1. What do you mean by clustering in machine learning? Honor is about to putoni ason

ans Clustering is an unasupercuised learning technique in machine learning used to group Similar douta Points tegether bossed on their features. It helps to Patteren discovery and Segmentation without medolined lebels.

2. What 95 the note of of feature selection to machine bearing?

one feature selection involves choosing the most relevant features from the dataset to improve model accuracy, reduce over fitting and Thoreten straining time.

8. What is the decision Tree algorithm? ams A Decision Tree is a superruised learning algorithm used for chassification and regression. It splits data into branches based on feature values, forming a tree structures that leads to decision outcomes.

4. What is Apache Spark 1981. ans Apache Sparck Melib is a Scalable Machine learning libruary built on Farek. It provides

\* Chico Sind atransaction and a said

1 - 1/2 x x 11 (4/2) distributed algorithms and tooks for Classification, regression, clustering, and Bion: making it suitable for handling Big Data, real businesseprency on of prisobered Sec-B of poen burning outpart in antique 1. How does forcest improve classification tasks? Emplain with an ensumple. si ans. Random forcest is an eassemble beauting a mostro d'thout builds multiple devision troos and combines theirs adoptes to improve accuracy and reduce overdrifting for enample, in a customer church prediction multiple trees consider different subsets of features and Samples. The final decision is based on the majority vote; leading to morre reliable modictions. 2. What are the key steps 3n evaluating a Machine Learning model? and key steps coulde: · Splitting data in to Accounting and testing sets. . ? someting contribute of ? in · choosing appropriate metrice leg. accurracy, Proceision, recall). · Percherening cross- Validation. courses library built on facts It movides

- · Analyzing confusion matrin.
- · fine tuning hyperparameters.
- · Validating with unseen doctor to theck generalization.

## Sec-c

1. Discuss the roll of Distributed computing in Machine learning.

learning tasks to be Persferenced across multiple nodes, enabling Parallel data Processing faster treatning, and handling of lourge doutasets. It is essential for Big Deuta environments where treaditional single-machine systems are insufficient.

2. How does spack support scalable machine learning?

are Apache Sparck Supports Scalable Machine bearing through!

- · In-memorry computation for speed.

  · melib for distrubuted algorithms.
- · Resilient Distributed Deutasets (RDPs)
  for fault to kerrance.

Integreation with Hadoop and various. Sources. The fulling typesquare garage " Scalability across clusters for larage datasets. madization. migratures the not of Districtanced computing in R of third be disting. SL 0 Cc McDistropated Conquerry allows tructure Ye analysh tosks to be thereforeway occoss. altiple redes combing tendelled dota goverssing when treatment out heardling of houseye. whaster H is escipted to Big 12814 william spring tonoilions and and and and the sist free and market 1 Grand Suggest Scillephe How does Hinche n 1 (Friends) gridge ang Aprilia Blanck Companie Scalable Machine 2 individe the centre : = Whin melanney consideration ince speed 1 2 · milly for distributed olderstans. ( (Regilient Distributed Tenaseis (RED) dunit to second