Exercise: SVM vs. Naïve Bayes - Detecting Sarcasm in Reviews

Objective:

In this exercise, you will train two AI models, Support Vector Machine (SVM) and Naïve Bayes, to detect sarcasm in text reviews for a restaurant. Follow the steps below to complete the activity.

Step 1: Prepare the Dataset

You will create a dataset of reviews labeled as genuine (1) or sarcastic (0).
Below is an example:

Review:
- "The food was delicious, and the service was excellent!" (1)
- "Oh wow, this place is amazing… if you love waiting 3 hours for food!" (0)
- "I absolutely love this phone. The battery lasts all day!" (1)
- "Best phone ever! Died after 10 minutes of use!" (0)

Task:
- Write additional genuine and sarcastic reviews and add them to your dataset.

Step 2: Convert Text to Numerical Data

Since machine learning models can't directly process text, transform the reviews into numerical features using TF-IDF (Term Frequency - Inverse Document Frequency).

Step 3: Train Two Models

You will train two classification models: Support Vector Machine (SVM) and Naïve Bayes (NB).

Task:
- What is the difference between how SVM and Naïve Bayes classify text?

Step 4: Test Your Models

Write two new reviews, one sarcastic and one genuine, and check how both models classify them.

Task:

- Run the predictions and compare results.
- Did both models agree on the classification?

Final Discussion:
- Which model had higher accuracy?
- Which model was faster?
- When would SVM be better than Naïve Bayes and vice versa?

Bonus Challenge:

- Extract actual reviews for a real restaurant, Modify the dataset, and see how the models react to shorter vs. longer reviews.
- Try a different feature extraction method, such as CountVectorizer.

have fun!