



OPEN DATA AYUNTAMIENTO DE MADRID

M30

GESTION INTELIGENTE DEL TRÁFICO

30 DE NOVIEMBRE DE 2017

MIKEL URANGA
IZASKUN LÓPEZ SAMANIEGO
CRISTINA PALOMO GARO

Contenido

Contenido	2
1 Análisis del contexto.....	3
2 Fuentes de datos	4
2.1 Datos de tráfico históricos	4
2.2 Datos de tráfico en tiempo real.....	4
2.3 Relación de días festivos	5
2.4 Datos metereológicos históricos	5
3 Tratamiento de la información.....	6
4 Modelo de predicción	10
4.1 Modelo 1: SVM.....	10
4.2 Modelo 2: Random Forest	10
4.3 Modelo 3: Serie temporal	11
4.4 Modelo 4: Regresión Múltiple.....	13
5 App.....	15
5.1 Servidor	15
5.2 Procesos	15
5.3 Interfaz App.....	16
6 Bibliografía	19

1 Análisis del contexto

Aunque “Smart City” nos pueda parecer un simple concepto en boga debemos de entender que dicho concepto engloba un movimiento que promueve la sostenibilidad urbana, basándose en el diseño de ciudades y acciones en las mismas que sean capaces de mejorar la calidad de vida de sus habitantes y de responder a las necesidades de los ciudadanos y las empresas.

Dentro del enfoque de las Smart Cities, e incluso anterior, nos encontramos con que la gestión del tráfico en las ciudades se ha convertido en un tema de suma importancia y habitualmente bastante polémico que ha venido a poner de manifiesto la necesidad de ofrecer al ciudadano soluciones en esta materia. Para poder detectar las necesidades reales y valorar el alcance de las mismas desde distintos organismos se ha puesto un fuerte énfasis en incrementar y mejorar la medición y obtención de datos de tráfico¹.

El proyecto M30 GESTIÓN INTELIGENTE DEL TRÁFICO recoge como base estas ideas enfocándolas en una ciudad concreta, MADRID, y dentro de esta ciudad se circunscribe a la vía de circunvalación M30, foco de algunos de los mayores problemas de tráfico de la ciudad.

En las siguientes secciones os mostraremos los pasos que hemos seguido con el objetivo de entender los patrones de tráfico en la M30 para llegar a predecir la carga en la misma a 30 minutos vista y mostrarlo al ciudadano en un entorno abierto. Y como para aportar una solución el primer paso es comprender el problema hemos basado el proyecto en la plataforma Open Data del Ayuntamiento de Madrid, donde disponemos de una ingente cantidad de datos en la que bucear para tratar de acotar correctamente la situación.

A partir de los datos obtenidos de la plataforma Open Data del Ayuntamiento de Madrid se han realizado acciones de extracción y tratamiento de las fuentes de datos existentes sobre el tráfico en la ciudad de Madrid, análisis de las mismas, diseño de modelos predictivos (buscando patrones y armonía) y finalmente la visualización de los resultados que permitan facilitar al usuario su movilidad en este aparente caos del tráfico de la capital.

2 Fuentes de datos

2.1 Datos de tráfico históricos

Tanto el Ayuntamiento de Madrid como el Gobierno de España ponen a nuestra disposición un poblado conjunto de datos con información referente al tráfico en la M-30 desde julio de 2013 en las siguientes direcciones:

<http://datos.madrid.es/sites/v/index.jsp?vgnextoid=33cb30c367e78410VgnVCM1000000b205>

<http://datos.gob.es/es/catalogo/101280796-traffic-intensity-of-traffic-since-july-2013-data-from-measuring-points>

La red de puntos de medida del tráfico del Ayuntamiento de Madrid consta de 3631 espiras de medición² (PM*****, identif) que proporcionan datos cada 15 minutos. En el presente trabajo nos centramos en los puntos de medida correspondientes a M30 y accesos (tipo_elem = M30).

Los datos que proporciona el ayuntamiento y que utilizaremos para los modelos son:

KPI	Unidad de medida
Intensidad	Vehículos/hora
Ocupación	% tiempo de ocupación
Carga	Tiene en cuenta la intensidad, ocupación y capacidad de la vía. Establece un grado de uso entre 0 y 100
Velocidad media	Km/h

2.2 Datos de tráfico en tiempo real

Una vez establecido y entrenado el modelo predictivo, la app se nutrirá de los datos de tráfico en tiempo real, servidos cada 5 minutos en el portal de datos abiertos del ayuntamiento en el siguiente enlace:

<http://datos.madrid.es/sites/v/index.jsp?vgnextoid=d5ec05dc4d1ab410VgnVCM2000000c205a0aRCDR&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCDR>

2.3 Relación de días festivos

Se obtienen los días festivos/inhábiles desde el Boletín Oficial de la Comunidad de Madrid (BOCM).

http://w3.bocm.es/boletin/CM_Orden_BOCM/2016/12/31/BOCM-20161231-5.PDF

2.4 Datos meteorológicos históricos

El portal de datos abiertos de la Agencia Estatal de Meteorología (AEMET) pone a nuestra disposición los valores climatológicos históricos en la dirección:

<https://opendata.aemet.es/centrodedescargas/inicio>

En el presente trabajo nos centraremos en los valores de precipitaciones (mm^3) diarios recogidos por la estación meteorológica de Retiro.

3 Tratamiento de la información

Para poder trabajar con los datos contamos con las mediciones del tráfico de la M30 recogidos cada 15 minutos además de otra información recopilada por el equipo de trabajo para cada uno de los puntos de medida.

El tratamiento y limpieza de los datos disponibles se realizó mediante el lenguaje de programación R³ en el entorno de Rstudio⁴.

El análisis del modelo se realiza sobre los 46 puntos de medida identificados como outliers, eliminando 3 por falta de información en las variables, y otros 46 puntos de medida escogidos con un muestreo aleatorio simple que nos permita tener la información balanceada en el modelo (código: DT_Seleccion_Outliers.Rmd).

La selección de los puntos outliers nos permite tener en cuenta los puntos más críticos de la M30.

Variables incluidas en los modelos:

Origen	Variable	Descripción
Datos abiertos de ayuntamiento de Madrid	Identif	Identificador del punto de medida en la M30
	Ds	Día y hora en la que se ha realizado la medición
	Carga	%carga de la vía
	Intensidad	Intensidad de la vía
	Ocupacion	Volumen de ocupación de la vía
	vmed	Velocidad media de la vía
Información AEMET	Prec	Volumen de precipitaciones en el día en que se realiza la medición
Calendario laboral	Día semana	Lunes/Martes/.../domingo
	Día mes	Número del día en el mes

	Mes	Número del mes en el año
	laborable...festivo...domingo.festivo	Laborable/Festivo/Domingo
	Tipo.de.Festivo	Nacional/autonómico/local
	Festividad	Nombre de la festividad

Para poder acercarnos a un modelo lo más ajustado posible se valora la hipótesis de que lo que ocurre en los 45 minutos previos al momento que estamos buscando tiene impacto sobre la carga de la M30 en dicho momento. Es por esto por lo que se incorpora esta información al data set (cada registro tiene como clave el punto de la M30 y el momento en que se toma la medida) para realizar el modelo de estimación.

El proceso seguido es (codigo: Buscar_fechas_previas.R):

1. Añadir a cada medida de cada punto lo que ocurrió en las medidas anteriores para poder ejecutar el modelo a 15 min, 30 min, 45 min
2. Analizar las variables a incluir en el modelo para trabajar con ellas del modo más óptimo:
 - a. Dimensionalidad: cada variable está en dimensiones diferentes por lo que se analizan las distribuciones de frecuencia para escalarlas con el método que mejor se ajusta a cada variable:
 - i. Carga: se comporta como una distribución uniforme, al estar valorada como un porcentaje la dividiremos entre 100 para estandarizarla
 - ii. Intensidad y ocupación: ambas se comportan como una distribución logarítmica por lo que le aplicaremos una función $\log()$ para escalarla.
 - iii. Velocidad: la distribución más parecida es la Normal aunque no está totalmente centrada, por la derecha la cola es muy corta. Aplicaremos la normalización
 - b. Correlación entre variables: Las variables de carga y velocidad en el tiempo están muy correlacionadas entre ellas por lo que el modelo no trabajaría con ellas de forma óptima. Para evitar esta correlación hemos trabajado con la variación entre dos momentos en un mismo punto tanto para la velocidad como para la carga.

Para poder trabajar con la información se realizan las siguientes transformaciones de la información vista anteriormente y en análisis realizado sobre las mismas:

Origen	Variable	Tranformación
Datos abiertos ayuntamiento de Madrid	vel.med	Media(vmed) por cada punto
	Carga.med	Carga(vmed) por cada punto
	Carga	%carga de la vía = Carga/100
	Intensidad	Log(intensidad)
	Ocupacion	Log(ocupación)
	vmed	Estandarización: (vmed-vel.med)/var(vmed)
Información AEMET	Prec_norm	Normalización: (max(prec)-prec)/(max(prec)-min(prec))
Calendario laboral	diaLunes	Booleana. Si diaSemana='Monday' entonces 1
	diaMartes	Booleana. Si diaSemana=' Tuesday' entonces 1
	diaMiercoles	Booleana. Si diaSemana='Wednesday' entonces 1
	diaJueves	Booleana. Si diaSemana='Thursday' entonces 1
	diaViernes	Booleana. Si diaSemana='Friday' entonces 1
	diaSabado	Booleana. Si diaSemana='Saturday' entonces 1
	diaDomingo	Booleana. Si diaSemana='Sunday' entonces 1
	n.festivo	Integer. Si laborable...festivo...domingo.festivo='laborable' entonces 0 Si laborable...festivo...domingo.festivo='sabado' entonces 1 Si laborable...festivo...domingo.festivo='domingo' entonces 2

		Si laborable...festivo...domingo.festivo=' festivo' entonces 3
--	--	---

4 Modelo de predicción

Tras analizar la información disponible decidimos que las variables con las que podríamos informar al usuario del estado del tráfico en la M30 son la carga y la velocidad. Para ello marcamos como objetivo del proyecto predecir de la manera más fiable la carga y la velocidad media en cada punto de la M30 para 15, 30 y 45 minutos después de la solicitud de información en la app.

Tras el análisis descriptivo entrenamos 4 tipologías de modelos (SVM, Random Forest, Serie temporal y regresión múltiple) y con los resultados de los entrenamientos y posteriores test cuando es posible decidimos cuál es el modelo más óptimo para poner en producción.

En general todos los modelos analizados avalan la hipótesis inicial de la importancia que tiene para la predicción lo que ha ocurrido en los momentos anteriores en el mismo punto de la M30, finalmente nos decantamos por el modelo de carga de Regresión Lineal Múltiple ya que es el que mejor se ajusta a los datos reales y además tiene un procesamiento menos complejo.

4.1 Modelo 1: SVM

El modelo no converge a una solución final por lo que se descarta dicho modelo.

4.2 Modelo 2: Random Forest

Se realiza el training con un random forest de 2 árboles con la librería “caret”⁵ y un random forest de 5 árboles con la librería “randomForest”⁶ los dos casos muestran un tiempo de procesamiento de más de 6 horas (llegando incluso a las 24 horas) y un RMSE superior a la regresión múltiple.

```

> set.seed(125)
>
> rf.M30 <- randomForest(carga ~ .,
+                         data = train.analisis,
+                         ntree = 5,
+                         do.trace = T)
  | Out-of-bag |
Tree | MSE %Var(y) |
1 | 0.005686 10.59 |
2 | 0.005293 9.85 |
3 | 0.005037 9.38 |
4 | 0.004804 8.94 |
5 | 0.004572 8.51 |

> rf.M30

Call:
randomForest(formula = carga ~ ., data = train.analisis, ntree = 5, do.trace = T)
Type of random forest: regression
Number of trees: 5
No. of variables tried at each split: 3

Mean of squared residuals: 0.004572317
% Var explained: 91.49

> error[!is.na(dif), list(error=sum(as.numeric(dif))^2/.N)]
error
1: 0.04818874

> error <- data.table(train.analisis$carga,

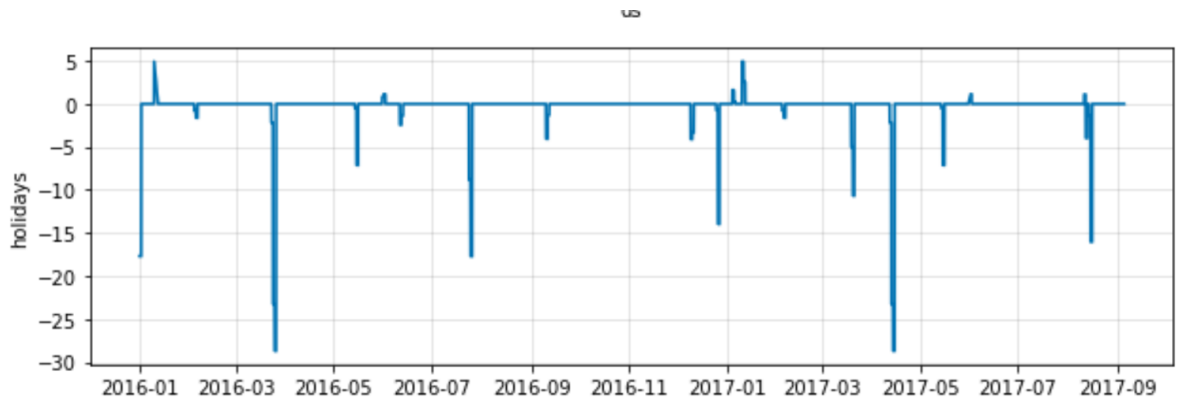
```

4.3 Modelo 3: Serie temporal

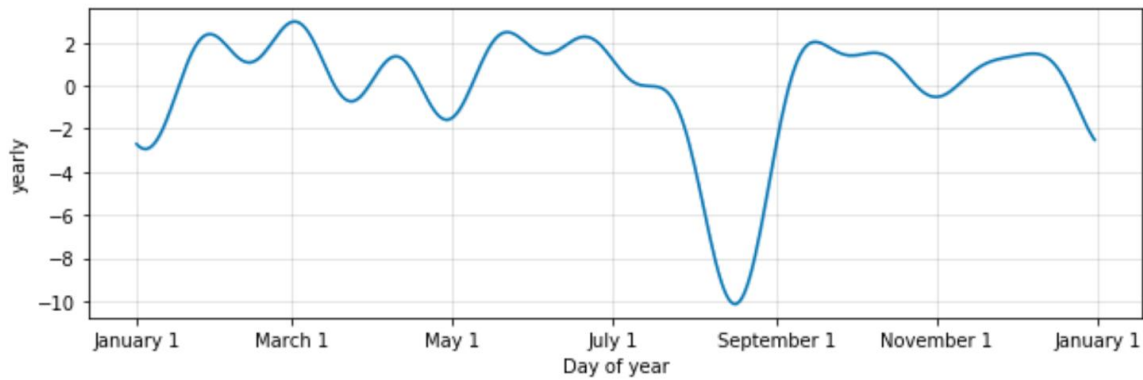
Realizamos una estimación de la serie temporal con la librería “prophet”⁷ que nos permite tener en cuenta el efecto de los festivos sobre la carga de la M30. Este modelo se realizó con Python⁸, porque la librería “prophet” es más completa para este lenguaje de programación, permitiéndonos llegar a la predicción horaria.

El resultado de la serie temporal no muestra unos resultados con un ajuste tan bueno como los obtenidos con la regresión múltiple, pero nos da mucha visibilidad sobre el estado del tráfico en la M30:

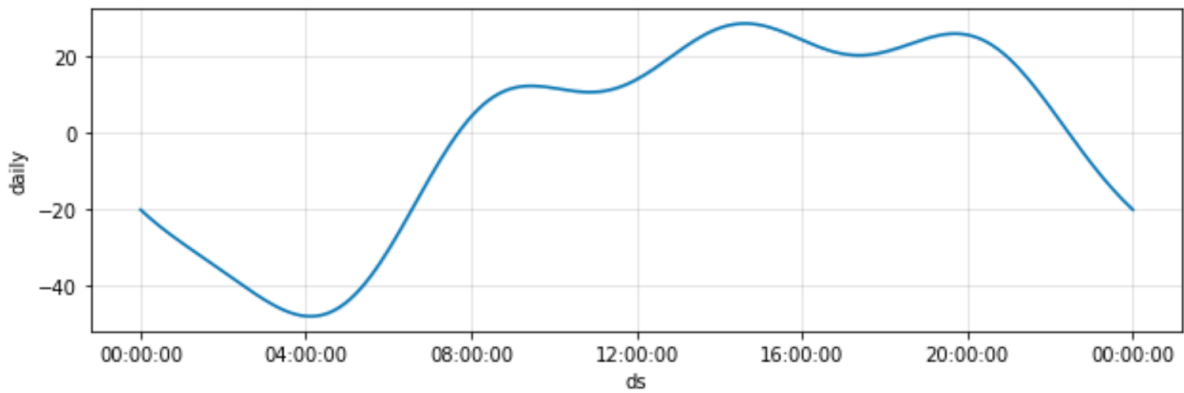
- i. La tendencia de la carga de la M30 en los últimos años es creciente, aunque existen puntos en la M30 donde parece haber descendido.
- ii. El efecto de los festivos hace que se reduzca el tráfico a excepción de los primeros días de enero que lo incrementan:



iii. Se visualiza claramente la bajada de tráfico en las temporadas de verano y navidades:



iv. En cuanto al horario diario podemos ver como hay un incremento a primera hora de la mañana con variaciones incrementales a la hora de la comida y de la vuelta del trabajo. Hay que tener en cuenta que cada punto es en una dirección concreta de la M30 por lo que este mismo gráfico para el punto en sentido contrario mostraría un mayor incremento en las primeras horas de la mañana y una caída posterior.



4.4 Modelo 4: Regresión Múltiple.

Los modelos de regresión lineal múltiple se realizan incluyendo todas las variables comentadas anteriormente, aunque luego se descartaran para cada modelo aquellas que no tengan un peso relevante en la predicción para simplificar el cálculo.

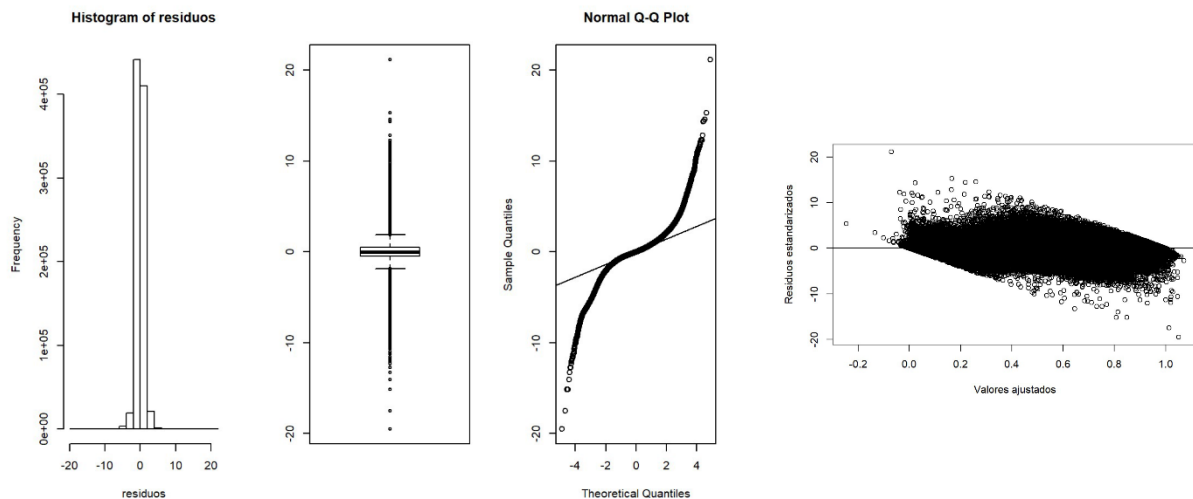
Se realizan tres modelos para predecir la carga (a 15 min, a 30 min y a 45 min) y tres modelos para predecir la velocidad (a 15 min, a 30 min y a 45 min). En un primer momento incluimos como variable predictiva las precipitaciones, pero finalmente se decide eliminar para la primera versión y visualización debido a que no es posible automatizar su extracción diaria gratuita. Eliminar esta variable del análisis no penaliza prácticamente la capacidad de predicción del modelo.

El modelo se entrena con el 70% de la información y posteriormente se testea con el 30% restante. El análisis del R^2 y el RMSE de los modelos nos indica que el modelo de velocidad no está lo suficientemente ajustado como para ponerlo en producción por el momento:

Modelo	Estimación	R^2	RMSE Training	RMSE Test
Carga (con precipitaciones)	15 min	0,9627	$1,76 \cdot e^{-24}$	0,003
	30 min	0,9206	$8,29 \cdot e^{-24}$	0,026
	45 min	0,8755	$3,60 \cdot e^{-23}$	0,042
Carga (sin precipitaciones)	15 min	0,9628	$1,66 \cdot e^{-24}$	0,002
	30 min	0,9207	$1,31 \cdot e^{-23}$	0,033
	45 min	0,8755	$3,53 \cdot e^{-23}$	0,009
Velocidad (con precipitaciones)	15 min	0,8665	63295,12	26444,68
	30 min	0,7933	62885,49	26829,01
	45 min	0,7284	62885,53	27186,51
Velocidad (sin precipitaciones)	15 min	0,8657	62622,68	27321,73
	30 min	0,7936	62937,32	26937,77

45 min	0,7283	63217,99	26673,72
--------	--------	----------	----------

El análisis de los residuos nos muestra cómo se ajustan a una normal, pero la tendencia de la última gráfica nos indica que existe algún efecto (variable) que no hemos podido identificar por el momento pero que afecta al modelo:



En el equipo se manejan dos hipótesis de variables que no hemos podido incluir:

- Los accidentes de tráfico: no hemos podido encontrar el histórico con los datos desagregados por día y localización espacial.
- Qué ha pasado en la última medición en puntos anteriores al que nos encontramos: mejora a implementar en próximos pasos.

5 App

La información se pone a disposición del usuario en la app [M30 Gestión Inteligente del Tráfico](#) donde el usuario final puede consultar información en tiempo real de la carga y velocidad de la M30 y las estimaciones a 30 minutos.

Además de esta información M30 Gestión Inteligente del Tráfico pone a disposición del usuario los datos en tiempo real, la tendencia del tráfico en las últimas 24 horas y un buscador para identificar el estado de la M30 en puntos concretos.

M30 - Gestión Inteligente del Tráfico



Introducción

Gracias a la plataforma Open Data del Ayuntamiento de Madrid podemos extraer y procesar las distintas fuentes de información disponibles sobre el tráfico en la ciudad. De esta forma podemos analizar, diseñar modelos predictivos (buscando patrones y armonía) y, finalmente, visualizar los resultados que ayudarán al usuario a tomar decisiones más informadas en relación a la movilidad en una ciudad un poco más smart.

5.1 Servidor

Se ha configurado un servidor que gestione los procesos necesarios para dar servicio a la app. Las tareas principales que se realizan en el servidor son:

- Recabar los datos de las APIs de los distintos servicios para almacenarlos.
- Tratamiento de los datos para alimentar los modelos
- Ejecución de los modelos necesarios para la app
- Tratamiento de los resultados para generar los ficheros que serán consumidos por la app.

5.2 Procesos

Un proceso Cron⁹ se encarga de lanzar los siguientes scripts cada 15 minutos:

- **get_traffic_data.R** : Llama a la API proporcionada por el ayuntamiento de Madrid para la lectura de los datos de los puntos de medida. Recoge los datos los trata y almacena.
 - Url API: <http://informo.munimadrid.es/informo/tmadrid/pm.xml>
 - Documentación API:
http://datos.madrid.es/FWProjects/egob/contenidos/datasets/ficheros/Transporte_trafico/PUNTOS%20MEDIDA%20TRAFICO_MADRID.pdf
- **get_weather_data.R**: Llama a la API proporcionada por AEMET para la obtención de la predicción meteorológica diaria de Madrid. Recoge los datos los trata y almacena.

- ✓ Url API:

<https://opendata.aemet.es/opendata/api/prediccion/especifica/municipio/horaria/28079>

<https://opendata.aemet.es/opendata/api/prediccion/especifica/municipio/horaria/28079>

- ✓ Documentación API:

<https://opendata.aemet.es/centrodedescargas/docs/FAQs130917.pdf>

- **calculate_prediction_XXXX.R:** Toma los datos almacenados y alimenta los modelos predictivos ya entrenados con ellos. Existe un proceso distinto para cada modelo seleccionado debido a la especificidad del formato de datos que necesitan consumir. Los resultados se almacenan para que puedan ser consumidos por la app.

5.3 Interfaz App

Una aplicación desplegada en Shiny¹⁰ para ofrecer una interfaz gráfica en la que consumir los resultados de los modelos predictivos, así como otros datos de interés:

- **ui.R:** El usuario dispone de distintas opciones para configurar el visualizado de los datos.
- **server.R:** Los datos se cargan mediante los ficheros generados por el proceso Cron del Servidor y se tratan para ofrecerlos de forma correcta al ui.

Mediante esta app, el ciudadano tendrá a su disposición información del tráfico en la M30 de manera actualizada. Las funcionalidades de la misma son:

- **Mapas:** El usuario podrá consultar en el mapa el grado de ocupación en cada punto de medición a partir de un código de colores (rojo, verde y amarillo) y seleccionar según sus necesidades el estado actual o la predicción a 30 minutos vista. Así como las tendencias estimadas.

M30 - Gestión Inteligente del Tráfico

ACTUALIZAR

Cada vez que cambies los ajustes de cualquier de los apartados o selecciones una opción pulsa el botón para reflejar los cambios.

[Actualizar Datos](#)

AJUSTES BÁSICOS

Introduce un lugar de interés:

Ejemplos: Madrid, Monasterio, etc.

MAPA

Selecciona los datos del mapa:

Selecciona tipo de Google Map:

☐ No y Negro?

CONTROL DE VERSIONES

Versión 2.2.2


Desplegado el 29-Nov-2017

[Código en GitHub](#)

AUTOR

Mikel Uranga

<https://github.com/perogrupo>



- **Camaras:** Se incluyen también las imágenes de las cámaras de tráfico situadas en los puntos de medición para facilitar al ciudadano una rápida evaluación visual de la situación.

M30 - Gestión Inteligente del Tráfico

ACTUALIZAR

Cada vez que cambies los ajustes de cualquier de los apartados o selecciones una opción pulsa el botón para reflejar los cambios.

[Actualizar Datos](#)

AJUSTES BÁSICOS

Introduce un lugar de interés:

Ejemplos: Madrid, Monasterio, etc.

MAPA

Selecciona los datos del mapa:

Selecciona tipo de Google Map:

☐ No y Negro?

CONTROL DE VERSIONES

Versión 2.2.2

Desplegado el 29-Nov-2017


[Código en GitHub](#)

AUTOR

Mikel Uranga

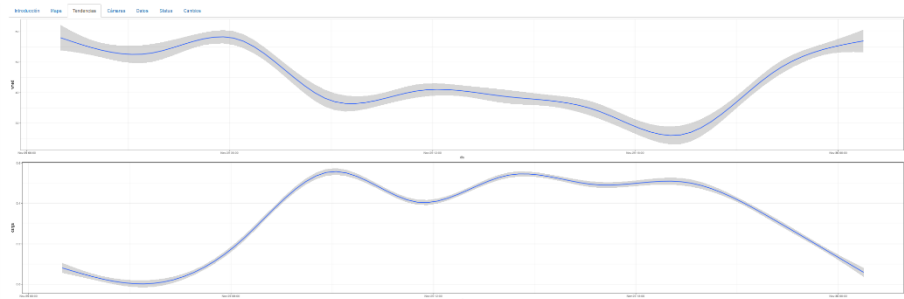
<https://github.com/perogrupo>

Otra Cámara:



- **Tendencias:** Estado del tráfico y la velocidad en las últimas 24 horas

M30 - Gestión Inteligente del Tráfico

[illegible]

- **Datos:** Detalle de la información diaria donde se puede seleccionar un punto concreto de la M30

M30 - Gestión Inteligente del Tráfico

[illegible][illegible]

6 Bibliografía

1 Eleni I. Vlahogianni, Matthew G. Karlaftis, John C. Golias, Short-term traffic forecasting: Where we are and where we're going, In Transportation Research Part C: Emerging Technologies, Volume 43, Part 1, 2014, Pages 3-19, ISSN 0968-090X,

2 http://datos.madrid.es/FWProjects/egob/contenidos/datasets/ficheros/Transporte_trafico/PUNTOS MEDIDA TRAFICO_MADRID.pdf

3 R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>

4 RStudio (2012). RStudio: Integrated development environment for R (Version 0.96.122) [Computer software]. Boston, MA. Retrieved May 20, 2012.
Available from <http://www.rstudio.org/>

5 Max Kuhn. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer and Allan Engelhardt (2012). caret: Classification and Regression Training. R package version 5.15-044. <http://CRAN.R-project.org/package=caret>

6 A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18–22.

7 S. J. Taylor and B. Letham. Forecasting at Scale. Facebook Research. 2017. https://facebookincubator.github.io/prophet/static/prophet_paper_20170113.pdf

8 G. van Rossum, Python tutorial, Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995. Python Software Foundation. Python Language Reference, version 2.7. Available at <http://www.python.org>

9 [https://es.wikipedia.org/wiki/Cron_\(Unix\)](https://es.wikipedia.org/wiki/Cron_(Unix))

10 Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2017). shiny: Web Application Framework for R. R package version 1.0.5. <https://CRAN.R-project.org/package=shiny>