MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY, BHOPAL

Department of Mathematics Bioinformatics and Computer Applications

Mid Term Examination, Sept. 2023 21212013

Course: MCA **Semester: V**

Subject: Natural Language Processing **Subject Code: CA 802**

Time: 90 Minutes **Maximum Marks: 20**

Note: Attempt All Questions. Each Question Carry Equal Marks.

Q.	Questions	Marks
1.	a) Justify your answer with examples How NLP is related or involved with AI systems?	5 (2+3)
	b) Explain Regular Expressions and Finite-State automata.	
2.	a) What do you mean lexical ambiguity and syntactic ambiguity in natural language? What are the way to resolve these ambiguities?	5 (2+3)
	b) Write and explain the minimum edit distance algorithm.	
3.	Given the following CFG grammar from ATIS System, USA. Perform syntactic analysis of the following sentence using any of the parsing method. "Book the flight through Houston." **Grammer** S → NP VP S → Aux NP VP S → VP NP → Pronoun NP → Proper-Noun NP → Det Nominal Nominal → Noun Nominal → Nominal Noun Nominal → Nominal PP VP → Verb VP → Verb NP VP → VP PP PP → Prep NP **Lexicon** Det → the | a | that | this Noun → Book | flight | meal | money Verb → Book | include | prefer Pronoun → I | he | she | me Proper-Noun → Houston | NWA Aux → does Prep → from | to | on | near | through	5
4.	a) What do you mean by stemming. Explain Porter's Stemming algorithm in detail.	5 (3+2)
	b) What is morphology. Why do we need to do Morphological analysis?	

Maulana Azad National Institute of Technology, Bhopal

Department of Mathematics, Bioinformatics and Computer Applications

Mid Term Examination, 2023

Program: MCA | **Semester: 5th**

Subject: Cloud Computing | **Code: CA-501**

Time: 1.5 hrs | **MM: 20**

Note: Attempt all questions. All questions carry equal marks.

Q.1	Differentiate between cloud computing and Grid computing
Q.2	What are cloud computing challenges?
Q.3	What are the benefits of using cloud computing?
Q.4	Explain cloud and Dynamic infrastructure

21

MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY, BHOPAL

Department of Mathematics, Bioinformatics and Computer Applications

Mid Term Examination, September 2023

Course: MCA | **Semester: V**

Subject: Big Data Analytics | **Subject Code: CA 503**

Time: 90 min | **Max. Marks: 20**

Note: Attempt all questions. Each question carries equal marks.

Q.	Questions	Marks
1.	Provide a detailed explanation of analytical architecture accompanied by an illustrative diagram.	4
2.	Compare and contrast the following: I. Bagging and Boosting techniques II. Random Forest algorithms and Single Decision Trees	4
3.	Explain the interplay between Apache Hadoop, HDFS, and the MapReduce framework, highlighting the core mechanisms that enable this distributed data processing system to function effectively?	4
4	Concisely explain the following topics: i) Pig ii) Hive iii) HBase iv) Mahout	4
5	What are some commonly used tools and techniques for visualizing Big Data, and how do these tools help in gaining insights from large and complex datasets?	4

MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY, BHOPAL

Department of Mathematics, Bioinformatics and Computer Applications

Mid Term Examination, September 2023

Course: MCA
Semester: V
Subject: Machine Learning
Subject Code: CA 502
Time: 90 min
Max. Marks: 20

Note: Attempt all questions. Each question carry equal marks.

Q.	Questions	Marks
1.	**a)** Define Machine learning. Explain with specific example.	5 (2+3)
	b) Distinguish supervised learning from unsupervised learning.	
2.	**a)** What are the key distinctions between regression and classification in machine learning?	5 (2+3)
	b) Explain the terms (i) Overfitting (ii) Underfitting	
3.	A bank wants to decide whether a customer can be given a loan, based on two features related to the monthly salary of the customer, and his/her account balance. For simplicity, we model the two features with two binary variables X_1, $X_2 \in \{0,1\}$ and the class $Y \in \{0,1\}$ where Y=1 indicates that the customer can be given loan, and Y=0 indicates otherwise. Consider the following dataset having four instances: (X_1=0, X_2=0, Y=0) (X_1=0, X_2=1, Y=0) (X_1=1, X_2=0, Y=0) (X_1=1, X_2=1, Y=1) (i) Which model is better for the said application- logistic regression or linear regression? Explain briefly. (ii) If we change the first instance to (X_1=0, X_2=0, Y=1), can there be any logistic regression classifier using X_1 and X_2 as features, that perfectly classifies the data?	5
4	Define decision tree learning. List and explain appropriate problems for decision tree learning.	5

MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY, BHOPAL

Department of Mathematics, Bioinformatics and Computer Applications

Mid Term Examination (30-Sep-2023)

Program: MCA | **Semester: 5th**

Subject: NextGen Network | **Code: CA 904**

Time: 90 mins | **MM: 20**

Note: Attempt all questions. All questions carry equal marks.

Q.1	Find the netid and the hostid for each addresses: - a. 4.23.145.90 b. 227.34.78.7 c. 246.7.3.8 d. 129.6.8.4 B e. 198.76.9.23 C
Q.2	Explain the concepts of Masks without subnetting, Masks with subnetting and Non-boundary Level Masking.
Q.3	Define the following: - SMTP (Simple Mail Transfer Protocol), TFTP (Trivial file transfer protocol) and UA (User Agent).
Q.4	What is the basic difference between IPv4 and IPv6? Explain the following points in IPv6 - traffic class, flow label and payload length.
Q.5.a	Illustrate the working process of dynamic host configuration protocol. DHCP
.b	Differentiate between fixed header and optional headers.

21/12/2023

Maulana Azad National Institute of Technology, Bhopal

Department of Mathematics, Bioinformatics and Computer Applications

End Term Examination, 2023

Program: MCA | **Semester: 5th**

Subject: Cloud Computing | **Code: CA-501**

Time: 3 Hrs | **MM: 50**

Note: Attempt all questions. All questions carry equal marks.

Q.1	a. What are the different types of fault that can occur in a cloud computing environment? b. How testing can be performed under cloud environment?
Q.2	a. What are the fundamental components included in Cloud computing Reference model? b. Explain the services provided by Amazon Infrastructure Cloud.
Q.3	a. What is the need of Process Virtual machines, Host virtual Machines and native VMM's? b. Describe different type of Hypervisors with block diagram.
Q.4	a. Describe the architecture of Hadoop b. Describe the top threats identified by CSA(Cloud Security Alliance)
Q.5	a. Describe the concept of Map reduce with example. b. What is a third party cloud service? MDB

2120131

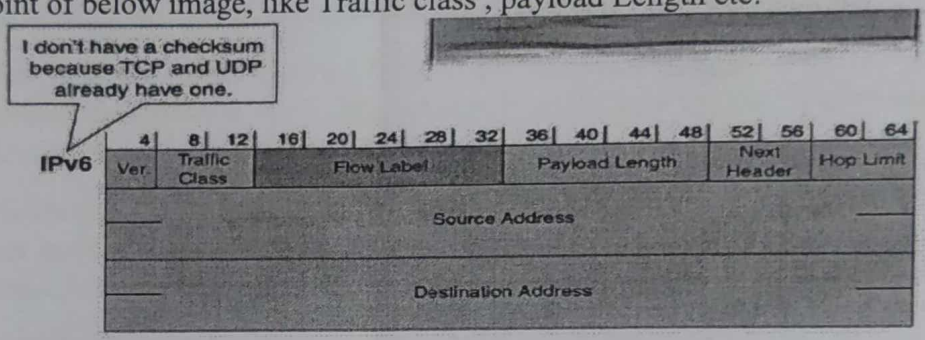
Name..... Roll No.....

MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY, BHOPAL

Department of Mathematics, Bioinformatics and Computer Applications

End Term Examination (November-2023)

Program: MCA | **Semester: 5th**
Subject: Next Generation Network | **Sub-code: CA 904**
Time: 3.00 Hrs | **Max. Marks: 50**
Note: Attempt all questions. All questions carry equal marks.

Q.1 .a	Illustrate the temporary measures did the IAB and IETF implement when they recognized the impending shortage of IPv4 addresses?	6
.b	What are the differences between OSPF for IPv4 and OSPF for IPv6?	4
Q.2 .a	Explain each point of below image, like Traffic class , payload Length etc: 	6
. b	Assume we have a host with a public type IPv4 address of 192.168.10.10. Give the following address for this host I. An IPv4 Embedded IPV6 Address II. An ISATAP Addresses	4
Q.3 .a	What is the value of the EtherType field in an Ethernet II frame, when the payload is an IPv6 Packet?	6
.b	Write about the following: Future scope, Applications & Mobility in IPv6.	4
Q.4	A company is granted the following block of address 2001:AB04:DB20::/52 from its provider and they have the following options, then explain the address distribution for all the 2 options individually I. Option A: 2 sites, 4 sub-sites (at each site), 7 subnets (at each sub-site) II. Option B: 4 sites, 4 sub-sites (at each site), 4 subnets (at each sub-site). Perform sub netting and assign addresses.	10
Q.5 .a	What is the EUI-64 based IPv6 interface identifier for the universally administered unicast IEEE 802 address of 00-02-b3-1e-83-29? What is the corresponding link-local address and Global unicast address?	4
.b	Write short notes on following (Any 3) I. Neighbor Discovery Mechanism In ICMP6 II. The IPv6 Reachability TLV (type 236) III. IPV6 Extension Header IV. Tunneling IPv6	6

Name of Student: Scholar No.:

MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY, BHOPAL

DEPARTMENT OF MATHEMATICS BIOINFORMATICS AND COMPUTER APPLICATIONS

END TERM EXAMINATION, NOVEMBER 2023

Course: MCA **Semester: V**

Subject: Machine Learning **Subject Code: CA-502**

Time: 3.00 Hours **Max. Marks: 50**

Note: Attempt all questions. Each question carry equal marks.

Q.1. **(a)** Define Machine Learning. Describe the steps involved in designing learning system. **(05)**

(b) Briefly describe the concept on model selection and generalization. **(05)**

Q.2. **(a)** What are the different methods for measuring classifier performance? **(05)**

(b) State the mathematical formulation of the SVM problem. Explain how Support Vector Machine can be used for classification of linearly separable data. **(05)**

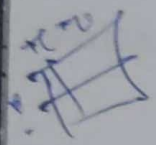
Q.3. **(a)** Compare Feature Extraction and Feature Selection techniques. Explain how dimensionality can be reduced using subset selection procedure. **(05)**

(b) Define Hidden Markov Model. What is meant by evaluation problem and how is this solved? **(05)**

Q.4. **(a)** The following data set contains factors that determine whether tennis is played or not. Using Naive Bayes classifier, find the play prediction for the day <Sunny, Cool, High, Strong> **(05)**

DAY	OUTLOOK	TEMP	HUMIDITY	WIND	PLAY
Day1	Sunny	Hot	High	Weak	NO
Day2	Sunny	Hot	High	Strong	NO
Day3	Overcast	Hot	High	Weak	YES
Day4	Rain	Mild	High	Weak	YES
Day5	Rain	Cool	Normal	Weak	YES
Day6	Rain	Cool	Normal	Strong	NO
Day7	Overcast	Cool	Normal	Strong	YES
Day8	Sunny	Mild	High	Weak	NO
Day9	Sunny	Cool	Normal	Weak	YES
Day10	Rain	Mild	Normal	Weak	YES
Day11	Sunny	Mild	Normal	Strong	YES
Day12	Overcast	Mild	High	Strong	YES
Day13	Overcast	Hot	Normal	Weak	YES
Day14	Rain	Mild	High	Strong	NO

(b) For the following set of training samples, find which attribute can be chosen as the root for decision tree classification. **(05)**



Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

Q.5. (a) Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For the healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute the precision and recall for the data. **(05)**

(b) Describe the random forest algorithm to improve classifier accuracy. **(05)**

****

Name.....

Roll No.....

MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY, BHOPAL

Department of Mathematics, Bioinformatics and Computer Applications

End Term Examination (November-2023)

Program: MCA
Subject: Big Data Analytics
Time: 3.00 Hrs
Note: Attempt all questions. All questions carry equal marks.

Semester: 5th
Code: CA 503
MM: 50

Q.1 .a	What is meant by big data? What is the key classification of big data types?	4
.b	What is HDFS (Hadoop distributed file system)? What is the primary purpose of HDFS in the Hadoop ecosystem, and how does it differ from traditional file systems?.	6
Q.2 .a	Describe the MapReduce execution steps with neat diagram. What decides number of mappers in a MapReduce job?	6
. b	What is Hive and Pig Latin. Define Why Hive is used instead of Pig Latin?	4
Q.3 .a	What tools and techniques are frequently employed for visualizing big data, and how do these tools contribute to extracting insights from extensive and intricate datasets?	6
.b	Explain the following terms: I. Bagging II. Random Forest III. k-means clustering IV. Dimensionality reduction	4
Q.4 .a	How is regression analysis used in machine learning, and what types of problems can it address?	4
.b	What is decision tree? Explain various terms used in Decision Tree.	3
.c	What is overfitting and underfitting in the context of supervised learning, and what techniques can be employed to prevent or mitigate it?	3
Q.5 .a	What is CAP theorem? What are the various database types as per the cap theorem.	4
.b	What is MangoDB? Explain characteristics of MangoDB.	3
.c	Discuss the NoSQL data stores and their characteristic features	3

2/12/2023

Name of Student:

Scholar No. :

MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY, BHOPAL

DEPARTMENT OF MATHEMATICS, BIOINFORMATICS AND COMPUTER APPLICATIONS

END TERM EXAMINATION, NOVEMBER 2023

Course: MCA — **Semester: V**

Subject: Natural Language Processing — **Subject Code: CA 802**

Time: 3.00 Hours — **Max. Marks: 50**

Note: Attempt all questions.

Q. No.	Questions	Marks
1.(a)	Describe the main differences between formal languages (such as logics or programming languages) and a natural language (such as English). What different requirements do these differences place on techniques for (i) syntactically parsing (ii) semantically interpreting natural, as opposed to formal, languages?	5
1.(b)	Explain the core concepts and challenges in Natural Language Processing (NLP).	5
2.(a)	What is a finite state transducer (FST), and what is it used for in computational linguistics? How does it differ from a finite state automaton?	4
2.(b)	**(i)** Write regular expressions for the following languages. *1.* the set of all alphabetic strings. *2.* the set of all lower case alphabetic strings ending in a *b.*	6
	(ii) What is lemmatization and stemming? How is lemmatization done?	
3.(a)	Explain the algorithm to edit one string X of length n to a string Y of length m. Show the steps of your algorithm for X = **INTENTION** and Y = **EXECUTION**.	6
3.(b)	Define the concept of an N-gram language model. Examine the influence of varying N-gram sizes on both the performance and complexity of the language model.	4
4.(a)	Consider the following context-free grammar: S -> NP VP N -> dog V -> sees NP -> Det N N ->cat V -> hates VP -> V N -> mouse V -> sneezes VP -> V NP Det -> the **(i)** Which of the following sentences are recognized by this grammar, and why? *1.* the dog sneezes the cat *2.* the mouse hates *3.* the cat the mouse hates **(ii)** Modify the grammar so that the following sentence is now accepted in addition: the dog the cat the mouse sees hates sneezes Your modification should express the linguistic phenomenon as efficiently and elegantly as possible. Justify your choice.	5

4.(b)	Discuss the evaluation of: (i) Stochastic part of speech (POS) tagging (ii) Word sense disambiguation	5
5.(a)	Define shallow parsing. What are the applications of shallow parsing?	4
5.(b)	Describe a real-world application where NLP has played a crucial role.	6