

ML Practical List

1 Import House prediction dataset from Kaggle and find out the total missing value from the dataset and fill the missing value and visualize the data in heat map.

2 Create a dummy database (.csv file) from below figure and Split the data into two parts(dependent and independent variable) ,find out the total missing value from above dataset and encode the total number of columns into categorical value into numeric value.

	Develoer	Age	Salary	Married
0	Karachi	42.0	78000.0	No
1	Lahore	32.0	48000.0	Yes
2	Karachi	36.0	60000.0	No
3	Multan	41.0	68000.0	No
4	Lahore	42.0	NaN	Yes
5	Multan	43.0	59000.0	Yes
6	Karachi	NaN	59000.0	No
7	Lahore	44.0	79000.0	Yes
8	Multan	52.0	99000.0	No
9	Lahore	32.0	69000.0	Yes

3 Write a python code to implement a decision tree for the below given dataset. Identify the root node and all subparts or children of the node and draw the tree.

	Item no	Age	Income	Student	Credit Reting	Buys-Computer
0	1	Youth	High	No	Fair	No
1	2	Youth	High	No	Excellent	No
2	3	Middle	High	No	Fair	Yes
3	4	Senior	Medium	No	Fair	Yes
4	5	Senior	Low	Yes	Fair	Yes
5	6	Middle	Low	Yes	Excellent	No
6	7	Senior	Low	Yes	Excellent	Yes
7	8	Youth	Medium	No	Fair	No
8	9	Youth	Low	Yes	Fair	Yes
9	10	Senior	Medium	Yes	Fair	Yes
10	11	Youth	Medium	Yes	Excellent	Yes
11	12	Middle	Medium	No	Excellent	Yes
12	13	Middle	High	Yes	Fair	Yes
13	14	Senior	Low	No	Excellent	No

4 Write a python code to implement a decision tree for Iris Dataset (from sklearn.dataset import load_iris) dataset. Identify the root node and all subparts or children of the node and draw the tree.

5 Write a python code to implement the K-nearest neighbourhood program for the given dataset.

	<table><tr><th>Item no</th><th>Temp</th><th>Humidity</th><th>Wind Speed</th><th>Play</th></tr><tr><td>1</td><td>85</td><td>85</td><td>12</td><td>No</td></tr><tr><td>2</td><td>80</td><td>90</td><td>9</td><td>No</td></tr><tr><td>3</td><td>83</td><td>86</td><td>4</td><td>Yes</td></tr><tr><td>4</td><td>70</td><td>96</td><td>3</td><td>Yes</td></tr><tr><td>5</td><td>68</td><td>80</td><td>5</td><td>Yes</td></tr><tr><td>6</td><td>65</td><td>70</td><td>20</td><td>No</td></tr><tr><td>7</td><td>64</td><td>65</td><td>2</td><td>Yes</td></tr><tr><td>8</td><td>72</td><td>95</td><td>12</td><td>No</td></tr><tr><td>9</td><td>69</td><td>70</td><td>5</td><td>Yes</td></tr><tr><td>10</td><td>75</td><td>80</td><td>2</td><td>Yes</td></tr><tr><td>11</td><td>75</td><td>70</td><td>3</td><td>Yes</td></tr><tr><td>12</td><td>72</td><td>90</td><td>4</td><td>Yes</td></tr><tr><td>13</td><td>81</td><td>75</td><td>5</td><td>Yes</td></tr><tr><td>14</td><td>71</td><td>91</td><td>15</td><td>No</td></tr></table>	Item no	Temp	Humidity	Wind Speed	Play	1	85	85	12	No	2	80	90	9	No	3	83	86	4	Yes	4	70	96	3	Yes	5	68	80	5	Yes	6	65	70	20	No	7	64	65	2	Yes	8	72	95	12	No	9	69	70	5	Yes	10	75	80	2	Yes	11	75	70	3	Yes	12	72	90	4	Yes	13	81	75	5	Yes	14	71	91	15	No															
Item no	Temp	Humidity	Wind Speed	Play																																																																																							
1	85	85	12	No																																																																																							
2	80	90	9	No																																																																																							
3	83	86	4	Yes																																																																																							
4	70	96	3	Yes																																																																																							
5	68	80	5	Yes																																																																																							
6	65	70	20	No																																																																																							
7	64	65	2	Yes																																																																																							
8	72	95	12	No																																																																																							
9	69	70	5	Yes																																																																																							
10	75	80	2	Yes																																																																																							
11	75	70	3	Yes																																																																																							
12	72	90	4	Yes																																																																																							
13	81	75	5	Yes																																																																																							
14	71	91	15	No																																																																																							
6	Write a python code to implement the K-nearest neighbourhood program for the Breast Cancer Wisconsin Dataset (from sklearn.dataset import load_breast_cancer) dataset.Obtain model accuracy,generate classification report.																																																																																										
7	<p>Write a python code to apply Naive Bayesian algorithm to classify that whether a person can buy computer or not based on given test data:</p> <table><tr><th>Item no</th><th>Age</th><th>Income</th><th>Student</th><th>Credit Rating</th><th>Buys-Computer</th></tr><tr><td>1</td><td>Youth</td><td>High</td><td>No</td><td>Fair</td><td>No</td></tr><tr><td>2</td><td>Youth</td><td>High</td><td>No</td><td>Excellent</td><td>No</td></tr><tr><td>3</td><td>Middle</td><td>High</td><td>No</td><td>Fair</td><td>Yes</td></tr><tr><td>4</td><td>Senior</td><td>Medium</td><td>No</td><td>Fair</td><td>Yes</td></tr><tr><td>5</td><td>Senior</td><td>Low</td><td>Yes</td><td>Fair</td><td>Yes</td></tr><tr><td>6</td><td>Middle</td><td>Low</td><td>Yes</td><td>Excellent</td><td>No</td></tr><tr><td>7</td><td>Senior</td><td>Low</td><td>Yes</td><td>Excellent</td><td>Yes</td></tr><tr><td>8</td><td>Youth</td><td>Medium</td><td>No</td><td>Fair</td><td>No</td></tr><tr><td>9</td><td>Youth</td><td>Low</td><td>Yes</td><td>Fair</td><td>Yes</td></tr><tr><td>10</td><td>Senior</td><td>Medium</td><td>Yes</td><td>Fair</td><td>Yes</td></tr><tr><td>11</td><td>Youth</td><td>Medium</td><td>Yes</td><td>Excellent</td><td>Yes</td></tr><tr><td>12</td><td>Middle</td><td>Medium</td><td>No</td><td>Excellent</td><td>Yes</td></tr><tr><td>13</td><td>Middle</td><td>High</td><td>Yes</td><td>Fair</td><td>Yes</td></tr><tr><td>14</td><td>Senior</td><td>Medium</td><td>No</td><td>Excellent</td><td>No</td></tr></table> <p>Test Data Age: Youth Income: LOW Student: No Credit Rating: Fair Buy Computer -??</p>	Item no	Age	Income	Student	Credit Rating	Buys-Computer	1	Youth	High	No	Fair	No	2	Youth	High	No	Excellent	No	3	Middle	High	No	Fair	Yes	4	Senior	Medium	No	Fair	Yes	5	Senior	Low	Yes	Fair	Yes	6	Middle	Low	Yes	Excellent	No	7	Senior	Low	Yes	Excellent	Yes	8	Youth	Medium	No	Fair	No	9	Youth	Low	Yes	Fair	Yes	10	Senior	Medium	Yes	Fair	Yes	11	Youth	Medium	Yes	Excellent	Yes	12	Middle	Medium	No	Excellent	Yes	13	Middle	High	Yes	Fair	Yes	14	Senior	Medium	No	Excellent	No
Item no	Age	Income	Student	Credit Rating	Buys-Computer																																																																																						
1	Youth	High	No	Fair	No																																																																																						
2	Youth	High	No	Excellent	No																																																																																						
3	Middle	High	No	Fair	Yes																																																																																						
4	Senior	Medium	No	Fair	Yes																																																																																						
5	Senior	Low	Yes	Fair	Yes																																																																																						
6	Middle	Low	Yes	Excellent	No																																																																																						
7	Senior	Low	Yes	Excellent	Yes																																																																																						
8	Youth	Medium	No	Fair	No																																																																																						
9	Youth	Low	Yes	Fair	Yes																																																																																						
10	Senior	Medium	Yes	Fair	Yes																																																																																						
11	Youth	Medium	Yes	Excellent	Yes																																																																																						
12	Middle	Medium	No	Excellent	Yes																																																																																						
13	Middle	High	Yes	Fair	Yes																																																																																						
14	Senior	Medium	No	Excellent	No																																																																																						
8	Write a python code to apply Naive Bayesian algorithm to classify whether a person can survive or not for Titanic Dataset (https://www.openml.org/d/42438) and obtain model accuracy.																																																																																										

9	Implement support vector machine approach to predictive modelling for Diabetes (from sklearn.dataset import load_diabetes) datasets. Also, evaluate the model.																																		
10	Implement support vector machine approach to predictive modelling for iris (from sklearn.dataset import load_iris) datasets. Also, evaluate the model.																																		
11	<p>Implement a python program that takes interest rate (x), finds the equation that best fits the data and is able to forecast out median home price for given interest rate using the data given below. (Use linear regression)</p> <table><tr><th>Interest rate (%) (x)</th><th>Median home price (y)</th></tr><tr><td>10.3</td><td>\$183,800</td></tr><tr><td>10.3</td><td>\$183,200</td></tr><tr><td>10.1</td><td>\$174,900</td></tr><tr><td>9.3</td><td>\$173,500</td></tr><tr><td>8.4</td><td>\$172,900</td></tr><tr><td>7.3</td><td>\$173,200</td></tr><tr><td>8.4</td><td>\$173,200</td></tr><tr><td>7.9</td><td>\$169,700</td></tr><tr><td>7.6</td><td>\$174,500</td></tr><tr><td>7.6</td><td>\$177,900</td></tr><tr><td>6.9</td><td>\$188,100</td></tr><tr><td>7.4</td><td>\$203,200</td></tr><tr><td>8.1</td><td>\$230,200</td></tr><tr><td>7</td><td>\$258,200</td></tr><tr><td>6.5</td><td>\$309,800</td></tr><tr><td>5.8</td><td>\$329,800</td></tr></table>	Interest rate (%) (x)	Median home price (y)	10.3	\$183,800	10.3	\$183,200	10.1	\$174,900	9.3	\$173,500	8.4	\$172,900	7.3	\$173,200	8.4	\$173,200	7.9	\$169,700	7.6	\$174,500	7.6	\$177,900	6.9	\$188,100	7.4	\$203,200	8.1	\$230,200	7	\$258,200	6.5	\$309,800	5.8	\$329,800
Interest rate (%) (x)	Median home price (y)																																		
10.3	\$183,800																																		
10.3	\$183,200																																		
10.1	\$174,900																																		
9.3	\$173,500																																		
8.4	\$172,900																																		
7.3	\$173,200																																		
8.4	\$173,200																																		
7.9	\$169,700																																		
7.6	\$174,500																																		
7.6	\$177,900																																		
6.9	\$188,100																																		
7.4	\$203,200																																		
8.1	\$230,200																																		
7	\$258,200																																		
6.5	\$309,800																																		
5.8	\$329,800																																		
12	<p>Using the Multiple Linear Regression approach solve the problem of predicting home price in monroe, new jersey (USA) Below is the table containing home prices in monroe twp, NJ. Here price depends on area (square feet), bed rooms and age of the home (in years). Given these prices we have to predict prices of new homes based on area, bed rooms and age.</p> <table><tr><th>area</th><th>bedrooms</th><th>age</th><th>price</th></tr><tr><td>2600</td><td>3.0</td><td>20</td><td>550000</td></tr><tr><td>3000</td><td>4.0</td><td>15</td><td>565000</td></tr><tr><td>3200</td><td>NaN</td><td>18</td><td>610000</td></tr><tr><td>3600</td><td>3.0</td><td>30</td><td>595000</td></tr><tr><td>4000</td><td>5.0</td><td>8</td><td>760000</td></tr><tr><td>4100</td><td>6.0</td><td>8</td><td>810000</td></tr></table> <p>Given these home prices find out price of a home that has,</p> <p>3000 sqr ft area, 3 bedrooms, 40 year old</p> <p>2500 sqr ft area, 4 bedrooms, 5 year old</p>	area	bedrooms	age	price	2600	3.0	20	550000	3000	4.0	15	565000	3200	NaN	18	610000	3600	3.0	30	595000	4000	5.0	8	760000	4100	6.0	8	810000						
area	bedrooms	age	price																																
2600	3.0	20	550000																																
3000	4.0	15	565000																																
3200	NaN	18	610000																																
3600	3.0	30	595000																																
4000	5.0	8	760000																																
4100	6.0	8	810000																																

13

Consider the below dataset and This database contains hiring statistics for a firm such as experience of candidate, his written test score and personal interview score. Based on these 3 factors, HR will decide the salary. Given this data, you need to build a machine learning model for the HR department that can help them decide salaries for future candidates.

experience	test_score(out of 10)	interview_score(out of 10)	salary(\$)
NaN	8.0	9	50000
NaN	8.0	6	45000
five	6.0	7	60000
two	10.0	10	65000
seven	9.0	6	70000
three	7.0	10	62000
ten	NaN	7	72000
eleven	7.0	8	80000

Using this predict salaries for following candidates,

2 yr experience, 9 test score, 6 interview score

12 yr experience, 10 test score, 10 interview score

Note:Solve the given problem using Multiple Linear Regression

14

Implement Logistic Regression approach to predictive modelling for Diabetes (from sklearn.dataset import load_diabetes) datasets. Also, evaluate the model

15

Build a prediction model using regression technique for Boston house-prices (from sklearn.dataset import load_boston) Also, evaluate the model

16

Write a python code to implement Apriori algorithm, apply join and prune method and find frequent item sets.

Sr#	Item no	Na me
1	T1	Bread, butter, milk, soda
2	T2	Coke, egg, milk
3	T3	Bread, butter, egg
4	T4	Break, coke, jam
5	T5	Bread, butter
6	T6	Potato chips, soda
7	T7	Coke, fruit, juice
8	T8	Bread, coke, milk
9	T9	Coke, soda, jam, milk
10	T10	Bread, butter, egg, milk, soda
11	T11	Bread, milk
12	T12	Bread, jam

17	Implement Apriori algorithm in python to find rules which explain association between different products for given transactions at a retail store. (The data is available at https://drive.google.com/file/d/1NUXoptUIHY8z4KcFKpFA6sQN5KnWzk3p/view?usp=sharing)																		
18	Implement unsupervised machine learning algorithm (Clustering – K Means) in python on Titanic dataset (https://www.openml.org/d/42438) to cluster data by removing the class label. Use elbow method to find the optimal number of clusters.																		
19	Implement unsupervised machine learning algorithm (Clustering – K Means) in python on Breast Tumour dataset to cluster data (use Breast Tumour dataset) by removing the class label.																		
20	<p>Implement an unsupervised machine learning algorithm (Clustering – K Means) in python on the below dataset to cluster data by removing the class label. Use elbow method to find the optimal number of clusters.</p> <table><thead><tr><th>Name</th><th>Age</th><th>Income(\$)</th></tr></thead><tbody><tr><td>Rob</td><td>27</td><td>70000</td></tr><tr><td>Michael</td><td>29</td><td>90000</td></tr><tr><td>Mohan</td><td>29</td><td>61000</td></tr><tr><td>Ismail</td><td>28</td><td>60000</td></tr><tr><td>Kory</td><td>42</td><td>150000</td></tr></tbody></table>	Name	Age	Income(\$)	Rob	27	70000	Michael	29	90000	Mohan	29	61000	Ismail	28	60000	Kory	42	150000
Name	Age	Income(\$)																	
Rob	27	70000																	
Michael	29	90000																	
Mohan	29	61000																	
Ismail	28	60000																	
Kory	42	150000																	