Name: **Enrolment No:**	UPES UNIVERSITY WITH A PURPOSE

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
School of Computer Science

End Semester Examination, December 2020

Course : Pattern and anomaly detection **Semester : VI**
Program : B.Tech CSE AIML **Time : 03 Hours**
Course Code : CSAI 3107 **Max. Marks : 100**

SECTION A

	SECTION A **1. Each Question will carry 5 Marks**	**Marks**	
Q1	Select all correct uses of anomaly detection: a. Fraud detection b. Customer segmentation c. Intrusion detection d. Retention detection	**05**	**CO3**
Q2	Which of the following algorithm is not used for pattern detection: a. Support vector machines b. Linear regression c. Decision trees d. Isolated random forest	**05**	**CO1**
Q3	Select all that applies: You trained a Deep learning model which gives very high accuracy on the training data, but much lower accuracy on validation data. Which of the following may be true? a) This is an instance of overfitting. b) This is an instance of underfitting. c) The training was not well regularized. d) The training and testing examples are sampled from different distributions.	**05**	**CO3**
Q4	Null hypothesis is the basic assumption that we trying to prove. – True or False	**05**	**CO2**
Q5	Support vector machine can have which of the following kernals: a. Polynomial b. Radial basis function	**05**	**CO4**

	c. Custom d. None of these		
Q6	Which of the followings are true: a. Support vector machines can be used regression b. Decision trees are used for regression c. Logistic regression cannot be used for classification d. None of these	**05**	**CO3**
	SECTION B		
SECTION B **1. Each question will carry 10 marks** **2. Instruction: Write short / brief notes**			
Q7	Explain the following in context to pattern detection: a. Data Acquisition b. Pre-processing c. Feature Extraction d. Classification	**2.5*4**	**CO2**
Q8	Explain the use and importance of following in ML: a. Confusion Matrix b. F1 score c. Precision d. Recall	**2.5 *4**	**CO1**
Q9	Differentiate between z-test and t-test.	**10**	**CO2**
Q10	Analyze the use of following in Pattern detection: a. Gaussian mixture models b. Bayesian networks	**5+5**	**CO3**
Q11	Explain k-means algorithm and the application of the same for product recommendation system for online marketplace. OR Explain the following in relation to feature extraction: a. Robustness assessment b. Information evaluation c. Prognostic performance evaluation d. Redundancy reduction	**10**	**CO3**

SECTION C			
SECTION C **1. Each Question carries 10 Marks. 10*2=20** **2. Instruction: Write long answer.**			
Q12	Explain Local Outlier factor. Analyze its use for outlier detection and provide code using sklearn as base lib for the following: **Data fields** record ID - The unique identifier for each connection record. duration_ This feature denotes the number of seconds (rounded) of the connection. For example, a connection for 0.17s or 0.3s would be indicated with a "0" in this field. src_bytes This field represents the number of data bytes transferred from the source to the destination (i.e., the amount of out-going bytes from the host). dst_bytes This feature represents the number of data bytes transferred from the destination to the source (i.e., the amount of bytes received by the host). **What to submit** Your detection result should be in the same format as described in the handout of project 2. Specifically, there should be only 2 columns separated by the comma: record ID - The unique identifier for each connection record. is_anomaly?_ This binary field indicates your detection result: 0 denotes the transmission is normal, 1 indicates anomalous. OR What is point anomaly detection? Analyze the following types of point anomaly detection: a. Classification Based b. Nearest neighbor based c. Clustering based d. Statistical based	**20**	**CO4**