SAMPLE QUESTION PAPER 1
CLASS X
ARTIFICIAL INTELLIGENCE (417)
TERM 2

Maximum Marks: 25 **Time: 60 Minutes**

General Instructions:

1. Please read the instructions carefully.
2. This Question Paper is divided into 3 sections, viz., Section A, Section B and Section C.
3. Section A is of 5 marks and has 6 questions on Employability Skills.
   (a) Questions numbers 1 to 4 are one mark questions. Attempt any three questions.
   (b) Questions numbers 5 and 6 are two marks questions. Attempt any one question.
4. Section B is of 12 marks and has 12 questions on Subject-specific Skills.
   (a) Questions numbers 7 to 12 are one mark questions. Attempt any four questions.
   (b) Questions numbers 13 to 18 are two marks questions. Attempt any four questions.
5. Section C is of 8 marks and has 3 competency-based questions.
   (a) Questions numbers 19 to 21 are four marks questions. Attempt any two questions.
6. Do as per the instructions given in the respective sections.
7. Marks allotted are mentioned against each section/question.

Section A

(3 + 2 = 5 Marks)

Answer any 3 questions out of the given 4 questions. **1 x 3 = 3**

1. Write two managerial functions of an entrepreneur. 1

Ans. (i) **Organizing:** It includes activities such as arranging, directing, guiding, coordinating and controlling the key business activities.

(ii) **Staffing:** IT includes activities related to human resource management, such as manpower planning, recruitment, selection, etc.

2. Enlist any 2 SDGs which are formulated to address problems related to conservation of natural resources. 1

Ans. (i) Responsible Consumption and Production

(ii) Sustainable Cities and Communities

3. List two points of how society plays a role in boosting entrepreneurship. 1

Ans. (i) Society creates the needs and demands of any product or service.

(ii) Society provides raw materials, manpower and capital supply.

4. Define the term 'E-waste'. 1

Ans. E-waste or Electronic waste refers to discarded electrical or electronic devices such as keyboards, smartphones, printers, refrigerators, batteries, etc.

Answer any 1 question out of the given 2 questions. **2 x 1 = 2**

5. "A lot of money is required to set up entrepreneurship." Justify this statement. 2

Ans. While a certain amount of capital is required for taking a business off the ground, it varies depending on the business. The key in entrepreneurship is to start with whatever resources available and grow steadily. It is not necessary that one should fear or deter from entrepreneurship due to lack of financial resources.

6. List two future aspects of 'Green Economy'. 2

Ans. (i) Green Economy will help maintain resources for present use as well as for future generations.

(ii) A green economy will observe fewer health concerns and promote a healthy lifestyle with the use of best technology to grow sustainably.

Section B

(4 + 8 = 12 Marks)

Answer any 04 questions out of the given 06 questions. **1 x 4 = 4**

7. What is the lemmatization and stemming of the word 'coming'? 1

Ans. Lemmatization: come

Stemming: com

8. How many tokens are there in the sentence given below? 1

Transparency is all about knowing who, why, what, how and how much. Seeking and receiving information is a human right that can act as a safeguard against corruption and increase trust in decision-makers and public institutions.

Ans. There are 43 tokens in the sentence.

9. What is a dictionary in NLP? 1

Ans. Dictionary is a list of all the unique words occurring in the corpus. While creating a dictionary, repeated words are written only once.

10. Identify any two stopwords in the given sentence: 1

Ozone is a gas that is naturally present in our atmosphere. Each ozone molecule contains three atoms of oxygen and is denoted chemically as O_3. Ozone is found primarily in two regions of the atmosphere.

Ans. Two stopwords in the sentence are: is, that

11. List two possible reasons for an AI model not being efficient. 1

Ans. (i) Lack of training data

(ii) Insufficient or wrong data

12. What type of error occurs when the actual value was positive but the model predicted a negative value? 1

Ans. Type 2 error

Answer any 4 questions out of the given 6 questions **2 x 4 = 8**

13. What is the significance of converting the text into a common case? Give example. 2

Ans. In text normalization, we go through various processes to normalize the text to a lower level. We transform the entire text to a similar case, preferably lower case, after removing stop words. This ensures that the case-sensitivity of the machine does not consider the same words as different just because of different cases. For example,

Apple, applE, aPPle, APPle, aPpLE are converted to apple.

14. Give an example of a situation where High Precision is not usable. 2

Ans. Let us consider a model that predicts a mail as Spam or Not Spam. Here, False Positive condition predicts that the mail is "spam" but it is "not spam" and the False Negative condition predicts that the mail is 'not spam' but it is "spam". In such cases, too many False Negatives will make the spam filter ineffective but False Positives may cause important mails to be missed and hence Precision is not always usable.

15. Write any two applications of Natural Language Processing. 2

Ans. (i) **Chatbots:** A chatbot is a computer program that is designed to simulate human conversation through voice commands or text chats or both.

(ii) **Sentiment and Emotion Analysis:** It is used to predict the general emotion of a text among several social media posts or even in the same post where emotion is not always explicitly expressed.

16. Explain the concept of Bag of Words. 2

Ans. Bag of Words is a Natural Language Processing model which helps in extracting features out of the text which can be helpful in machine learning algorithms. In bag of words, we get the occurrences of each word and construct the vocabulary for the corpus and create a set of vectors containing the frequency of these words in the document.

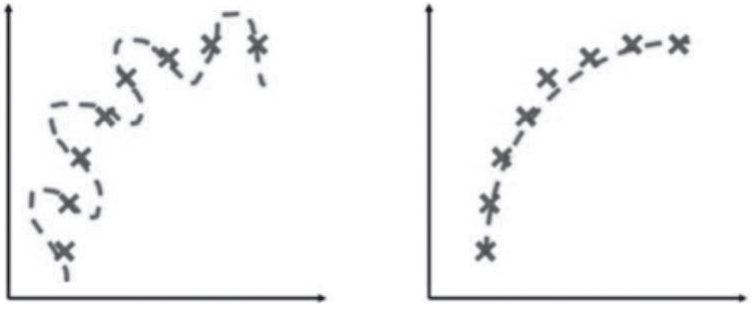
17. Explain which evaluation metric is more important for any case? 2

Ans. The F1 evaluation metric is more important in any case. F1 score sort maintains a balance between the precision and recall for the classifier. If the precision is low, the F1 is low and if the recall is low again F1 score is low. The F1 score is a number between 0 and 1 and is the harmonic mean of precision and recall

$$\text{F1 Score} = 2 * \frac{\text{Precision x Recall}}{\text{Precision + Recall}}$$

When we have a value of 1 (that is 100%) for both Precision and Recall, the F1 score would also be an ideal 1 (100%). It is known as the perfect value for F1 Score. As the values of both Precision and Recall range from 0 to 1, the F1 score also ranges from 0 to 1.

18. Observe the graphs carefully and classify them according to how well the model's output (dashed lines) matches the data samples (crosses): 2



(i) (ii)

Ans. (i) In the first graph, the model's performance tries to cover all data samples, even if they aren't aligned with the true function. This model is thought to be overfitting, with decreased accuracy.

(ii) In the second graph, the model's performance equals the true function, indicating that the model has optimal accuracy and is referred to as a perfect fit.

Section C
(Competency-based Questions)
(4 x 2 = 8 Marks)

Answer any 2 questions out of the given 3 questions. **4 x 2 = 8**

19. Through a step-by-step process, calculate TFIDF for the given corpus: 4

Document 1: To the swinging and the ringing

Document 2: of the bells, bells, bells

Document 3: Of the bells, bells, bells, bells

Document 4: Bells, bells, bells

Document 5: To the rhyming and the chiming of the bells.

Ans. Step 1: Create document vectors for the given documents (Term Frequency Table)

To	the	swinging	and	ringing	of	bells	rhyming	chiming
1	1	1	1	1	0	0	0	0
0	1	0	0	0	1	3	0	0
0	1	0	0	0	1	4	0	0
0	0	0	0	0	0	3	0	0
1	1	0	1	0	1	1	1	1

Step 2: Record the occurrence of the word in the document using the term frequency table (Document Frequency Table)

To	the	swinging	and	ringing	of	bells	rhyming	chiming
2	4	1	2	1	3	11	1	1

Step 3: Draw the inverse document frequency table wherein, we need to put the document frequency in the denominator while the total number of documents is the numerator. Here, the total number of documents are 5, hence inverse document frequency becomes:

To	the	swinging	and	ringing	of	bells	rhyming	chiming
5/1	5/4	5/1	5/2	5/1	5/3	5/11	5/1	5/1

Step 4: The formula of TFIDF for any word W becomes: TFIDF(W) = TF(W) * log (IDF(W))

To	the	swinging	and	ringing	of	bells	rhyming	chiming
1*log(5/2)	1*log(5/4)	1*log(5/1)	1*log(5/2)	1*log(5/1)	0*log(5/3)	0*log(5/11)	0*log(5/1)	0*log(5/1)
0*log(5/2)	1*log(5/4)	0*log(5/1)	0*log(5/2)	0*log(5/1)	1*log(5/3)	3*log(5/11)	0*log(5/1)	0*log(5/1)
0*log(5/2)	1*log(5/4)	0*log(5/1)	0*log(5/2)	0*log(5/1)	1*log(5/3)	4*log(5/11)	0*log(5/1)	0*log(5/1)
0*log(5/2)	0*log(5/4)	0*log(5/1)	0*log(5/2)	0*log(5/1)	0*log(5/3)	3*log(5/11)	0*log(5/1)	0*log(5/1)
1*log(5/2)	1*log(5/4)	0*log(5/1)	1*log(5/2)	0*log(5/1)	1*log(5/3)	1*log(5/11)	1*log(5/1)	1*log(5/1)

20. Students frequently experience difficulties receiving accurate alerts at appropriate time, including critical information such as campus interviews, training and placement activities, vacations and special announcements. It is just impossible for any institution to interact with all of its leads in real-time, especially after office hours or on weekends. Delays and impersonal responses cause applicants to lose interest and institute to lose credibility. 4

For the situation given above,

1. Create a 4W Project Canvas for this situation.
2. Can you create a model for managing such a situation?
3. Write two types of chatbots.

Ans. Who Canvas—Who has the problem?

Who are the stakeholders?	Students and their institutions are the stakeholders.
What do we know about them?	Students for not receiving timely notifications about important information and institutions for having a negative impact on enrolments.

Where Canvas—What is the nature of the problem?

What is the problem?	Due to lack of timely information, students frequently miss campus interviews, training and placement programs and other information regarding holidays and special announcements.
How do you know it is a problem?	Students should receive timely information.

Where Canvas—Where does the problem arise?

What is the context/situation in which the stakeholders experience this problem?	• When students miss important information regarding their careers. • Institutions lose credibility and reputation due to the lack of communication process.

Why Canvas—Why do you think it is a problem worth solving?

What would be of key value to the stakeholders?	• Students get timely information and even they can share their concerns related to any facility. • Institutions get a medium to help students with admissions, student support and providing immediate access to information.
How would it improve their situation?	• Students would be able to get every information they need and connect with the institute round the clock. • Institutions can attract potential students in an effective and engaging manner.

2. The model to manage the above situation would be deploying a chatbot. Chatbots provide automated responses to prospective student queries and help them carry out basic tasks.

3. (i) Rule-based chatbot

(ii) Smart bot using ML

21. What is a Confusion Matrix? Explain in detail with the help of an example. 4

Ans. A Confusion Matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

Evaluation of the performance of a classification model is based on the counts of test records correctly and incorrectly predicted by the model. Therefore, Confusion Matrix provides a more insightful picture which is not only the performance of a predictive model but also which classes are being predicted correctly and incorrectly, and what type of errors are being made. Confusion Matrix is useful for measuring Recall (also known as Sensitivity), Precision, Accuracy and F1 Score.

The following Confusion Matrix table illustrates how the 4-classification metrics are calculated (TP, FP, FN, TN), and how is our predicted value compared to the actual value in a confusion matrix.

Confusion Matrix	Actual Value:1	Actual Value: 0
Prediction: 1	True Positive (TP)	False Positive (FP)
Prediction: 0	False Negative (FN)	True Negative (TN)

The target variable has two values: Positive or Negative

The columns represent the actual values of the target variable.

The rows represent the predicted values of the target variable—True Positive, True Negative, False Positive and False Negative in a Confusion Matrix.

The values of the Confusion matrix are:

True Positive (TP): The predicted value matches the actual value.

The actual value was positive and the model predicted a positive value.

True Negative (TN): The predicted value matches the actual value.

The actual value was negative and the model predicted a negative value.

False Positive (FP): The predicted value was falsely predicted.

The actual value was negative but the model predicted a positive value.

Also known as Type 1 error.

False Negative (FN): The predicted value was falsely predicted.

The actual value was positive but the model predicted a negative value also known as the Type 2 error.

Example Case: Loan (Good loan & Bad loan)

The Confusion Matrix for loan will be:

TP—Bad loans are correctly predicted as bad loans.

TN—Good loans are correctly predicted as good loans.

FP—Good loans are incorrectly predicted as bad loans.

FN—Bad loans are incorrectly predicted as goods loans.

The banks would lose a lot of money if the actual bad loans were predicted as good loans as a result of loans not being repaid. Banks, on the other hand, will not be able to generate revenue if good loans are projected to be bad loans. As a result, False Negatives are more expensive than False Positives.

SAMPLE QUESTION PAPER 2
CLASS X
ARTIFICIAL INTELLIGENCE (417)
TERM 2

Maximum Marks: 25 **Time: 60 Minutes**

General Instructions:

1. Please read the instructions carefully.
2. This Question Paper is divided into 3 sections, viz., Section A, Section B and Section C.
3. Section A is of 5 marks and has 6 questions on Employability Skills.
   (a) Questions numbers 1 to 4 are one mark questions. Attempt any three questions.
   (b) Questions numbers 5 and 6 are two marks questions. Attempt any one question.
4. Section B is of 12 marks and has 12 questions on Subject Specific Skills.
   (a) Questions numbers 7 to 12 are one mark questions. Attempt any four questions.
   (b) Questions numbers 13 to 18 are two marks questions. Attempt any four questions.
5. Section C is of 8 marks and has 3 competency-based questions.
   (a) Questions numbers 19 to 21 are four marks questions. Attempt any two questions.
6. Do as per the instructions given in the respective sections.
7. Marks allotted are mentioned against each section/question.

Section A
(3 + 2 = 5 Marks)

Answer any 3 questions out of the given 4 questions. **1 x 3 = 3**

1. Write any two positive impacts of entrepreneurship on society. 1

Ans. (i) Creates jobs and employment opportunities

(ii) Promotes welfare of the society

2. Write two man-made disruptions that cause ecological balance. 1

Ans. (i) Deforestation

(ii) E-waste generation

3. Explain the role of an entrepreneur as an 'Innovator'. 1

Ans. Entrepreneurs as innovators bring new products or services to the market or improve existing products or services.

4. What are the four 4Rs used to manage waste in a Green Economy? 1

Ans. REFUSE, REDUCE, REUSE and RECYCLE

Answer any 1 question out of the given 2 questions. **2 x 1 = 2**

5. "Entrepreneurship is a process of constant learning." Justify this statement. 2

Ans. Entrepreneurs grow when they are constantly learning new things. Starting a business needs a lot of energy, creativity and the willingness to step outside one's comfort zone. It is not always about how much you know but how quickly you learn. So, entrepreneurs need to constantly learn in order to grow the business and stay competitive.

6. What do you mean by Green Economy? 2

Ans. According to UNEP (United Nations Environment Programme), a green economy is defined as low carbon, resource-efficient and socially inclusive. In a green economy, growth in employment and income is driven by public and private investment into such economic activities, infrastructure and assets that allow reduced carbon emissions and pollution, enhanced energy and resource efficiency, and prevention of loss to biodiversity and ecosystem services.

Section B
(4 + 8 = 12 Marks)

Answer any 4 questions out of the given 6 questions. **1 x 4 = 4**

7. What will be the output of the word 'cities' after the lemmatization and stemming process? 1

Ans. Lemmatization: city

Stemming: citi

8. Write two types of data that can be used in Natural Language Processing applications. 1

Ans. In Natural Language Processing applications, the data in the form of written words and/or spoken words can be used.

9. What is the meaning of (a) Syntax and (b) Semantics in NLP? 1

Ans. (a) Syntax: The grammatical structure of a sentence is called Syntax.

(b) Semantics: The meaning of the sentence is called Semantics.

10. What are stop words? 1

Ans. Stop words are the most common words in a language that do not carry important meanings and are usually removed from the text. For example, "the", "a", "on", "is", "all".

11. Which two parameters are considered for evaluation of a model? 1

Ans. Prediction and Reality are the two parameters considered for the evaluation of a model.

12. Write the formula to calculate F1 score. 1

Ans. F1 Score = $2*\frac{\text{Precision * Recall}}{\text{Precision + Recall}}$

Answer any 4 questions out of the given 6 questions. **2 x 4 = 8**

13. Differentiate between Script bot and Smart bot. 2

Script bot	Smart bot
Script bots are based on scripts and FAQs without any AI.	Smart bots are based on ML.
Script bots are easy to make.	Smart bots are difficult to make.
Script bots have limited functionality and are related to the closed domain.	Smart bots have wide functionality and are related to the open domain.

14. What are the reasons for an AI model not being efficient? 2

Ans. The reasons for an AI model not being efficient:

1. **Lack of training data:** If the data is not sufficient for developing an AI model or if the data is missed while training the model, it will not be efficient.
2. **Unauthenticated data:** If the data is not authenticated and correct, then the model will not give good results.
3. **Inefficient coding or wrong algorithms:** If the writen algorithms are not correct and relevant, the model will not give the desired routput.
4. **Not tested:** If the model is not tested properly, it will not be efficient.
5. **Less accuracy:** If the model gives less accuracy scores in production or test data or if it is not able to generalize well on unseen data, the model is not efficient.

15. How does text normalization reduce the vocabulary of a corpus? 2

Ans. In text normalization, the text is reduced to a minimum vocabulary by—

(i) eliminating stop words, special characters and numbers from the vocabulary of a corpus and

(ii) deleting the affixes and converting the words to their base form using stemming and lemmatization.

16. What is TFIDF? Write its formula. 2

Ans. TFIDF stands for Term Frequency–Inverse Document Frequency. It helps us in identifying the value for each word in a corpus. It is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. The number of times a word appears in a document is divided by the total number of words in the document. Every document has its own term frequency. The formula of TFIDF for any word W is:

TFIDF(W) = TF(W) * log(IDF(W))

17. What are True Positive and False Positive situations in the confusion matrix regarding an AI model? 2

Ans. **True Positive (TP):** The predicted value matches the actual value, *i.e.*, the actual value was positive and the model predicted a positive value.

False Positive (FP): The predicted value was falsely predicted, i.e., the actual value was negative but the model predicted a positive value. It is also known as Type1 error.

18. Explain the role of NLP in the sentiment analysis of human beings. 2

Ans. We can use sentiment analysis to predict the general emotion of a text. The goal of sentiment analysis is to identify sentiment among several posts or even in the same post where emotion is not always explicitly expressed. Companies use Natural Language Processing applications, such as sentiment analysis, to identify opinions and sentiment online to help them understand what customers think about their products and services.

For example, I love this car, or I don't like this product and will never purchase it in the future.

Sentiment analysis understands sentiment in context to help better understand what's behind an expressed opinion, which can be extremely relevant in understanding and driving purchasing decisions.

Section C
(Competency-Based Questions)
(4 x 2 = 8 Marks)

Answer any 2 questions out of the given 3 questions. **4 x 2 = 8**

19. Through a step-by-step process, calculate TFIDF for the given corpus: 4

Document 1: Row, row, row your boat

Document 2: Gently down the stream

Document 3: Merrily, merrily, merrily, merrily

Document 4: Life is but a dream

Ans. **Step 1:** Create document vectors for the given documents (Term Frequency Table)

row	your	gently	down	the	stream	merrily	life	is	but	a	dream
3	1	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	0	0	0	0	0	0
0	0	0	0	0	0	4	0	0	0	0	0
0	0	0	0	0	0	0	1	1	1	1	1

Step 2: Record the occurrence of the word in the document using the term frequency table (Document Frequency Table)

row	your	gently	down	the	stream	merrily	life	is	but	a	dream
3	1	1	1	1	1	4	1	1	1	1	1

Step 3: Draw the inverse document frequency table wherein we need to put the document frequency in the denominator while the total number of documents is the numerator. Here, the total number of documents are 4, hence inverse document frequency becomes:

row	your	gently	down	the	stream	merrily	life	is	but	a	dream
4/3	4/1	4/1	4/1	4/1	4/1	4/4	4/1	4/1	4/1	4/1	4/1

Step 4: The formula of TFIDF for any word W becomes: TFIDF(W) = TF(W) * log (IDF(W))

row	your	gently	down	the	stream	merrily	life	is	but	a	dream
3*log (4/3)	1*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/4)	0*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/1)
0*log (4/3)	0*log (4/1)	1*log (4/1)	1*log (4/1)	1*log (4/1)	1*log (4/1)	0*log (4/4)	0*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/1)
0*log (4/3)	0*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/1)	4*log (4/4)	0*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/1)
0*log (4/3)	0*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/1)	0*log (4/4)	1*log (4/1)	1*log (4/1)	1*log (4/1)	1*log (4/1)	1*log (4/1)

20. All the companies want their clients to know about them and they can get more projects or sell their products or services. So, the most primary way to do this is to have a good interactive website in the advanced world of digital marketing. But in this advanced world, just having a website is not enough as it does not cover all the details and cannot solve the queries of the clients. Maybe a comment section is an answer, but it's tedious work, and 'time is money'. Your client will not wait too long for your response. And you cannot assign humans to answer everyone at the same time 24 x 7. 4

For the situation given above:

1. Write the problem statement template.
2. Why do we need to collect data?
3. What does an NLP-based AI model require?

Ans. 1. The problem statement template for the given scenario would be:

Our	companies who are selling products and services.	Who?
have a problem that	they are not being able to expand their business to a broader audience.	What?
while	they need a platform to accelerate their business and provide instant and real-time assistance.	Where?
An ideal solution would be	to implement a chatbot to cater to customer complaints and generic requests.	Why?

2. Data collection is essential in order to understand the customer's queries and requirements, after which the computer can interpret the words and their meanings.
3. An NLP-based AI model requires normalized text with a minimum vocabulary, which was obtained as a result of data cleaning and pre-processing during the data exploration phase.

21. Calculate Accuracy, Precision, Recall and F1 Score for the following Confusion Matrix on Water Shortage in Schools: Also suggest which metric would not be a good evaluation parameter here and why? 4

The Confusion Matrix (Water shortage in school)	Reality: 1	Reality: 0
Prediction: 1	68	8
Prediction: 0	8	16

Ans. **Accuracy:** Accuracy is defined as the percentage of correct predictions out of all the observations.

$$Accuracy = \frac{Correct\ Prediction}{Total\ Cases} \times 100\%$$

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100\%$$

Where

TP is True Positive,

TN is True Negative

FP is False Positive

FN is False Negative

$$Accuracy = \frac{(68 + 16)}{(68 + 16 + 8 + 8)} \times 100\%$$

$$= \frac{84}{100} \times 100\%$$

$$= 0.84$$

Precision: Precision is defined as the percentage of true positive cases versus all the cases where the prediction is true.

$$\text{Precision} = \frac{True\ Positive}{All\ Predicted\ Positives}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$= 68/(68+8)$$

$$= 0.89$$

Recall: It is defined as the ratio of positive cases that are correctly identified.

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$= 68/(68+8)$$

$$= 68/76$$

$$= 0.89$$

F1 Score: F1 score is defined as the measure of balance between precision and recall.

$$\text{F1 Score} = 2 \times \frac{Precision\ x\ Recall}{Precision + Recall}$$

Therefore,

$$= 2x((0.89x0.89)/(0.89+0.89))$$

$$= 0.89$$

Accuracy = 0.84%

Precision = 0.89%

Recall = 0.89%

F1 Score =0.89

Here Accuracy, Precision, Recall and F1 score all are same.