**Assignment Submission for 2020AIML016**

**Q.1 The following questions are to enhance your understanding of basic concepts and definitions. You are expected to answer these questions on your own without referring to internet**

**a**. **State any four business applications of machine learning.**

**Ans**. 4 business applications of machine learning are :

1. CyberSecurity – Machine learning can be used to make cybersecurity defence stronger as machine learning based systems can constantly analyze network traffic and build stronger models which can identify abnormal traffic and proactively act to stop imminent threats . Some of the cybersecurity products which use machine learning are Microsoft Advanced Threat Protection , Microsoft Azure Sentinel etc.
2. Voice and Translation Services – Automatic machine learning based language services are used to aid conversational interfaces and are embedded in many modern business collaboration platforms like Microsoft Teams etc.
3. Healthcare Advisory – Modern healthcare systems leverage machine learning to do preemptive diagnosis of potential diseases in patients before the disease aggravates .
4. Financial Services – Machine learning has major applications in modern fintech services especially in the areas of fraud detection. Many modern fintech companies use ML based systems which can systematically detect fraudulent transactions and probably save institutions from financial loss .

**b**. **Explain the difference between supervised learning and unsupervised learning with example.**

**Ans.** Supervised learning uses labeled data to systematically train different machine learning models and predict outcomes based on certain inputs . The whole training lifecycles includes creation of labeled datasets which can be fed to niche algorithms to create trained models . Broadly supervised learning is of 2 types regression & classification. Example – Predicting the type of fruit by analyzing a set of attributes relating to the fruit

Unsupervised learning doesn’t use any labeled data to create refined models . Instead it allows itself to work on its own to discover patterns and other information for a given set of attributes . Broadly unsupervised learning is of 2 types , clustering & association. Example - Creating groups of different kinds of fruit based on their attributes .

**c. Provide differences between regression, classification and clustering problems. Give an example where each of these models could be used.**

Ans. Regression is a supervised learning technique used to predict a single non categorical value using the training data . Here the model is trained on different labeled attributes which have a single value as an output . Example – Regression can be used to predict the price of a house from training data . The attributes can be location of the house , size of the house , amenities available etc.

Classification is a supervised learning technique used to group different outputs in one or many classes . Classification can be binary , multi label or multi class classification. Example – Classification can be used to predict the type of fruit based on input attributes , i.e for a given model trained on data which predicts between an apple or an orange , classification can be used to predict the appropriate output for unknown data and categorize the input as an apple or an orange

Clustering is an unsupervised learning technique used to find groups or patterns in a set of unlabeled data . Clustering will process data and find groupings if any exist within the data that is provided . No of clusters to identify etc can be set as a hyper parameter within the training algorithm . Example – If randomly a dataset is provided which has different attributes of different kinds of flowers then clustering algorithm can be used to segregate and categorize the data into different clusters where each cluster might represent a particular type of flower

**Q.3** **The following dataset of size n = 63 provides total payment (in thousands) for all the claims of Swedish Kronor for geographical zones in Sweden. The variables are y = total payment in thousands and x = number of claims. Find the intercept, coefficient value and error term for this regression problem. How would you convey these numbers in a business context?**

**Ans**. From the attached notebook and code , the values calculated are as follows :

coefficient is 3.4138235600663664 intercept is 19.99448575911481

Error: 35.366

Now from a business perspective , the total payments in thousands (y) will depend on the variable x which is total number claims . The intercept tells that if the number of claims is 0 then what is the amount to be paid . The coefficient tells the change in the value of the payment if the claims changes by 1 . Error value tells us how much the predicted payment might vary from the actuals , it can be +-35.366 in this case on an average