

AI Academy, ML Final Project

Autor: [@Nohemi Limones](#) [@Angel J Aguilar Argaez](#) [@Christopher H Orea](#) [@Alejandro I Jimenez](#)

Summary

The AI Academy is an EY GDS Mexico initiative to level up the skills for the new joiners of the firm and enable cross-skilling of existing ones. The project is designed to evaluate the successful implementation of real-world AI/ML system capable of providing value to multiple stakeholders.

Objectives

- Implement and evaluate traditional or deep learning models for classification, regression, NLP, or recommendation tasks.
- Implement a cloud-based MLOps workflow for training, deploying, monitoring, and CI/CD automation.
- Generate reports with the expected business value added through this data initiative.

Instructions

1. Pick one of the following data related business situations and develop a ML/AI solution to add business value.
2. Develop an MLOps cloud base framework that allows users to consume the model in an easy way.
3. Prepare a pitch deck to showcase your methodology, results, and next steps if any.

Multiclass Classification

[Telco Customer Churn](#)

[Customer Personality Analysis](#)

Anomaly Detection

 [Fraud Detection](#)  [EDA + ML](#)

[Large-scale Energy Anomaly Detection \(LEAD\) | Kaggle pump_sensor_data](#)

Recommendation Engine

[Crop Recommendation Dataset](#)

[Amazon Product Reviews](#)

Time Series

Rubric

The following table describes the criteria that will be used to grade the project submissions.

	Criteria	Exemplary (100%)	Proficient (75%)	Developing (50%)	Needs Improvement (25%)
1	Problem Definition	Identifies the business problem that needs to be solved. Identify 3 or more KPIs relevant to the business. Identifies 1 or more KRIs relevant to the business.	Identifies the business problem that needs to be solved. Identify 3 or more KPIs, but they are not relevant to the business.	Identifies the business problem that needs to be solved. Identify at least 1 KPI for the business.	The problem identified is not complete or missing several relevant parts.
	Data Exploration	An EDA is presented using accepted statistical validation techniques, including, but not limited to, correlation, causality, independence, scaling, etc.	An EDA is presented using accepted statistical validation techniques, including, but not limited to, correlation, causality, independence, scaling, etc.	An EDA is presented using accepted statistical validation techniques, including, but not limited to, correlation, causality, independence, scaling, etc.	An EDA is presented but does not include relevant statistical methods.

		<p>At least 3 relevant business-related hypotheses are presented or tested.</p> <p>At least 1 relevant business-related insight is presented or extracted.</p>	At least 1 relevant business-related hypotheses are presented or tested.		
	Model Selection	<p>Clearly identify the type of problem presented (regression, classification, time series, unsupervised) and select a mathematical framework accordingly.</p> <p>Present 3 or more appropriate evaluation metrics for the model with high degree of confidence</p>	<p>Selects a relevant mathematical framework but does not clarify nor explain assumptions.</p> <p>Select at least 3 evaluation metrics but does not exhibit a high degree of confidence.</p>	<p>The selected mathematical framework is shallow on the assumptions or does not generalize well enough for production purposes.</p> <p>The model presents at least 1 evaluation metric with some degree of confidence.</p>	<p>The mathematical framework selected does not correspond to the nature of the problem, or the proposed solution is unfeasible.</p> <p>The model does not present a relevant evaluation metric.</p>
2	Explainable AI	Uses AI/ML explainability frameworks, including, but not limited to, Partial Dependence Plots SHAP, LIME, ICE, DeepLIFT, LRP, CEM, Weight of Evidence Coherence, etc.	Attempts to explain model results based on extreme scenarios, graphs, and use cases.	Provides some insights into model behavior with basic explanations.	No attempt to explain model insights is presented.
	CI/CD Pipeline	<p>The model uses a champion model framework where challengers compete on a fixed set of scoring metrics.</p> <p>The model uses an AI automation tool to experiment, and test lagged results.</p>			
	Scalability	The model is available to consume in batch or rest API through Databricks endpoint, MLFlow model or	The model is available to consume in batch or rest API through Databricks endpoint, MLFlow model or	The model is stored in a standard file including, but not limited to, pickle, joblib, ONNX, TensorFlow, etc.	The model is only accessible on local files or Jupyter notebooks.

		docker container webapp in ACR. A detailed plan for scalability is presented considering the costs of hosting and cybersecurity.	docker container webapp in ACR.		
3	Communication and story telling	A pitch deck is presented with an organized content: summary, problem statement, objective, methodology, experiments, results, next steps and appendix. The oral presentation uses formal and scientific language. Engages the audience effectively with clear, intuitive explanations and compelling storytelling.	A pitch deck is presented with an organized content: summary, problem statement, objective, methodology, experiments, results, next steps and appendix. The oral presentation uses formal and scientific language.	A pitch deck is presented but lacks some of the following sections: summary, problem statement, objective, methodology, experiments, results, next steps and appendix. The oral presentation uses informal and scientific language.	A pitch deck is presented but lacks most of the following sections: summary, problem statement, objective, methodology, experiments, results, next steps and appendix. The oral presentation uses informal and non-scientific language.

Resources

Churn

<https://www.salesforce.com/es/blog/churn-rate-tasa-abandono-clientes/>

<https://stripe.com/mx/resources/more/how-to-build-a-customer-churn-model-a-guide-for-businesses>

<https://learn.microsoft.com/en-us/fabric/data-science/customer-churn>

Recommendation system

<https://www.nvidia.com/en-us/glossary/recommendation-system/>

<https://developers.google.com/machine-learning/recommendation>

<https://www.ibm.com/think/topics/recommendation-engine>

Anomaly detection

<https://aws.amazon.com/what-is/anomaly-detection/>

<https://www.datacamp.com/tutorial/introduction-to-anomaly-detection>

<https://www.mathworks.com/discovery/anomaly-detection.html>