Assignment No :A6

Roll No.4431

1 Title:

Business Analytics and Intelligence for profit maxim.

2 Problem Definition

Select an Industrial sector and write an application using BIA tool for maximizing the profit.

- Design an application using tool for clustering the iris flowers dataset.
- Design an application using tool to predict whether to play tennis or not.

3 Learning Objectives

1. To understand the concept of BAI.
2. To understand the profit maximisation in Business Analytics and intelligence.
3. To be able to synthesize the BAI tools for customised problems.

4 Learning Outcomes

1. Ability to analyse real world problems and create BAI tools for the same.
2. Understanding of deep and profound theoretical aspects of the BAI systems.

5 Related Mathematics

Mathematical model is given as below,

S=fs,e,X,Y,Fme,DD,NDD,Mem_shared

Where,
s = Initial State
e = End State
X=Input
Datasets of iris flowers and tennis.

Y=Osutput

Clustering of iris flowers and prediction whether to play tennis or not, using the interactive table, scatterplot, histogram, etc.

Fme=Algorithms used (K-means and NaÃ¯ve-Bayes)

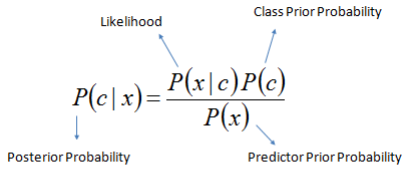
- K-means Algorithm : (Used for iris dataset)

  1. Randomly select 'c' cluster centers.
  2. Calculate the distance between each data point and cluster centers.
  3. Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers..
  4. Recalculate the new cluster center using:

  $$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

  where, 'c' represents the number of data points in ith cluster.
  5. Recalculate the distance between each data point and new obtained cluster centres.
  6. If no data point was reassigned then stop, otherwise repeat from step 3.

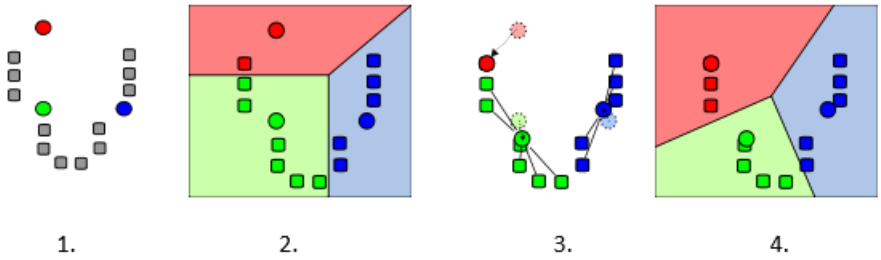
- NaÃ¯ve-Bayes Algorithm : (Used for tennis dataset)



$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \cdots \times P(x_n|c) \times P(c)$$

Fig : Bayes Rule

K-means example :



1. k initial "means" (in this case k=3) are randomly generated within the data domain (shown in color)
2. k clusters are created by associating every observation with the nearest mean. The partitions here represent the diagram generated by the means
3. The centroid of each of the k clusters becomes the new mean.
4. Steps 2 and 3 are repeated until convergence has been reached.

Naive-Bayes Example :

Outlook	Tempreature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

Figure 1: Given a training set, we can compute the probabilities

Outlook	P	N		Humidity	P	N
sunny	2/9	3/5		high	3/9	4/5
overcast	4/9	0		normal	6/9	1/5
rain	3/9	2/5				
Tempreature				Windy		
hot	2/9	2/5		true	3/9	3/5
mild	4/9	2/5		false	6/9	2/5
cool	3/9	1/5				

$X =< sunny, mild, high, true >$
$Pr(X|"no").Pr("no") = (3/5.2/5.4/5.3/5).5/14 = 0.04$
$Pr(X|"yes").Pr("yes") = (2/9.4/9.3/9.3/9).9/14 = 0.007$

DD=Deterministic Data
Iris dataset,Tennis Dataset.
NDD=Non -Deterministic Data
Number of clusters, etc.
Mem-shared=memory shared by the applications.

6 Concepts related theory

6.1 Introduction to Business Analytics and Intelligence :

Business Analytics :
Business analytics (BA) is the practice of iterative, methodical exploration of an organization's data with emphasis on statistical analysis.
Business analytics is used by companies committed to data-driven decision making.

Business Intelligence is querying, reporting, OLAP, and alert tools can answer questions such as what happened, how many, how often, where the problem is, and what actions are needed. Business analytics can answer questions like why is this happening, what if these trends continue, what will happen next (that is, predict), what is the best that can happen (that is, optimize).

Some examples of BA are :

- Exploring data to find new patterns and relationships (data mining)
- Forecasting future results (predictive modeling, predictive analytics)
- Explaining why a certain result occurred (statistical analysis, quantitative analysis)

Business Intelligence :

Business intelligence (BI) is often described as "the set of techniques and tools for the transformation of raw data into meaningful and useful information for business analysis purposes".

BI technologies are capable of handling large amounts of unstructured data to help identify, develop and otherwise create new strategic business opportunities. The goal of BI is to allow for the easy interpretation of these large volumes of data. Identifying new opportunities and implementing an effective strategy based on insights can provide businesses with a competitive market advantage and long-term stability.

BI technologies provide historical, current and predictive views of business operations. Common functions of business intelligence technologies are reporting, online analytical processing, analytics, data mining, process mining, complex event processing, business performance management, benchmarking, text mining, predictive analytics and prescriptive analytics.

BI can be used to support a wide range of business decisions ranging from operational to strategic.

KNIME tool :

KNIME, the Konstanz Information Miner, is an open source data analytics, reporting and integration platform. KNIME integrates various components for machine learning and data mining through its modular data pipelining concept. A graphical user interface allows assembly of nodes for data pre-processing $ETL : Extraction, Transformation, Loading$, for modelling and data analysis and visualization. Since 2006, KNIME has been used in pharmaceutical research,[2] but is also used in other areas like CRM customer data analysis, business intelligence and financial data analysis.

Innovative organizations use our open-source, enterprise-grade analytics platform to discover the potential hidden in their data, mine for fresh insights or predict new futures. It is quick to deploy, easy to scale, and intuitive, KNIME is used in over 60 countries on data of every kind: from numbers to images, molecules to humans, signals to complex networks, from kilo- to petabytes, or simple reports to complex analyses.

7 Output

- Input: Training and Input Data:

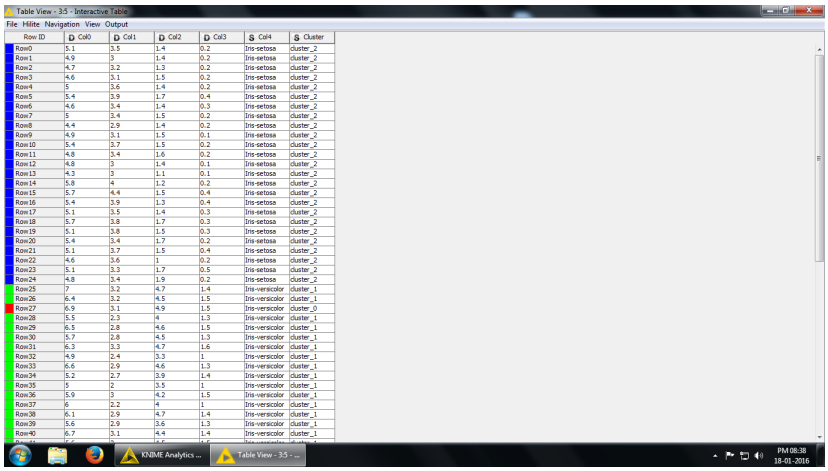


table1 kmeans.png

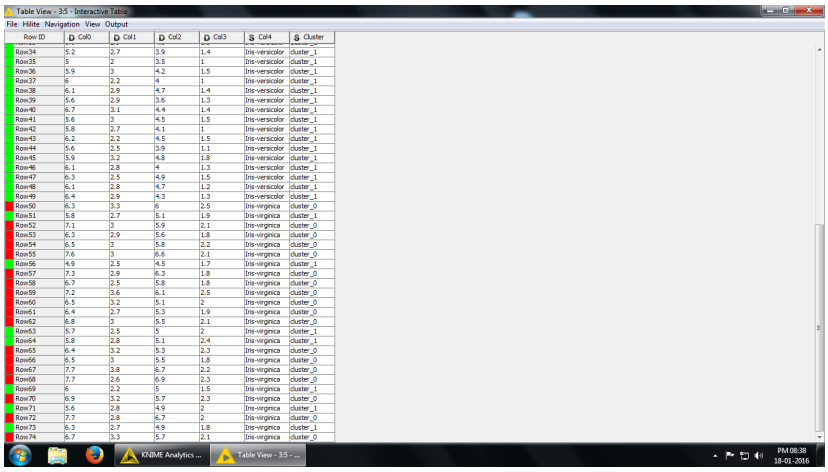
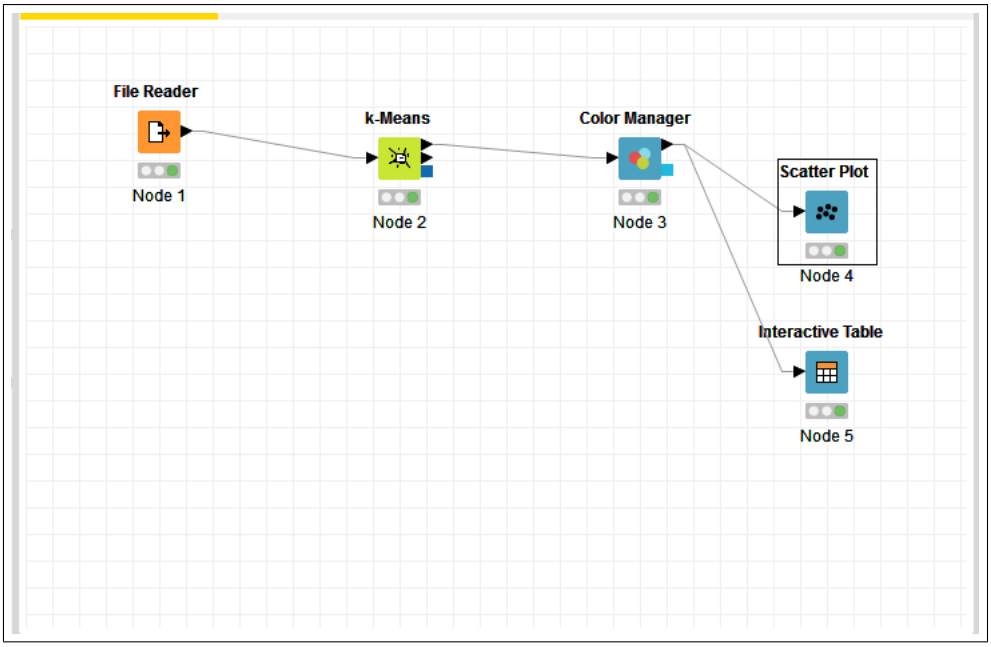
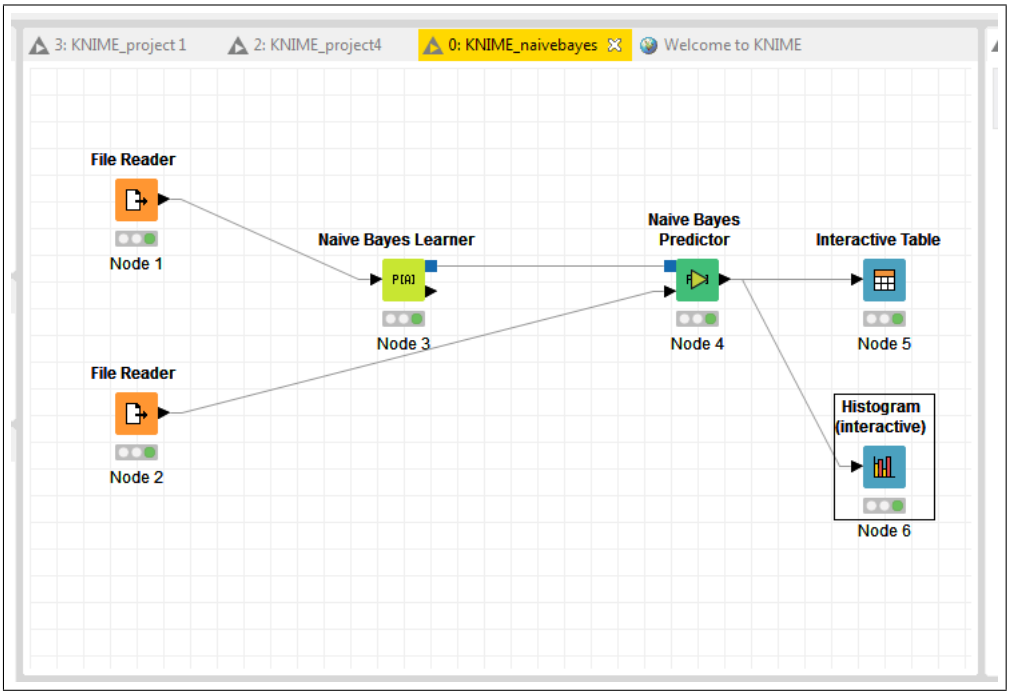


table2 kmeans.png

- Knime Design:

  Training and Input Data:





- Output

plot k-means.png

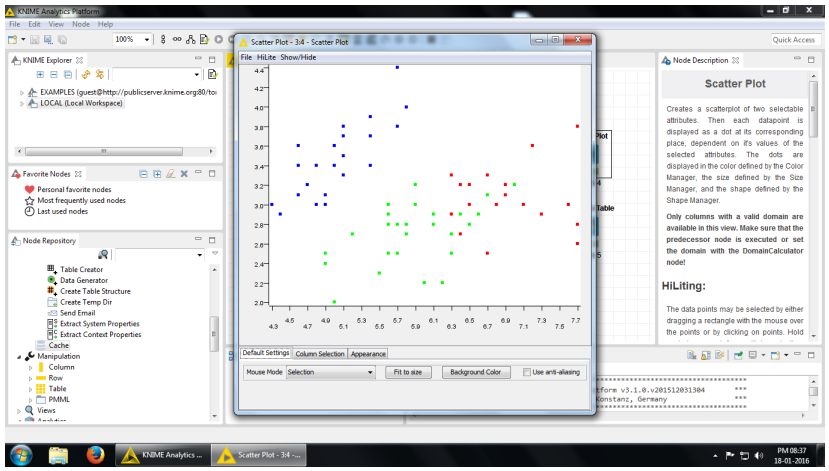
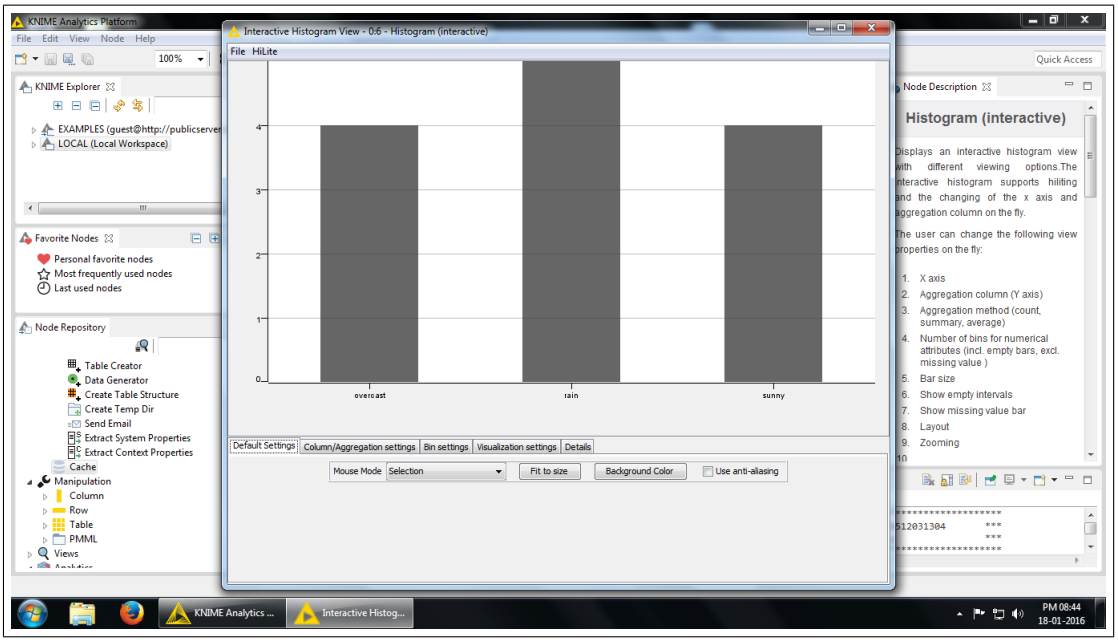


Table View - 0:5 - Interactive Table

File Hilite Navigation View Output

Row ID	sunny	hot	high	weak	Predicti...
D2	sunny	hot	high	strong	yes
D3	overcast	hot	high	weak	yes
D4	rain	mild	high	weak	yes
D5	rain	cool	normal	weak	yes
D6	rain	cool	normal	strong	yes
D7	overcast	cool	normal	strong	yes
D8	sunny	mild	high	weak	yes
D9	sunny	cool	normal	weak	yes
D10	rain	mild	normal	weak	yes
D11	sunny	mild	normal	strong	yes
D12	overcast	mild	high	strong	yes
D13	overcast	hot	normal	weak	yes
D14	rain	mild	high	strong	yes



8 CONCLUSION :

Hence, we have designed an application using tool for categorizing the iris flowers dataset and an application to predict whether to play tennis or not.