



KEEPING MOBILE PHONE/SMART WATCH, EVEN IN 'OFF' POSITION, IS TREATED AS EXAM MALPRACTICE

Answer ALL Questions

(10 X 10 = 100 Marks)

1. How Sources of Data related to Information Commons will choose for Health Insurance Premium prediction and explain with all necessary process model steps with illustrate diagram by giving the Data Science Project Life Cycle framework.
 2. Why Probability models are necessary for adopting most data science projects design during the development process? Solve the following probability problem, A family has two children. We ask the father, "Do you have at least one daughter named Lilia?" He replies, "Yes!" What is the probability that both children are girls? In other words, we want to find the probability that both children are girls, given that the family has at least one daughter named Lilia. Here you can assume that if a child is a girl, her name will be Lilia with probability $\alpha \ll 1$ independently from other children's names. If the child is a boy, his name will not be Lilia.
 $\log\left(\frac{1}{4+1}\right) \rightarrow \log\left(\frac{1}{5}\right) \rightarrow \log\left(\frac{1}{5}\right) \rightarrow \log\left(\frac{1}{5}\right) \rightarrow \log\left(\frac{1}{5}\right)$
 3. Compare Linear Regression and Logistic Regression. What are the assumptions made through the Logistic Regression? Create own dataset for recently launched SUV cars and build a Machine Learning model using the Logistic regression algorithm to relate the features.
 4. Explain the different important key trends (common and Uncommon) involved in K Nearest Neighbors and Decision Trees. Demonstrate a Decision Tree model with assumed dataset of independent variables namely Outlook, Temperature, Humidity, and Wind. The dependent variable is whether to play Cricket or not.
 5. Compile the important functions used in Reading and Writing Datafile in R. Demonstrate the following problems in R Dataframe.
 - i. To create a data frame from four given vectors.
 - ii. To get the structure of a given data frame.
- When will you use a histogram and when will you use a bar chart in R? Explain with an own real time dataset example by leveraging R package. Can plots be exported as image files or other file formats in R? Explain briefly.
- Analyze ggplot2 heatmap with customization of reproducible with
- (a) basic plot
 - (b) Control color palette
 - (c) wide Input format,
 - (d) Interactive with Plotly.
- Suppose you want to develop a machine learning algorithm that predicts the number of views on the articles in a blog. Your data analysis is based on features like author name, number of articles written by the same author, etc. Justify which evaluation metrics would you choose? Explain briefly with an example.
- Organize the steps involved in Model selection depending on the following Evaluation criteria: F1 Score, Confusion Matrix, Area Under the ROC Curve in Machine learning process.

10. Assume that your business relies on computing services where the power consumed by your machines varies throughout the day. You do not know the actual cost of the electricity consumed by the machines throughout the day, but the organization has provided you with historical data of the price of the electricity consumed by the machines. Below is the information of the data we have for the task of forecasting electricity prices:

Date Time, Holiday, Holiday Flag, Day Of Week where 0 is Monday, Week Of Year, Day, Month, Year, Period Of Day, Forecast Wind Production, System Load EA forecasted national load forecasted price, actual national system load, the actual price of the electricity consumed (labels or values to be predicted)

So your task here is to use this data to train a machine learning model to predict the price of electricity consumed by the machines.

