

Revisión Práctico 3- Aprendizaje Supervisado

Integrantes:

- Rinaudo, Eleana
- Sulca, Anahí

Descripción

Los datos corresponden a información recopilada correspondiente a los llamados de todo el país atendidos por los y las profesionales del 137 sobre casos de Violencia Familiar desde Enero de 2017 hasta Febrero del 2022 organizados de manera unificada.

Pueden consultar en:

<http://datos.jus.gob.ar/dataset/linea-137-victimas-de-violencia-familiar>

El dataset contiene información de id, descripción del llamante, género del llamante, tipo de violencia, género de la víctima, cantidad de víctimas, cantidad de agresores, género del agresor, relación con la víctima, fecha y hora de la llamada.

repositorio con los datos listos para procesar:

<https://github.com/IsabelRivadero/ir-mentoría2022>

Correcciones:

- *Importante: compartir los enlaces de la resolución del tp3 en los repositorios personales de cada una ya que me lo solicitan así.*
- En el tratamiento de datos inicial para *semanas del año* usan:
`linea137_df['semana_del_anio'] = linea137_df['anio'].astype("str") + "-" + [sem.zfill(2) for sem in linea137_df['dia_semana'].astype("str")]`
En vez de: 'dia_semana' debería usarse 'semana'. Después de esto victimas por semana arrastraría error, pero veo que no lo vuelven a usar así que no es necesario modificar, solo tengan en cuenta eso.
- Gráfico de Holt Winters Single Exponential:
Una observación puede ser que elegir a mano un *alpha* sin dejar que el modelo que ajuste puede ser una de las causas por las que no obtenemos el mejor ajuste.
- Holt Winters Triple Exponential Smoothing Additive and Multiplicative Seasonality
Sería útil usar las métricas(MSE, MAE, RMSE o alguna de ellas) para comparar los resultados de las distintas pruebas de los modelos, esto para no decidir a ojo cual es más representativo.
Está bien lo que hicieron de calcular la métrica del train-test, pero además sería útil usar estos resultados(métricas) para comparar las pruebas de simple, doble y triple.
- Por Mes de cada año:
las predicciones parecen estar desplazadas en ambos casos, no solo el multiplicativo.

El trabajo práctico está Aprobado. A continuación le dejo información importante sobre los próximos pasos a seguir.

En términos generales sólo resta el último video, y entregar el último TP.

Están en la recta final, Ánimos!

Información importante para la segunda y final presentación de la mentoría:

1- Detalles sobre video de presentación final:

- **Fecha de entrega: 23/09**
- **MÁXIMO 10 min**
- Manteniendo la forma en que trabajaron la primera presentación, y como continuación del mismo dando una breve introducción para conectar ambas instancias contarán que trabajaron en el tp3 con qué problemas se encontraron, y las limitaciones de los datos. Si llegan con el TP4, entonces también podría agregar conclusiones o lo que hayan llegado a probar del TP4.
- Sobre los temas a tratar en la segunda presentación: pueden contar dificultades que tuvieron, cambios de estrategia, limitaciones de los datos, problemas que se resolvieron y como, modelos aplicados, resultados y cómo se podrían usar esos resultados. También cómo lo mejorarían si tuvieran más tiempo, cuáles serían los próximos pasos o trabajo a futuro.

Jornadas de 11 y 12 de noviembre. Presentación de mentorías.

Dos jornadas en donde cada grupo **NO EXPONE TODO EL TRABAJO NUEVAMENTE** sino que, luego de una breve introducción de 2 minutos (preparada para ese tiempo), **responde preguntas/comentarios que los demás grupos les hacen.**

Se asignará a cada grupo al menos 1 video obligatorio para que vean y analicen en profundidad.

Para tener una idea: las jornadas del año pasado

<https://sites.google.com/unc.edu.ar/proyectosdiplodatos2021/programa>

IMPORTANTE para el video: Si están muy atrasadas en el último TP priorizaremos el video.

Objetivo para el 23/septiembre, si no llegan con todo para esa fecha, primero hacer el video y luego cerrar el trabajo pendiente.