

Proyecto #1 Series de tiempo

Otoño 2024

Profesor: Daniel Nuño daniel.nuno@iteso.mx

Fecha de entrega y presentación: 14 octubre, 2024

Introducción:

El propósito del primer proyecto es mostrar sus conocimientos adquiridos para analizar, pronosticar y compartir sus resultados de una forma profesional, convincente y clara. El trabajo y presentación es individual.

Selecciona una serie de tiempo (de una de las dos series de tiempo que habías seleccionado con anterioridad, la que no estés usando para las tareas y actividades).

Tienes que responder implícitamente (indirectamente), lo siguiente:

- Define el problema, descripción de la serie de tiempo. ¿Cuál es el motivo o la necesidad de pronosticar? ¿Cuánto tiempo en el futuro es necesario pronosticar y por qué? ¿Estos datos, en el mundo real que representan?,
- Recaba la información. ¿Cómo se recopilan?, ¿Cada cuanto se recopilan?, ¿Desde cuándo existen o desde cuándo vas a utilizar?, ¿De dónde se obtienen?
- Análisis exploratorio. ¿Cuáles son las características de la serie de tiempo? ¿tiene patrones? ¿Cuáles son las medidas estadísticas de tendencia central y de dispersión?
- Ajusta diferentes modelos y compara. En base a las métricas de error, análisis de residuales, generalización del modelo, cross validation, interpretación de los modelos.
- Selecciona un modelo y pronostica. Da tus conclusiones.

Entregables:

Su proyecto debe ser reproducible y entendible por terceros.

- Presentación:
 - En PDF
 - 5 minutos para presentar.
 - Explica ¿qué estás pronosticando?, ¿por qué lo estás pronosticando?, ¿Cuánto tiempo en el futuro lo estás pronosticando? ¿qué modelo(s) se ajusta mejor?, ¿Por qué decidiste usar este modelo?, ¿Cuál es el pronóstico puntual e intervalos de predicción?
 - Da tus conclusiones
 - Sin código, ni ecuaciones en la presentación.

- Asume que le estas presentando a tu jefe, inversionistas, políticos, clientes o población en general. Es decir, a quien está interesado en saber el pronóstico de tu serie de tiempo para tomar decisiones, pero no está interesado en saber Python o en matemáticas.

- Notebook en Python:

Sigue como ejemplo el archivo de remesas. Tú código debe ser reproducible y tus comentarios deben estar en línea.

- Definición del problema y datos recabados.
- Exporta los datos
- Limpieza de los datos
- Inspección visual de la serie con gráficas:
 - Dispersión.
 - líneas.
 - caja y bigotes.
 - gráficas estacionales.
 - histogramas, densidades, correlaciones.
 - autocorrelaciones.
- Detección de atípicos u observaciones faltantes e investigación de ellos.
 - ¿Cuál es el motivo?, ¿es un dato real o error de medición?
 - Si consideran necesario quitarlos, justificarlo y realizarlo.
- Divide tus datos en entrenamiento y prueba. Tu medición de error para la selección del modelo con el que vas a pronosticar tiene que ser con los datos de prueba.
- Revisar si la serie requiere algún ajuste o transformación matemática. Justifica tu decisión.
- Hacer la descomposición de la serie de tiempo. ¿presenta estacionalidad o no?
- Realizar el flujo de pronóstico completo, incluyendo el mejor (o mejores) modelos de referencia y ponerlos a competir con la regresión lineal y
 - Escoge un modelo de referencia y justifica porque lo escogiste en lugar de otros (mean, naive, seasonal naive, drift).
 - Realiza varios modelos de regresión lineal con al menos un variable exógena correlacionada (causal) y variables que representen rezagos, estacionalidad, tendencia o atípicos. Justifica porque estas usando esta variable cómo predictora. Interpreta tus resultados del ajuste de la regresión lineal y de los errores. ¿cumple con las recomendaciones de características de los errores?
 - Aplica alguna técnica de regularización o selección de variables. Stepwise, Ridge, Lasso.
 - Evalúa tus modelos analizando los errores, una métrica de error y los datos de prueba.
 - Selecciona el mejor modelo.

- Realiza los pronósticos y grafica en el horizonte de tiempo qué habías decidido en la definición del problema. Las fechas y horizonte en la definición y esta sección tienen que cuadrar.
- Datos
 - En caso de aplicar, un archivo con los datos usados (xlsx, csv, etc).

Criterios de evaluación

Cada criterio se califica con una escala de 1 (peor) a 5 (mejor). Todos tienen la misma ponderación.

1. Contenido
 - a. ¿Cuenta con la teoría suficiente?
 - b. ¿Las tablas, gráficas son relevantes al análisis de la serie de tiempo?
 - c. ¿Muestra dominio de los temas presentados?
 - d. ¿Los modelos propuestos se ajustan correctamente a los datos?
 - e. ¿Son relevantes los modelos para las series elegidas?
 - f. ¿El código está optimizado, utilizando las paqueterías vistas en clase? (¿es escalable, no se duplican operaciones de modelado, etc.?)
2. Claridad y presentación
 - a. ¿Se lleva una buena estructura de la presentación?
 - b. ¿Las gráficas, tablas, etc. tienen sus correspondientes ejes, títulos bien marcados?
 - c. ¿El análisis está presentado de forma útil para la empresa/situación de negocios correspondiente?