

Descripción General: el proyecto consiste en resolver un problema sobre un conjunto de datos. Se debe analizar el contexto, hacer un análisis exploratorio de datos, y seleccionar el mejor modelo al evaluar diferentes técnicas de análisis de series temporales. Se debe evaluar si el conjunto de datos es estacionario, realizar diferentes modelos de predicción (forecasting) cómo: AR, MA, ARIMA, VAR, SARIMA, VAR-MA, modelos aditivos (prophet) y modelos tradicionales de aprendizaje automático vistos en clase (creando features temporales adecuadas, y usando modelos potentes cómo RandomForest y Boosted Machines). Se debe escoger el mejor modelo mediante validación cruzada para series temporales (`TimeSeriesSplit`) y evaluar el desempeño en un test set independiente.

Directrices: el proyecto consistirá en

1. Estudiar el contexto del problema a resolver.
2. Una primera parte del proyecto, consiste en realizar un análisis exploratorio de los datos, en donde se realicen conclusiones, mediante gráficas y cálculos estadísticos básicos (aquí pueden involucrar técnicas de reducción de la dimensión para visualización).
3. Realizar un Jupyter Notebook en el que se describan claramente el problema, el desarrollo, análisis, metodología y los procedimientos desarrollados y las pertinentes conclusiones que obtiene en el contexto del problema considerado. El código del mismo debe estar bien documentado y organizado.
4. Se realizarán unas preguntas de defensa del proyecto, programadas individualmente por cada grupo.

Requerimientos: el proyecto debe:

1. Desarrollarse en Python.
 2. Depurar y preprocesar el conjunto de datos.
 3. Análisis exploratorio de datos. Consiste en utilizar herramientas simples cómo graficar y resumir estadísticas para entender mejor los datos. Repasar las gráficas y los resúmenes observando estructuras temporales cómo tendencias, anomalías (datos perdidos, corrupción y valores extremos) y cualquier otra estructura que pueda impactar las previsiones.
 4. Elegir y ajustar los modelos. Evaluar un conjunto de modelos diferentes sobre el problema. Los modelos se pueden escoger en función de los supuestos que hacen y si el conjunto de datos se ajusta a los mismos. Los modelos se pueden configurar y ajustar a los datos históricos.
 5. Usar y evaluar el modelo. El modelo se debe usar para realizar pronósticos, evaluando su desempeño sobre los mismos. Esto implica realizar pruebas posteriores con datos históricos o esperar a que estén disponibles nuevas observaciones. En este punto, también es interesante realizar validaciones cruzadas.
 6. Concluir con las observaciones generadas a través del análisis, creación y evaluación de los modelos construidos.
-

7. El trabajo deber estar bien escrito, sin errores gramaticales y ortográficos.
8. El trabajo puede entregarse en un notebook de Jupyter organizado, como un informe que posee ilustraciones y código, pero que se pueda leer de forma amena.

Entregables y Fechas:

Entrega: Notebook de Jupyter con la descripción del problema considerado, el análisis exploratorio de datos, el conjunto de técnicas utilizadas y los procedimientos desarrollados, una discusión y análisis de los resultados y las conclusiones obtenidas.

Fecha de entrega: a más tardar el **18 de noviembre de 2021**.

Puntaje: 80 puntos

Descripción del problema, análisis exploratorio y calidad de las conclusiones: 20 puntos

Descripción del procedimiento, análisis y discusión realizados: 20 puntos.

Claridad argumentativa, ortografía, presentación: 10 puntos.

Código Python **en notebook y documentado** que respalde todos los análisis/procedimientos y permita replicar todos los resultados: 10 puntos

Desarrollo del proyecto durante la clase: 10 ó 30 puntos. **Si este punto se logrará alcanzar con éxito, el puntaje de éste ítem exonerará y sustituirá la sustentación del proyecto .**

Defensa: Sustentación del trabajo desarrollado mediante preguntas.

Fecha: Se acordará una sesión del grupo con el profesor, posterior a la entrega del documento.

Puntaje: 20 puntos

Otros:

Grupos: el proyecto se realizará en los grupos designados al inicio del semestre.

Caso especial: si se detecta que un estudiante no participó en la realización de su proyecto, su nota será cero.