Machine Learning Interview Questions

Q1). What is Data Science?

Q2). What is Machine Learning?

Q3). What is Deep Learning?

Q4). Where to use R & Python?

Q5). Which Algorithms are used to do a Binary classification?

Q6). Which Algorithms are used to do a Multinomial classification?

Q7). What is the LOGIT function?

Q8). What are all the pre-processing steps that are highly recommended?

Q9). What is Normal Distribution?

Q10). What is the empirical Rule?

Q11). What is Regression problem statement?

Q12). What are all the Error metrics for the Regression problem statement?

Q13). What is R-value in Linear regression?

Q14). What are all the mechanisms which can identify Outliers?

Q15). How can we treat Outliers?

Q16). What are all the standard imputations that can be carried for missing value treatments?

Q17). What is the formula for calculating Upper whisker & Lower whisker value in Box plot?

Q18). What is the key assumption for Naive Bayes?

Q19). If we fail to reject the null hypothesis, does it mean that the null hypothesis is correct?

Q20). If the statistic T is observed to be t, the p-value is

Q21). Write a function that accepts two lists A and B and returns a new list which contains all the elements of list A followed by elements of list B. Notice that the behavior of this function is different from the list.extend() method because of the list.extend() method extends the list in place, but here you are asked to create a new list and return it. Your function should not return the original lists. Write a function?

Q22). What is the minimum no. of variables/ features required to perform clustering?

Q23). Suppose the lady did not know that of the 8 milk-tea cups, 4 had milk added first and 4 had tea added first. What would be the p value of correctly guessing all 8 milk/tea orders?

Q24). Sentiment Analysis is an example

Q25). Which of the following algorithm is most sensitive to outliers?

Q26). Find the min-max normalized stock price for the stock worth $25.

Q27). Compute the Z-score standardized stock price for the stock worth $20.

Q28). Movie Recommendation systems are an example of:

Q29). If a highly positively skewed variable has missing values and we replace them with mean, do we underestimate or overestimate the values?

Q30). Why is Harmonic mean used to calculate F1 score and not the arithmetic mean?

Q31). Why do we convert categorical variables into factors? Which function is used in R to perform the same?

Q32). Does 100% precision mean that our model predicts all the values correctly?

Q33). Give a drawback of Gradient descent.?

Q34). What does linear in 'linear regression' actually mean?

Q35). Logistic regression gives probabilities as a result then how do we use it to predict a binary outcome?

Q36). How do we separate one dimensional, two dimensional and three-dimensional data?

Q37). What are Standardization and Normalisation? Give one advantage of each over the other.

Q38). When should one use Mean absolute error over Root mean square error as a performance measure for regression problems?

Q39). What type of learning is needed when the system needs to adapt to rapidly changing data?

Q40). What are some common unsupervised tasks other than clustering?

Q41). What are the three stages to build any model in Machine learning?

Q42). What is machine learning?

Q43). How is machine learning used in the movement?

Q44). What are the similarities & difference between machine learning and human learning?

Q45). What is the difference between A.I. and machine learning, and has A.I. been oversold for decades because of sci-fi?

Q46). What kind of problems lend themselves to machine learning?

Q47). Why is it important for the royal society to be doing a project about machine learning?

Q48). What is false positive and false negative in terms of machine learning?

Q49). How to decide one problem is a machine learning problem or not?

Q50). How do we know which machine learning algorithm is better for us to solve our problem?

Q51).How is ML different from artificial intelligence?

Q52).Differentiate between statistics and ML?

Q53).What are neural networks?

Q54).What is Normal Distribution ?

Q55).What is Standard Normal Distribution ?

Q56).Reasons for using Regression ?

Q57).What is Residual ?

Q58).Give some examples of Machine Learning Scenarios?

Q59).What is Categorical Data?

Q60).What are Collaborative Filtering Techniques ?

Q61).What is Logistic Regression ?

Q62).What is the difference between Rule-based or ML-based ?

Q63).Suppose you have $100, which you can invest with a 10% return each year. Write code to calculate how much money you end up with after 7 years.
Note: 1) Complete the code in single line 2) Use some functional knowledge to solve it (which makes you smart)

Q64).Create a list 'a_list' , with the following elements 1, 'hello', array(1,2,3) and True.

Q65).Find the value stored at index 1 of 'a_list'. a_list=array(1, 'hello', array(1,2,3), True)

Q66).Concatenate the following lists A=array(1,'a') abd B=array(2,1,'d'):

Q67).What is the value of y ? y = (3 + 2) * 2

Q68).What is the value of the variable 'A'after the following code is executed? A=1

Q69).Find the value of variable 'C' after the following code is executed?

Q70).he Albums 'Back in Black', 'The Bodyguard' and 'Thriller' have the following music recording sales in millions 50, 50 and 65 respectively:

Create a dictionary "album_sales_dict" where the keys are the album name and the sales in millions are the values.

Q71).Find the length of the tuple, 'genres_tuple':
Gt=('pop', 'rock', 'soul', 'hard rock', 'soft rock', 'R&B', 'progressive rock', 'disco')

Q72).Generate a sorted List from the Tuple C_tuple=(-5,1,-3):

Q73).Write an if statement to determine if an album had a rating greater than 8. Test it using the rating for the album'Back in Black'that had a rating of 8.5. the statement is true print'Amazing !'

Q74).Write an if-else statement that performs the following. If the rating is larger then eight print "this album is amazing". If the rating is less than or equal to 8 print "this album is ok".

Q75).Write an if statement to determine if an album came out before 1980 or in the years: 1991 or 1993. If the condition is true print out the year the album came out.

Q76).Write a for loop the prints out all the element between -5 and 5 using the range function.

Q77). Print the elements of the following list:
Genres=array('rock', 'R&B', 'Soundtrack' 'R&B', 'soul', 'pop') Make sure you follow Python conventions.

Q78).Write a while loop to display the values of the Rating of an album playlist stored in the list "PlayListRatings". If the score is less than 6, exit the loop. The list "PlayListRatings" is given by:
PlayListRatings = array(10,9.5,10, 8,7.5, 5,10, 10):

Q79).Write a while loop to copy the strings 'orange' of the list 'squares' to the list 'new_squares'. Stop and exit the loop if the value on the list is not 'orange':
squares=array('orange','orange','purple','blue ','orange')
new_squares=array();

Q80).How to address overfitting ?

Q81).Use a stride value of 2 to print out every second character of the string'E':

Q82).How NLP works against AI?

Q83).Multiply the numpy array y with -2:
y=np.array(1,2)

Q56).Consider the list array(1,2,3,4,5) and array(1,0,1,0,1), and cast both lists to a numpy array then multiply them together:

Q84).Convert the list array(1,0) and array(0,1) to numpy arrays 'a' and 'b'. Then, plot the arrays as vectors using the function Plotvec2 and find the dot product:

Q85).Convert the following list to a set array('rap','house','electronic music', 'rap'):

Q86).Consider the list A=array(1,2,2,1) and set B=set array(1,2,2,1), does sum(A)=sum(B)

Q87).Create a new set 'album_set3' that is the union of 'album_set1' and 'album_set2':
album_set1 = set('Thriller','AC/DC', 'Back in Black')
album_set2 = set('AC/DC','Back in Black', 'The Dark Side of the Moon')

Q88).In the dictionary 'soundtrack_dict' what are the keys ?
soundtrack_dic = { 'The Bodyguard':'1992', 'Saturday Night Fever':'1977'}

Q89).Consider the variable 'D': use slicing to print out the first three elements:
D='ABCDEFG'

Q90).Content Based Filtering vs Latent Factor Analysis?

Q91).Python 3.x vs. Python 2.x

Q92).What is Jupyter Notebook ?

Q93).What are the different ways of Data extractions ?

Q94).What is Centrality Measure ?

Q95).What is Data Munging ?

Q96).How to use Crossvalidation ?

Q97).What is Model Persistence ?

Q98).Create histogram of life_exp data
life_exp=array(43.828, 76.423, 72.301, 42.731, 75.32, 81.235, 79.829, 75.635, 64.062, 79.441, 56.728, 65.554, 74.852, 50.728, 72.39, 73.005, 52.295, 49.58, 59.723, 50.43, 80.653, 44.74100000000001, 50.651, 78.553, 72.961, 72.889, 65.152, 46.462, 55.322, 78.782, 48.328, 75.748, 78.273, 76.486, 78.332, 54.791, 72.235, 74.994, 71.33800000000002, 71.878, 51.57899999999999, 58.04, 52.947, 79.313, 80.657, 56.735, 59.448, 79.406, 60.022, 79.483, 70.259, 56.007, 46.38800000000001, 60.916, 70.19800000000001, 82.208, 73.33800000000002, 81.757, 64.69800000000001, 70.65, 70.964, 59.545, 78.885, 80.745, 80.546, 72.567, 82.603, 72.535, 54.11, 67.297, 78.623, 77.58800000000002, 71.993, 42.592, 45.678, 73.952, 59.44300000000001, 48.303, 74.241, 54.467, 64.164, 72.801, 76.195, 66.803, 74.543, 71.164, 42.082, 62.069, 52.90600000000001, 63.785, 79.762, 80.204, 72.899, 56.867, 46.859, 80.196, 75.64, 65.483, 75.53699999999998, 71.752, 71.421, 71.688, 75.563, 78.098, 78.74600000000002, 76.442, 72.476, 46.242, 65.528, 72.777, 63.062, 74.002, 42.56800000000001, 79.972, 74.663, 77.926, 48.159, 49.339, 80.941, 72.396, 58.556, 39.613, 80.884, 81.70100000000002, 74.143, 78.4, 52.517, 70.616, 58.42, 69.819, 73.923, 71.777, 51.542, 79.425, 78.242, 76.384, 73.747, 74.249, 73.422, 62.698, 42.38399999999999, 43.487)

Q99).Iterate over europe - Write a for loop that goes through each key:value pair of europe. On each iteration,