Assignment-3: Numerical Problems

Introduction to Machine Learning Techniques

1. Given a dataset of product reviews and their corresponding sentiments (positive or negative):

Review	Sentiment
This product exceeded my expectations.	Positive
The item is of poor quality and not worth the price.	Negative
Highly recommended, a great purchase!	Positive
Disappointing product, terrible functionality.	Negative
Exceptional quality, worth every penny.	Positive

You receive a new review for classification:
Review: "This item is exceptional, and its functionality is impressive."
What sentiment would you assign to this review?

2. Suppose you're working with a dataset of restaurant reviews and their associated sentiments (positive or negative):

Review	Sentiment
The food was delicious, and the service exceptional.	Positive
Terrible experience, the staff was rude.	Negative
A fantastic dining experience, top-notch service.	Positive
Mediocre food and slow service.	Negative
Exceptional ambiance, and the food was exquisite.	Positive

You receive a new review for classification: **Review:** "The service is exceptional, and the food is exquisite." What sentiment would you assign to this review?

3. Imagine you're working on a dataset of hotel reviews and their associated sentiments (positive or negative):

Review	Sentiment
A wonderful stay, highly recommended.	Positive
Unpleasant experience, noisy surroundings.	Negative
Excellent service and beautiful accommodations.	Positive
Disappointing stay, uncomfortable beds.	Negative
Magnificent views, worth every penny.	Positive

You receive a new review for classification: **Review:** "The accommodations are beautiful, and the service is excellent." What sentiment would you assign to this review?

4. Predicting App Subscription Continuation: You are analyzing users of a subscription-based app to predict their likelihood of continuing the subscription. The dataset consists of usage

patterns and subscription history of customers. Your goal is to classify users as either "likely to continue the subscription" or "likely to churn."

Customer	Usage Patterns	Subscription History	Classification
1	Medium	Continual	Likely
2	Low	Churned	Unlikely
3	High	Continual	Likely
4	Low	Churned	Likely
5	High	Continual	Likely

Given the provided dataset, utilize the K-Nearest Neighbors (KNN) algorithm with K=3 to classify a new user with moderate usage and continual subscription history as either "likely to continue the subscription" or "likely to churn."

5. Online Retail Customer Retention: Your task is to classify online retail customers into "likely to repurchase" or "likely to churn" based on their purchase frequency and order history. The dataset comprises customers' purchase frequency and consistency of orders.

Customer	Purchase Frequency	Order History	Classification
1	Moderate	Consistent	Likely
2	Infrequent	Inconsistent	Unlikely
3	High	Consistent	Likely
4	Infrequent	Inconsistent	Unlikely
5	High	Consistent	Likely

Utilizing the K-Nearest Neighbors (KNN) algorithm with K=3, predict whether a new customer with infrequent purchases but a consistent order history is "likely to repurchase" or "likely to churn."

6. Social Media User Engagement: You're tasked with categorizing social media users as "engaged" or "inactive" based on their posting frequency, likes, and activity consistency. The dataset includes information on users' posting frequency, likes, and activity consistency.

User	Posts	Likes	Activity Frequency	Classification
1	High	High	Consistent	Engaged
2	Low	Low	Inconsistent	Inactive
3	High	High	Consistent	Engaged
4	Low	Low	Inconsistent	Inactive
5	High	High	Consistent	Engaged

Using the K-Nearest Neighbors (KNN) algorithm with K=3, predict the engagement level for a new user with high activity and frequent interaction - categorizing as either an engaged user or an inactive user.

7. Health and Fitness App User Retention: You're analyzing users of a health app based on their activity levels and consistency to predict their likelihood of staying or leaving the app. The dataset comprises details about users' activity levels and consistency.

User	Activity Level	Regularity	Classification
1	Moderate	Consistent	Likely
2	Low	Inconsistent	Unlikely
3	High	Consistent	Likely
4	Low	Inconsistent	Unlikely
5	High	Consistent	Likely

Using the K-Nearest Neighbors (KNN) algorithm with K=3, predict whether a new user with high activity and regular engagement is "likely to stay" or "likely to leave" the app.

8. E-commerce Cart Abandonment: Your task is to classify online shoppers into "likely to complete purchase" or "likely to abandon cart" based on their browsing history and purchase behavior. The dataset includes details of shoppers' browsing history and purchase behavior.

Shopper	Browsing History	Purchase Behavior	Classification
1	Extensive	Infrequent	Likely
2	Moderate	Frequent	Unlikely
3	Extensive	Infrequent	Likely
4	Moderate	Frequent	Unlikely
5	Extensive	Infrequent	Likely

Using the K-Nearest Neighbors (KNN) algorithm with K=3, classify a new shopper with extensive browsing history but infrequent purchases as "likely to complete the purchase" or "likely to abandon the cart."

9. Retail Sales Prediction: Consider the datset given below.

Field	Irrigation Level	Fertilizer Usage	Crop Quality
1	Moderate	Low	Low
2	High	High	High
3	Low	Low	Low
4	Moderate	High	High
5	High	High	High
6	Low	Low	Low

Predict the "Sales Performance" for a new day with the following observations:

- Daily Sales: $3500
- Foot Traffic: 50 customers

Utilize the KNN algorithm with K=3:

i. Compute the Euclidean distances between the new day and the existing dataset.
ii. Identify the three nearest days.
iii. Predict the new day's "Sales Performance" based on the majority class of the three nearest days.

10. Case Study: Crop Yield Prediction. Consider the following dataset about crop yield prediction.

Field	Irrigation Level	Fertilizer Usage	Crop Quality
1	Moderate	Low	Low
2	High	High	High
3	Low	Low	Low
4	Moderate	High	High
5	High	High	High
6	Low	Low	Low

Predict the "Crop Quality" for a new field with the following observations:

- Irrigation Level: Moderate
- Fertilizer Usage: High

Apply the KNN algorithm with K=3:

i. Calculate the distances between the new field and the existing dataset.
ii. Determine the three nearest fields.
iii. Predict the new field's "Crop Quality" based on the majority class of the three nearest fields.

11. Case Study: consider the Energy Consumption dataset given below.

Hour	Hourly Consumption (kW)	Temperature (°C)	Consumption State
1	900	30	High
2	700	25	Low
3	800	28	High
4	600	22	Low
5	950	32	High
6	750	27	Low

Forecast the "Consumption State" for a new hour with the following observations:

- Hourly Consumption: 800 kW
- Temperature: 28°C

Utilize the KNN algorithm with K=3:

i. Measure the Euclidean distances between the new hour and the existing dataset.
ii. Find the three nearest hours.
iii. Predict the new hour's "Consumption State" based on the majority class of the three nearest hours.

12. Hospital Patient Classification: Consider the dataset given below.

Customer_ID	Income Level	Loan Repayment Status	Outcome
1	High	On-Time	Positive
2	Low	Delayed	Negative
3	High	On-Time	Positive
4	Low	Delayed	Negative
5	Low	On-Time	Positive
6	High	On-Time	Positive

Suppose the dataset contains patient information such as "Blood Pressure," "Cholesterol Level," and their respective "Outcome" categorized as "Positive" or "Negative."

i. Calculate the entropy (H) for the "Outcome" variable within this dataset. Show your step-by-step calculations for entropy determination.

ii. Compute the Information Gain (IG) for both "Blood Pressure" and "Cholesterol Level" regarding the "Outcome" variable. Display your step-by-step calculations for information gain.

13. Consider the following dataset about Financial Risk Assessment.

Customer_ID	Income Level	Loan Repayment Status	Outcome
1	High	On-Time	Positive
2	Low	Delayed	Negative
3	High	On-Time	Positive
4	Low	Delayed	Negative
5	Low	On-Time	Positive
6	High	On-Time	Positive

Consider a dataset involving customer details, "Income Level," "Loan Repayment Status," and their "Outcome" classified as "Positive" or "Negative."

i. Compute the entropy (H) associated with the "Outcome" variable using this dataset. Display your detailed entropy calculations.
ii. Determine the Information Gain (IG) for both "Income Level" and "Loan Repayment Status" concerning the "Outcome" variable. Illustrate the step-by-step information gain calculations for both features.

14. Consider the following dataset about E-commerce Product Categorization

Product_ID	Price Range	Customer Ratings	Outcome
1	High	Positive	Positive
2	Low	Negative	Negative
3	High	Positive	Positive
4	Low	Negative	Negative
5	Low	Positive	Positive
6	High	Positive	Positive

Imagine a dataset capturing product details like "Price Range," "Customer Ratings," and their "Outcome" categorized as "Positive" or "Negative."

i. Calculate the entropy (H) for the "Outcome" variable within this dataset. Present your step-by-step entropy calculations.
ii. Compute the Information Gain (IG) for both "Price Range" and "Customer Ratings" with respect to the "Outcome" variable. Show your step-by-step information gain calculations for both features.