

ITM-618

Assignment-2 (10%): Clustering, Similarity, and confusion matrix

[Mandatory: You can do the calculation on paper, but please attach the calculations for all three questions along with your submission, without the calculation, the assignment will not be graded]

Please answer the following questions:

Q1. (Item similarity calculation)

4 Points

The following table represents the attributes and their values for three Persons (PersonA, PersonB, and PersonC)

Attribute	Person A	Person B	Person C
Age	35	50	55
Credit Score	730	820	680
Income	95K	120K	75K
Loan	100K	40K	100K

Using Euclidean distance, calculate the similarities between the three persons (PersonA, PersonB, and PersonC) and complete the following table with their similarity scores:

	PersonA	PersonB	PersonC
PersonA			
PersonB			
PersonC			

Now, use the table to find the most similar person to PersonA?

Q2. (Classification using k-NN)

7 Points

The following 7- customers have applied for credit cards to a credit card company. The column 'Approved' shows the credit card application status: 'No' means the application was not approved, and 'Yes' means the application was approved.

Customer	Age	Income	Loan	Credit Score	Approved
John	35	95K	300K	730	No
Rachel	22	50K	10K	830	Yes
Hannah	63	180K	20K	800	Yes
Tom	59	50K	80K	650	No
Nellie	25	60K	250K	680	No
James	45	120K	40K	850	Yes
Robin	42	70K	40K	780	Yes

Now, a new customer 'David' applies for a credit card, and David has the following profile:

Age: 37

Income: 85K

Loan: 45K

Credit Score: 790

Apply the k-NN algorithm to determine whether David's application will be Approved or Not!

Complete the following table with the approval status (Yes/No) for the given values of k:

k	Approve
1	
2	
3	
4	
5	
6	
7	

Q3. (Confusion matrix)

14 Points

Given this confusion matrix for machine learning classifier, answer the following questions:

- 3.1 how many classes are there in this data set?
- 3.2 What is the data set size?
- 3.3 How many samples were misclassified?
- 3.4 How many cases belong to class (C1)?
- 3.5 The accuracy of the classifier is _____
- 3.6 Compute the precision of the classifier and interpret it.
- 3.7 Compute the recall of the classifier and interpret it.

		Predicted labels		
		C1	C2	C3
Actual labels	C1	25	5	1
	C2	6	24	3
	C3	2	4	26