

Tarea 4 – Aprendizaje Automático

1. Utilizando la herramienta Jupyter Notebook y el lenguaje de programación Python, realiza la siguiente actividad:

Considera la función: $f(x,y) = 3x^2 + 2xy + 15x + 6y^2 - y$, comenzando desde el punto inicial $(x, y) = (-2, 3)$, encuentra el mínimo valor estimado para el cual esta función converge.

2. Es muy común identificar situaciones donde se deben realizar optimizaciones para encontrar el óptimo de una función dado un conjunto de restricciones iniciales. Identifica una problemática que cumpla con estas condiciones en tu entorno laboral y desarrolla un ejemplo completo en donde obtengas la solución utilizando el lenguaje de programación Python. Considera que dicho ejemplo se compartirá con el resto de tus compañeros a modo de presentación digital.
3. Los cuatro pilares del aprendizaje automático son la regresión, la reducción de la dimensionalidad, la estimación de la densidad y la clasificación. Investiga sobre los modelos matemáticos que más se utilizan para representar cada una de estas situaciones y elabora una tabla-resumen sobre cada uno de ellos.

Al menos de los modelos que más he utilizado o conocido, están los siguientes:

Pilar	Descripción	Modelos ML	Ejemplo
Regresión	Predicción de un valor continuo basado en distintas variables	Regresión lineal, Lasso, Ridge.	Usualmente se utiliza para asignar valores continuos.
Reducción de dimensionalidad	Simplificación de datos manteniendo la estructura inicial.	PCA, t-SNE, LDA.	Busca combinaciones en maneras más sencillas (1 dimensión menos)
Estimación de la densidad	Determinación de la distribución de probabilidad de un	KDE, GMM.	Se suele usar en funciones kernel para suavizar la estimación
Clasificación	Técnica para asignar etiquetas a datos en	SVM (Support Vector Machine),	Separación de clases.

	función de sus características.	XGBoost, NN (Neural Network)	
--	---------------------------------	------------------------------	--

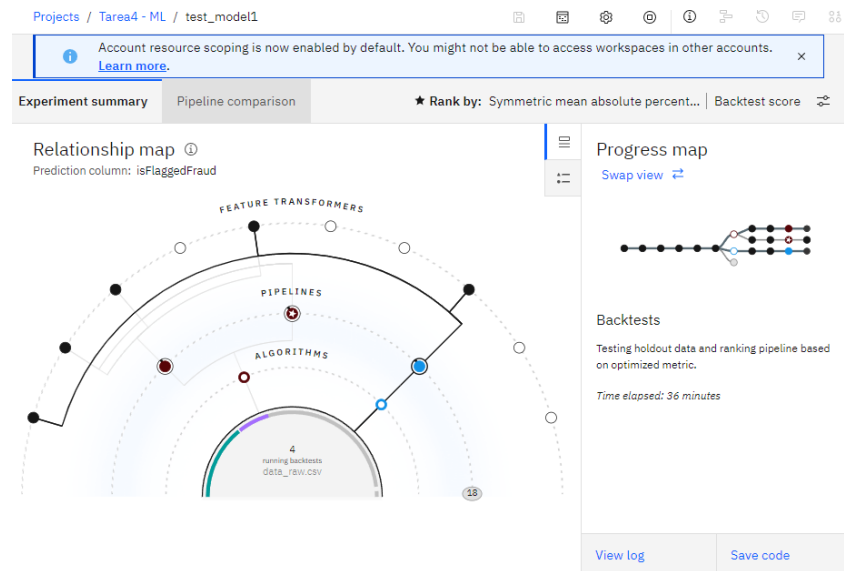
4. Lee el siguiente artículo y responde las siguientes preguntas:

http://bibliotecadigital.udea.edu.co/bitstream/10495/20164/1/CarmonaMaricela_2021_DeteccionFraudeFinanciero.pdf

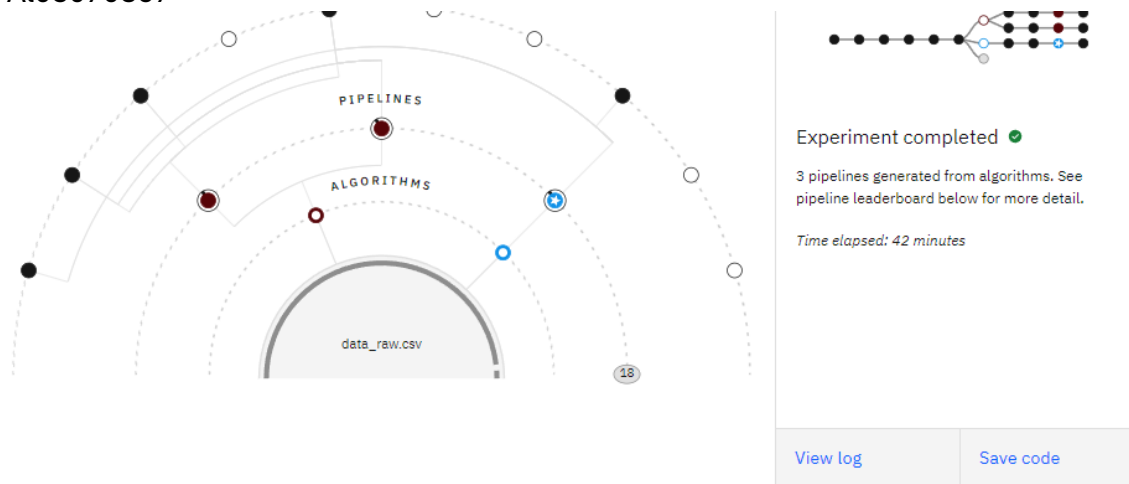
Con base en el artículo de la pregunta 4, utiliza la herramienta AutoAI de la plataforma IBM Watson Studio y genera un modelo de aprendizaje automático que evalúe el mismo conjunto de datos analizado en el material. Compara los criterios de selección del modelo que fueron considerados por el autor del artículo y las métricas entregadas por la aplicación AutoAI. Incluye tus conclusiones en el informe electrónico de la actividad.

El proceso para realizar un entrenamiento en la herramienta de IBM Watson Studio es bastante sencillo, siempre y cuando se tenga un buen conocimiento de los datos, dentro del mismo notebook anclado en esta tarea.

Inicialmente, se tiene que crear un proyecto nuevo, el cual le otorgué el nombre de Tarea4 – ML, dentro, se crea un nuevo modelo, el cual tiene el nombre de tarea4_test2, se carga la información (previamente mencionado), la aplicación preguntará si se desea realizar un forecasting en los datos, en un test1 lo



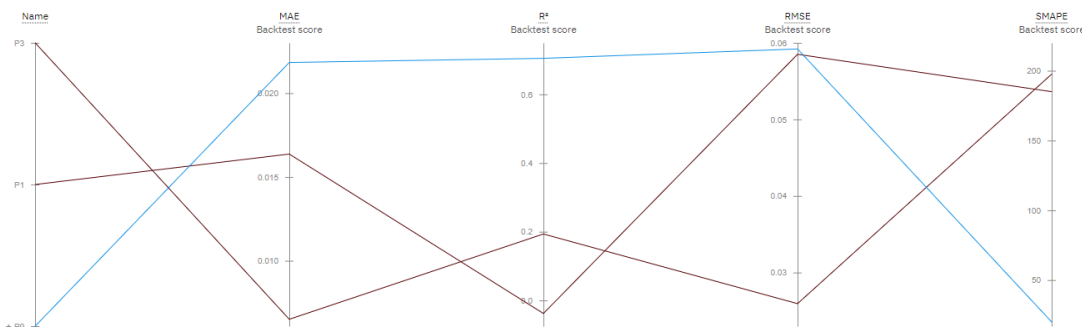
realicé con el fin académico de ver qué hacía, siendo una revisión sobre todos los posibles modelos que se podrían utilizar para determinar el mejor modelo a utilizar (Figura 1). Una vez elegido el modelo, se procede a seleccionar el set de entrenamiento, validación y holdout, además de mostrar las métricas elegidas y comparándose:



Pipeline leaderboard ⌵

	Rank ↑	Name	Algorithm	SMAPE (Optimized) Validation	SMAPE (Optimized) Holdout	SMAPE (Optimized) Backtest
★	1	Pipeline 9	● Ensembler	1.765	0.000	20
	2	Pipeline 1	● Random Forest	198.073	200	185.000
	3	Pipeline 3	● Random Forest	0.407	200	197.577

Tal como se muestra en la imagen, el modelo ganador fue un Ensembler, seguido de un random forest, la diferencia es que el modelo número 1 dio un resultado más bajo de SMAPE (Symmetric Mean Absolute Percentage Error) el cual es una métrica utilizada para medir la precisión de las predicciones en modelos de series temporales y otros tipos de regresión. Esta métrica penaliza los errores de manera similar, independientemente de si las predicciones son sobreestimadas o subestimadas. El resultado dio una predicción perfecta, por ende, es notable el triunfo de ese modelo.



Aquí otra comparación de las métricas de los pipelines revisados.

Miguel Angel Reséndiz Tinoco
Aprendizaje Automático
AI05070367