**SVKM’S NMIMS Deemed-to-be-University Mukesh Patel School of Technology Management & Engineering Department of Computer Engineering**

|  |  |  |  |
| --- | --- | --- | --- |
| Course Code |  | Program |  |
| Semester | V | Year | III |
| Name of the Faculty | Prachi Natu | Class |  |
| Course Title | Data Mining | Academic Year | 2020-21 |
|  |  |  |  |

LAB

Manual

PART

A

(PART A : TO BE REFFERED BY

STUDENTS)

**Experiment**

**No.06**

**A.1**

**Aim:**

Write a program for Naive Bayes Classification.

**A.2**

**Prerequisite:**

Refer the Lab manual for the steps and any programming language.

**A.3**

**Outcome:**

**After successful completion of this experiment students will be able to**

1. To understand the concept of Data Mining by implementing some data mining algorithm.
2. To understand the various Classification techniques in Mining.
3. To understand Naive Bayes Classification Algorithm.

**A.4 Theory**

**Classification Problem**

* Given a database D={t1,t2,…,tn} and a set of classes C={C1,…,Cm}, the ***Classification*** ***Problem*** is to define a mapping f:DgC where each tiis assigned to one class.
* Actually divides D into ***equivalence*** *classes*.
* ***Prediction*** is similar, but may be viewed as having infinite number of classes.

**SVKM’S NMIMS Deemed-to-be-University Mukesh Patel School of Technology Management & Engineering Department of Computer Engineering**

|  |  |  |  |
| --- | --- | --- | --- |
| Course Code |  | Program |  |
| Semester | V | Year | III |
| Name of the Faculty | Prachi Natu | Class |  |
| Course Title | Data Mining | Academic Year | 2020-21 |
|  |  |  |  |

**Classification Examples**

* Teachers classify students’ grades as A, B, C, D, or F.
* Identify mushrooms as poisonous or edible.
* Predict when a river will flood.
* Identify individuals with credit risks.
* Speech recognition
* Pattern recognition

**Bayesian Classification**

* Given training data D, posterior probability of a hypothesis h, P(h|D) follows the Bayes theorem

|  |  |
| --- | --- |
|  | P(D|h) P(h) |
| P(h|D) = | P(D) |

* P(h) : independent probability of h: prior probability
* P(D) : independent probability of D
* P(D|h) : conditional probability of D given h: likelihood
* P(h|D) : conditional probability of h given D: posteriori probability

**A.5 Procedure/Algorithm:**

**A.5.1 TASK 1:**

D : Set of tuples

\_ Each Tuple is an ‘n’ dimensional attribute vector

\_ X : (x1,x2,x3,…. xn)

Let there be ‘m’ Classes : C1,C2,C3…Cm

Naïve Bayes classifier predicts X belongs to Class Ci iff \_ P (Ci/X) > P(Cj/X) for 1<= j <= m

**SVKM’S NMIMS Deemed-to-be-University Mukesh Patel School of Technology Management & Engineering Department of Computer Engineering**

|  |  |  |  |
| --- | --- | --- | --- |
| Course Code |  | Program |  |
| Semester | V | Year | III |
| Name of the Faculty |  | Class |  |
| Course Title | Data Mining | Academic Year | 2020-21 |
|  |  |  |  |

Maximum Posteriori Hypothesis

\_ P(Ci/X) = P(X/Ci) P(Ci) / P(X)

\_ Maximize P(X/Ci) P(Ci) as P(X) is constant

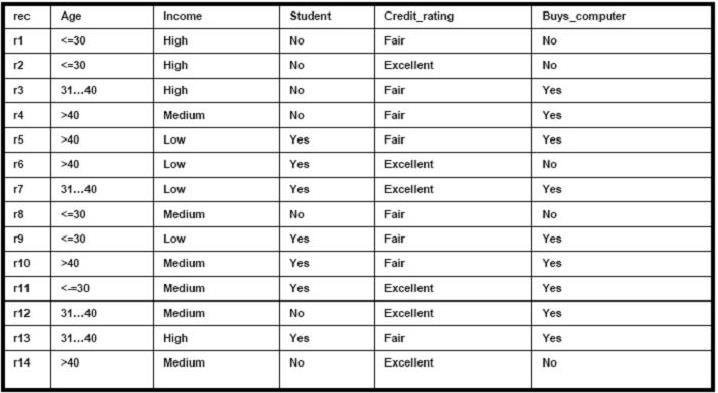
With many attributes, it is computationally expensive to evaluate P(X/Ci).

Naïve Assumption of “class conditional independence”

P(X/Ci) = P(x1/Ci) \* P(x2/Ci) \*…\* P(xn/ Ci)

X = ( age= youth, income = medium, student = yes, credit\_rating = fair)

A person belonging to tuple X will buy a computer?



P(C1) = P(buys\_computer = yes) = 9/14 =0.643

P(C2) = P(buys\_computer = no) = 5/14= 0.357

P(age=youth /buys\_computer = yes) = 2/9 =0.222

P(age=youth /buys\_computer = no) = 3/5 =0.600

P(income=medium /buys\_computer = yes) = 4/9 =0.444

P(income=medium /buys\_computer = no) = 2/5 =0.400

P(student=yes /buys\_computer = yes) = 6/9 =0.667

P(student=yes/buys\_computer = no) = 1/5 =0.200

P(credit rating=fair /buys\_computer = yes) = 6/9 =0.667

P(credit rating=fair /buys\_computer = no) = 2/5 =0.400

**SVKM’S NMIMS Deemed-to-be-University Mukesh Patel School of Technology Management & Engineering Department of Computer Engineering**

P(X/Buys a computer = yes) = P(age=youth /buys\_computer = yes) \* P(income=medium /buys\_computer = yes) \* P(student=yes /buys\_computer = yes) \* P(credit rating=fair /buys\_computer = yes) = 0.222 \* 0.444 \* 0.667 \* 0.667 = 0.044 P(X/Buys a computer = No) = 0.600 \* 0.400 \* 0.200 \* 0.400 = 0.019

Find class Ci that Maximizes P(X/Ci) \* P(Ci)

=>P(X/Buys a computer = yes) \* P(buys\_computer = yes) = 0.028 =>P(X/Buys a computer = No) \* P(buys\_computer = no) = 0.007

Prediction : Buys a computer for Tuple X