# Invent.ai – ML Engineer Case Study

This case study involves developing a machine learning model to analyze customer reviews of products on an e-commerce platform. You will design a complete machine learning pipeline, incorporating data analysis, model development, and deploying the model as an inference service to simulate a production environment.

## Dataset

The dataset provided includes customer reviews, product names, prices, discount_rates, categories, and user ratings made on the product. This dataset is intended to help you build and evaluate your solution. Below is a description of the dataset fields:

| column | description |
| --- | --- |
| review_date | Date the review has been made |
| product_id | Unique identifier of the product |
| product_name | Name of the product |
| product_category | Category information of the product (e.g., Sports & Outdoors) |
| user_id | Unique identifier of the customer |
| price | Resale price of the product in USD |
| review_text | Review of the customer on the product |
| discount_rate | Discount rate applied to the resale price, expressed as a value between 0 and 1 (e.g. 0.25 represents a 25% discount) |
| rating | Rating of the customer to the product, from 1 (worst) to 10 (best). |

This dataset is generated using a sample of Amazon Reviews Dataset and curating it according to needs of this case study.

# Problem Statement

The aim of this case study is to develop a predictive model for predicting a review rating assigned by a customer to a product based on information such as product price, product category, customer review.

Once a new customer review has arrived, your model should predict the rating the customer has given to the product. Predicting customer review ratings can be formulated as a Classification problem or a Regression problem depending on your design choice:

1. **Regression**: Predict the rating value (e.g., 8.3)
2. **Classification**: Predict the exact rating (e.g., 8)

Clearly explain your design choice and the reason behind it.

## Question 1: Exploratory Data Analysis (EDA)

In this question, you're asked to conduct an Exploratory Data Analysis (EDA) on the dataset to discover patterns, trends, and potential issues that will guide you through the model development process.

**Deliverable**: Prepare a summary of your findings using detailed observations. Highlight key takeaways that will guide the modeling process.

## Question 2: Model Development

After the EDA step, you are asked to develop a machine learning model of your choice to predict customer review ratings using the insights of the EDA step. Incorporate appropriate feature engineering steps, including text processing for review data.

Depending on your solution approach (Regression or Classification), choose an appropriate machine learning model algorithm.

You should also evaluate the model using relevant metrics (e.g., RMSE for regression, F1-score for classification), and justify your selected machine learning model.

**Deliverable**: Provide an explanation of your approach, including preprocessing steps, model architecture, and evaluation results.

## Question 3: Serve the model

In this question, you are asked to deploy your trained model as an inference service to simulate a production environment. Using the machine learning model developed in Question 2, design and develop an endpoint that predicts customer review ratings given a customer rating.

The review can be made on an existing product or a totally new product, and the reviewing customer can be an existing customer or a new customer. The prediction endpoint should at least take the customer review as an input, and return the predicted rating as an output. In addition to customer review, the endpoint can receive additional parameters based on your design.

**Deliverable:** Inference endpoint and guidance for using your endpoint.

**Examples**

- Customer review: "Great product! Totally Satisfied" → Predicted Rating: 9.0
- Customer review: "This product is a waste of time and money." → Predicted Rating 3.5

## Bonus Question: Model Registration and Versioning

In this question, you are asked to develop a robust system for managing machine learning models in production with their versions, registry steps, metadata etc. Develop a model registry that can serve models with versions. Your endpoint developed in question 3 should take the model version parameter and return predictions based on the specified version.

**Deliverable**: Updated inference endpoint with model versioning support

# General Instructions

- Your solution should be developed using Python. You are allowed to use any library and framework of your choice.
- You are allowed to use LLM API integrations or pretrained models only for generating text embeddings.
- (Bonus) Containerize your solution.
- (Bonus) Develop unit tests and provide it with your submission.

# Submission

- Do not publish your solution in public repositories.
- Submit all source code, configuration files, datasets, and scripts as a zip file.
- Write a README on how to use your solution.