Learn Research
Innovate

UNITED INSTITUTE OF TECHNOLOGY

(An Autonomous Institution)
(Approved by AICTE | Affiliated to Anna University |
Accredited by NAAC with A+ Grade | Certified by ISO 9001:2015)
Periyanaickenpalayam, Coimbatore – 641020

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

QUESTION BANK

II YEAR

EVEN SEMESTER

ACADEMIC YEAR 2024 – 2025

INDEX

HEAD OF THE DEPARTMENT

ACOE **PRINCIPAL** **CHAIRMAN**

MA3391
PROBABILITY AND STATISTICS

UNIT – 1

PROBABILITY AND RANDOM VARIABLES

Axioms of probability – Conditional probability – Baye's theorem - Discrete and continuous random variables – Moments – Moment generating functions – Binomial, Poisson, Geometric, Uniform, Exponential and Normal distributions – Functions of a random variable.

Q. NO	QUESTION	CO	BTL	Marks
	PART–A			
1.	State the axioms of probability.	1	1	2
2	Find C, if a continuous random variable X has the density function $f(x) = \frac{c}{1+x^2}, \ -\infty < x < \infty.$	1	2	2
3.	A continuous random variable X that can assume any value between x=2 and x=5 has a density function given by f(x)= (2/27)(1+x). Find P(X<4).	1	2	2
4.	Show that $M_{X+Y}(t) = M_X(t) \ . \ M_Y(t)$ where X and Y are two independent random variables.	1	1	2
5.	If the moment generating function of a random variable X is, $\frac{1}{1-t}, \ \lvert t \rvert < 1,$ find $E(X)$ and $E(X^2)$.	1	2	2
6.	The mean of Binomial distribution is 20 and standard deviation is 4. Find the parameters of this distribution.	1	2	2
7.	What are the limitations of Poisson distribution?	1	1	2
8.	A random variable X is uniformly distributed between 3 and 15. Find the variance of X.	1	2	2

PART-B

1.	A random variable X has the following probability distribution x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 P(X=x) | 0 | K | 2K | 2K | 3K | K^2 | $2K^2$ | $7K^2+K$ Determine (i) the value of K (ii)P(X<6), P(1<X<5), P(X≥6) and (iii)If P[X≤C]>1/2, then find the minimum value of C.	1	5	16
2.	(i) Estimate the moment generating function of a random variable X whose probability function $P(X = x) = \frac{1}{2^x}$, $x = 1,2,\ldots$ Hence find its mean. (ii) A continuous random variable X has the density function f(x) given by $P(X = x) = \frac{k}{x^2+1}, -\infty < x < \infty$. Find the value of 'k' and the cumulative distribution of X.	1	5	16
3.	(i)Out of 800 families with 4 children each, how many families would be expected to have (i) 2 boys and 2 girls (ii) at least 1 boy (iii)at most 2 girls and (iv) children of both genders. Assume equal probabilities for boys and girls. (ii)State and prove memory less property of Geometric distribution.	1	3	16
4	(i)The mileage which car owners get with a certain kind of radial tire is a random variable having an exponential distribution with mean 40,000 km. Find the probability that one of these tires will last atleast 20,000 km. Also find the probability that one of these tires will last at most 30,000 km. (ii) An electrical firm manufactures light bulbs that have a life before burn-out that is normally distributed with mean equal to 800 hrs and a S.D. of 40 hrs. Determine (a) the probability that a bulb burns more than 834 hrs(b)the probability that a bulb burns between 778 and 834 hrs.	1	3	16

UNIT – 2

TWO- DIMENSIONAL RANDOM VARIABLES

Joint distributions – Marginal and conditional distributions – Covariance – Correlation and linear regression – Transformation of random variables – Central limit theorem (for independent and identically distributed random variables).

Q. NO	QUESTION	CO	BTL	Marks

PART–A

1.	Let X and Y be two independent R.V.s with Var(X)=9 and Var(Y)=3. Find Var(4X-2Y+6).	2	2	2

2	The joint probability mass function of (X, Y) is given by $p(x,y) = k(2x+y), x = 1,2$ and y=1,2 where k is a constant. Find the value of k.	2	2	2
3.	The joint p.d.f. of the random variable X and Y is defined as $f(x,y) = \begin{cases} 25e^{-5y}, & 0 < x < 0.2,\ y > 0 \\ 0 & otherwise \end{cases}$. Find the marginal p.d.f.'s of X and Y.	2	2	2
4.	The joint p.d.f. of the random variable X and Y is defined as $f(x,y) = \begin{cases} 1/4, & 0 \le x, y \le 2 \\ 0 & otherwise \end{cases}$. Find P(x+y≤1).	2	2	2
5.	If X and Y are independent random variables prove that Cov(X, Y) = 0.	2	2	2
6.	The regression equations are x+9y = 7 and y+4x = 49/3. Find the correlation coefficient between X and Y.	2	2	2
7.	Can $y = 5 + 2.8x$ and $x = 3 - 0.5y$ be the estimated regression equation of y on x respectively explain your answer.	2	2	2
8.	State central limit theorem.	2	1	2

PART-B

1.	The joint probability mass function of (X, Y) is given by $p(x,y) = k(2x+3y)$ $x = 0,1,2;\ y = 1,2,3$. Determine all the marginal and conditional probability distributions. Also find the probability distribution of (X+Y) and P(X+Y>3).	2	3	16
2.	The joint p.d.f. of the R.V. (X,Y) is given by $f(x,y) = \begin{cases} \frac{1}{8}(6-x-y), & 0 < x < 2,\ 2 < y < 4 \\ 0 & otherwise \end{cases}$. Estimate (i) P(X < 1 ∩ Y < 3) (ii) P(X + Y < 3) (iii) P (X < 1 / Y < 3).	2	5	16

3. From the following data estimate (a) the two regression equations (b) the coefficient of correlation between the marks in economics and statistics. (c) the most likely marks in statistics when marks in economics are 30.

x	25	28	35	32	31	36	29	38	34	32
y	43	46	49	41	36	32	31	30	33	39

2 5 16

4 The equation of two regression lines are $8x - 10y + 66 = 0$ and $40x - 18y - 214 = 0$. Find $\bar{x}$, $\bar{y}$ and the correlation co-efficient between X and Y. 2 3 16

UNIT – 3

ESTIMATION THEORY

Unbiased estimators - Efficiency - Consistency - Sufficiency - Robustness - Method of moments - Method of maximum Likelihood - Interval estimation of Means - Differences between means, variations and ratio of two variances.

Q. NO	QUESTION	CO	BTL	Marks

PART–A

1.	Write the difference between Point Estimate and Interval Estimate of population parameters.	3	1	2
2	What are the characteristics that should be satisfied by a good estimator?	3	1	2
3.	What are the commonly used methods of Point Estimation?	3	1	2
4.	Define Population Estimation.	3	1	2
5.	What is called interval estimation?	3	1	2
6.	Define unbiased estimator.	3	1	2
7.	Define maximum likelihood estimator of θ.	3	1	2
8.	Mention the properties of the maximum likelihood estimator.	3	1	2

PART-B

1.	(i) Show that the sample mean $\overline{x}$ is an unbiased estimator for the population mean μ. (ii) If $\{x_1, x_2, \dots, x_n\}$is a random sample of size n, drawn from a geometric distribution, the probability mass function of which is given by $P(x = r) = pq^{r-1}$; $r = 1,2,3, \dots, \infty$. Prove that the mean of the sample is a consistent estimator of the population mean.	3	2	16
2.	Find the maximum likelihood estimator for the parameter P of the binomial distribution$B(N, P)$, where N is very large but finite, on the basis of sample size n. Also find its variance.	3	3	16
3.	A research worker wants to determine the average time it takes a mechanic to rotate the tyres of a car and he wants to be able to assert with 95% confidence that the mean of his sample is of by atmost 0.5 minutes. If he can presume from past experience that $\sigma = 1.6$ minutes, how large a sample will have to take?	3	3	16
4	(i)The average monthly electricity consumption for a sample of 100 families is 1250 units. Assuming the standard deviation of electric consumption of all families is 150 units, construct a 95% confidence interval estimate of the actual mean electric consumption. (ii)In a test given to two groups of students the marks obtained were as follows:	3	3	16

I group	18	20	36	50	49	36	34	49	61
II group	29	28	26	35	30	44	46		

Construct a 95% confidence interval on the mean marks secured by students of the above two groups.

UNIT – 4

NON- PARAMETRIC TESTS

Introduction - The Sign test - The Signed - Rank test - Rank - sum tests - The U test - The H test - Tests based on Runs - Test of randomness - The Kolmogorov Tests .

Q. NO	QUESTION	CO	BTL	Marks

PART–A

1.	Name four non – parametric tests used in statistical study.	4	1	2
2	Write any two advantages of non – parametric methods over parametric methods.	4	1	2

3.	State the limitations of non – parametric test.	4	1	2
4.	Distinguish between the Mann-Whitney U-test and the Kruskal-Wallis.	4	1	2
5.	Write the formula for Kruskal-Wallis test.	4	1	2
6.	Explain run test with an example	4	1	2
7.	What are the uses of run test?	4	1	2
8.	What are the assumptions of the Mann-Whitney?	4	1	2

PART-B

1. (i)In a factory, 20 observations of the factors that could heat up a conveyor belt yielded the following results : 0.36, 0.41, 0.25, 0.34, 0.28, 0.26, 0.39, 0.28, 0.40, 0.26, 0.35, 0.38, 0.29, 0.42, 0.37, 0.37, 0.39, 0.32, 0.29 and 0.36. Use the sign test at the 0.01 level of significance to test the null hypothesis $\mu = 0.34$ against the alternative hypothesis $\mu \neq 0.34$.
(ii)The following are the average weekly losses of worker-hours due to accidents in 10 industrial plants before and after a certain safety program was put into operation

Before :	45	73	46	124	33	57	83	34	26	17
After :	36	60	44	119	35	51	77	29	24	11

Use the sign test at the 0.05 level of significance to test whether the safety program is effective.

4 4 16

2. Drop in diastolic blood pressure (in mm mercury)

Drug D_1	10	16	10	4	2	14	4
Drug D_2	33	34	41	36	42	42	32

Test whether there is any difference in the effectiveness of drugs at $\alpha = 0.05$. Using Wilcoxon's signed rank test. (Table value for n=7 is 2)

4 4 16

3. Two classes of students are tested using a certain competitive exam. The scores of a sample of students from each class is given below:

Class A	45	44	47	48	55	53	55	63			
Class B	65	67	77	65	56	67	78	55	66	65	58

4 4 16

Use Mann Whitney – U test to test whether both classes have similar scholastic levels.

4 (i)An experiment designed to compare three preventative methods against corrosion yielded the following maximum depths of pits (in thousands of an inch) in pieces of wire subjected to the respective treatments.

Method A	77	54	67	74	71	66	
Method B	60	41	59	65	62	64	52
Method C	49	52	69	47	56		

4 4 16

Use the Kruskal-Wallis test at the 5% level of significance to test the null hypothesis that the three samples come from identical populations.
(ii)A technician is asked to analyze the results of 22 items made in a preparation run. Each item has been measured and compared to engineering specifications. The order of acceptance 'a' and rejections 'r' is

aa rrr a rr aaaaa rr a rr aa r a

Determine whether it is a random sample or not. (Use $\alpha = 0.05$).

UNIT – 5

STATISTICAL QUALITY CONTROL

Control charts for measurements (X and R charts) – Control charts for attributes (p, c and np charts) – Tolerance limits - Acceptance sampling.

Q. NO	QUESTION	CO	BTL	Marks

PART–A

1. Define Statistical Quality Control and state its objectives. 5 1 2

2 State the benefits of statistical Quality Control. 5 1 2

3. Define control charts. 5 1 2

4.	Classify control charts.	5	1	2
5.	Define defect and defective.	5	1	2
6.	Define mean chart.	5	1	2
7.	Define R Chart.	5	1	2
8.	Write the control limits for the R chart.	5	2	2

PART-B

1. The following are the sample means and ranges for ten samples, each of size 5. Construct the control chart for mean and range and comment on the nature of control.

Sample No.	1	2	3	4	5	6	7	8	9	10
Mean	12.8	13.1	13.5	12.9	13.2	14.1	12.1	15.5	13.9	14.2
Range	2.1	3.1	3.9	2.1	1.9	3.0	2.5	2.8	2.5	2.0

5 4 16

2. The values of sample mean $\overline{X}$ and sample standard deviation s for 15 samples, each of size 4, drawn from a production process are given below. Draw the appropriate control charts for the process average and process variability. Comment on the state of control.

Sample No.	1	2	3	4	5	6	7	8	9
Mean	15	10	12.5	13	12.5	13	13.5	11.5	13.5
S.D.	3.1	2.4	3.6	2.3	5.2	5.4	6.2	4.3	3.4

Sample No.	10	11	12	13	14	15
Mean	13	14.5	9.5	12	10.5	11.5
S.D.	4.1	3.9	5.1	4.7	3.3	3.3

5 4 16

3. (i)20 pieces of cloth out of different rolls contained respectively 1, 4, 3, 2, 4, 5, 6, 7, 2, 3, 2, 5, 7, 6, 4, 5, 2, 1, 3 and 8 imperfections. Ascertain whether the process is in a state of statistical control.
(ii)15 tape – recorders were examined for quality control test. The number of defects in each tape – recorder below. Draw the appropriate control chart and comment on the state of control.

Unit No.(i)	1	2	3	4	5	6	7	8	9
No. of defects	2	4	3	1	1	2	5	3	6
Sample No.	10	11	12	13	14	15			
No. of defects(c)	7	3	1	4	2	1			

5 4 16

4 The data given below are the number of defectives in 10 samples of 100 items each. Construct a p-chart and an np-chart and comment on the results :

Sample No.	1	2	3	4	5	6	7	8	9	10
No. of defectives	6	16	7	3	8	12	7	11	11	4

5 4 16

-----END----

AL3452
OPERATING SYSTEMS

UNIT I
INTRODUCTION

Computer System - Elements and organization; Operating System Overview - Objectives and Functions - Evolution of Operating System; Operating System Structures – Operating System Services - User Operating System Interface - System Calls – System Programs - Design and Implementation - Structuring methods.

PART - A

Q.NO	QUESTION	CO	BTL	MARK
1.	What is an operating system?	1	1	2
2	What are the objectives of operating systems?	1	1	2
3.	How are multiple interrupt dealt with?	1	2	2
4.	Difference b/w Monolithic & Microlithic.	1	2	2
5.	Define system call.	1	1	2
6.	What are the disadvantages of multiprocessor systems?	1	1	2
7.	What is the purpose of system programs?	1	1	2
8.	Why API s need to be used rather than system calls?	1	2	2

PART - B

1.	Explain Operating System Structure and components.	1	1	16
2.	i) Describe the differences between symmetric and asymmetric multiprocessing. i) What are three advantages and one disadvantage of multiprocessor systems?	1	4	16
3.	i) Explain Cache memory and its mapping. ii) Describe evolution of operating system.	1	2	16
4	i) Describe the differences between symmetric and asymmetric multiprocessing. What are three advantages and one disadvantage of multiprocessor systems? ii) How could a system be designed to allow a choice of operating systems from which to boot? What would the bootstrap program need to do?	1	3	16

UNIT 2
PROCESS MANAGEMENT

Processes - Process Concept - Process Scheduling - Operations on Processes - Inter-process Communication; CPU Scheduling - Scheduling criteria - Scheduling algorithms: Threads - Multithread Models – Threading issues; Process Synchronization - The critical-section problem - Synchronization hardware – Semaphores – Mutex - Classical problems of synchronization - Monitors; Deadlock - Methods for handling deadlocks, Deadlock prevention, Deadlock avoidance, Deadlock detection, Recovery from deadlock

PART - A

Q.NO	QUESTIONS	CO	BTL	MARK
1.	Define Process?	2	2	2
2	What is turnaround time?	2	1	2
3.	Under What circumstances CPU scheduling decision takes place.	2	4	2
4.	What are the difference b/w user level threads and kernel level threads?	2	4	2
5.	What is preemptive and non-preemptive scheduling?	2	1	2
6.	What is critical section problem?	2	1	2

7.	What are conditions under which a deadlock situation may arise?	2	2	2
8.	What is semaphore? Mention its importance in operating system.	2	1	2

PART - B

1.	Explain the FCFS, preemptive and non-preemptive versions of Shortest-Job First and Round Robin (time slice = 2) scheduling algorithms with Gantt charts for the four Processes given. Compare their average turnaround and waiting time.	2	5	16

Process	Arrival Time	Waiting Time
P1	0	8
P2	1	4
P3	2	9
P4	3	5

2.	Write in detail about several CPU scheduling algorithms.	2	4	16
3.	i) Explain the Banker algorithm for deadlock avoidance in detail with an example. ii) Discuss in detail the critical section problem and also write the algorithm for Readers-Writers Problem with semaphores.	2	4	16
4	Consider the following set of processes, with the length of the CPU-burst time in given ms:	2	5	16

Process	Burst Time	Arrival Time
P1	8	0.00
P2	4	1.001
P3	9	2.001
P4	5	3.001
P5	3	4.001

Draw four Gantt charts illustrating the execution of these processes using FCFC,SJF, Priority and RR (Quantum=2) scheduling. Also calculate waiting time and turnaround time for each scheduling algorithms

UNIT 3
MEMORY MANAGEMENT

Main Memory - Swapping - Contiguous Memory Allocation – Paging - Structure of the Page Table - Segmentation, Segmentation with paging; Virtual Memory - Demand Paging – Copy on Write - Page Replacement - Allocation of Frames –Thrashing.

PART - A

Q.NO	QUESTION	CO	BTL	MARK
1.	What is address binding?	3	1	2
2	Difference between internal and external fragmentation	3	2	2
3.	Define Demand paging and write advantages.	3	1	2
4.	What do mean by page fault?	3	1	2
5.	Differentiate between Global and Local page replacement algorithms.	3	2	2
6.	Define lazy swapper.	3	1	2
7.	How do you limit the effects of thrashing?	3	2	2
8.	How does the system detect thrashing?	3	2	2

PART - B

1.	i) Given memory partitions of 100KB, 500KB, 200KB, 300KB and 600KB(in order), how would each of the first-fit, best-fit and worst-fit algorithms place processes of212KB, 417KB, 12KB and 426KB(in order)? Which algorithm makes the most efficient use of memory?	3	3	16
2.	i) What is the copy-on-write feature, and under what circumstances is its use beneficial? Why hardware support is required to implement this feature? ii) Explain the difference between internal and external fragmentation.	3	5	16
3.	Consider the following page reference string 1,2,3,4,2,1,5,6,2,1,3,7,6,3,2,1,3,6. How many page faults would occur for the following replacement algorithms, assuming one, two, three and four frames? i)LRU replacement ii) FIFO replacement iii) Optimal replacement	3	3	16
4	(i) Describe a mechanism by which one segment could belong to the address space of two different processes. (8) (ii) Why are segmentation and paging sometimes combined into one scheme? Explain them in detail with example.(8)	3	4	16

UNIT 4
STORAGE MANAGEMENT

Mass Storage system – Disk Structure - Disk Scheduling and Management; File-System Interface - File concept - Access methods - Directory Structure - Directory organization - File system mounting - File Sharing and Protection; File System Implementation - File System Structure - Directory implementation - Allocation Methods - Free Space Management; I/O Systems – I/O Hardware, Application I/O interface, Kernel I/O subsystem

PART - A

Q.NO	QUESTION	CO	BTL	MARK
1.	What are the various file operations?	4	1	2
2	What are the different accessing methods of a file?	4	1	2
3.	Define seek time and latency time.	4	1	2
4.	What is the content of a typical file control block?	4	1	2
5.	What is meant by polling?	4	1	2
6.	Define buffering.	4	1	2
7.	How does DMA increase system concurrency?	4	2	2
8.	Differentiate between file and directory	4	2	2

PART - B

1.	i)Explain the different disk scheduling algorithms with examples. ii) Explain the allocation of frames in detail.	4	2	16
2.	i) What do you mean by directory structure? Also discuss Tree-Structure Directories and Acyclic-Graph Directories. (ii) Describe in details about file system implementation and file allocation method	4	2	16
3.	On a disk with 1000 cylinders, numbers 0 to 999, Compute the number of tracks the disk arm must move to satisfy the entire request in the disk queue. Assume the last received was at track 345 and the head is moving towards	4	4	16

	track 0. The queue in FIFO order contains requests for the following tracks. 123,874,692,475,105 and 376.Find the seek length for the following scheduling algorithm. (1) SSTF (2) LOOK (3) CSCAN			
4	i) Write a brief note on the steps involved in DMA transfer. ii) Explain the data structures supported by kernel I/O system	4	2	16

UNIT 5
VIRTUAL MACHINES AND MOBILE OS

Virtual Machines – History, Benefits and Features, Building Blocks, Types of Virtual Machines and their Implementations, Virtualization and Operating-System Components; Mobile OS - iOS and Android.

PART - A

Q.NO	QUESTION	CO	BTL	MARK
1.	What are the components of kernel mode?	5	1	2
2	What is process Identity?	5	1	2
3.	What is virtualization?	5	1	2
4.	Write the purpose of using virtualization	5	1	2
5.	Enlist the advantages of using kernel modules in Linux	5	2	2
6.	What is handle? How does a process obtain a handle?	5	2	2
7.	Which layer of iOS contains fundamental system services for apps?	5	1	2
8.	List the advantages and disadvantage of writing an operating system in high level language such as C.	5	2	2

PART - B

1.	i)Discuss three advantages of dynamic (shared) linkage of libraries compared with static linkage. Describe two cases in which static linkage is preferable (8) (ii) Explain the step by step procedure for setting up a local network services. (8)	5	2	16
2.	i)Why is live migration possible in virtual environments but much less possible for a native operating system? (ii) What are the primary goals of the conflict-resolution mechanism used by the Linux kernel for loading kernel modules.	5	4	16
3.	i) How to install and configuring network services in LINUX. (ii)Describe the benefits of virtualization in LINUX OS.	5	2	16
4	With frame work explain the working function of android operating system architecture. Compare the feature of IoS and android.	5	4	16

----- END ----

AL3451
MACHINE LEARNING

UNIT I
INTRODUCTION TO MACHINE LEARNING

Review of Linear Algebra for machine learning; Introduction and motivation for machine learning; Examples of machine learning applications, Vapnik-Chervonenkis (VC) dimension, Probably Approximately Correct (PAC) learning, Hypothesis spaces, Inductive bias, Generalization, Bias variance trade-off.

PART A

S. No	Question	CO	BTL	Marks
1	Describe vector, and how is it used in machine learning	1	2	2
2	Discuss matrix, and why is it significant in machine learning	1	2	2
3	Define machine learning	1	1	2
4	Differentiate between supervised and unsupervised learning	1	4	2
5	Describe reinforcement learning, and how does it differ from supervised learning	1	2	2
6	Discuss overfitting in machine learning.	1	2	2
7	Define hypothesis space	1	1	2
8	Explain how machine learning is used in fraud detection	1	2	2

PART B

1	Describe the Vapnik-Chervonenkis (VC) dimension and its significance in machine learning and explain the concept of Probably Approximately	1	2	16
2	Discuss inductive bias in machine learning? Explain its role with suitable examples.	1	2	16
3	Explain the importance of hypothesis spaces in machine learning and how they impact learning performance	1	2	16
4	Discuss bias-variance trade-off? Explain its mathematical formulation and impact on machine learning models	1	2	16

UNIT II
SUPERVISED LEARNING

Linear Regression Models: Least squares, single & multiple variables, Bayesian linear regression, gradient descent, Linear Classification Models: Discriminant function – Perceptron algorithm, Probabilistic discriminative model - Logistic regression, Probabilistic generative model – Naive Bayes, Maximum margin classifier – Support vector machine, Decision Tree, Random Forests

PART A

S. No	Question	CO	BTL	Marks
1	State the primary goal of linear regression	2	1	2
2	Discuss least squares method in linear regression	2	2	2
3	Discuss how do Bayesian linear regression differ from ordinary linear	2	2	2
4	Summarize the objective of a Support Vector Machine (SVM).	2	2	2
5	Compare the Perceptron algorithm with logistic regression	2	2	2
6	Discuss gini index, and how is it used in decision trees	2	2	2
7	Explain the main objective of the Maximum Margin Classifier	2	2	2
8	Describe dual formulation in the context of SVM	2	2	2

PART B

1	Compare and contrast Bayesian linear regression and least squares regression	2	2	16
2	Illustrate the logistic regression model and derive its cost function	2	4	16
3	Explain the Naive Bayes classifier and its assumptions. Discuss its application in text classification	2	2	16
4	Distinguish decision trees with random forests in terms of accuracy, overfitting, and interpretability	2	4	16

UNIT III
ENSEMBLE TECHNIQUES AND UNSUPERVISED LEARNING

Combining multiple learners: Model combination schemes, Voting, Ensemble Learning - bagging, boosting, stacking, Unsupervised learning: K-means, Instance Based Learning: KNN, Gaussian mixture models and Expectation maximization.

PART A

S. No	Question	CO	BTL	Marks
1	Explain the concept of ensemble learning	3	2	2
2	Describe the bagging technique and its advantages	3	2	2
3	Discuss boosting and how does it improve model performance.	3	2	2
4	Explain the concept of stacking in ensemble learning	3	2	2
5	Explain stacked generalization	3	2	2
6	Discuss K-means clustering algorithm and how does it work.	3	2	2
7	Describe the working of the K-Nearest Neighbors (KNN) algorithm	3	2	2
8	Discuss Expectation-Maximization (EM) algorithm and how is it used in GMMs	3	2	2

PART B

1	What is voting in ensemble learning? Differentiate between majority voting and weighted voting with examples	3	2	16

2	Analyze the impact of overfitting in ensemble methods and how it is addressed by techniques like bagging and boosting	3	4	16
3	Illustrate the K-means clustering algorithm with an example. Discuss its limitations and solutions	3	4	16
4	Illustrate working of Gaussian Mixture Models (GMM) for clustering. Compare GMM with K-means	3	4	16

UNIT IV
NEURAL NETWORKS

Multilayer perceptron, activation functions, network training – gradient descent optimization – stochastic gradient descent, error backpropagation, from shallow networks to deep networks –Unit saturation (aka the vanishing gradient problem) – ReLU, hyperparameter tuning, batch normalization, regularization, dropout**.**

PART A

S. No	Question	CO	BTL	Marks
1	Define Multilayer Perceptron (MLP).	4	1	2
2	Explain learning rate important in gradient descent.	4	2	2
3	Discuss backpropagation in neural networks.	4	2	2
4	Deduce why deep networks often more powerful than shallow networks	4	4	2
5	Explain vanishing gradient problem.	4	2	2
6	Define hyperparameter tuning in machine learning. Name two common methods for hyperparameter tuning.	4	1	2
7	Define regularization in the context of neural networks. Name two common regularization techniques.	4	1	2
8	Deduce how does dropout improve the performance of neural networks.	4	4	2

PART B

1	Describe the forward propagation process in a multilayer perceptron. Provide an example with two hidden layers	4	2	16
2	Explain the gradient descent optimization process for training neural networks.	4	2	16
3	Explain hyperparameter tuning in neural networks. Discuss different methods for optimizing hyperparameters	4	2	16
4	Describe the concept of regularization in neural networks. Compare L1 and L2 regularization techniques	4	2	16

UNIT V
DESIGN AND ANALYSIS OF MACHINE LEARNING EXPERIMENTS

Guidelines for machine learning experiments, Cross Validation (CV) and resampling – K-fold CV, bootstrapping, measuring classifier performance, assessing a single classification algorithm and comparing two classification algorithms – t test, McNemar's test, K-fold CV paired t test

PART A

S. No	Question	CO	BTL	Marks
1	Describe why is it important to have a well-defined experimental setup in machine learning	5	2	2
2	Describe cross-validation? Name two common resampling techniques used in cross-validation	5	2	2
3	Define bootstrapping in the context of machine learning	5	2	2
4	Analyse the purpose of measuring classifier performance? Name two metrics commonly used to measure classifier performance	5	4	2
5	Explain the importance of assessing a single classification algorithm	5	2	2
6	Why is it important to compare two classification algorithms	5	4	2
7	Discuss what does a p-value in a t-test indicate.	5	2	2
8	Describe why McNemar's test used for in machine learning	5	2	2

PART B

1	Explain the key guidelines for designing machine learning experiments	5	2	16
2	Explain the various methods of measuring Classifier Performance with suitable examples.	5	2	16
3	Discuss the K-fold CV paired t-test method for comparing two machine learning models. Provide a detailed example	5	4	16
4	Explain the importance of statistical significance testing in evaluating machine learning models	5	2	16

---------- END ---------

AD3491
FUNDAMENTALS OF DATA SCIENCE AND ANALYTICS

UNIT 1
INTRODUCTION TO DATA SCIENCE

Need for data science – benefits and uses – facets of data – data science process – setting the research goal – retrieving data – cleansing, integrating, and transforming data – exploratory data analysis – build the models – presenting and building applications.

PART - A

Q.NO	QUESTION	CO	BTL	MARK
1.	What is the need for data science in modern industries?	1	2	2
2.	What is meant by structured data in data science?	1	2	2
3.	Explain the role of data cleansing in data science.	1	2	2
4.	What is data transformation, and why is it necessary?	1	2	2
5.	What is the purpose of building models in data science?	1	2	2
6.	What is the role of visualization in presenting data science results?	1	2	2
7.	Mention two types of data often analyzed in data science.	1	1	2
8.	List two techniques used in exploratory data analysis.	1	1	2

PART - B

1.	Discuss the importance of data science in modern industries and explain its key benefits and uses.	1	2	16
2.	Describe the various stages of the data science process. Provide a detailed explanation of each stage.	1	2	16
3.	Explain the process of data cleansing and transformation. How do these steps contribute to the quality of the data?	1	2	16
4	What is exploratory data analysis (EDA)? Discuss the techniques involved and how it helps in understanding the data.	1	2	16

UNIT 2
DESCRIPTIVE ANALYTICS

Frequency distributions – Outliers –interpreting distributions – graphs – averages - describing variability – interquartile range – variability for qualitative and ranked data - Normal distributions – z scores –correlation – scatter plots – regression – regression line – least squares regression line – standard error of estimate – interpretation of r2 – multiple regression equations – regression toward the mean.

PART - A

Q.NO	QUESTION	CO	BTL	MARK
1.	What is a frequency distribution, and why is it important in statistical analysis?	2	2	2
2.	Define outliers. How can they affect the analysis of data?	2	2	2
3.	Define variability in statistics. Why is it important in data analysis?	2	2	2
4.	What is a normal distribution? List its key characteristics.	2	2	2

5.	What is the interquartile range, and how is it calculated?	2	1	2
6.	Define a scatter plot and its use in showing relationships between two variables.	2	2	2
7.	What is the difference between simple and multiple regression?	2	1	2
8.	What are the different types of graphs used in frequency distributions?	2	2	2

PART - B

1.	Describe the measures of central tendency (mean, median, and mode) and discuss their relevance in interpreting frequency distributions. Provide examples.	2	2	16
2.	Define interquartile range. How is it calculated, and why is it considered a robust measure of variability? Compare it with the range and standard deviation.	2	2	16
3.	What are scatter plots, and how are they used to visually represent the relationship between two variables? Explain how scatter plots can be used to identify patterns, trends, and correlations.	2	2	16
4.	What is meant by describing variability in data? Explain the importance of measures such as variance, standard deviation, and interquartile range in statistical analysis.	2	2	16

UNIT 3
INFERENTIAL STATISTICS

Populations – samples – random sampling – Sampling distribution- standard error of the mean - Hypothesis testing – z-test – z-test procedure –decision rule – calculations – decisions – interpretations - one-tailed and two-tailed tests – Estimation – point estimate – confidence interval – level of confidence – effect of sample size.

PART - A

Q.NO	QUESTION	CO	BTL	MARK
1.	Explain the concept of random sampling.	3	2	2
2	What is the difference between a population and a sample?	3	2	2
3.	What is the purpose of hypothesis testing?	3	2	2
4.	Define the null hypothesis and alternative hypothesis.	3	2	2
5.	What is point estimation in statistics?	3	2	2
6.	Define a confidence interval and explain its significance.	3	2	2
7.	Define one-tailed and two-tailed tests with an example.	3	2	2
8.	Describe the significance of the p-value in hypothesis testing.	3	2	2

PART – B

1.	Explain the concept of hypothesis testing. Outline the steps involved in hypothesis testing and discuss their importance.	3	2	16
2.	Discuss the procedure for performing a z-test. Include the formulation of null and alternative hypotheses, significance level, and decision rule.	3	2	16
3.	Define point estimation and explain how it is used in statistical analysis. How does it relate to the estimation of population parameters?	3	2	16
4	Differentiate between one-tailed and two-tailed tests in hypothesis testing. Provide examples of when each test is used.	3	2	16

UNIT 4
ANALYSIS OF VARIANCE

t-test for one sample – sampling distribution of t – t-test procedure – t-test for two independent samples – p-value statistical significance – t-test for two related samples. F-test – ANOVA – Two-factor experiments – three f-tests two-factor ANOVA –Introduction to chi-square tests.

PART - A

Q.NO	QUESTION	CO	BTL	MARK
1.	Define t-test for one sample.	4	2	2
2.	Explain the t-test for two independent samples.	4	2	2
3.	How does the t-test for two independent samples differ from the t-test for one sample?	4	2	2
4.	Describe the basic concept of Analysis of Variance (ANOVA).	4	2	2
5.	How does ANOVA differ from the t-test?	4	2	2
6.	Define chi-square test.	4	2	2
7.	What is the null hypothesis in a one-way ANOVA?	4	2	2
8.	What is the difference between the t-distribution and the normal distribution?	4	2	2

PART – B

1.	Discuss the concept of statistical significance in hypothesis testing. How is the p-value interpreted in the context of a t-test for one sample?	4	2	16
2.	Describe the procedure for a t-test for two independent samples. What assumptions are made, and how is the p-value used to assess statistical significance?	4	2	16
3.	What is the F-test and how is it used to compare variances in statistical analysis? Provide examples of its application.	4	2	16
4.	Provide an introduction to chi-square tests. How are chi-square tests used for testing goodness of fit and for testing independence? Discuss with examples.	4	2	16

UNIT 5
PREDICTIVE ANALYTICS

Linear least squares – implementation – goodness of fit – testing a linear model – weighted resampling. Regression using StatsModels – multiple regression – nonlinear relationships – logistic regression – estimating parameters – Time series analysis – moving averages – missing values – serial correlation – autocorrelation. Introduction to survival analysis.

PART - A

Q.NO	QUESTION	CO	BTL	MARK
1.	What is the concept of linear least squares in regression analysis?	5	2	2
2	What does the term "weighted resampling" mean in statistical analysis?	5	2	2
3.	What is the primary function of StatsModels in regression analysis?	5	2	2
4.	Describe a method for identifying nonlinear relationships in data.	5	2	2
5.	Define the goodness of fit in the context of linear regression.	5	2	2
6.	What is the challenge of modeling nonlinear relationships in regression?	5	2	2
7.	What is survival analysis and why is it important in statistics?	5	2	2
8.	What is the purpose of using moving averages in time series analysis?	5	2	2

PART – B

1.	Explain the concept of Linear Least Squares and describe the steps involved in its implementation. Discuss its significance in regression analysis.	5	2	16
2.	Using StatsModels, explain the process of performing Multiple Regression. Discuss the assumptions and limitations of multiple regression models.	5	2	16
3.	Discuss the methods for handling Missing Values in time series data. Explain why it is important to address missing data and how imputation techniques can be applied.	5	2	16
4.	Explain the concept of Logistic Regression. How does it differ from linear regression, and what are the key steps in estimating parameters for logistic models?	5	2	16

---------- END ----------

CS3591
COMPUTER NETWORKS

UNIT 1

INTRODUCTION AND APPLICATION LAYER

Data Communication - Networks – Network Types – Protocol Layering – TCP/IP Protocol suite – OSI Model – Introduction to Sockets - Application Layer protocols: HTTP – FTP – Email protocols (SMTP - POP3 - IMAP - MIME) – DNS – SNMP

PART - A

Q.NO	QUESTION	CO	BTL	Marks
1.	List the three fundamental characteristics determine the effectiveness of a data communication system.	1	1	2
2	Define the transmission modes.	1	2	2
3.	For n devices in a network, what is the number of cable links required for a mesh, ring, bus and star topology?	1	2	2
4.	Differentiate between Transport layer and Network layer.	1	2	2
5.	Define socket.	1	2	2
6.	List the features of HTTP.	1	1	2
7.	State the difference between a User Agent(UA) and a Mail Transfer Agent(MTA).	1	2	2
8.	Why is an application such as POP3 needed for electronic messaging?	1	2	2

PART - B

1.	Draw the OSI network architecture and explain the functionalities of each layer in detail.	1	2	16
2.	Explain in detail about HTTP and its operations.	1	2	16
3.	Describe in detail about the SMTP protocol and FTP protocol.	1	2	16
4	Explain in detail about DNS and its frame format.	1	2	16

UNIT 2
TRANSPORT LAYER

Introduction - Transport-Layer Protocols: UDP – TCP: Connection Management – Flow control - Congestion Control - Congestion avoidance (DECbit, RED) – SCTP – Quality of Service

PART - A

Q.NO	QUESTION	CO	BTL	Marks
1.	Define segmentation and reassembly of transport layer.	2	2	2
2.	List the services provided by the transport layer protocol.	2	1	2
3.	Differentiate connection oriented and connectionless services.	2	2	2
4.	Draw the header format of TCP.	2	2	2
5.	What is the value of the receiver window (rwnd) for host X. If the receiver host Y, has a buffer size of 5000 bytes and 1000 bytes of received and unprocessed data?	2	3	2
6.	Define a window size of sliding window in TCP.	2	2	2
7.	Difference between flow control and congestion control.	2	2	2
8.	State the use of SCTP Multiple stream service.	2	2	2

PART - B

1.	Explain in detail about the sliding window flow control of TCP with neat diagrams.	2	2	16
2.	Explain in detail congestion avoidance technquies in TCP.	2	2	16
3.	Describe with examples the three mechanism by which congestion control is achieved in TCP.	2	2	16
4	Explain the various approaches to improve Quality of Service (QoS) in a data transmission network.	2	2	16

UNIT 3

NETWORK LAYER

Switching : Packet Switching - Internet protocol - IPV4 – IP Addressing – Subnetting - IPV6, ARP, RARP, ICMP, DHCP

PART – A

Q.NO	QUESTION	CO	BTL	Marks
1.	List the advantages of packet switching over circuit switching.	3	1	2
2	Define classful addressing in IPv4.	3	2	2
3.	Find the subnetwork address for the following: a) IP address:125.34.12.56 Mask :255.255.0.0 b) IP address:141.181.14.16 Mask :255.255.224.0	3	3	2
4.	An organization is assigned the block 2000:1456:2474/48. What is the CIDR notation for the blocks in the first and second subnets in the organization?	3	3	2
5.	State the advantages of IPv6 over IPv4.	3	1	2
6.	Define subnetting.	3	2	2
7.	Draw the packet format for IPv6.	3	2	2
8.	Expand ICMP and write its functions.	3	2	2

PART – B

1.	An ISP is given a block of addresses beginning with 190.100.0.0/16. The ISP needs to distribute these addresses to 3 groups of customers as follows: a) Group 1 has 64 customers each needs 256 addresses. b) Group 2 has 128 customers each needs 128 addresses. c) Group 3 has 128 customers each needs 64 addresses. Design the sub-blocks and give the slash notation for each sub-block. How many addresses are still available after these allocations?	3	3	16
2.	Explain about IPv6 datagram format with suitable diagram.	3	2	16
3.	Describe in detail about ARP operation, packet format and the situations when the ARP services are used.	3	2	16
4	Discuss in detail about the Internet Control Message Protocol (ICMP) with a neat diagram.	3	2	16

UNIT 4
ROUTING

Routing and protocols: Unicast routing - Distance Vector Routing - RIP - Link State Routing – OSPF – Path-vector routing - BGP - Multicast Routing: DVMRP – PIM.

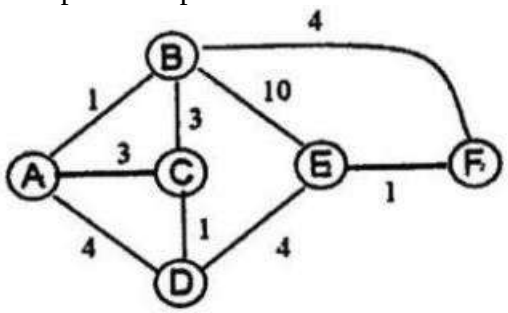
PART - A

Q.NO	QUESTION	CO	BTL	Marks
1.	Define routing protocol.	4	2	2
2	Differentiate Distance Vector Routing and Link State Routing.	4	2	2
3.	State the advantages of RIP over OSPF.	4	2	2
4.	Mention the functions of area border routers.	4	1	2
5.	Define PIM.	4	2	2
6.	List the functions of BGP.	4	1	2
7.	What does a router do when it receives a packet with a destination address that it does not have an entry for, in its routing table?	4	2	2

8.	How are packets routed using Link State Routing protocol?	4	2	2

PART – B

1.	i)Discuss in detail about Distance Vector Routing and demonstrate how distance table gives routing table.	4	2	8
	ii) Explain the working of Link State Routing in detail.	4	2	8
2.	Consider two routers exchanging information via BGP protocol. Explain BGP sessions and packet formats in detail between two routers.	4	3	16
3.	Consider the network shown below. Show the operation of Dijikstra's (Link State) algorithm for computing the least cost path from router F to all destinations. Explain the process.	4	3	16



4	Discuss in detail about Open Shortest Path First (OSPF) with neat diagrams.	4	2	16

UNIT 5
DATA LINK AND PHYSICAL LAYERS

Data Link Layer – Framing – Flow control – Error control – Data-Link Layer Protocols – HDLC – PPP - Medi Access Control – Ethernet Basics – CSMA/CD – Virtual LAN – Wireless LAN (802.11) - Physical Layer: Data an Signals - Performance – Transmission media- Switching – Circuit Switching.

PART - A

Q.NO	QUESTION	CO	BTL	Marks
1.	List the responsibilities of data link layer.	5	1	2
2	Differentiate fixed and variable size framing.	5	2	2

3.	Define bit stuffing.	5	2	2
4.	Compare error detection and correction.	5	2	2
5.	Define CSMA/CD.	5	2	2
6.	List the parameters used to measure network performance.	5	1	2
7.	Give the format for ethernet address.	5	2	2
8.	Name the authentication protocols of PPP.	5	1	2

PART – B

1.	Elaborate on the CSMA protocols. What is the main downfall of the Carrier Sense Multiple Access (CSMA) method? How does Collision Detection (CD) methods will help to alleviate this problem?	s	2	16
2.	Discuss in detail abour wireless LAN (802.11).	5	2	16
3.	Describe in detail about HDLC.	5	2	16
4	Consider the code word M(X)=111100010101 is transmitted in an unreliable medium .X^2+1 is the divisor commonly used for both sender and receiver. Calculate the CRC at the sender side as T(X). An error occurred at the bit position 6 (Note: from left to right) in T(X). Prove that the error occurs at the receiver side using CRC	5	3	16

---------- END ----------

GE3451
ENVIRONMENTAL SCIENCE AND SUSTAINABILITY

UNIT 1
ENVIRONMENT, ECO SYSTEM AND BIO-DIVERSITY

Definition, scope and importance of environment – need for public awareness. Eco-system and Energy flow– ecological succession. Types of biodiversity: genetic, species and ecosystem diversity– values of biodiversity, India as a mega-diversity nation – hot-spots of biodiversity – threats to biodiversity: habitat loss, poaching of wildlife, man-wildlife conflicts – endangered and endemic species of India – conservation of biodiversity: In-situ and ex-situ.

PART - A

Q.NO	QUESTION	CO	BTL	Marks
1.	State the significance of environmental education.	1	1	2
2	What are biotic and abiotic components of an ecosystem?	1	1	2
3.	How does a biome differ from ecosystem?	1	1	2
4.	Identify any two major threats to biodiversity	1	1	2
5.	Define Ecological succession.	1	1	2
6.	Distinguish between endangered and endemic species	1	1	2
7.	Define primary succession and secondary succession	1	2	2
8.	Mention the disadvantages of Exsitu Conservation..	1	2	2

PART - B

1.	Analyse the structure and function of an ecosystem with real examples.	1	5	16
2.	Explain the values of the biodiversity(8) “India is a mega diversity nation”–Discuss.(8)	1	5	16
3.	Identify the major threats to biodiversity(10) Write a note on Hot Spot Biodiversity (6)	1	5	16
4	Explain the conservation plan of biodiversity by integrating In-Situ and Ex-Situ.	1	5	16

UNIT 2
ENVIRONMENTAL POLLUTION

Causes, Effects and Preventive measures of Water, Soil, Air and Noise Pollutions. Solid, Hazardous and E-Waste management. Case studies on Occupational Health and Safety Management system (OHASMS). Environmental protection, Environmental protection acts.

PART – A

Q.NO	QUESTION	CO	BTL	Marks
1.	Classify air pollutants into primary and secondary category.	2	1	2
2	State the pollutant responsible for photo chemical smog.	2	1	2
3.	Compare point and non point sources of water pollutants.	2	2	2
4.	Define e-Waste Management	2	1	2
5.	Write the objectives of environmental acts.	2	1	2
6.	Define hazardous waste management	2	1	2
7.	What are the effects of noise pollution	2	2	2
8.	Identify the major cause of soil pollution.	2	1	2

PART - B

1.	Describe the role of individual in the prevention of pollution. Explain the sources, effects and control methods of noise pollution.	2	3	16
2.	Design a flow sheet and explain the steps involved in Solid Waste Management.	2	5	16
3.	Examine the Occupational Health and Safety Management System(OHASMS) with a relevant case study.	2	5	16
4.	Develop a schematic representation of an Industrial Waste Water treatment technique.	2	5	16

UNIT 3
RENEWABLE SOURCES OF ENERGY

Energy management and conservation, New Energy Sources: Need of new sources. Different types new energy sources. Applications of- Hydrogen energy, Ocean energy resources, Tidal energy conversion. Concept, origin and power plants of geothermal energy.

PART - A

Q.NO	QUESTION	CO	BTL	Marks
1.	Mention the objectives of energy management.	3	1	2
2	Assess the significance of OTE.	3	1	2
3.	Investigate the role of Artificial intelligence in energy sector.	3	1	2
4.	What is Bio-mass energy?	3	1	2
5.	Give any five applications of tidal energy conservation.	3	1	2
6.	Specify some important applications of GTE	3	2	2
7.	State the term DESS? Mention its components	3	1	2
8.	Mention the applications of hydrogen energy	3	1	2

PART - B

1.	Explain the applications of Ocean energy and Tidal Energy..	3	5	16
2.	Elaborate the principle and various steps involved in the energy management.	3	5	16
3.	Analyse the role of new energy sources in global energy demands.	3	5	16
4	Develop a detail report on the origin, concept and advantage and disadvantages Geo-Thermal energy.	3	5	16

UNIT 4
SUSTAINABILITY AND MANAGEMENT

Development GDP Sustainability- concept, needs and challenges-economic, social and aspects of sustainability-from unsustainability to sustainability-millennium development goals, and protocols Sustainable Development Goals-targets, indicators and intervention areas Climate change- Global, Regional and local environmental issues and possible solutions-case studies. Concept of Carbon Credit, Carbon Footprint. Environmental management in industry-A case study.

PART – A

Q.NO	QUESTION	CO	BTL	Marks
1.	Define the term GDP.	4	1	2
2	Identify any four millennium development goals.	4	1	2
3.	What is meant by carbon credit?	4	1	2
4.	What are the sources of carbon foot print?	4	2	2
5.	Define environmental management.	4	2	2
6.	Mention any five important needs of sustainability.	4	2	2
7.	Write some advantages of carbon credits.	4	1	2
8.	Mention some effects of climate change.	4	1	2

PART B

1.	Summarise the Millennium Development Goals and Sustainabilty protocols.	4	5	16
2.	Explain the various steps of environmental management.	4	6	16
3.	Elaborate the goal and aim of sustainable development.	4	4	16
4	Describe the causes, effects and possible solutions of climate change?	4	5	16

UNIT 5
SUSTAINABILITY PRACTICES

Zero waste and R concept, Circular economy, ISO 14000 Series, Material Life cycle assessment, Environmental Impact Assessment. Sustainable habitat: Green buildings, Green materials, Energy efficiency, Sustainable transports. Sustainable energy: Non-conventional Sources, Energy Cycles carbon cycle, emission and sequestration, Green Engineering: Sustainable urbanization- Socio economical and technological change.

Q.NO	QUESTION	CO	BTL	Marks
1.	What is zero waste and R concept?	5	1	2
2	Define circular economy.	5	2	2
3.	Identify key elements of ISO14000.	5	1	2
4.	Mention the objectives of EIA	5	1	2
5.	State the term green engineering.	5	1	2
6.	What is meant by energy cycles?	5	1	2
7.	Give an idea about the concept sustainable urbanization.	5	1	2
8.	Define carbon sequestration.	5	1	2

PART B

1.	Explain the various steps to achieve zero waste? Mention advantages and disadvantages of zero waste	5	2	16
2.	Analyse the 3R concept and suggest improvements for better sustainability	5	5	16
3.	Assess the role of green building materials and sustainable transport in urban planning.	5	5	16
4	Brief the rules to develop sustainable urbane.	5	3	16

----------END----------