DELHI PUBLIC SCHOOL FIROZABAD

(UNDER THE AEGIS OF DELHI PUBLIC SCHOOL SOCIETY EAST OF KAILASH NEW DELHI)

(A SENIOR SECONDARY SCHOOL)

AFFILIATED TO CBSE, AFFILIATION NO. 2133064 SCHOOL NO: 61225

Session– 2021-22

Subject- Artificial Intelligence

Class-X

Worksheet-1

Roll No._____ Name _____ Date_____

A. Tick the correct answers:-

1. In the Hold-Out approach, we divide our data into 3 parts:
Ans._____, _____, _____

2. Ink-Fold Cross-Validation, what does 'k' denote?
Ans. _____

3. A Confusion Matrix is used to solve which problem in Machine Learning?
Ans. _____

4. A Binary Classification problem has how many classes of data?
Ans. _____

5. The ratio of correctly predicted positive values to the total predicted positive values is known as:
Ans. _____

6. What do we call commonly used words in NLP?

Ans. _____

7. Autoencoder is a:
Ans. _____

8. What is the full form of PCA?
Ans. _____

9. Which of the following is not a Dimensionality Reduction technique?
Ans. _____

10. Which is the most important component of a Chatbot?
Ans. _____

B. Write (T) for True and (F) for False against the statements.

1. In k-Fold Cross-Validation, we divide our input dataset into k subsets of the same size.
2. A Confusion Matrix is a matrix of dimensions N*N which is used for evaluating the performance of a Regression Model in Machine Learning.
3. True Positive (TP) indicates the predicted value matched the actual value. Actual value was Positive and the model predicted a Positive value.
4. False Positive is also known as a Type 2 error.
5. Accuracy is the most important parameter to evaluate the performance of a Machine Learning model.
6. Bag-of-Words is used to count the number of occurrences of all Stop Words within a document.
7. Inverse Data Frequency (IDF)gives us the frequency of the words present in a document.
8. Stemming and Lemmatization are the same operations in NLP.

C. Select the suitable words and fill in the blanks.

1. _____is the ratio of correctly predicted positive values to all observations in the class.

2. F1 Score is the weighted average of_____and recall.

3. A_____compares the predicted values of a target variable to the actual values of the target variable for measuring the accuracy with which it is calculating output.

4. Whenever we train a Machine Learning model, be it on CV or NLP concepts, we need to have a sense of the_____with which it is calculating output.

5. _____is the data used to analyze the performance of a ML model during training.

6. TF-IDF stands for_____.

7. The process of splitting a sentence into words is known as_____.

8. Lemmatization is similar to_____but it brings context to the words, i.e. it link words with similar meaning to one root word.

9. But a machine does not understand our language and looks at a word in the form of a _____ which is a set of real numbers.

D. Answer the following:

1. Why do we evaluate Machine Learning models?

---

2. What is Cross-Validation?

---

3. What is a Confusion matrix used for?

---

4. What are True Positives, True Negatives, False Positives and False Negatives in a Confusion Matrix?

---

5. What is the difference between Precision and Recall? How do we calculate each?

---

6. What is Natural Language Processing?

---

7. What is a Chatbot?

---

8. What are Support Vector Machines?