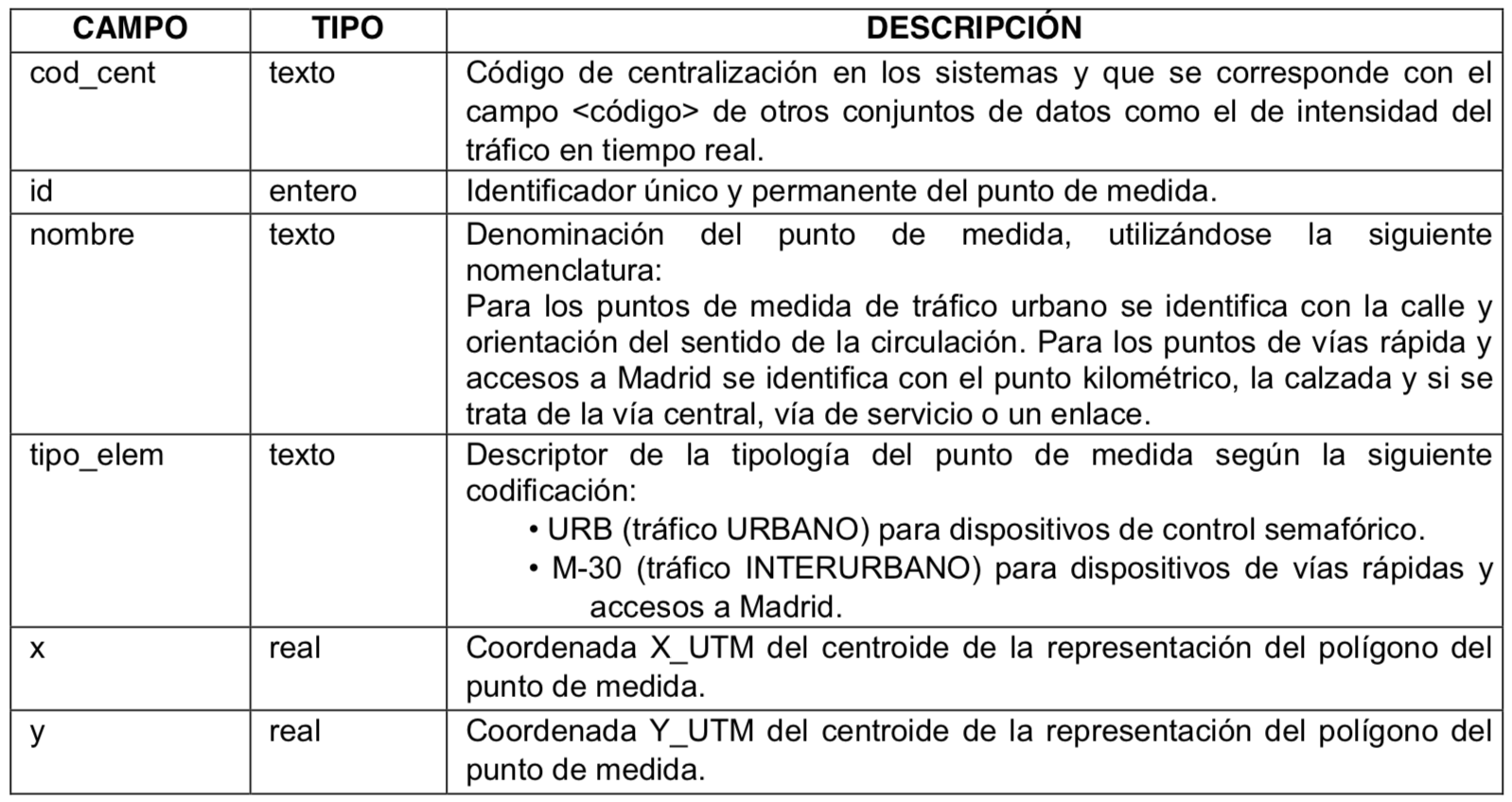
Los datos se estructuran de la siguiente forma:



Para este trabajo, se han usado datos correspondientes a los meses de Enero, Febrero, Marzo, Abril, Mayo, Junio, Julio, Agosto y Septiembre de 2018.

De la misma que los datos de Intensidad del tráfico, tenemos los datos relativos a cada punto de medida. En este caso, en el portal de datos existen varios ficheros, clasificados por meses, con la información relativa a los puntos de medida operativos a final de dicho mes. En nuestro caso, al contar con datos hasta Septiembre, sólo se ha usado el archivo de datos correspondiente. De esta forma nos aseguramos que los puntos de medida sobre los que trabajamos están operativos de acuerdo a la fecha de los datos con los que vamos a trabajar.

Para los puntos de medida, los datos proporcionados tienen la siguiente estructura:



De la misma forma que los datos de intensidad de tráfico, estos datos se presentan para la descarga en formato CSV.

1. METODOLOGÍA

Una vez obtenidos los datos con los que se va a trabajar, en este apartado se describe el tratamiento que se ha aplicado a estos. El proceso parte de la obtención de los datos en crudo hasta la obtención de un modelo de predicción que sea capaz de estimar, con un cierto error, la carga de tráfico en un itinerario que vendrá dado.

**1. Data Acquisition, Preparations and Transformations.**

**(**Notebook: Data Acquisition and Preparation.ipynb)

***1.1. Formato de las variables.***

Transformaciones de variables tipo carácter en tipo fecha y tipo entero y viceversa.

*1.2. Obtención de la Latitud y la Longitud de cada sensor partiendo de las coordenadas UTM X e Y.*

*1.3. Missing Values.*

*1.4. Homogenización de las características tras concatenar varios registros de varios meses.*

*1.5. Tratamiento de registros con información errónea.*

*1.6. Outliers.*

*1.7. Codificación de variables categóricas.*

*1.8. Adecuación de los datos de entrada a algoritmos de ML.*

*1.8. Guardado de los datos procesados.*

***2. Predicción con algoritmos de ML.***

*2.1. Algoritmos adecuados para el tipo de datos usados.*

*2.2. Algoritmos basados en Árboles de decisión.*

*2.3. Random Forest Regressor*

*2.4. Predicción y error*

***3. Front-End. Dashboard interactivo.***

**Esto incluye tanto una descripción genérica del proceso de datos que se ha realizado (sin excesivos detalles porque para los detalles ya está el código escrito), como una descripción del método estadístico o de machine learning que se use.**

1. RESULTADOS

Descripción de los resultados tras el análisis

1. CONCLUSIONES

Conclusiones sobre el trabajo, problemas encontrados, puntos de interés…

Sugerencias de mejora, mejoras para futuros trabajos…