**📄 PDF Conversion Report**

Original File: Unit-4 DM himani.pdf  
Conversion Method: Enhanced Text Extraction  
Date: 30/8/2025, 11:36:17 am

--- Page 1 ---  
1  
set. This is done to ensure that the model generalizes well .  
6. Model Evaluation : It involves assessing the performance of the trained model on a test  
using the training set, and its performance is evaluated on validation set.  
the data. The data is divided into a training set and a validation set. The model is trained  
5. Model Training: It i nvolves the selected classification algorithm to learn the patterns in  
classify inputs into predefined categories.  
layers of interconnected nodes (neurons) that learn complex patterns from data to  
Neural Networks is a computational model inspired by the human brain, consisting of  
hyper plane to separate data points of different classes with the maximum margin.  
Support Vector Machine (SVM) is a supervised learning algorithm that finds optimal  
Theorem to predict probability of class based on prior knowledge and observed features.  
Bayesian classification is a probabilistic classification t echnique that applies Bayes'  
creating a tree - like structure to predict the target class by following decision rules.  
D ecision T ree is hierarchical model that splits data into subsets based on feature values,  
4. Model Selection: It involves selecting the appropriat e classification algo rithm. Thes e are :  
dataset. It identifies most important features in dataset and removes redundant ones.  
Principal Component Analysis is a technique used to reduce the dimensionality of the  
classification. Features with high information gain are selected for class ification.  
Information Gain is a measure of the amount of information that a feature provides for  
variables to identify patterns, dependencies, and associations in the data.  
Correlation Analysis measures the statistical relationship b etween two or more  
classification. This can be done using various techniques, such as :  
3. Feature Selection: It involves identifying the most relevant attributes in the dataset for  
suitable form for analysis.  
This involves handling missing values, dealing with outliers, a nd transforming the d ata into  
2. Data Pre - processing: The collected data needs to be pre - processed to ensure its quality.  
various sources, such as surveys, questionnaires, websites, and databases.  
necessary attributes and labels needed for classification. The data can be collect ed from  
1. Data Collection: In this step, relevant data is collected. The data should contain all the  
Process of Classification:  
unseen data bas ed on past observations. Example: email spam detection.  
predefined groups or classes. The goal is to build a model that can accurately classify new,  
Classification is a supervised learning technique in data mining used to categorize data into  
CLASSIFICATION:  
UNIT - 4 CLASSIFICATION AND PREDICTION  
  
--- Page 2 ---  
2  
performance after training.  
 Used to evaluate the model’s  
Test Data:  
 Used to build and train the model.  
Training Data:  
and retrain it as needed to maintain accuracy over time.  
it to make predictions on real - world data. Continuously monitor the mode l's performance  
6. Deployment and Monitoring: Once we a re satisfied with the model's performance, deploy  
actual outcomes. It f ine - tune s the model to improve its accuracy and generalization ability.  
data) that the model hasn't seen before. It evaluate that how the model's predictions mat ches  
5. Model Evaluation: It evaluate model's performance using separate portion of data (testing  
that it can use to make predictions on new data.  
relationships between the features and the outcomes. The mode l learns patterns and rules  
4. Model Training: It uses a portion of our data (training data) to teach the model the  
o Time Series Analysis: Predicts values over time (e.g, sales forecasts, weather patterns)  
o Classification: Predicts categories or classes ( e.g., spam/not spam )  
o Regression: Predicts continuous values (e.g., house prices, stock values)  
the prediction task. Some common models include:  
3. Mod el Selection: Choose s an appropriate prediction model based on the type of data and  
outcome you want to predict. This step helps simplify the model and improve its accuracy.  
2. Feature Selection: It identifies the most important variables (features) that influence the  
the data into a suitable format for analysis.  
and pre - process the data to handle missing values, inconsistencies, and noise. It t ransform s  
1. Data Collection and Preparation: It gathers relevant data from various sources. It cleans  
Process of Prediction:  
 Estimating customer purchases based on past shopping behaviour.  
 Forecasting stock market trends using historical s tock data.  
 Predicting house prices based on location, size, and facilities .  
decisions by forecasting trends, behaviours , or numerical values. For example :  
patterns found in historic al data. It helps businesses and researchers make data - driven  
Prediction is a technique used to estimate unknown values or future outcomes based on  
PREDICTION:  
  
--- Page 3 ---  
3  
DECISION TREE INDUCTION:  
area, including - 1 to 1 or 0 to 1 .  
Normalization includes scaling all values for a given attribute in side a smal l specified  
 Since generalization reduces the initial training data, t he data can also be normalized .  
street, can be generalized to highe r - level concepts, such as the city.  
including low, medium, and high. Similarly , nominal - va lued attributes , such as the  
numeric values for the attribute income can be generalized to the discrete field  
be used . This is especially helpful for continuous - valued attributes. For example ,  
 The data can be generalized to a higher - level ap proach. Here, Concept hierarchies can  
3. Data transformation :  
efficiency at the time of load increment without reducing the performance.  
referred to as feature selection. Hence the main objective is to improve classification  
redundant attributes from the learning procedure. In machine learning, this step is  
 Therefore, Relevance Analysis is performed on the data to delete some irrelevant or  
some other attributes can be redundant.  
a bike was fil ed i s unlikely to be relevant to the success of the application. Moreover,  
prediction task. For example , data recording the day of the week on which a dealer sold  
 There are various attributes in the data that can be irrelevant to the classification or  
2. Relevance analysis :  
reducing confusion during learning.  
have some structures for managing noisy or missing information, this step can supp ort  
methods and the operation of missing values. Although various classification algorithms  
 Th is defines the pre - processing of data to eliminate or reduce noise by using smoothing  
1. Data cleaning :  
Issues in Classification and Prediction:  
  
--- Page 4 ---  
4  
3. For more class labels, the computational complexity of the decision tree may increase.  
2. It may have over - fitting issue, which can be resolved using Random Forest algorithm.  
1. The decision tree contains lots of layers, which makes it complex.  
Disadvantages of the Decision Tree:  
3. It helps to think about all the possible outcomes for a problem.  
2. It can be very useful for solving decision - related problems.  
1. It is simple to understand.  
Advantages of the Decision Tree :  
5. Parent/Child node: The root node of tree is parent node, & other nodes are child nodes.  
4. Branch/Sub Tree: A tr ee formed by splitting the tree.  
according to the given conditions.  
3. Splitting: Splitting is the process of dividing the decision node/root node into subnodes  
further after getting a leaf node.  
2. Leaf Node : Leaf nodes are the final output node, and the tree cannot be segregated  
dataset, which further gets divided into two or more homogeneous sets.  
1. Root Node: Root node is from where the decision tree starts. It represents the entire  
Decision Tree Terminologies:  
with the root node, which expands on further branches and constructs a tree - like structure.  
Regression Tree algorithm. It is called a decision tree because, similar to a tree, it starts  
 In order to build a tree, we use the CART algorithm, which stands for Classification and  
numeric data.  
(Yes/No), it further splits tree into sub - trees. It can contain cate gorical data (YES/NO) and  
based on given conditions. A decision tree simply asks a question, and based on the answer  
 It is a graphical representation for getting all the possible solutions to a problem/decision  
 The decisions or the test are performed on the basis of features of the given dataset .  
b) Leaf nodes are output of those decisions and do not contain any further branches.  
a) Decision nodes are used to make any decision and have multiple branches.  
 In a Decision tree, there are two nodes, which are the Decision Node a nd Leaf Node.  
branches represent the decision rules and each leaf node represents the outcome.  
 It is a tree - structured classifier, where internal nodes represent the features of a dataset,  
and Regression problems, but mostly it is preferred for solving Classification p roblems.  
 Decision Tree is a supervised learning technique that can be used for both classification  
  
--- Page 5 ---  
5  
 S is the dataset.  
aft er splitting by A.  
S is the dataset, A is the attribute and Sv are subsets of S  
 p i is the probability of class i .  
Where,  
Where,  
| 퐒 |  
Entropy(S) = − ∑ p i log 2 p i  
IG(S,A) = Entropy(S) − ∑ ⋅ Entropy(Sv)  
| 퐒퐯 |  
Entropy of the entire dataset: Information Gain of each attribute:  
measures are used for attribute selection, such as:  
 This algorithm selects t he attribute that best divides data into distinct groups. Various  
1. Selecting the Best Attribute for Splitting:  
STEPS IN DECISION TREE INDUCTION :  
in diagnosis.  
to make decision tree. Tree classifies patients into diabetic or non - diabetic, assisting doctors  
based on clinical test results. Features like glucose levels, BMI, and blood pressure are used  
2. Medical Diagnosis: A healthcare provider wants to predict whet her a patient has diabetes  
make quick and reliable decisions.  
and loan history. The decision tree predicts loan approval or rejection, helping the bank  
based on customer profi les. Input features include income, credit score, employment status,  
1. Loan Approval in Banking : A bank needs to decide whether to approve a loan application  
Applications of Decision Trees:  
Declined offer ).  
into two leaf nodes ( Accepted offers and  
one leaf node. Finally, decision node splits  
into one decision node ( Cab facilit y ) and  
 The next decision node further gets split  
corresponding labels.  
and one leaf node based on the  
decision node ( distance from the office )  
 The root node splits further into the next  
t he root node ( Salary attribute by ASM ).  
this problem, the decision tree starts with  
should accept the offer or Not. So, to solve  
offer and wants to decide whether he  
 Suppose there is a candidate who has a job  
For Example:  
  
--- Page 6 ---  
6  
14 Rainy Mild High Strong No  
13 Overcast Hot Normal Weak Yes  
12 Overcast Mild High Strong Yes  
11 Sunny Mild Normal Strong Yes  
10 Rainy Mild Normal Weak Yes  
9 Sunny Cool Normal Weak Yes  
8 Sunny Mild High Weak No  
7 Overcast Cool Normal Strong Yes  
6 Rainy Cool Normal Strong No  
5 Rainy Cool Normal Weak Yes  
4 Rainy Mild High Weak Yes  
3 Overcast Hot High Weak Yes  
2 Sunny Hot High Strong No  
1 Sunny Hot High Weak No  
ID Outloo k Temperature Humidity Wind Play Tennis  
 Features: Outlook, Temperature, Humidity, Wind  
 Target Variable: Play Tennis (Yes/No)  
tennis based on the weather conditions.  
We will use the "Play Tennis" dataset where the goal is to predict whether a person will play  
EXAMPLE DATASET:  
on majority voting or averaging (for regression tasks).  
4. Assigning Class Labels : Once splitting stops, each leaf node is assigned a class label based  
 The tree reaches a predefined depth.  
 No remaining attributes to split.  
 All i nstances in a node belong to the same class.  
 The process is repeated for each child node until a stopping condition is met, such as:  
3. Recursion:  
values. Each subset forms a child node.  
2. Splitting the Data: Once the best attribute is chosen, the dataset is divided based on its  
  
--- Page 7 ---  
7  
 IG(S, Outlook) = 0.25  
 IG(S, Outlook) = 0.94 – 0.69  
14 14 14  
I nformation G ain (S, Outlook) = 0.94 – [ ×0.97 + ×0 + ×0.97 ]  
5 4 5  
Now, calculate Information G ain of Outlook  
 Entropy(Sunny) = 0.97  
 Entropy(Sunny) = − { − 0.97 }  
 Entropy(Sunny) = − { ( 0.6 × − 0.74 ) + (0.4 × − 1.32 ) }  
5 5 5 5  
Entropy( Rainy ) = − ( log 2 + log 2 )  
3 3 2 2  
Entropy for Rainy :  
4 4 4 4  
Entropy( Overcast ) = − ( log 2 + log 2 ) => Entropy( Overcast ) = 0  
4 4 0 0  
Entropy for Overcast : (Since all are yes, so entropy is 0)  
 Entropy(Sunny) = 0.97  
 Entropy(Sunny) = − { − 0.97 }  
 Entropy(Sunny) = − { (0.4 × − 1.32 ) + ( 0.6 × − 0.74 ) }  
5 5 5 5  
Entropy(Sunny) = − ( log 2 + log 2 )  
2 2 3 3  
Entropy for Sunny :  
 Rainy: 5 instances → (3 Yes, 2 No) → Entropy = 0.97  
 Overcast: 4 instances → (4 Yes, 0 No) → Entropy = 0  
 Sunny: 5 instances → (2 Yes, 3 No) → Entropy = 0.97  
1. Information Gain for "Outlook" : We split the dataset based on the Outlook attribute:  
| 퐒 |  
IG(S,A) = Entropy(S) − ∑ ⋅ Entropy(Sv)  
| 퐒퐯 |  
Step 2: Compute Information Gain for Each Attribute  
 Entropy(S) = 0.94  
 E ntropy(S) = − { − 0.94 }  
 Entropy(S) = − { ( 0.642 × − 0.64 ) + ( 0.357×−1.48) }  
14 14 14 14  
Entropy(S) = − ( log 2 + log 2 )  
9 9 5 5  
Total Yes = 9 ; Total No = 5 ; Total Samples = 14  
Step 1: Calculate Entropy of the entire Data set Entropy(S) = − ∑ p i log 2 p i  
  
--- Page 8 ---  
8  
If Normal → Play Tennis = Yes If Weak → Play Tennis = Yes  
If High → Play Tennis = No If Strong → Play Tennis = No  
Play Tennis = Yes Check Humidity: Check Wind:  
I f Outlook is Overcast: If Outlook is Sunny : If Outlook is Rainy :  
the highest IG (0 .25), it becomes the root.  
The attribute with the highest Information Gain is selected as the root. Since "Outlook" has  
Step 3: Construct the Decision Tree  
 IG(S, Wind) = 0.0 5  
 IG(S, Wind) = 0.94 – 0.89  
14 14  
I n formation G ain (S, Wind ) = 0.94 – [ ×0. 81 + × 1.00 ]  
8 6  
 Strong: 6 instances → (3 No, 3 Yes) → Entropy = 1.00  
 Weak: 8 instances → (2 No, 6 Yes) → Entropy = 0.81  
3. Information Gain for "Wind" : We split the dataset based on the Wind attribute:  
 IG(S, Humidity) = 0.01  
 IG(S, Humidity) = 0.94 – 0.93  
14 14  
I nformation G ain (S, Humidity ) = 0.94 – [ ×0.9 8 + ×0 .86 ]  
7 7  
 Normal: 7 instances → (2 No, 5 Yes) → Entropy = 0.86  
 High: 7 instances → (3 No, 4 Yes) → Entropy = 0.98  
2. Information Gain for "Humidity" : We split the dataset based on the Humidity attribute:  
  
--- Page 9 ---  
9  
P(¬S)=0.6 (60% of emails are not spam) P(X ∣ ¬S)=0.2 (20% of non - spam emails contain "free")  
P(S)=0.4 (40% of emails are spam) P(X ∣ S)=0.8 (80% of spam emails contain "free")  
Prior Probabilities: Likelihood (Probability of "free" given class):  
Step 1: Given Data , (X means word “free”)  
P(X)  
P(C|X) =  
P(X|C) \* P(C)  
"free" using Bayes’ Theorem :  
We will classify an email as Spam (S) or Not Spam (¬S) based on the presence of the word  
Numerical Example: Spam Email Classification  
acts as a normalizing constant.  
 P(X): Evidence - The probability of observing features X, regardless of the class. This  
before considering the features X.  
 P(C): Prior probability - The prior probability of the data point belonging to class C,  
belongs to class C.  
 P(X|C): Lik elihood - The probability of observing features X, given that the data point  
given the observed features X. This is what we want to calculate.  
 P(C|X): Posterior probability - The probability of the data point belonging to class C,  
Where:  
P(X)  
Bayes' Theorem: P(C|X) =  
P(X|C) \* P(C)  
assign the data point to the class with the highest posterior probability.  
In classification, we use Bayes' theorem to calculate P(C|X) for each possible class C. We then  
helps us calculate the probability of a data point belonging to a certain class, given its features.  
Bayes' theorem provides a way to update our beliefs about an event given some evidence. It  
Bayes' Theorem :  
distribution P(X, Y) of the features X and labels Y.  
 Bayesian classifiers belong to generative models , which learn the joint probability  
most likely class for a given input by updating pri or beliefs with new evidence.  
theorem, which calculates the probability of a class given observed data. It determines the  
 Bayesian classification is a probabilistic approach to classification based on Bayes'  
BAYESIAN CLASSIFICATION:  
  
--- Page 10 ---  
10  
3. Fraud detection: Identifies fraudulent transactions based on past data.  
2. Medical diagnosis: Predicts diseases based on symptoms.  
1. Spam filtering: Classifies emails as spam or not spam.  
Applic ations:  
3. Poor priors can lead to biased results.  
2. Performance drops if features are dependent.  
1. It Assumes features are independent, which is often unrealistic.  
Disadvantages:  
3. They works well with limited data.  
2. They can handle large num ber of features suitable for text classification .  
1. They are easy to implement, computationally inexpensive & suitable for large datasets.  
Advantages of Bayesian Classification:  
Final Answer: The email is classified as Spam .  
Since , P(S ∣ X) = 0.727 is greater than P(¬S ∣ X)=0.273, the email is more likely to be Spam .  
Step 5: Classification Decision  
0.44  
 ≈ 0.273  
0.12  
0 . 44  
  
(0.2×0.6)  
P(X)  
P(¬S ∣ X) =  
P(X ∣ ¬S) \* P(¬S)  
Step 4: Compute P(¬S ∣ X)  
0 . 44  
 ≈ 0.727  
0 . 32  
0 . 44  
  
( 0 . 8 × 0 . 4 )  
P(X)  
P(S ∣ X) =  
P(X ∣ S) \* P(S)  
Step 3: Compute Posterior P robability P(S ∣ X)  
 0.44  
 0.32 + 0.12  
 (0.8×0.4) + (0.2×0.6)  
P(X) = [ P(X ∣ S) \* P(S) ] + [ P(X ∣ ¬S) \* P(¬S) ]  
Step 2: Compute Evidence P(X) - The total probability of an email containing "free" is:  
  
--- Page 11 ---  
11  
likely classified as spam .  
"free," "win," "prize" appear frequently in spam emails, an email containing them is more  
spam is calc ulated based on word frequencies in spam vs. non - spam emails. If words like  
In spam filtering , words in an email are treated as features. The probability of an email being  
Example: Multinomial Naïve Bayes for Text Classification  
presence determines class probability.  
word doesn’t appear = 0). Example: Document classification , where each word’s  
classification where words are represented as binary indicators (word appears = 1,  
 It is u sed for binary features (presence or absence of a feature). It is s uitable for text  
3. Bernoulli Naïve Bayes (BNB):  
emails.  
treated as features with probabilities based on their occurrence in spam vs. non - spam  
classification (e g, spam detection ). Example: Email classification , where words are  
 It is u sed for discrete features (word counts, frequencies). It is c ommon in text  
2. Multinomia l Naïve Bayes (MNB):  
distribution. Example: Classifying iris flowers based on petal width and length.  
 It is u sed for continuous numerical features. It a ssumes data follows a normal (Gaussian)  
1. Gaussian Naïve Bayes (GNB):  
Types of Naïve Bayes Classifiers :  
4. Works w ell with c ategorical & b inary d ata : - s uitable for problems like medical diagnoses.  
3. Handles High - Dimensional Data Well means Performs well in text classification .  
2. Works Well with Small Data  
1. Fast and Efficient  
Benefits of the Naïve Assumption :  
in spam emails .  
example, in email classification, words like "free" and "offer" frequently appear together  
 This assumption is "naïve" because, in real - world data, features are often correlated. For  
even with simple assumptions.  
 Naïve Bayes is widely used in text classification because it is fast, scalable, and effective  
affect the probability of another feature occurring, given the class.  
independent given the class label. This means that the presence of one feature does not  
 The Naïve Bayes classifier makes a naïve assumption that all features are conditionally  
Naive Bayes Classifier :  
  
--- Page 12 ---  
12  
 Lift: 1.5 (Buying bread increases the likelihood of buying milk by 1.5 times).  
 Confidence: 80% (80% of people who buy bread also buy milk).  
 Support: 30% (30% of transactions include both bread and milk).  
Example of an Association Rule : "Customers who buy bread also t end to buy milk"  
 Lift < 1: Negative correlation (Buying X reduces the chance of buying Y).  
 Lift = 1: No correlation .  
 Lift > 1: X and Y are positively correlated (Buying X increases chance of buying Y).  
Supp ort(Y)  
Lift(X ⇒ Y) =  
Confidence(X ⇒ Y)  
random chance.  
4. Lift : Measures how much more likely Y is bought when X is bought, compared to  
{bread} → {milk} is 0.8 (80%) .  
Example: If 80% of people who buy bread also buy milk, the confidence of the rule  
Support(X)  
Confidence(X ⇒ Y) =  
Support(X ∪ Y)  
purchased.  
3. Confidence : Measures the likelihood that if an item X is purchased, item Y is also  
Example: If 3 out of 10 transactions include {bread, milk}, the support is 0.3 (30%) .  
Total number of transactions  
Support(X) =  
Number of transactions co ntaining X  
2. Support : Measures how frequently an itemset appears in the dataset.  
milk, eggs} is an itemset in a shopping basket.  
1. Itemsets : A collection of items bought together in a transaction. Example: {bread,  
Key Concepts :  
product placement, and recommendation syste ms.  
transaction data, helping businesses make data - driven decisions, such as cross - selling,  
 The goal of association rule mining is to discover interesting and useful rules from  
analysis, where business analyse purchasing patterns to recommend related products.  
correlations between items in large datasets. It is commonly applied in market basket  
 Association Rule Mining is a technique used to identify relationships, patterns, and  
Introduc tion to Association Rule Mining :  
CLASSIFICATION BASED ON CONCEPTS FROM ASSOCIATION RULE MINING :  
  
--- Page 13 ---  
13  
ranking mechanism (like confidence - based selection) is u sed.  
and assigns the corresponding class label. If there are multiple rules appl ied , a  
 When a new instance needs to be classified, the classifier finds the most relevant rule  
4. Classification:  
confidence, support, and rule length .  
based on measures like support, confidence, and lift . The rules are sorted based on  
 Since many rules can be generated, Only the best rules are taken for classification  
3. Rule Pruning and Selection:  
 Example - 2 : IF (Age=Young AND Income=Low), THEN Buy=No.  
 Example - 1 : IF (O utlook=Sunny AND Temperature=Hot), THEN Play=No.  
 From frequent itemsets, association rules are generated in form of IF - THEN rules.  
2. Rule Generation:  
 Example - 2 : {Age=Young, Income=Low} → {Buy=No}.  
 Example - 1 : {Outlook=Sunny, Temperature=Hot} → {Play=No}.  
Apriori algorithm to find frequent itemsets.  
 The dataset is analyzed to find frequent itemsets that appear to gether. It u ses the  
1. Frequent Itemset Mining:  
How CBA Works?  
keep only useful ones.  
multiple database scans. It m ay generate too many rules , req uiri ng extra pruning to  
 Its Li mitations are : It is c omputationally expensive due to Apriori is need ed for  
association rule mining with classification by generating rules from frequent itemsets .  
 CBA is one of th e earliest associative classification algorithms. It integrates  
1. CBA (Classification Based on Associations) :  
Algorithms for Associative Classification: ( CBA and CMA R)  
between attributes.  
It is particularly useful when traditional classifiers struggle with complex relationships  
 The goal is to build accurate and interpretable classifiers by discovering patterns in the data.  
Naïve Bayes, it derives classification rules from frequent patterns in the dataset.  
rule mining with classification. Instead of using traditional classifiers like decision trees or  
 Associa tive Classification (AC) is a classification technique that integrates association  
ASSOCIATIVE CLASSIFICATION:  
  
--- Page 14 ---  
14  
management complex.  
3. Large Number of Rules : It can generate large number of rules, making rule selection and  
rules, whic h adds complexity to the model - building process.  
2. Rule Pruning Complexity : Effective rule pruning is required to remove redundant or weak  
large datasets, as it involves mining frequent itemsets and generating multiple rules.  
1. High Computational Cost : Generating association rules is expensive , especially with  
Disadvantages of Associative Classification :  
it acts as a feature selection method, reducing noise.  
4. Effective Feature Selection : Since it selects strong rules based on support and confidence ,  
attributes, making it flexible across different datasets.  
3. Handles Numeric and Categorical Data : It can work with both categorical and numeric  
classifiers (e.g., decision trees or Naïve Bayes) because it considers multiple rules .  
2. High Accuracy : It achieves better classification accuracy compared to traditional  
1. Interpretability and Transparency : The generated rules are easy to understand .  
Advantages of Associative Classification :  
multiple rules predict different class labels, it chooses one with highest weight .  
 CMAR assigns different weigh ts to rules based on support and confidence . If  
3. Rule Weighting and Conflict Resolution:  
CMAR applies multiple strong rules for classification.  
( CR - tree ) , a tree - structured rule database. Instead of using just one rule like CBA,  
 CMAR stores classification association rules (CARs) in a classification rule tree  
2. Rule Selection and Classification:  
scans , improving speed and memory usage.  
frequent patterns more efficiently. The FP - tree reduces the number of database  
 Instead of Apriori, CMAR uses an FP - tree (Frequent Pattern Tree) to mine  
1. Frequent Patt ern Mining with FP - tree:  
H ow CMAR Works?  
conflicts bette r by considering multiple rules instead of just one.  
 It is m ore efficient than CBA due to FP - tree’s ability to compress data. It h andles rule  
multiple rules to improve classification accuracy.  
 CMAR improves limitations of CBA by making rule mining more efficient and using  
2. CMAR (Classification based on Multiple Association Rules)  
  
--- Page 15 ---  
15  
customer C6 is classified as a High Spender .  
 Final Classification: Since {Laptop, Mouse} strongly indicates High Spender ,  
o {Notebook} → No rule found  
o { Laptop, Mouse} → High Spender  
 Matching Rules:  
A new customer C6 purchases: Lap top, Mouse, Notebook .  
S tep 4: Classifying a New Customer .  
 Rule 3: IF (Laptop AND Headphones) → High Spender  
 Rule 2: IF (Notebook AND Pen) → Low Spender  
 Rule 1: IF (Laptop AND Mouse) → High Spender  
From fre quent itemsets, we derive classification rules :  
Step 3: Generating Classification Rules .  
 {Laptop, Headphones} → High Spender (Support = 40%, Confidence = 100%)  
 {Notebook, Pen} → Low Spender (Support = 40%, Confidence = 100%)  
 { Laptop, Mouse} → High Spender (Support = 40%, Confidence = 100%)  
Using association rule mining , we find frequently occurring patterns:  
Step 2: Finding Frequent Itemsets .  
C5 Notebook, Pen, Mouse Low Spender  
C4 Laptop, Headphones High Spender  
C3 Notebook, Pen Low Spender  
C2 Laptop, Mouse High Spender  
C1 Laptop, Mouse, Headphones High Spender  
Customer Items Purchased Spender Class  
The store collects data on customer purchases:  
S tep 1: Transaction Data .  
on their purchasing behavior.  
 A retail store wants to classify customers as "High Spender" or "Lo w Spender" based  
 Classifying Customers Based on Purchase History .  
Example Associative Classification :  
  
--- Page 16 ---  
16  
in the training data.  
We'll use Euclidean distance to calculate the distance betwee n the new fruit and each fruit  
2. Calculating Distances (k=3):  
We want to classify this fruit as either apple or orange using k - NN.  
 Texture: 1 (Smooth)  
 Weight: 160 grams  
1 . The New Data Point: Suppose we have a new fruit with:  
140 1 Apple  
170 0 Orange  
180 0 Orange  
150 1 Apple  
Weight (grams) Texture (1 = Smooth, 0 = Rough) Fruit Label  
based on weight and texture .  
 Suppose w e have a dataset where we classify whether a fruit is an apple or an orange  
Exampl e of k - N earest N eighbour Classification :  
4. Assign the majority class label among the k neighbo u rs to the new data point.  
3. Select the k nearest neighbo u rs (smallest distance values).  
2. Compute Euclidean distance between the new data point and all training samples.  
1. Cho ose the number of neighbo u rs ( k ).  
Working of k - N earest N eighbour Algorithm :  
k and proper feature scaling for better results.  
recommendation systems, and medical diagnosis. However, it requires careful selection of  
 This is an effective, easy - to - understand classification method used in pattern recognition,  
nearest neighbo u rs in the feature space.  
classification technique. It classifies a new data point based on the majority class of its k  
 The k - Nearest Neighbo u rs algorithm is a simple, non - parametric, and instance - based  
k - NEAREST NEIGHBO U RS ( k - NN) ALGORITHM FOR CLASSIFICA TION :  
OTHER METHODS OF CLASSIFICATION:  
  
--- Page 17 ---  
17  
3. No training phase (lazy learner) 3. Choosing the optimal k value is challenging  
2. Works well for small data sets 2. Sensitive to irrelevant features and noise  
1. Simple to implement 1. Computationally expensive for large datasets  
Advantages: Disadvantages:  
Advantages and Disadvantages of k - N earest N eighbour Algorithm :  
the new fruit is classified as an apple .  
Out of the 3 nearest neighbors, 2 are apples and 1 is an orange. Therefore, by majority vote,  
4. Classification by Majority Vote:  
3. Apple 2 (distance = 20)  
2. Orange 2 (distance ≈ 10.05)  
1. Apple 1 (distance = 10)  
We're using k=3, so we select the three closest neighbors:  
3. Selecting the k - Nearest Neighbors (k=3):  
 √ ( ( 160 − 140 ) ² + ( 1 − 1 ) ² ) = √ ( 400 + 0 ) = 20  
 Distance to Apple 2 (140, 1)  
 √ ( ( 160 − 170 ) ² + ( 1 − 0 ) ² ) = √ ( 100 + 1 ) = √ 101 ≈ 10. 05  
 D istance to Orange 2 (170, 0):  
 √ ( ( 160 − 180 ) ² + ( 1 − 0 ) ² ) = √ ( 400 + 1 ) = √ 401 ≈ 20.02  
 D istance to Orange 1 (180, 0):  
 √ ( ( 160 − 150 ) ² + ( 1 − 1 ) ² ) = √ ( 100 + 0 ) = 10  
 Distance to Apple 1 (150, 1) :  
y2: te xture of new data point  
y1: texture of given data  
x2: weight of new data point  
x1: weights of given data  
Here,  
√ ( ( x2 − x1 ) ² + ( y2 − y1 ) ² )  
The formula for Euclidean distance between two points (x1, y1) and (x2, y2) is:  
  
--- Page 18 ---  
18  
Actual Not Spam 10 30  
Actual Spam 50 10  
Predicted Spam Predicted Not Spam  
Confusion Matrix:  
not spam (TN), and 10 were actually spam but were missed by the model (FN).  
were 40 emails that were not predicted as spam. Out of those, 30 were correctly identified as  
predicts 60 as spam. Out of those 60, 50 were actually spam (TP), and 10 were not (FP). There  
Example: Let's say we're building a spam email classifier. We have 100 emails, and the model  
calculated as: F1 - score = 2 \* (Precision \* Recall) / (Precision + Recall)  
 F1 - score: The harmonic mean of precision and recall, providing a balanced measure. It's  
calculated as: Recall = (True Positives) / (True Positives + False Negatives)  
 Recall: Measures how many of the actual positiv e instances were correctly predicted. It's  
calculated as: Precision = (True Positives) / (True Positives + False Positives)  
 Precision: Measures how many of the positive predictions were actually correct. It's  
Other Important Metrics:  
other metrics to provide a more evaluation.  
class B. This highlights the limitation of accuracy in imbalanced scenarios. Therefore, we need  
have an accuracy of 95%, which seems excellent. However, it completely fails to recognize  
instances of class A and 5 instances of class B. A mo del that always predicts class A would  
Limitations of Accuracy and the Need for Other Metrics: Imagine a dataset with 95  
( Total number of instances )  
Accuracy =  
( Number of correctly classified instances )  
correctly classifies 80 out of 100 instances, its accu racy is 80%. It's calculated as:  
model's predictions. Accuracy is typically expressed as a percentage. For example , if a model  
Classifier Accuracy: It is a single metric that quantifies overall correctness of a classification  
of data . The quality of predictions is assessed using various metrics and simple accu racy.  
 The goal is to create a model that generalizes well and can accurately predict class labels  
predefined categories.  
unseen data. For classificati on, this involves assigning a data point to on e of the  
 P rediction refers to the process of using a trained model to estimate the output for new,  
Prediction:  
machine learning models, particularly in classification.  
Prediction and classifier acc uracy are fundamental concepts in evaluating the performance of  
PREDICTION AND CLASSIFIER ACCURACY:  
  
--- Page 19 ---  
19  
 Out of all actual not spam emails, 75% were correctly identified.  
(for Not Spam)  30 / (30 + 10) = 30/40 = 0.75 (approx. 75%)  
Recall  TN / (TN + FP)  
 Out of all emails predicted as not spam, 75% were actually not spam.  
(for Not Spam)  30 / (30 + 10) = 30/40 = 0.75 (approx. 75% )  
Precision  TN / (TN + FN)  
 Out of all actual spam emails, 83% were correctly identified.  
(for Spam)  50 / (50 + 10) = 50/60 = 0.83 ( approx. 83% )  
Recall  TP / (TP + FN)  
 Out of all emails predicted as spam, 83% were actually spam.  
(for Spam)  50 / (50 + 10) = 50/60 = 0.83 ( approx. 83% )  
Precision  TP / (TP + FP)  
 (50 + 30) / 100 = 80/100 = 0.8 (approx. 80%)  
Accuracy  (TP + TN) / Total  
Calculations:  
False Negatives (FN ): 10 (Incorrectly predicted not spam - actually spam)  
Spam: 40  
Predicted Not True Negatives (TN): 30 (Correctly predicted not spam)  
False Positives (FP): 10 (Incorrectly predicted spam - actually not spam)  
Spam: 60  
Predicted True Positives (TP): 50 (Correctly predicted spam)  
Total Ema ils: 100  
Given Information: