

# CBSE | DEPARTMENT OF SKILL EDUCATION

## ARTIFICIAL INTELLIGENCE (SUBJECT CODE 417)

### CLASS X (SESSION 2021-2022) SAMPLE QUESTION PAPER FOR TERM - II

Max. Time Allowed: 1 Hour (60 min)

Max. Marks: 25

#### General Instructions:

1. Please read the instructions carefully
2. This Question Paper is divided into 03 sections, viz., Section A, Section B and Section C.
3. Section A is of 05 marks and has 06 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 08 marks and has 03 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. Each question is of the mark.		1 x 3 = 3
Q.1	<p><b>Write any two qualities for being a good entrepreneur.</b> <b>Ans:</b> <b>Any two of the following points-</b></p> <ul style="list-style-type: none"><li>• They are confident. They believe in themselves and their abilities.</li><li>• They keep trying new ideas in their business.</li><li>• They are patient.</li><li>• They are creative and think differently about business ideas.</li><li>• They take responsibility for their actions.</li><li>• They make decisions after thinking about them.</li><li>• They work hard.</li><li>• They do not give up when they face a difficulty</li></ul> <p>(½ mark for each point; ½ x 2=1)</p>	1
Q.2	<p><b>What is sustainable development?</b> <b>Ans:</b> Sustainable development is the development that satisfies the needs of the present without compromising the capacity of future generations, guaranteeing the balance between economic growth, care for the environment and social well-being. (1 mark for correct answer/explanation)</p>	1
Q.3	<p><b>Entrepreneurship has a positive impact on society. Write down any two.</b> <b>Ans:</b></p> <ol style="list-style-type: none"><li>1. Some of them work towards saving the environment.</li><li>2. Some of them give money to build schools and hospitals.</li></ol> <p>(½ mark for each point; ½ x 2=1)</p>	1
Q.4	<p><b>How many sustainable development goals were formulated by the United Nations?</b> <b>Ans:</b> There are 17 sustainable development goals formulated by the United Nations. (1 mark for correct answer)</p>	1

<b>Answer any 1 question out of the given 2 questions. Each question is of mark.</b>		<b>2 x 1 = 2</b>
<b>Q.5</b>	<p><b>“Entrepreneurs are born, not made.” Do you agree with this statement? Justify your answer.</b></p> <p><b>Ans:</b>          No, this is a myth/misconception about entrepreneurship.          Being an entrepreneur starts with a way of thinking. One must believe that anything is possible and it shall be achieved. It starts with thinking of an idea that you want to work on, making it different.          (1 mark for the option(NO);1 mark for correct explanation)</p>	<b>2</b>
<b>Q.6</b>	<p><b>Enlist any 2 SDGs which are formulated to address the problems related to water?</b></p> <p><b>Ans:</b>          Clean water and sanitation          Life below water          Responsible consumption, and production          (any 2 SDG s related to water; 1 mark for each SDG)</p>	<b>2</b>

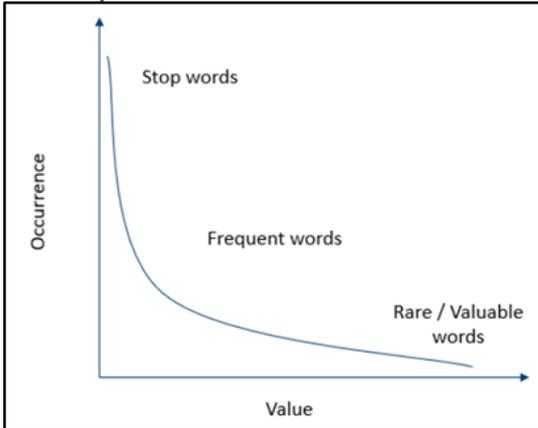
## SECTION B

**(4 + 8 = 12 marks)**

<b>Answer any 04 questions out of the given 06 questions</b>		<b>1 x 4 = 4</b>
<b>Q.7</b>	<p><b>What will be the output of the word “<i>studies</i>” if we do the following:</b></p> <p>a. Lemmatization          b. Stemming</p> <p><b>Ans:</b>          The output of the word after <b>lemmatization</b> will be <b>study</b>.          The output of the word after <b>stemming</b> will be <b>studi</b>.          (½ mark for for lemmatization, ½ mark for stemming)</p>	<b>1</b>
<b>Q.8</b>	<p><b>How many tokens are there in the sentence given below?</b>  <i>Traffic Jams have become a common part of our lives nowadays. Living in an urban area means you have to face traffic each and every time you get out on the road. Mostly, school students opt for buses to go to school.</i></p> <p><b>Ans:</b>          46 tokens are there in the given sentence          (1 mark for correct answer)</p>	<b>1</b>
<b>Q.9</b>	<p><b>What is a corpus?</b></p> <p><b>Ans:</b>          The term used to describe the whole textual data from all the documents altogether is known as corpus.          (1 mark for any correct explanation)</p>	<b>1</b>
<b>Q.10</b>	<p><b>Identify any 2 stopwords in the given sentence:</b>  <i>Pollution is the introduction of contaminants into the natural environment that cause adverse change.The three types of pollution are air pollution, water pollution and land pollution.</i></p> <p><b>Ans:</b>          Stopwords in the given sentence are: <b>is, the, of, that, into, are, and</b>          (any two correct answers; ½ mark each)</p>	<b>1</b>
<b>Q.11</b>	<p><b>Why should we avoid using the training data for evaluation?</b></p> <p><b>Ans:</b>          This is because our model will simply remember the whole training set, and will therefore always predict the correct label for any point in the training set.          (1 mark for any correct explanation)</p>	<b>1</b>
<b>Q.12</b>	<p><b>What should be the value of F1 score if the model needs to have 100% accuracy?.</b></p> <p><b>Ans:</b>          The model will have an F1 score of 1 if it has to be 100% accurate.          (1 mark for correct answer)</p>	<b>1</b>

<b>Answer any 04 questions out of the given 06 questions</b>		<b>2 x 4 = 8</b>
<b>Q.13</b>	<p><b>“Automatic summarization is used in NLP applications”. Is the given statement correct? Justify your answer with an example.</b></p> <p><b>Ans:</b> Yes, the given statement is correct. Automatic summarization is relevant not only for summarizing the meaning of documents and information, but also to understand the emotional meanings within the information, such as in collecting data from social media. Automatic summarization is especially relevant when used to provide an overview of a news item or blog post, while avoiding redundancy from multiple sources and maximizing the diversity of content obtained.</p> <p>(1 mark for explanation, 1 mark for example)</p>	<b>2</b>
<b>Q.14</b>	<p><b>Give an example of a situation wherein false positive would have a high cost associated with it.</b></p> <p><b>Ans:</b> Let us consider a model that predicts that a mail is spam or not. If the model always predicts that the mail is spam, people would not look at it and eventually might lose important information. Here False Positive condition (Predicting the mail as spam while the mail is not spam) would have a high cost.</p> <p>(2 marks for any correct example with explanation; 1 marks can be given if only explanation is written without example)</p>	<b>2</b>
<b>Q.15</b>	<p><b>Write any two applications of TFIDF</b></p> <p><b>Ans:</b></p> <ol style="list-style-type: none"> <li>1. Document Classification Helps in classifying the type and genre of a document.</li> <li>2. Topic Modelling It helps in predicting the topic for a corpus.</li> <li>3. Information Retrieval System To extract the important information out of a corpus.</li> <li>4. Stop word filtering Helps in removing the unnecessary words out of a text body.</li> </ol> <p>(1 mark for each application name/explanation)</p>	<b>2</b>
<b>Q.16</b>	<p><b>Write down the steps to implement bag of words algorithm.</b></p> <p><b>Ans:</b> The steps to implement bag of words algorithm are as follows:</p> <ol style="list-style-type: none"> <li>1. Text Normalisation: Collect data and pre-process it</li> <li>2. Create Dictionary: Make a list of all the unique words occurring in the corpus. (Vocabulary)</li> <li>3. Create document vectors: For each document in the corpus, find out how many times the word from the unique list of words has occurred.</li> <li>4. Create document vectors for all the documents.</li> </ol> <p>(½ mark for each step)</p>	<b>2</b>
<b>Q.17</b>	<p><b>What is a confusion matrix? What is it used for?</b></p> <p><b>Ans:</b> The confusion matrix is used to store the results of comparison between the prediction and reality. From the confusion matrix, we can calculate parameters like recall, precision, F1 score which are used to evaluate the performance of an AI model.</p> <p>(1 mark for definition, 1 mark for use)</p>	<b>2</b>

Explain from the given graph, how the value and occurrence of a word are related in a corpus?



Q.18

2

**Ans:**

As shown in the graph, occurrence and value of a word are inversely proportional. The words which occur most (like stop words) have negligible value. As the occurrence of words drops, the value of such words rises. These words are termed as rare or valuable words. These words occur the least but add the most value to the corpus.

(complete explanation 2 marks)

## SECTION C

(2 x 4 = 8 marks)

### (COMPETENCY-BASED QUESTIONS)

Answer any 02 questions out of the given 03 questions

Q.19

Through a step-by-step process, calculate TFIDF for the given corpus

**Document 1:** *Johny Johny, Yes Papa,*

**Document 2:** *Eating sugar? No Papa*

**Document 3:** *Telling lies? No Papa*

**Document 4:** *Open your mouth, Ha! Ha! Ha!*

**Ans:**

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

Johny	Yes	Papa	Eating	Sugar	No	Telling	Lies	Open	your	Mouth	Ha
2	1	1	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	0	0	0	0	0	0
0	0	1	0	0	1	1	1	0	0	0	0
0	0	0	0	0	0	0	0	1	1	1	3

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

Johny	Yes	Papa	Eating	Sugar	No	Telling	Lies	Open	your	Mouth	Ha
1	1	3	1	1	2	1	1	1	1	1	1

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:

Johny	Yes	Papa	Eating	Sugar	No	Telling	Lies	Open	your	Mouth	Ha
4/1	4/1	4/3	4/1	4/1	4/2	4/1	4/1	4/1	4/1	4/1	4/1

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

Johny	Yes	Papa	Eating	Sugar	No	Telling	Lies	Open	your	Mouth	Ha
$2 * \log(4/1)$	$1 * \log(4/1)$	$1 * \log(4/3)$	$0 * \log(4/1)$	$0 * \log(4/1)$	$0 * \log(4/2)$	$0 * \log(4/1)$					
$0 * \log(4/1)$	$0 * \log(4/1)$	$1 * \log(4/3)$	$1 * \log(4/1)$	$1 * \log(4/1)$	$1 * \log(4/2)$	$0 * \log(4/1)$					
$0 * \log(4/1)$	$0 * \log(4/1)$	$1 * \log(4/3)$	$0 * \log(4/1)$	$0 * \log(4/1)$	$1 * \log(4/2)$	$1 * \log(4/1)$	$1 * \log(4/1)$	$0 * \log(4/1)$	$0 * \log(4/1)$	$0 * \log(4/1)$	$0 * \log(4/1)$
$0 * \log(4/1)$	$0 * \log(4/1)$	$0 * \log(4/3)$	$0 * \log(4/1)$	$0 * \log(4/1)$	$0 * \log(4/2)$	$0 * \log(4/1)$	$0 * \log(4/1)$	$1 * \log(4/1)$	$1 * \log(4/1)$	$1 * \log(4/1)$	$3 * \log(4/1)$

(1 mark for each correct table)

**Q.20**

***The world is competitive nowadays. People face competition in even the tiniest tasks and are expected to give their best at every point in time. When people are unable to meet these expectations, they get stressed and could even go into depression. We get to hear a lot of cases where people are depressed due to reasons like peer pressure, studies, family issues, relationships, etc. and they eventually get into something that is bad for them as well as for others. So, to overcome this, Cognitive Behavioural Therapy (CBT) is considered to be one of the best methods to address stress as it is easy to implement on people and also gives good results. This therapy includes understanding the behaviour and mindset of a person in their normal life. With the help of CBT, therapists help people overcome their stress and live a happy life.***

For the situation given above,

1. Write the problem statement template
2. List any two sources from which data can be collected.
3. How do we explore the data?

**Ans**

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

Our	people undergoing stress	Who?
have a problem that	they are not being able to share their feelings	What?
while	they need help to vent out their emotions	Where?
An ideal solution would be	To provide a platform to share their thoughts anonymously and suggest help whenever required.	Why?

2. Data can be collected from one of the following sources:
  - a. surveys
  - b. observing therapist's sessions
  - c. databases available on the internet
  - d. interviews
3. Once the textual data has been collected, it needs to be processed and cleaned so that an easier version can be sent to the machine. Thus, the text is normalised through various steps and is lowered to minimum vocabulary since the machine does not require grammatically correct statements but the essence of it.

(2 marks for problem statement template; ½ marks for each data sources; 1 mark for correct explanation of data exploration)

**4**

Q.21

Take a look at the confusion matrix:

4

The Confusion Matrix		Reality	
		Yes	No
Prediction	Yes	True Positive (TP)	False Positive (FP)
	No	False Negative (FN)	True Negative (TN)

How do you calculate F1 score?

**Ans:**

We begin the calculation by first using the formula to calculate Precision  
Precision is defined as the percentage of true positive cases versus all the cases where the prediction is true. That is, it takes into account the True Positives and False Positives.

$$\text{Precision} = \frac{\text{True Positive}}{\text{All Predicted Positives}} \times 100\%$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\%$$

Next, we calculate recall as the fraction of positive cases that are correctly identified.

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

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

Finally, we calculate the F1 score as the measure of balance between precision and recall.

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

(1 mark for precision formula; 1 mark for recall formula; 1 mark for F1 score formula; 1 mark for explanation)