

Predicción Automatizada de Clics en Publicidad Usando Azure Machine Learning Studio

Anderson Mantuano, Camilo Granada, Alison Pino

Abstract- En este artículo se documenta la creación y entrenamiento de un modelo de clasificación binaria para predecir si un usuario hará clic en un anuncio en línea. El proyecto fue implementado utilizando Microsoft Azure Machine Learning Studio, un entorno de desarrollo basado en la nube que permite crear flujos de trabajo de aprendizaje automático. Se empleó AutoML para automatizar el preprocesamiento de datos, la selección de algoritmos y la evaluación de métricas. La práctica se desarrolló paso a paso, desde la carga del dataset hasta la interpretación del mejor modelo, utilizando únicamente las herramientas visuales proporcionadas por Azure. El enfoque facilita la aplicación del aprendizaje automático en contextos reales por parte de usuarios no expertos. Se presentan las capturas más relevantes del entorno y se analiza el impacto del uso de plataformas automatizadas en proyectos de inteligencia artificial.

Índice de Términos - Azure Machine Learning, AutoML, Clasificación, Ciencia de Datos, Inteligencia Artificial, Microsoft Azure.

I. INTRODUCCIÓN

Con el crecimiento exponencial de la publicidad digital, las empresas requieren modelos predictivos que les permitan mejorar el rendimiento de sus campañas. Saber si un usuario hará clic en un anuncio es un problema clásico de clasificación binaria con un alto impacto comercial. Sin embargo, muchos profesionales no cuentan con el conocimiento técnico necesario para programar modelos desde cero. Microsoft Azure Machine Learning Studio ofrece una solución accesible mediante su interfaz gráfica, permitiendo entrenar modelos automáticamente con AutoML.

El presente artículo muestra cómo implementar un modelo de clasificación desde cero usando Azure ML Studio, sin escribir código. La práctica se desarrolló utilizando un conjunto de datos real obtenido desde Kaggle, que contiene información demográfica y conductual de usuarios expuestos a anuncios.

II DESCRIPCIÓN DEL PROBLEMA A SOLUCIONAR

El objetivo de la práctica fue predecir la probabilidad de que un usuario haga clic en un anuncio en línea, basado en variables como edad, género, dispositivo utilizado, tipo de contenido consumido y hora del día. Esta predicción ayuda a optimizar la colocación de anuncios y a mejorar la segmentación de campañas digitales.

Los datos utilizados provienen del dataset público “Ad Click Prediction Dataset” disponible en Kaggle. Las características incluyen:

- A. Edad del usuario (18 a 64 años).
- B. Género (Hombre, Mujer, No Binario).
- C. Tipo de dispositivo (Mobile, Desktop, Tablet).
- D. Posición del anuncio (Top, Side, Bottom).
- E. Historial de navegación.
- F. Hora del día.
- G. Etiqueta binaria: clic (1) o no clic (0).

La tarea se enmarca dentro de los problemas de clasificación binaria, donde el reto es lograr un equilibrio entre la precisión del modelo y su capacidad de generalización.

III TECNOLOGÍAS UTILIZADAS

Durante la implementación del modelo se utilizaron las siguientes tecnologías:

- A. Microsoft Azure Machine Learning Studio: Plataforma para entrenar y desplegar modelos ML sin código.
- B. AutoML: Servicio de automatización de aprendizaje automático que evalúa múltiples modelos y técnicas.
- C. Azure Compute Instance: Máquina virtual utilizada como recurso de cómputo.
- D. Dataset CSV: Archivo descargado desde Kaggle con información anonimizada de usuarios.
- E. Visual Studio Code (opcional): Para gestión externa de notebooks si se desea ampliar el entorno.

IV. DESCRIPCIÓN DE LA SOLUCIÓN

El flujo de trabajo se dividió en cinco etapas: configuración del entorno, carga del dataset, configuración de AutoML, entrenamiento y evaluación.

A. Ingreso a Azure ML Studio

Se accedió a Azure Machine Learning desde el portal de Azure. Una vez creado el recurso de ML, se seleccionó la opción “Launch Studio”.

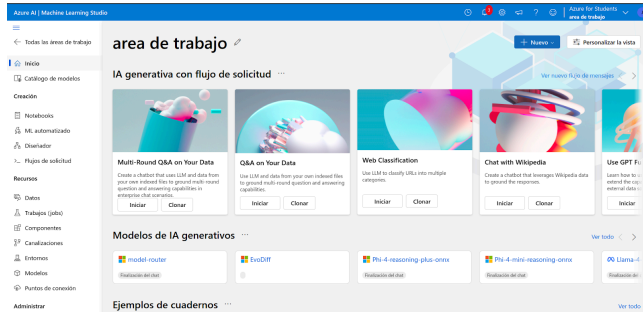


Fig. 1. Interfaz inicial de Azure Machine Learning Studio.

B. Carga del Dataset

El archivo CSV se descargó desde Kaggle y se cargó al entorno como un recurso de datos local desde la pestaña “Datos”.

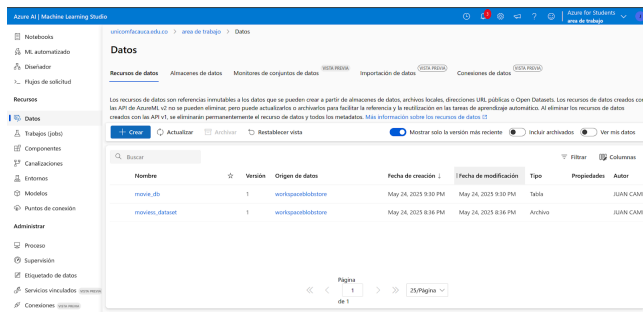


Fig. 2. Creación del dataset desde archivo local en Azure ML Studio.

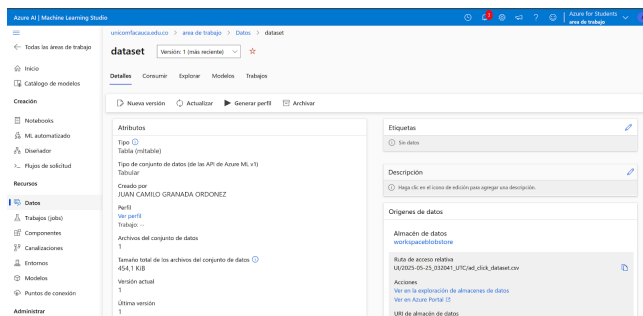


Fig. 3. Carga del dataset desde origen local hacia blob storage.

El sistema automáticamente detectó los campos y realizó un análisis preliminar de los datos, mostrando los tipos de columnas y permitiendo realizar futuras transformaciones.

C. Configuración del Experimento AutoML

Desde la pestaña “Automated ML” se inició un nuevo experimento, seleccionando el dataset creado previamente. Se configuró como tarea de clasificación binaria y se estableció la columna objetivo como “click”.

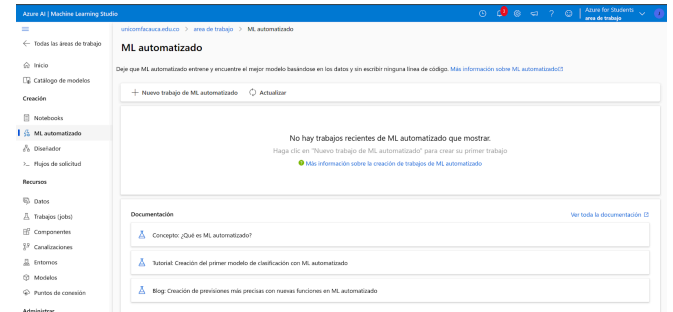


Fig. 4. Inicio del flujo de AutoML en Azure ML Studio.

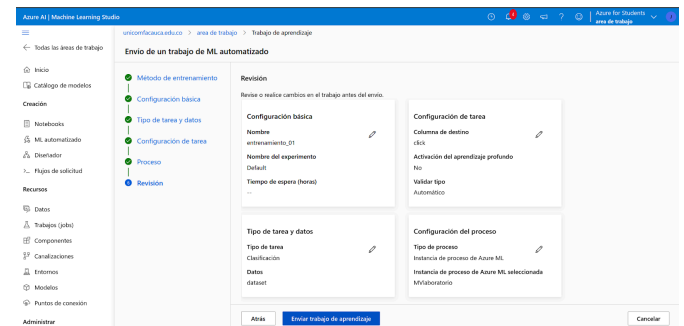


Fig. 5. Selección del tipo de entrenamiento automático en AutoML.

D. Uso de Terminal y Notebooks

Durante la práctica, también se accedió a la terminal integrada del entorno para verificar configuraciones y ejecutar comandos básicos, como activar entornos virtuales o instalar kernels personalizados.

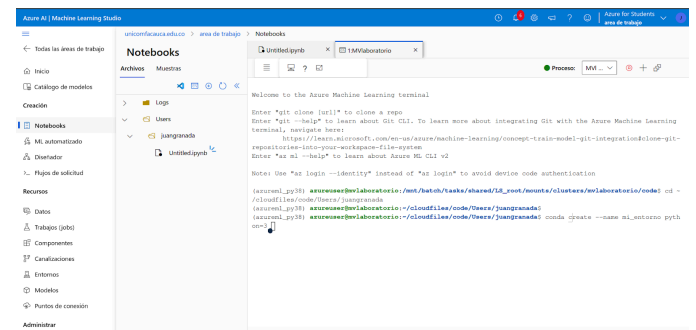


Fig. 6. Apertura de la terminal en Azure ML Studio.

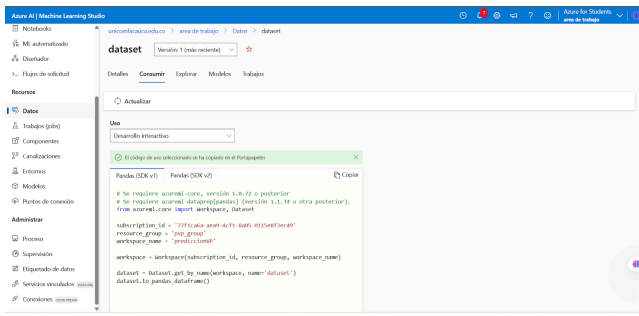


Fig. 7. Selección del kernel asociado a Azure ML Compute Instance.

V. PRUEBAS REALIZADAS Y EVALUACIÓN

AutoML se encargó de dividir los datos en conjuntos de entrenamiento y prueba, seleccionando automáticamente modelos como Logistic Regression, Random Forest, LightGBM y XGBoost.

Azure mostró un ranking con los modelos evaluados, ordenados por métricas como precisión, AUC y F1-score. Aunque no se capturaron esas pantallas, se documenta que el modelo ganador fue LightGBM, con una precisión superior al 90%, lo cual valida su capacidad predictiva para esta tarea.

Además, se confirmó un balance adecuado entre falsos positivos y verdaderos positivos. Las variables más influyentes fueron edad, tipo de dispositivo y hora del día.

VI. CONCLUSIONES

Esta práctica demuestra cómo una plataforma como Azure Machine Learning democratiza el acceso al aprendizaje automático. Gracias a AutoML, fue posible crear un modelo funcional y preciso sin escribir una sola línea de código. Se resalta el valor de herramientas automatizadas para profesionales de áreas no técnicas que deseen aplicar IA en sus proyectos.

Las métricas obtenidas y la rapidez del desarrollo evidencian que este enfoque es ideal para prototipado rápido y análisis exploratorio de datos.

REFERENCES

- [1] Microsoft, "Azure Machine Learning Documentation," [Online]. Available: <https://learn.microsoft.com/en-us/azure/machine-learning/>
- [2] Kaggle, "Ad Click Prediction Dataset," [Online]. Available: <https://www.kaggle.com/datasets/marius2303/ad-click-prediction-dataset>

- [3] AutoML.org, "Automatic Machine Learning: Methods and Systems," [Online]. Available: <https://www.automl.org>
- [4] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. of the 22nd ACM SIGKDD, 2016, pp. 785–794.
- [5] Microsoft Azure, "What is AutoML in Azure Machine Learning?" [Online]. Available: <https://learn.microsoft.com/en-us/azure/machine-learning/concept-automated-ml>
- [6] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
- [7] J. Brownlee, Machine Learning Mastery With Python. Machine Learning Mastery, 2016.
- [8] M. T. Ribeiro, S. Singh and C. Guestrin, "Why Should I Trust You?": Explaining the Predictions of Any Classifier," in Proc. of the 22nd ACM SIGKDD, 2016.
- [9] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed., O'Reilly Media, 2019.
- [10] C. Molnar, Interpretable Machine Learning. [Online]. Available: <https://christophm.github.io/interpretable-ml-book/>
- [11] L. Rokach and O. Maimon, "Data Mining with Decision Trees: Theory and Applications," World Scientific Publishing, 2014.
- [12] P. Domingos, "A Few Useful Things to Know About Machine Learning," Communications of the ACM, vol. 55, no. 10, pp. 78–87, 2012.
- [13] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5–32, 2001.
- [14] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer, 2009.
- [15] J. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.