



IIC2115 – Programación como Herramienta para la Ingeniería (II/2018)

Laboratorio 5

Objetivos

- Aplicar lo aprendido en el material de *Álisis de datos* a un problema real de ingeniería

Entrega

- **Lenguaje a utilizar:** Python 3.6
- **Lugar:** GitHub
- **Fecha límite de entrega:** Miércoles 1 de Noviembre, a las 23:59 horas.
- **Formato de entrega:** dos archivos con la solución propuesta, uno en formato **.py** y el otro en **.ipynb**. Deben ser exactamente iguales (el ayudante revisará cualquiera de ellos). Adicionalmente, incluya un archivo **README.md** con información útil para el ayudante corrector.

Introducción

eTrans es una empresa de transporte que desarrolla aplicaciones tecnológicas para ayudar a agencias y empresas operadoras de transporte público en sus procesos de gestión de flota, asistencia de conducción, planificación, operación y comunicación con usuarios. Dentro de sus servicios, se encuentra el análisis en vivo de camiones con información de telemetría. Durante el último tiempo, los camiones han sufrido distintos problemas, tales como: fallas, robo de gasolina, etc. Te han encomendado la misión de analizar la información para determinar en que momentos se robó y además dar de baja algunos sensores que no estén funcionando correctamente.

Datasets

Tienen a su disposición un set de datos, en formato .db que deberán abrir con pandas, que contiene información segundo a segundo sobre el estado de distintos vehículos. El set de datos se ve así:

	gps_utc_time		latitude		longitude		odometer_kms
0	2018-06-06 09:26:00		-33.522903		-70.610610		119422.101562
1	2018-06-06 09:26:00		-33.522903		-70.610610		119422.101562
2	2018-06-06 09:26:03		-33.522895		-70.610354		119422.101562
3	2018-06-06 09:26:06		-33.522891		-70.610134		119422.101562
4	2018-06-06 09:26:09		-33.522922		-70.609849		119422.140625

	fuel_tank_level_pct		speed_kmh		vehicle_license_plate		fleet_id
0	100		0		HKHK12		7
1	100		0		HKHK12		7
2	100		0		HKHK12		7
3	100		0		HKHK12		7
4	100		0		HKHK12		7

- **gps_utc_time:** corresponde al tiempo en que se realizó la medición. *hint: para trabajar con este dato utilicen la librería datetime*
- **latitude:** latitud que registra el gps del vehículo.
- **longitude:** longitud que registra el gps del vehículo.
- **odometer_kms:** el odómetro es el sensor que registra la información de los kilómetros totales recorridos por los vehículos.
- **fuel_tank_level_pct:** cantidad de combustible que lleva el vehículo en el estanque (porcentaje).
- **speed_kmh:** velocidad en kilómetros por hora del vehículo en un instante dado.
- **vehicle_license_plate:** patente del vehículo. Es única y puede utilizarse como id.
- **fleet_id:** id de la flota a la que pertenece el vehículo en cuestión.

Detección de robos

Dentro de tus objetivos está la creación de un detector de robos. Para esto, deberán trabajar con los datos (de la forma que usted desee) y así poder detectar o predecir el momento de un robo en base a nueva información inputada (la información inputada tendrá los mismos tipos de atributos de la base de datos entregada). En base a una nueva serie de tiempo (conjunto de datos) deberán mostrar el rango en que se robó gasolina.

Detección de fallas

Otro objetivo importante es la detección de fallas en el sensor de nivel de combustible. Nuevamente deberá trabajar los datos de la forma que estime conveniente (pero que tenga sentido ya que deberá justificar todo lo realizado) para determinar si durante un periodo de tiempo, un sensor de combustible se encuentra averiado.

Reporte

Deberá entregar la tarea en un archivo **.ipynb** en el cual se encuentre tanto el desarrollo de su trabajo como la justificación de cada una de las decisiones tomadas en él. Deberá incluir:

- Análisis de datos: Deberá mostrar información importante sobre los datasets entregados y justificar su elección y el comportamiento de estos (ruido en los datos, datos extremos, medias, percentiles, etc).
- Detección de robos: Deberá incluir el método utilizado para detectar los robos de gasolina. Además, se espera una justificación fundada de por qué se escogió ese método y por qué funciona (a través de gráficos, datos, etc).
- Detección de fallas: Deberá incluir el método utilizado para detectar las fallas en los sensores que miden el nivel de combustible. Además, se espera una justificación fundamentada de por qué se escogió ese método y por qué funciona (a través de gráficos, datos, etc).

Recursos adicionales

Además del material disponible en el sitio del curso, en este **link** podrán descargar gratuitamente (desde la red de la Universidad) el libro **Data Science for Transport**, que cubre muchos de los tópicos que se tocan en este laboratorio.

Avance parcial

Para la revisión de avance, se espera que las parejas tengan desarrollado al menos lo siguiente:

- Datos cargados en pandas.
- Visualización de datos, justificando por qué las visualizaciones desarrolladas son relevantes.
- Limpieza de los datos.
- Identificación y justificación de variables relevantes.

Política de Integridad Académica

Los alumnos de la Escuela de Ingeniería deben mantener un comportamiento acorde al Código de Honor de la Universidad:

“Como miembro de la comunidad de la Pontificia Universidad Católica de Chile me comprometo a respetar los principios y normativas que la rigen. Asimismo, prometo actuar con rectitud y honestidad en las relaciones con los demás integrantes de la comunidad y en la realización de todo trabajo, particularmente en aquellas actividades vinculadas a la docencia, el aprendizaje y la creación, difusión y transferencia del conocimiento. Además, velaré por la integridad de las personas y cuidaré los bienes de la Universidad.”

En particular, se espera que mantengan altos estándares de honestidad académica. Cualquier acto deshonesto o fraude académico está prohibido; los alumnos que incurran en este tipo de acciones se exponen a un procedimiento sumario. Ejemplos de actos deshonestos son la copia, el uso de material o equipos no

permitidos en las evaluaciones, el plagio, o la falsificación de identidad, entre otros. Específicamente, para los cursos del Departamento de Ciencia de la Computación, rige obligatoriamente la siguiente política de integridad académica en relación a copia y plagio: Todo trabajo presentado por un alumno (grupo) para los efectos de la evaluación de un curso debe ser hecho individualmente por el alumno (grupo), sin apoyo en material de terceros. Si un alumno (grupo) copia un trabajo, se le calificará con nota 1.0 en dicha evaluación y dependiendo de la gravedad de sus acciones podrá tener un 1.0 en todo ese ítem de evaluaciones o un 1.1 en el curso. Además, los antecedentes serán enviados a la Dirección de Docencia de la Escuela de Ingeniería para evaluar posteriores sanciones en conjunto con la Universidad, las que pueden incluir un procedimiento sumario. Por “copia” o “plagio” se entiende incluir en el trabajo presentado como propio, partes desarrolladas por otra persona. Está permitido usar material disponible públicamente, por ejemplo, libros o contenidos tomados de Internet, siempre y cuando se incluya la cita correspondiente.