Class Test #1 Solution

Programme Name: B.Tech (CSE- AIML) **Semester : 6th**
Course Name : Cognitive Analytics
Course Code : CSBA 3009 **Max. Marks : 25**
Nos. of page(s) : 01
Instructions : Answer all the questions.

S. No.		Marks	CO
Q1	What is a cognitive system? Describe the characteristics of a cognitive system. **Solution:** The explosion of data, mainly unstructured data, over the past few years led to the development of a new type of computer system known as a cognitive system. Unlike the programmable computers that preceded it, the focus of cognitive systems is not about doing fast calculations on large amounts of data through traditional computer programs. Cognitive systems are about exploring the data, finding new correlations, and new context in that data to provide new solutions. Cognitive systems aim at expanding the boundaries of human cognition rather than replacing or replicating the way the human brain works. **Characteristics of cognitive systems** (i) An important concept to understand is that the first key element of cognitive systems is to expand the boundaries of human cognition rather than replace or replicate the way the human brain works. (ii) The second key element is to have a more natural interaction between computers and humans. (iii) A third key element of cognitive systems is the use of learning, specifically machine learning (iv) The intent is to broaden the potential for learning and the ability of a to adapt over time with use, which is a fourth key element of cognitive systems.	5	CO1
Q2			

Name	Mobile_Sales	TV_Sales
Ahmed	2540	2200
Omar	1370	1900
Ali	1320	2150
Ziad	2000	1850
Salwa	2100	1770
Lila	2150	2000

(i) Write a Python script to create a data frame for the above table.	3	CO2
Solution:	1	
	1	

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
salesMen = ['Ahmed', 'Omar', 'Ali', 'Ziad', 'Salwa', 'Lila']
Mobile_Sales = [2540, 1370, 1320, 2000, 2100, 2150]
```

```
TV_Sales = [2200, 1900, 2150, 1850, 1770, 2000]
df = pd.DataFrame()
df ['Name'] =salesMen
df ['Mobile_Sales'] = Mobile_Sales
df['TV_Sales']=TV_Sales
df.set_index("Name",drop=True,inplace=True)
```

(ii) Create a bar plot of the sales volume.

Solution:

```
df.plot.bar( figsize=(20, 10), rot=0).legend(bbox_to_
anchor=(1.1, 1)) plt.xlabel('Salesmen') plt.ylabel('Sales')
plt.title('Sales Volume for two salesmen in \nJanuary and April
2017')
plt.show()
```

(iii) Create a box plot of item sales.

Solution:

```
df.plot.box()
```

Q3

Consider the problem of binary classification using the Naive Bayes classifier. You are given two dimensional features (X_1, X_2) and the categorical class conditional distributions in the tables below. The entries in the tables correspond to $P(X_1 = x_1|C_i)$ and $P(X_2 = x_2|C_i)$ respectively. The two classes are *equally likely*.

$X_1 =$ \ Class	C_1	C_2
-1	0.2	0.3
0	0.4	0.6
1	0.4	0.1

$X_2 =$ \ Class	C_1	C_2
-1	0.4	0.1
0	0.5	0.3
1	0.1	0.6

Given a data point $(-1, 1)$, calculate the following posterior probabilities:

$P(C_1|X_1 = -1, X_2 = 1) =$ Using Bayes' Rule and conditional independence assumption of Naive Bayes

Solution:

$$\frac{P(X_1=-1,X_2=1|C_1)P(C_1)}{P(X_1=-1,X_2=1)} = \frac{P(X_1=-1|C_1)P(X_2=1|C_1)P(C_1)}{P(X_1=-1|C_1)P(X_2=1|C_1)P(C_1)+P(X_1=-1|C_2)P(X_2=1|C_2)P(C_2)} = 0.1$$

$$P(C_2|X_1 = -1, X_2 = 1) = 1 - P(C_1|X_2 = -1, X_1 = 1) = 0.9$$

5

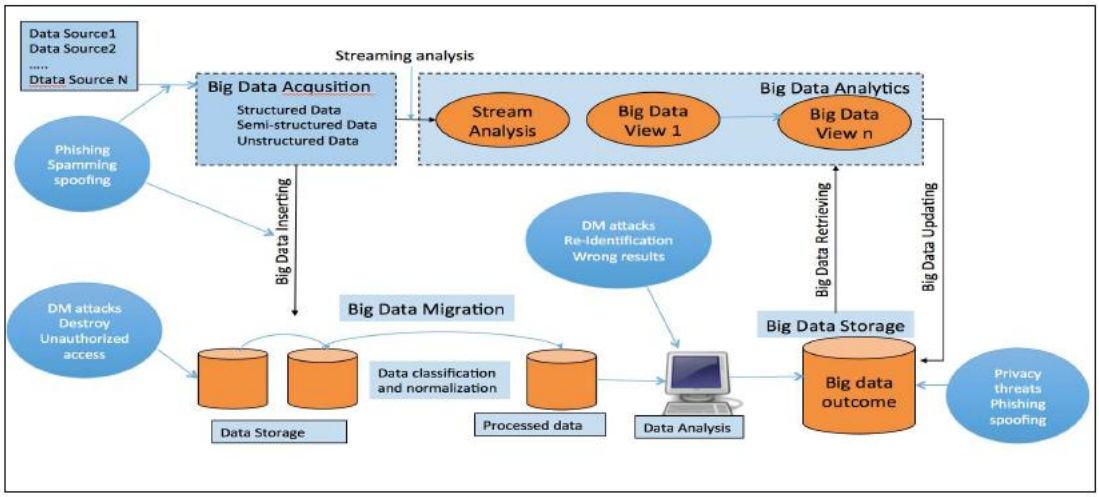
CO2

Q4

Suppose there are two models M1 and M2.
For M1: TP=6954, FN=46, FP=412 and TN=2588
For M2: TP=6800, FN=134, FP=566 and TN=2500
Calculate Accuracy, Recall, Precision and F1-score. Among M1 and M2 which one is more

5

CO2

	preferable model? **Solution:** **accuracy (ACC)** $$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{P} + \mathrm{N}} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$ **sensitivity, recall, hit rate, or true positive rate (TPR)** $$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} = 1 - \mathrm{FNR}$$ **precision or positive predictive value (PPV)** $$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} = 1 - \mathrm{FDR}$$ **F1 score** is the harmonic mean of precision and sensitivity: $$\mathrm{F}_1 = 2 \times \frac{\mathrm{PPV} \times \mathrm{TPR}}{\mathrm{PPV} + \mathrm{TPR}} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$		
Q5	Describe big data life cycle phases in the perspective of security and privacy. Also, discuss the threats involved in each phase. **Solution:** Four types of users' role in big data environment: data provider, data collector, data miner, and decision maker. However, our model addresses the phases of the big data lifecycle. The model consists of four phases in big data framework consists of data collection phase, data storage phase, data processing and analysis, and knowledge creation. Figure 1 presents the main elements in big data lifecycle.  **Figure 1: Big Data Lifecycle Threat Model**	5	CO1