



Universidad Europea Miguel de Cervantes

Escuela Politécnica Superior

Máster en gestión y análisis de grandes volúmenes de datos: BIG DATA

DESARROLLO DE ALGORITMOS METAHEURISTICOS PARA OPTIMIZACION DE REDES NEURONALES ARTIFICIALES APLICADOS A LA EFICIENCIA ENERGÉTICA BAJO SPARK

Autor:

Rubén Correoso Campillo

Tutor:

M^a del Carmen Pegalajar Jiménez

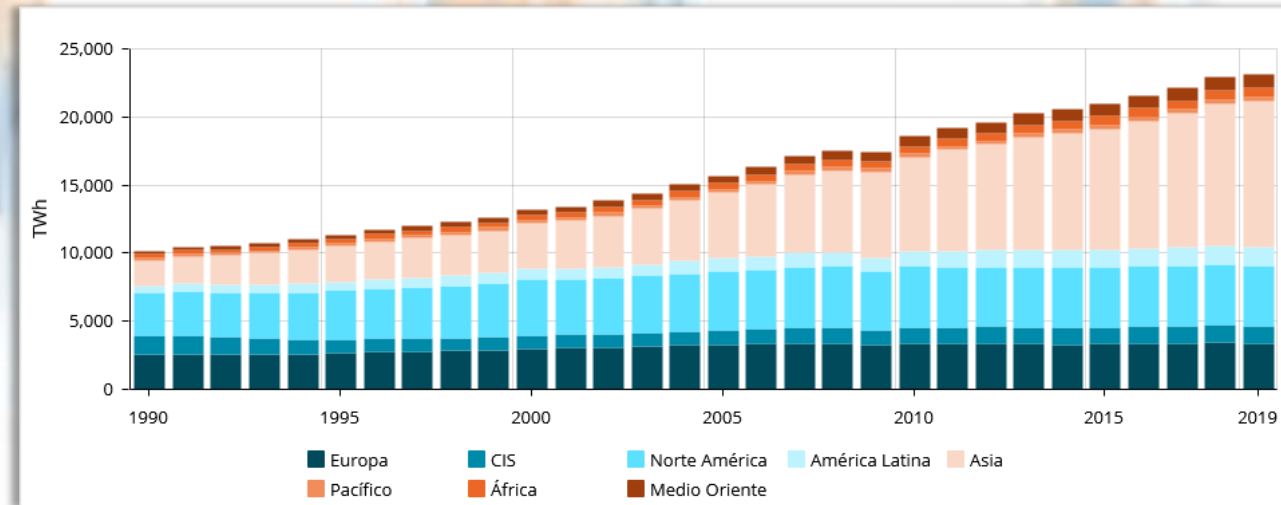
Índice

1. Introducción
2. Objetivos
3. Conclusiones
4. Planes de mejora

Introducción: Contextualización

Contextualización:

- Crecimiento de la población
- Consumo energético a nivel mundial
- Eficiencia energética
- Anticiparnos a las necesidades de la población



Introducción: Entorno y conjunto de datos

Entorno:

- Repositorio GIT
- Notebook en colab
- Entrenamiento en GPU desde Colab

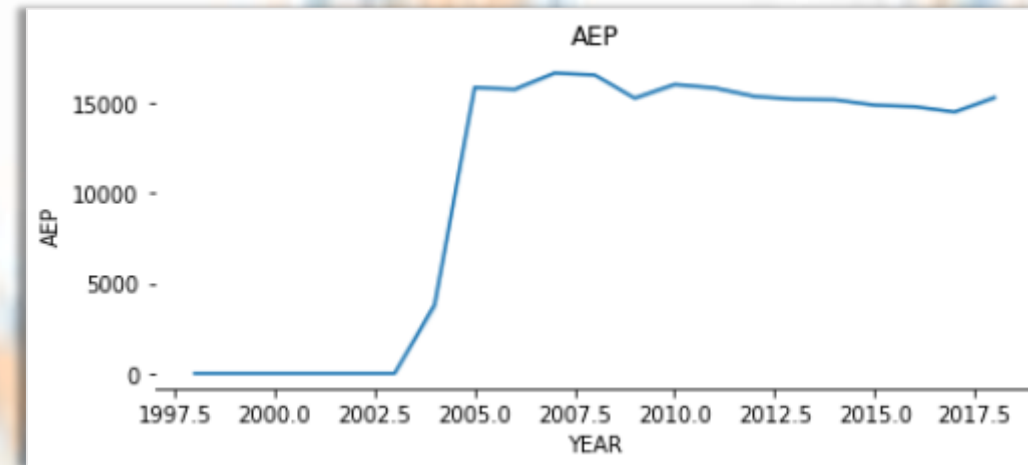
Conjunto de datos “*Hourly Energy Consumption*”:

- Kaggle → Rob Mulla Kernel Master
- PJM Interconnection LLC durante 20 años
- Diferentes zonas a lo largo de histórico

Introducción: Series temporales

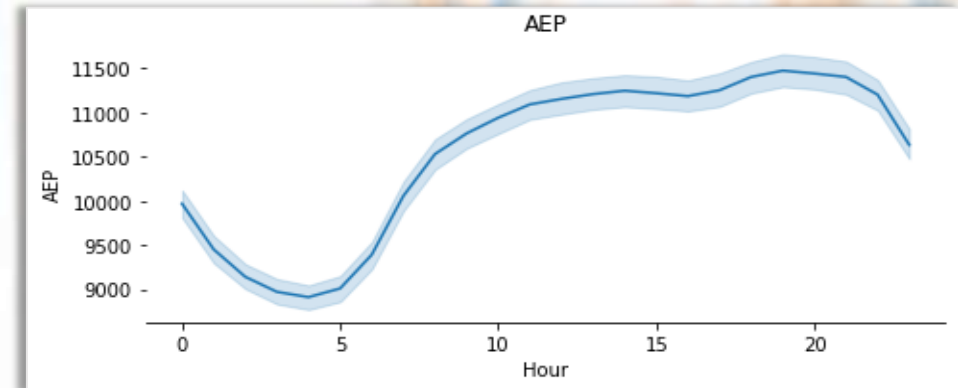
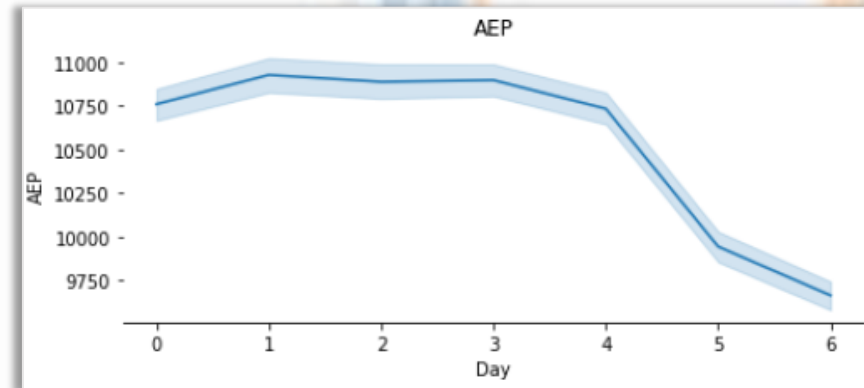
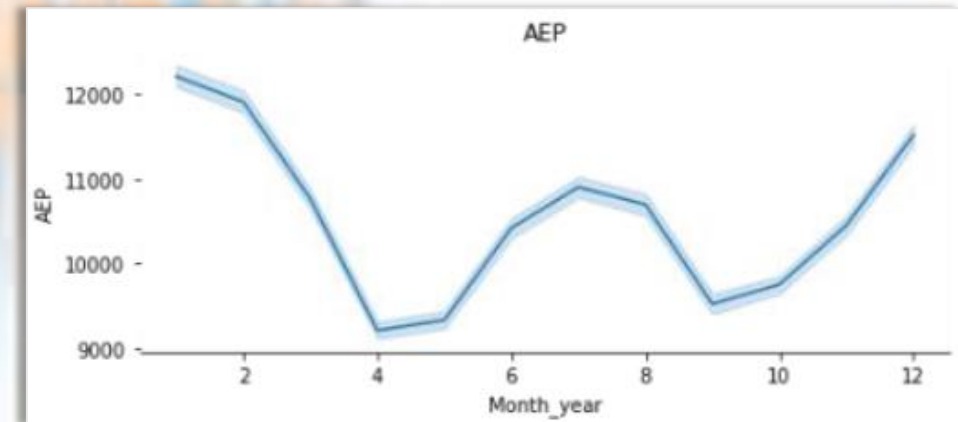
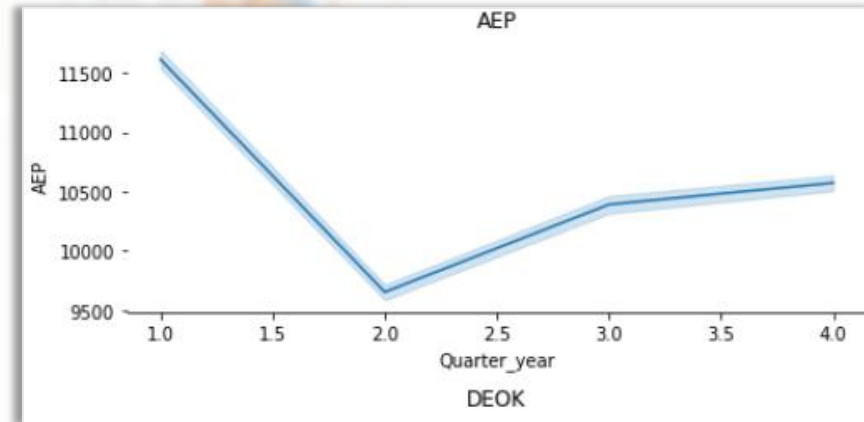
Series temporales:

- Estacionariedad del consumo energético
- Patrones y tendencias temporales (Horas, días de la semana, meses, estaciones, eventos vacacionales, ...)



Introducción: Series temporales

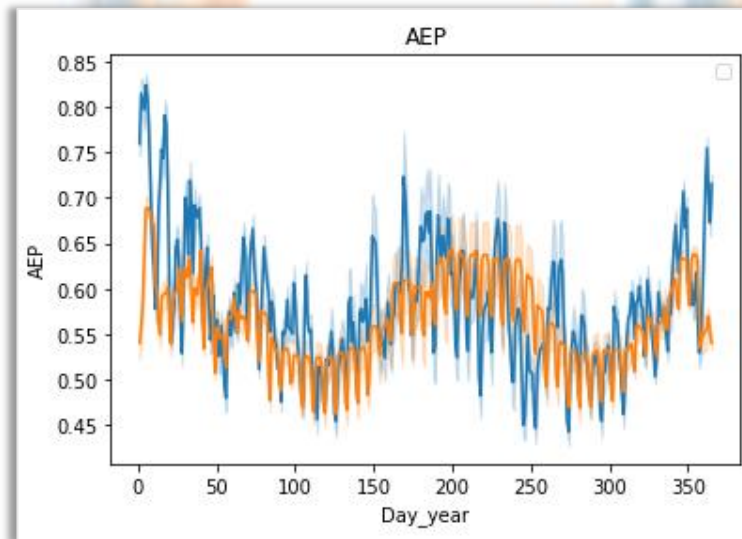
Ejemplos de estacionariedad:



Objetivos: Modelo predictivo

Desarrollo de un modelo predictivo:

- Predecir el consumo del último año del dataset y evaluar el error frente a los datos reales.
- Mediante algoritmos clásicos
 - XGBoost



Región	MSE	MAE
AEP	0.0042789	0.0503417
COMED	0.0037985	0.0428231
DAYTON	0.0039913	0.0483792
DEOK	0.0061520	0.0587638
DOM	0.0068216	0.0597597
DUQ	0.0035656	0.0445785
EKPC	0.0078098	0.0634716
FE	0.0032543	0.0413205

Objetivos: Modelo predictivo

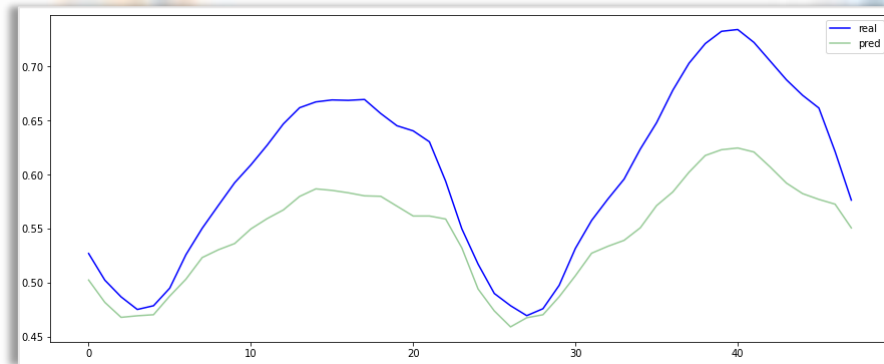
Desarrollo de modelo mediante redes neuronales:

- RNN
 - Cada capa recibe información de un estado de tiempo anterior (Matriz de pesos U).
 - La matriz de pesos U también se actualiza en la fase de BackPropagation.
- LSTM
 - Redes con memoria.
 - Pueden encontrar patrones en instantes alejados de tiempo.
 - Puertas de entrada, de salida y de olvidar.
- GRU
 - Mismo principio que LSTM pero simplificadas.
 - Rendimiento similar pero más eficientes.
 - Puertas reset y update

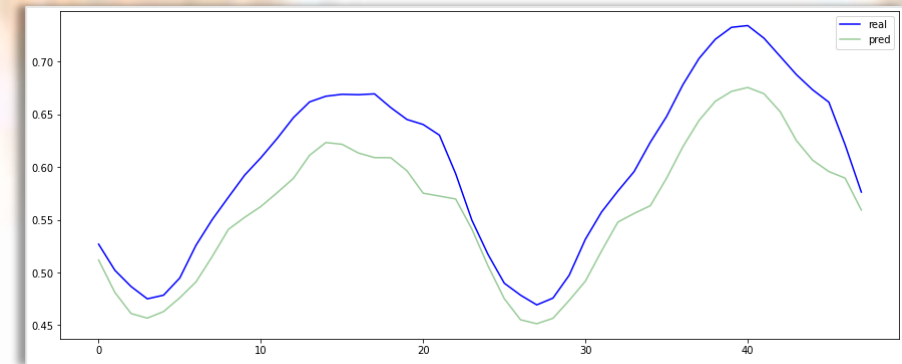
Objetivos: Modelo predictivo

Comparativa de predicción de los modelos

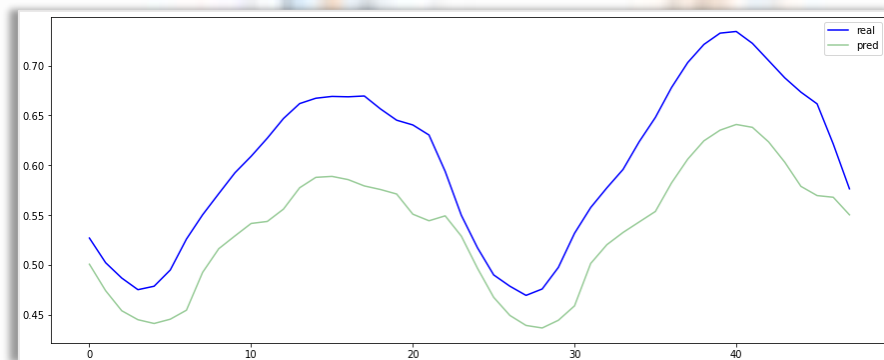
- RNN



- GRU



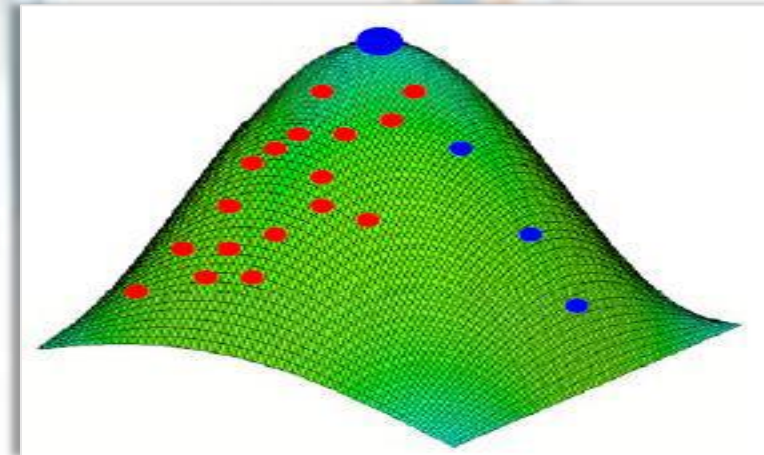
- LSTM



Objetivos: Algoritmos Metaheurísticos

Algoritmos Metaheurísticos:

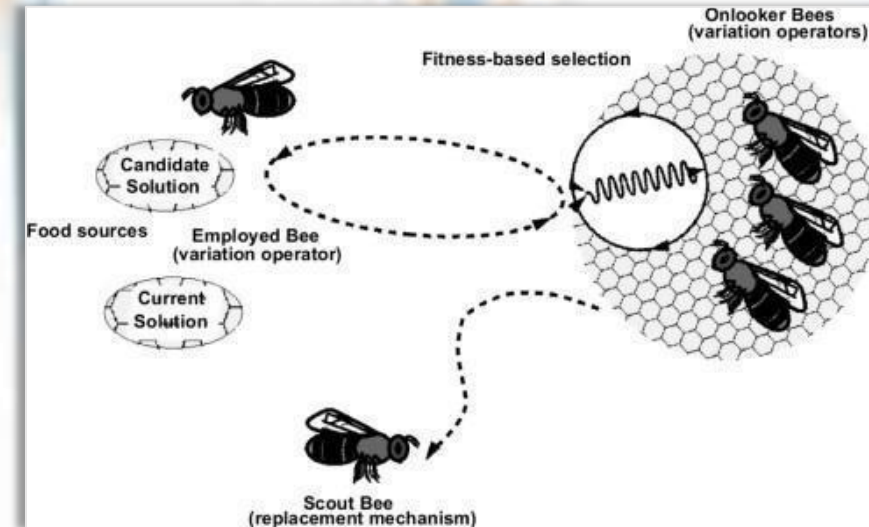
- ¿Qué son?
- Optimización.
- Múltiples mínimos locales en el espacio de soluciones.
- Algoritmos aproximados e iterativos que guían una heurística subordinada.
- Teorema No-Free-Lunch



Objetivos: Algoritmos Metaheurísticos

Algoritmo Ant Bee Colony (ABC):

- Swarm Intelligence
- ABC simula una colonia de abejas obreras, exploradoras, espectadoras.
- Exploración y explotación.
- ¿Cómo pueden mejorar el modelo predictivo?
 - Número de capas y neuronas
 - Entrenamiento de la red neuronal



Objetivos: Algoritmos Metaheurísticos

Desarrollo del algoritmo metaheurístico ABC:

- Desarrollo en Python

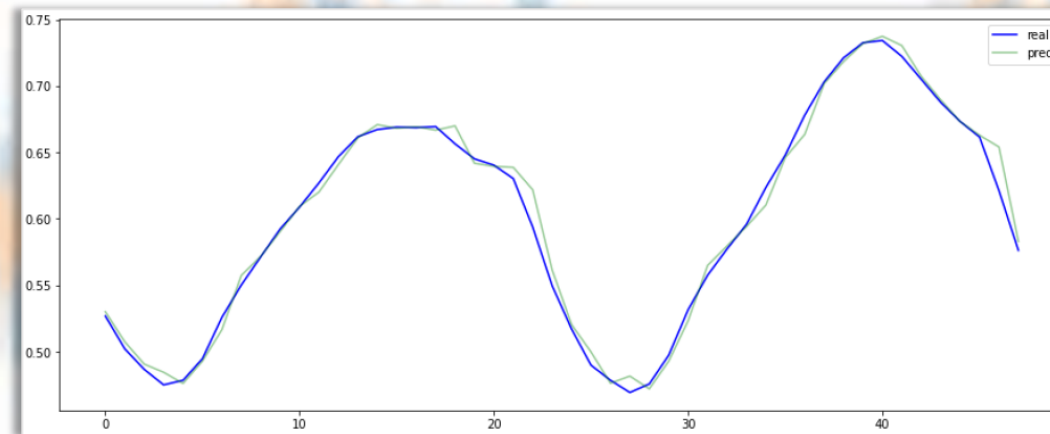
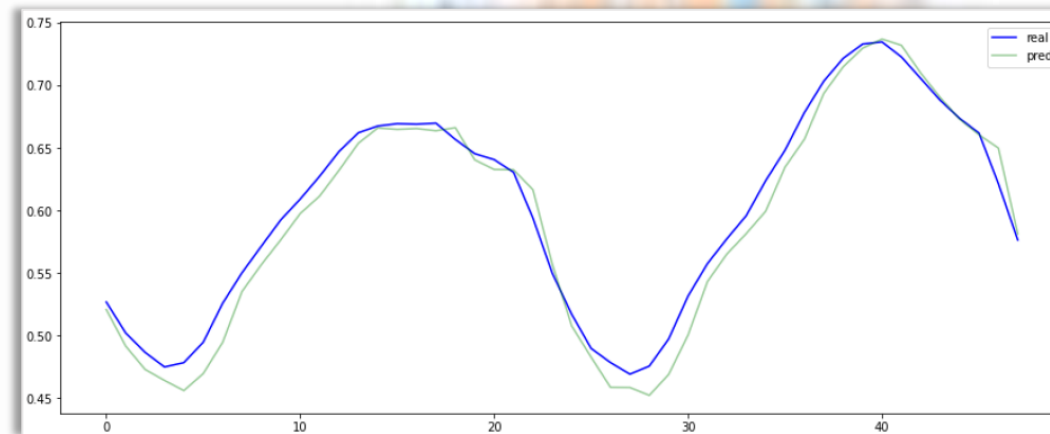
Para $i \leq N$

- Las abejas exploradoras buscan fuentes de comida (epochs=1)
- Vuelven a la colmena y las abejas obreras explotan cada una de las fuentes de comida (epochs=Nobreras)
- Una vez explotada la fuente de comida las abejas espectadoras deciden cual es la mejor solución.
- Integración con el entrenamiento de la red neuronal LSTM
 - Explotación y exploración.

Objetivos: Algoritmos Metaheurísticos

Resultados al aplicar el algoritmo ABC:

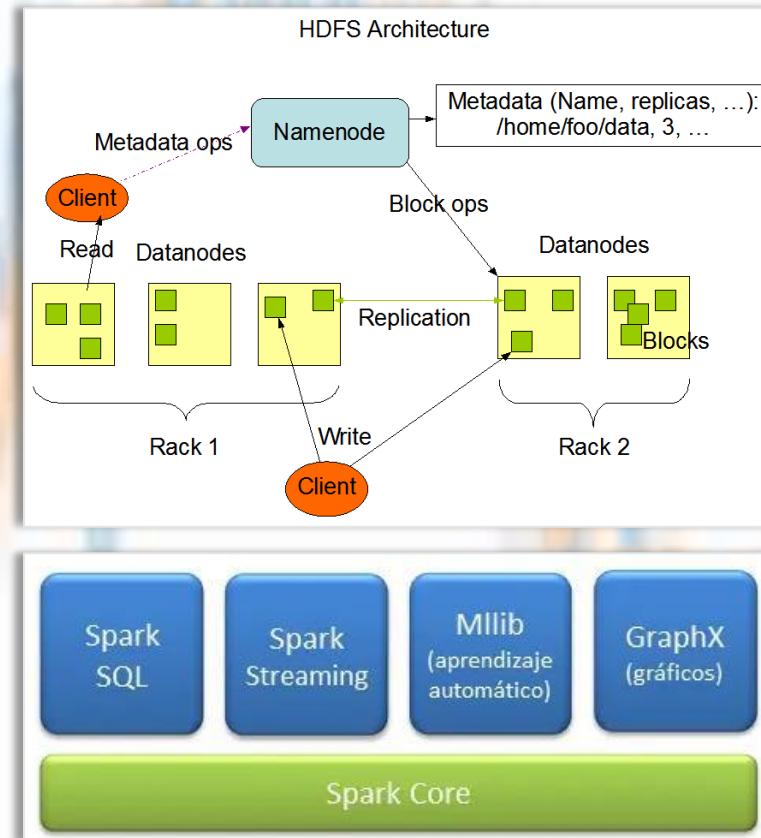
- LSTM
- LSTM + ABC



Objetivos: Spark

Ecosistema Apache Hadoop:

- Apache Hadoop
 - Escalabilidad horizontal
 - YARN
 - NameNode y Data Nodes
 - Map Reduce
- Apache Spark
 - Ventajas frente a Map reduce
 - Lenguajes soportados



Objetivos: Spark

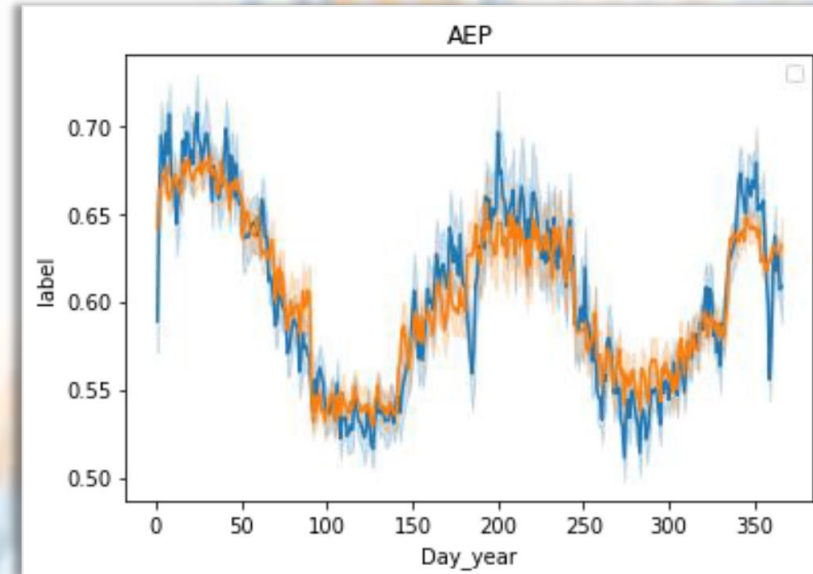
Modelo predictivo en Spark:

- Spark
 - PySpark
 - Integración con Google Colab
- MLlib
 - Ventajas
 - Limitaciones
 - Gradient Boosting Trees

Objetivos: Spark

Evaluación del modelo predictivo en Spark:

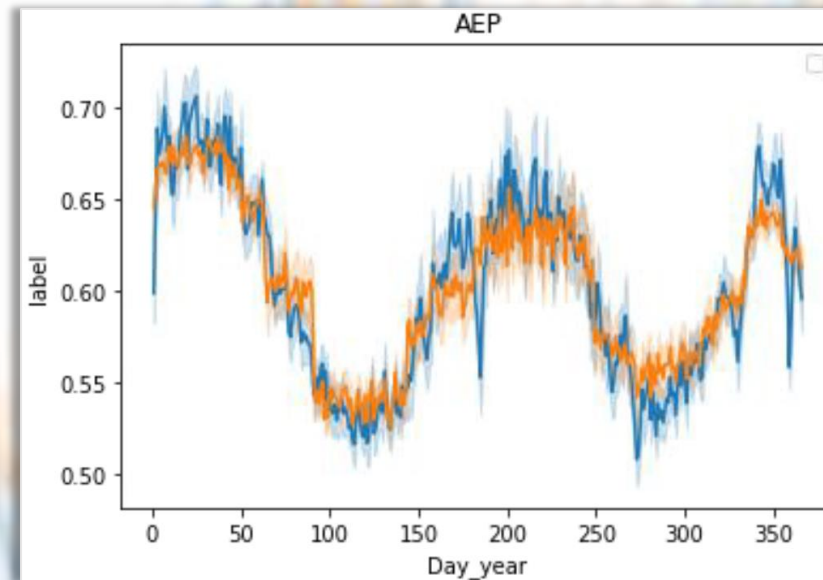
- Último año del conjunto de datos
- Gradient Boosting Trees
- Paralelización



Objetivos: Spark

Integración del algoritmo ABC en PySpark:

- Ventajas frente al modelo LSTM en Google Colab
- Cambios frente al modelo LSTM en explotación y exploración
epochs===maxiter



Conclusiones

Conclusiones:

- Tras evaluar múltiples modelos de regresión se ha observado que la red neuronal LSTM es la que ha aportado mejor resultado en cuanto a entrenamiento y precisión para el conjunto de datos de estudio.
- Se ha conseguido implementar el algoritmo ABC mejorando la precisión en el entrenamiento de los modelos tanto de redes neuronales como en Spark.
- Se ha creado un modelo de regresión en Spark utilizando la librería Mllib, consiguiendo un modelo fácilmente escalable horizontalmente sobre un clúster Hadoop.

Conclusiones

Conclusiones:

- Finalmente se ha integrado el algoritmo ABC en Spark para optimizar el entrenamiento del modelo.
- Todo lo anterior nos aporta técnicas fiables y precisas de predicción del consumo energético de la serie temporal, siendo este uno de los principales problemas a nivel mundial.

Planes de mejora

Mejoras:

- Implementar entrenamiento de redes neuronales en Spark.
- Optimizar la función objetivo del algoritmo ABC para mejorar el entrenamiento en tiempo y en precisión.
- Integración del algoritmo ABC para el dimensionado de redes neuronales, número de capas, neuronas, función de activación...
- Ejecutar el modelo PySpark en un clúster Hadoop para estudiar el comportamiento en varios nodos.
- Añadir nuevas características al conjunto de datos que puedan afectar al consumo energético, como datos geográficos, de periodos vacacionales, datos climáticos, densidad de población.
- Posible integración de otros algoritmos metaheurísticos.



Gracias por su atención