10.00		2/10 17.1	11 9 17	ospi okas	1/160 0 Ret
RN					



PES University, Bengaluru

(Established under Karnataka Act No. 16 of 2013)

UE20CS933

October 2024: END SEMESTER ASSESSMENT (ESA)
M TECH DATA SCIENCE AND MACHINE LEARNING_SEMESTER II

UE20CS933 - NATURAL LANGUAGE PROCESSING

	me:	3 Hrs Answer All Q	uestions Max Marks	: 100				
		INSTRU	CTIONS					
		All questions are compulsory. Section A should be handwritten in the ans Section B and C are coding questions which	wer script provided. h have to be answered in the system.					
		SECTION A	- 20 MARKS	5				
1	What do you mean by NLP? list 4 real-world applications of NLP. (Marks)							
	(a)	2)						
	b)	2) What are Large Language Models (LLMs)? List any four limitations/drawbacks of LLMs (Marks 1 + 4) Discuss RNN cell and its drawback. How LSTM overcomes RNN drawbacks?						
			사용보다 : '이용 이상 10 Act (1974 - 1984) - 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.	5				
1	d)	(Marks 3+2) Explain Named Entity Recognition with an example.						
	177		선택은 유럽 전에 맞춰 아래를 살아 있다면 하면 하면 아들이 어린다면 하는데 아들은 그들은 그 것이 그리고 하는 것이 그리고 있다면 하는데 되었다면 하는데 되었다면 하는데 되었다면 하는데					
		SECTION B -	40 MARKS	20				
_		SECTION B -	40 MARKS e following preprocessing steps:	20				
		SECTION B - Given the dataset reviews.csv, perform the	40 MARKS e following preprocessing steps:	20				
		Given the dataset reviews.csv, perform th	e following preprocessing steps:	20				
	a)	Given the dataset reviews.csv, perform the Load the dataset and display the first 5 ro	e following preprocessing steps: ws.	20 5				
-	a)	Given the dataset reviews.csv, perform the Load the dataset and display the first 5 ro	e following preprocessing steps: ws.	20510				
	a) b)	Given the dataset reviews.csv, perform the Load the dataset and display the first 5 room Remove any duplicate reviews. Clean the text by removing punctuation, or the second sec	e following preprocessing steps: ws.	5				
	a) b)	Given the dataset reviews.csv, perform the Load the dataset and display the first 5 room Remove any duplicate reviews. Clean the text by removing punctuation, or the load of	ws. converting to lowercase, and removing	5				
-	a) b)	Given the dataset reviews.csv, perform the Load the dataset and display the first 5 rown Remove any duplicate reviews. Clean the text by removing punctuation, of stopwords.	ws. converting to lowercase, and removing dings using TF-IDF.	5				
	a) b)	Given the dataset reviews.csv, perform the Load the dataset and display the first 5 room Remove any duplicate reviews. Clean the text by removing punctuation, or the load of	ws. converting to lowercase, and removing dings using TF-IDF.	5				

	_	_	_								_	_
	15.50	Win	1170	Ph.	YAY.	W.		MA	74.00	1.00	104	111
RN	10/77	100	1000	1010	1133	13 72	4年	1.88	100		7107	

		SECTION C -40 MARKS	
4		Model building using Naive Bayes	20
	a)	Split the dataset into training and testing sets.	5
	b)	Train a Naive Bayes classifier on the training set.	10
	c)	Evaluate the model on the testing set and display the accuracy.	5
5		Model Building using LSTM	20
	a)	Convert the cleaned text into sequences using Tokenizer.	5
	b)	Build and compile an LSTM model for sentiment analysis.	10
	c)	Train the model and evaluate its performance on the testing set.	5