

Ciencia de Datos y Reportes de Movilidad Google para Modelizar la Demanda de Combustible

Irma Noemí No, Julián E. Tornillo, Guadalupe Pascal, Leandro Rabbione

Instituto de Investigaciones en Ingeniería Industrial - I4,
Facultad de Ingeniería, Universidad Nacional de Lomas de Zamora
(1832) Buenos Aires, Argentina
{ino, jtornillo, gpascal}@ingenieria.unlz.edu.ar,
leandro.rabbione@gmail.com

Resumen. La reciente situación de pandemia mundial impulsó la generación de reportes abiertos de movilidad, iniciativa creada por la empresa Google en apoyo a las políticas sanitarias asociadas al COVID-19. El cambio significativo en el flujo de vehículos durante la situación de pandemia y la variación en el consumo de combustible asociado al transporte y a diversas actividades productivas, requirió la creación de nuevos modelos predictivos relacionados con un conjunto de datos inusuales (por ejemplo, la geolocalización de los conductores). La manipulación y el análisis adecuado de estos datos proporcionan un pronóstico que mejora la previsión de la demanda de combustible asociada al consumo real. En este trabajo analizamos las bases de datos de la venta de combustibles (nafta y diésel), disponibles y abiertas en sitios web oficiales e información de la empresa YPF. Los resultados muestran una correlación positiva entre las variables relacionadas a la demanda de estos combustibles y los registros de movilidad de Google, con ciertas particularidades. El lenguaje de programación utilizado para el desarrollo del código de visualización, geoestadística, cálculo predictivo y reportes de la investigación es “R”.

Palabras claves: Demanda de combustible, Movilidad Google, Series de tiempo, Transporte y Logística, Aprendizaje Automático.

1 Introducción

La empresa YPF y su departamento de analítica avanzada [1], ante el cambio en el comportamiento del mercado interno de combustibles relacionado con una logística restringida impuesta por el decreto 260/2020 de Aislamiento Social y Preventivo y Obligatorio (ASPO) en Argentina [2], rápidamente observaron un importante desajuste entre el modelo predictivo de regresión [3] (tradicionalmente utilizado por la empresa para estimar la demanda de combustible) y la realidad vigente en el mercado interno de transportación. Los indicadores utilizados por la fórmula deducida a través de series temporales históricas (con uso del utilitario SAS [4]), basadas en variables como la cantidad de Estaciones de servicio por habitante, el ingreso per cápita y la

cantidad de vehículos por habitante, entre otras [5], dejaban de tener el peso predictivo esperado y era necesaria la generación de un nuevo esquema que pudiera anticipar la demanda de combustible por actividades logísticas y productivas, con un margen de aproximación aceptable dado el contexto de transportación restringido por las medidas sanitarias vigentes. La creación de un escenario predictivo adecuado requirió hallar un nuevo conjunto de datos de entrada, que pudieran representar la realidad atípica generada por la nueva situación de circulación restringida. Luego de ensayar diferentes fuentes de datos asociados a la logística de transporte (órdenes de compra, ventas, permisos de circulación) se distingue a la movilidad (rastreado por geolocalización) como el input más ajustado a la demanda de combustible por transportación en situación de ASPO (ver Figura 1.).



Fig. 1. Ajuste temporal "Demanda de combustible - Movilidad Google". (YPF – 2020).

Estos resultados pueden validarse y ampliarse mediante el uso de bases de datos oficiales argentinas disponibles en abierto [6], realizando los recortes correspondientes a nuevos criterios, productos o empresas de interés, considerando los datos de los Reportes de Movilidad Google [7] como una componente principal del modelo predictivo.

2 Metodología y Resultados

En base a los antecedentes mencionados se desarrolla un proyecto de investigación sobre la siguiente hipótesis principal: *La demanda regional de combustible en Argentina puede modelizarse predictivamente mediante técnicas de aprendizaje automático con indicadores de geo-movilidad sobre la logística terrestre.* La metodología implementada es de tipo cuantitativa con manipulación de bases de datos reales, no experimentales, con énfasis en las fases empírica, analítica y de difusión. Se realizan

tareas de exploración de datos en bases oficiales argentinas [6] y abiertas de Google [7], manipulación de modelos de aprendizaje automático en el entorno R, explotación de la información correspondiente al modelo desarrollado por YPF para predecir la demanda de combustible en situación ASPO, pre y post-Covid, finalizando con la generación de informes y publicaciones creados con Rmarkdown desde el entorno de desarrollo integrado RStudio.

El filtrado y el manejo estadístico descriptivo de los datos oficiales demuestran que en los meses de abril, mayo y junio del año 2020 se establece un importante salto cuantitativo en los comportamientos de la demanda real de combustible en situación de ASPO. Se observó además una distribución heterogénea con respecto a la diferencia: “venta real versus estimación” entre diversas provincias y departamentos (condicionadas por diferentes políticas regionales de aislamiento).

Los pronósticos estándar creados con las herramientas de Power BI y la librería Prophet de R, permitieron visualizar el desfase entre la estimación de la demanda versus las ventas reales registradas durante la situación de una movilidad restringida.

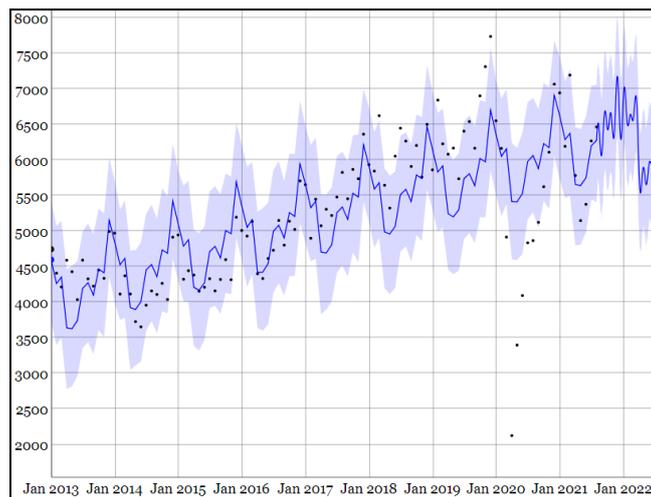


Fig. 2 - Pronóstico con herramienta Prophet de R (azul) y ventas reales de Nafta.

La manipulación de las bases de datos de movilidad Google también requirieron de un gran trabajo de adecuación, compatibilización de denominaciones de Departamentos y Provincias y la generación de valores promedio (por fecha) que posibilitaran su utilización en el pronóstico de la demanda de combustible en Argentina. Hallamos una correlación positiva (de 0.8) entre las variables movilidad Google y la venta de nafta consideradas. Aportamos a continuación una visualización conjunta para la comprensión de este comportamiento de similitud (Figura 3). Con respecto al Gasoil, la correlación entre la variable movilidad y demanda es baja (~ 0.3), una posible interpretación de este valor es la no desagregación del destino final del producto (conside-

rando que las ventas al agro y a la industria se realizan a granel y no se ven entonces correlacionadas fuertemente con la movilidad registrada en los reportes de Google).

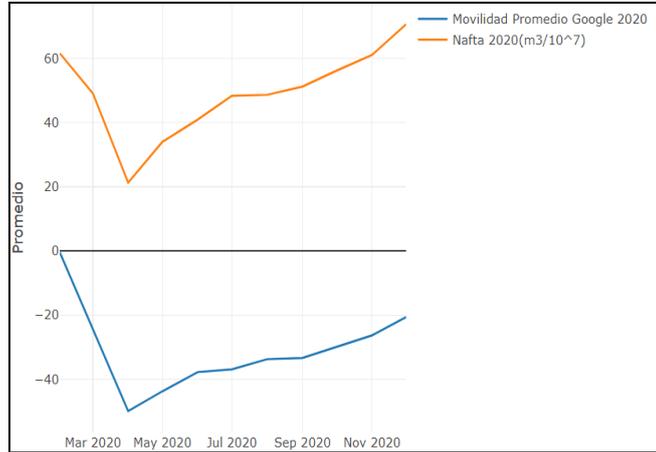


Fig. 3 -Comportamiento de la Movilidad promedio Google y la demanda de Nafta 2020.

3 Conclusiones y Trabajo en Proceso

Verificado el ajuste de la movilidad Google a la demanda de nafta, nos encontramos trabajando en la corroboración de una primera fórmula predictorora que considera la influencia del promedio de movilidad Google de manera diferenciada en los días de semana, fines de semana y feriados (se ha estudiado la incidencia del día de la semana y la demanda de combustible asociada – ver Figura 4). Finalmente se publicará el trabajo completo en abierto [8], sumando además inclusión de mapas (ya codificados en R) para facilitar una visualización regional de las características señaladas en los puntos anteriores.

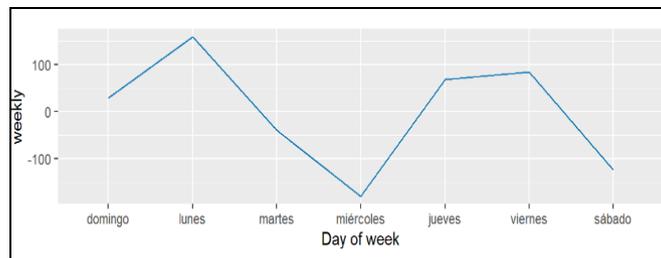


Fig. 4. - Comportamiento de la demanda de Nafta según el día de la semana.

4 Agradecimientos

Agradecemos a la Facultad de Ingeniería y a la Universidad Nacional de Lomas de Zamora por brindarnos los medios y el apoyo para desarrollar nuestras tareas de investigación en los proyectos acreditados de los cuales el presente trabajo es parte, y particularmente a nuestro Director del Programa de Mejora de Procesos FI-UNLZ, Dr. Andrés Redchuk.

Referencias

1. YPF. (2020). YPF-Ruta. Sitio Web. Consultado el 01/06/2022 de: <https://www.ypf.com/productosyservicios/YPF-Ruta/Paginas/home.html>
2. Ministerio de Salud. (2020). Decreto de Necesidad y Urgencia (PEJ). *DNU 260/2020*. Recuperado el 01/06/2022 de: <https://www.argentina.gob.ar/coronavirus/dnu>
3. Hernández F., Mazo M. (2020). Modelos de Regresión con R. Edición digital Bookdown [s.n]. Recuperado el 01/06/2022 de: https://fhernanb.github.io/libro_regresion/index.html
4. SAS. (s.f.). Forecast Server. Sitio Web. Consultado el 01/06/2022 de: https://www.sas.com/es_ar/software/forecast-server.html
5. Coloma, G. (1998). Análisis de comportamiento del mercado argentino de combustibles líquidos. *XXXIII Reunión de la Asociación Argentina de Economía Política*, Mendoza. Recuperado el 01/06/2022 de: <https://aaep.org.ar/anales/works/works1998/COLOMA.PDF>
6. Datos Argentina (s.f.). *Jefatura de Gobierno, Argentina Unida*. Sitio Web. Consultado en el portal de origen (ME -Secretaría de Energía) el 01/06/2022 desde: <https://datos.gob.ar/>
7. Google. (s.f.). Informes de Movilidad. Sitio Web. Consultado el 01/06/2022 de: <https://www.google.com/covid19/mobility/>
8. GitHub Inc. (s.f.). Plataforma de desarrollo colaborativo. Repositorio propio recuperado el 01/06/2022 desde: <https://github.com/irmanoemino/demandacombustible1>