CS3244 Machine Learning – Final (New Topics) Sample Questions

Section 1: Model Evaluation

Q1-2: A car repair company is using an audio AI system to diagnose whether car engines are faulty or normal (Prediction). For comparison, a senior mechanic also examined the engines and rated them (Actual). These are the results from 10 cars.

Car ID	Prediction	Actual
1	Normal	Normal
2	Normal	Normal
3	Normal	Normal
4	Faulty	Normal
5	Normal	Normal
6	Faulty	Normal
7	Normal	Normal
8	Faulty	Faulty
9	Normal	Faulty
10	Faulty	Faulty

Q1) [4 Marks] MRQ: What metric(s), which is not misleading, should you use to report how well the model is doing?

a) Cosine Distance
b) Accuracy
c) Recall of Faulty prediction
d) Precision of Faulty prediction

Answer: ____

Justification / Working:

Q2) [4 Marks] Calculation: Calculate the F_1 score for the model.

Justification / Working:

Section 2: Data Processing and Feature Engineering

Q3) [2 Marks] T/F: PCA provides better features for regression than LDA.

a) True
b) False

Answer: ____

Justification / Working:

Q4) [6 Marks] Calculation: Using Tokenization, Stop Word Filtering, and Bag-of-Words, calculate the distance between the numeric representations of the two sentences $x^{(1)}$ and $x^{(2)}$.

Stop words: [“a”, “any”, “as”, “by”, “just”, “other”, “that”, “we”, “which”]

$x^{(1)}$: “A rose by any other name smells just as sweet.”
$x^{(2)}$: “That which we call a rose by any other word would smell as sweet.”

Justification / Working:

Section 3: Explainable AI

Q5) [2 Marks] T/F: Explaining a logistic regression model using LIME with a sparse linear model will produce identical weights.

a) True
b) False

Answer: ____

Justification / Working:

--- End --