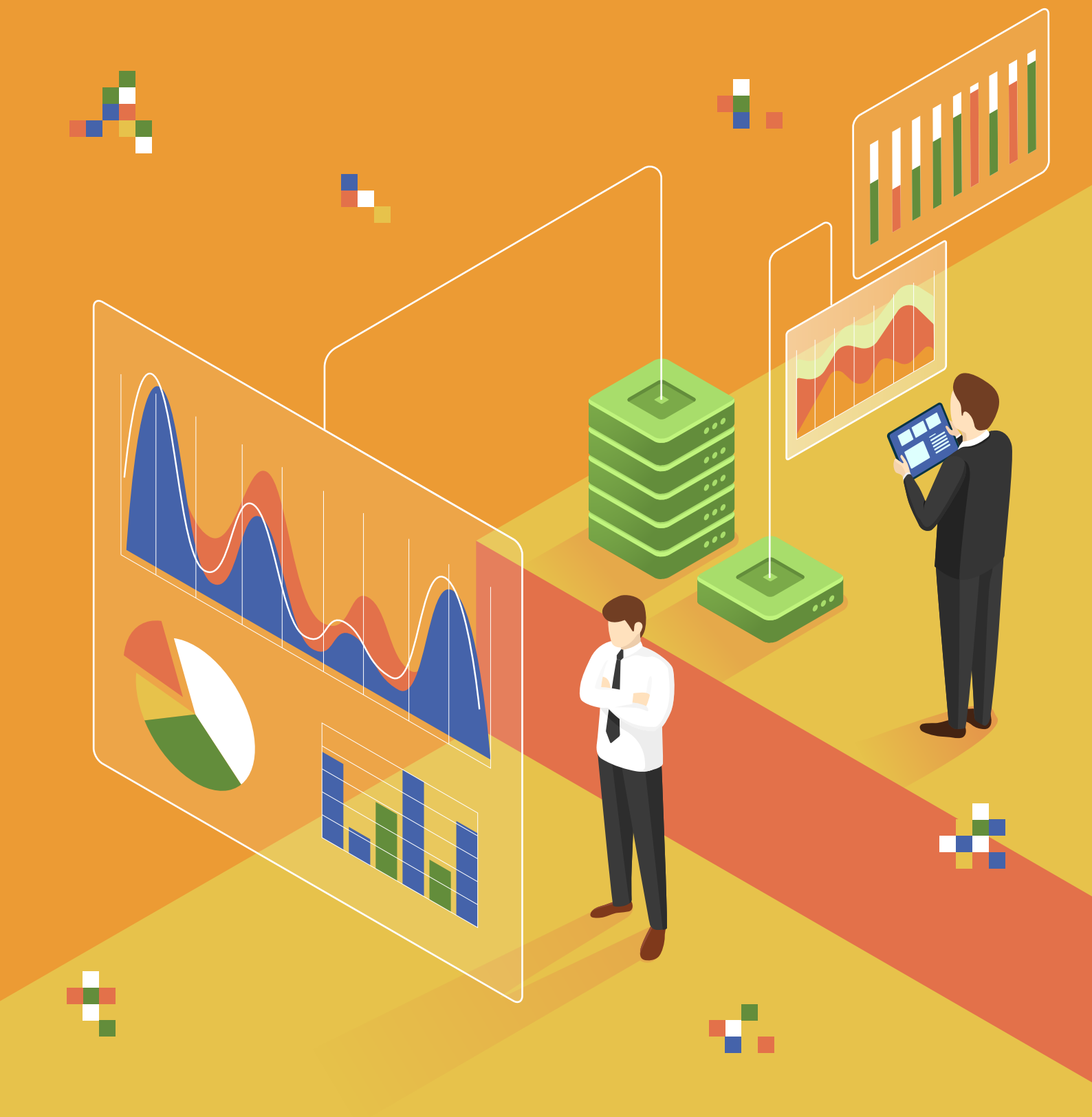
SuperDataScience

AWS Machine Learning
Practical Exams Reward



1. **You work at a car dealership and you have been tasked to develop a Machine Learning model to predict the purchasing power of your potential customers. You are given access to data that contains earning per year, net worth, number of cars they own and number of family members. Initial analysis shows there are no outliers in net worth and earnings. But after training your ML model, the model seems to perform quite poorly. How can you improve your model?**

   A Standardize the data of net worth and earnings
   B Apply categorical encoding
   C Apply one-hot encoding
   D Normalize the data of net worth and earnings

2. **You have been given access to 3 student marks data containing nominal categorical values like John, Smith and Jake. What transformation can be done to feed the data into ML regression model?**

   A Use one-hot encoding to get the data in numeric format
   B Remove the names column
   C Use normalization to get the data in numeric format
   D Use custom encoding to get the data in numeric format

3. **You are working on house price prediction model. You have given access to data containing features like area of the house, rooms in the house, name of the owner and price of the house. What are the processing steps that should be performed to prepare the data before running ML model?**

   A Drop the area column and perform PCA on the remaining data
   B Drop the owner name column and normalize the remaining data
   C Drop the price column and normalize the remaining data
   D Drop the rooms column and perform PCA on the remaining data

4. **You want to perform trigrams on the following text { Hi ! What are you doing ?}. You want apply lower case transformation and remove all the punctuations. What are the unique bigrams that are produced?**

   A {'hi what', 'What are', 'are you', 'you doing'}
   B {'hi !' ,'! what', 'what are', 'are you', 'you doing', 'doing ?'}
   C {'hi what', 'what are', 'are you', 'you doing'}
   D {'hi ! what', 'what are', 'are you', 'you doing?'}

5. **Which of the following cases you recommend the use of one-hot encoding technique?**

   A When the data contains 10000 nominal categorical values and the model need numeric value
   B When the data continuous values and the model need numeric value
   C When the data contains 3 nominal categorical values and the model need numeric value
   D When the data contains 3 ordinal categorical values and the model needs numeric value

6. **You have been given access to 3 student marks data containing nominal categorical values like John, Smith and Jake. What transformation can be done to feed the data into ML regression model?**

   A Remove the names column
   B Use normalization to get the data in numeric format
   C Use custom encoding to get the data in numeric format
   D Use one-hot encoding to get the data in numeric format

7. **You are building a model to predict whether the home loan would be approved or not. Data that you want to train contains the following features, number of rooms, area, pool size which is represented by small, medium and large and the type of the house like condo, apartment and house. How would you prepare this for training?**

   A Map different values for type and use one hot encoding to each pool size
   B Use one hot encoding for type and pool size
   C Map different values to each pool size and use one hot encoding for type
   D Map different values to each pool size and each type

8. **You have given access to a company's data and after analysis you find that there are missing values along with outliers. How would you fill in the missing values and how would you take care of the outliers?**

   A Drop the outliers and fill the missing values using mean values
   B Drop the outliers and the missing values
   C Keep the outliers and fill the missing values using mean values
   D Keep the outliers and drop missing values

9. **What are the pre-processing steps involved in processing unstructured text data before feeding through NTM?**

   A Use One-hot encoding and tokenize the data
   B Use Categorical encoding and form bag-of-words
   C Use vector transformation and one hot encoding
   D Tokenize the data and form Bag-of-words

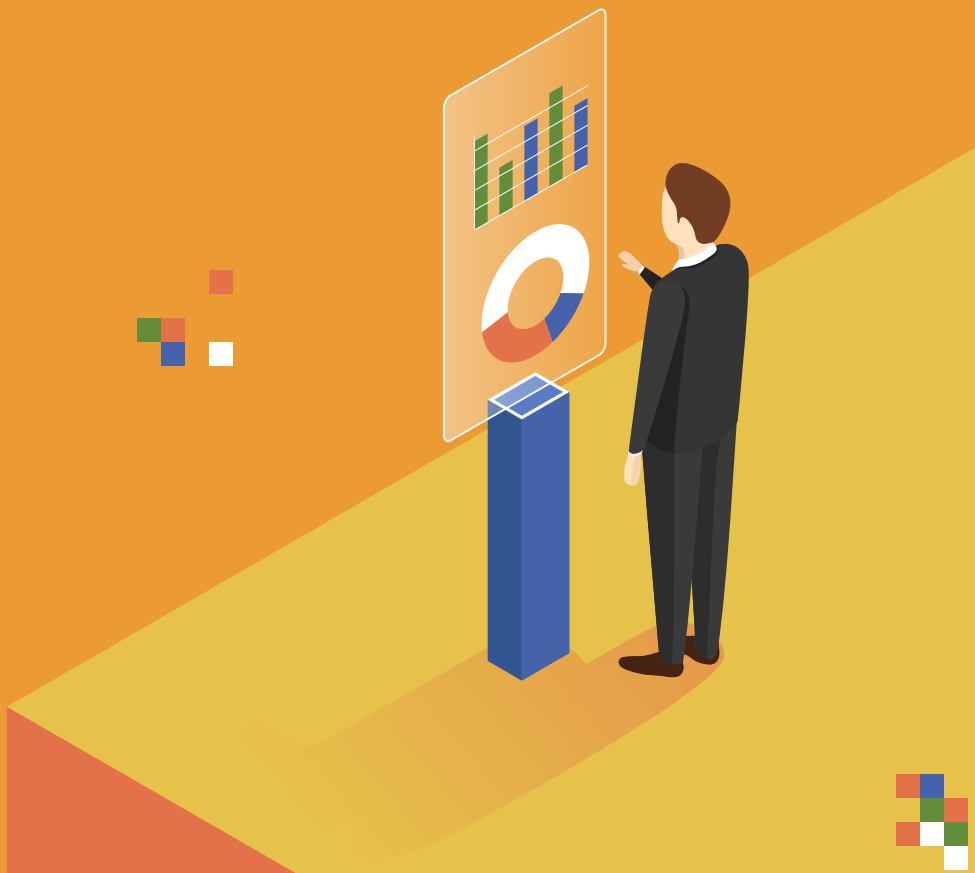
10. You have given access to a customer data containing their name, address, age, income and credit score. You want to create a model to predict the interest rate for the loan that the person would receive. What are the necessary preprocessing steps?

A Drop name and age column and normalize the remaining data

B Drop address and age column and normalize the remaining data

C Drop name and address column and normalize the remaining data

D Encode the categorical data and normalize the remaining data



ANSWER KEY:

1. D	2. A	3. B	4. C	5. C
6. D	7. C	8. A	9. D	10. C