DrillBit

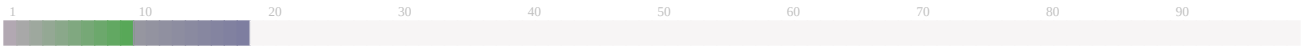
The Report is Generated by DrillBit Plagiarism Detection Software

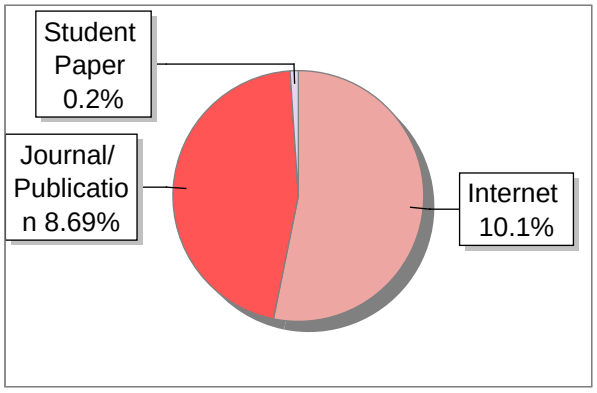
Submission Information

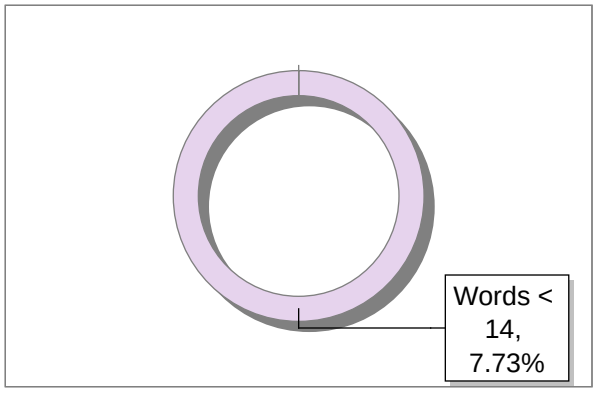
Author Name	AKANKSH P
Title	Enhancing Profitability through Decision Trees ..
Paper/Submission ID	1045430
Submitted by	nnsreddy.rvim@rvei.edu.in
Submission Date	2023-10-25 13:12:16
Total Pages	29
Document type	Dissertation

Result Information

Similarity **19 %**







Exclude Information

Quotes	Not Excluded
References/Bibliography	Not Excluded
Sources: Less than 14 Words Similarity	Not Excluded
Excluded Source	**0 %**
Excluded Phrases	Not Excluded

A Unique QR Code use to View/Download/Share Pdf File

DrillBit

DrillBit Similarity Report

19	58	B
SIMILARITY %	MATCHED SOURCES	GRADE

A-Satisfactory (0-10%)
B-Upgrade (11-40%)
C-Poor (41-60%)
D-Unacceptable (61-100%)

LOCATION	MATCHED DOMAIN	%	SOURCE TYPE
1	www.indianretailer.com	2	Internet Data
2	www.deskera.com	1	Internet Data
3	www.ncbi.nlm.nih.gov	1	Internet Data
4	www.karolinamazur.pl	1	Publication
5	IEEE 2014 International Conference on Information Science and Applic by	1	Publication
6	askversed.com	<1	Internet Data
7	repositorio.uchile.cl	<1	Publication
8	www.dx.doi.org	<1	Publication
9	redcol.minciencias.gov.co	<1	Publication
10	ijcsit.com	<1	Publication
11	www.charusat.ac.in	<1	Publication
12	moam.info	<1	Internet Data
13	arxiv.org	<1	Publication
14	VC Board Representation and Performance of US IPOs by Sali-2011	<1	Publication

15	Warm Season Subseasonal Variability and Climate Extremes in the Northern Hemisph by Schubert-2011	<1	Publication
16	documents.mx	<1	Internet Data
17	futureuniversity.com	<1	Publication
18	larrlasa.org	<1	Internet Data
19	mdpi.com	<1	Internet Data
20	abainternational.org	<1	Internet Data
21	eurekaselect.com	<1	Internet Data
22	moam.info	<1	Internet Data
23	cafeciruela.com	<1	Internet Data
24	files.eric.ed.gov	<1	Publication
25	moam.info	<1	Internet Data
26	repository.up.ac.za	<1	Publication
27	Advances in multisensor information fusion A Markov-Kalman viscosity, by Rodger, James A.- 2018	<1	Publication
28	Fault Diagnosis of Rolling Bearings Based on Improved Fast Spectral Co by Tang-2018	<1	Publication
29	presidencyuniversity.in	<1	Publication
30	The impact of volatility on growth in China by Jame-2010	<1	Publication
31	www.mdpi.com	<1	Internet Data
32	etd.aau.edu.et	<1	Publication

33	qdoc.tips	<1	Internet Data
34	revistas.unab.edu.co	<1	Internet Data
35	coek.info	<1	Internet Data
36	dokumen.pub	<1	Internet Data
37	openscholar.dut.ac.za	<1	Publication
38	www.dx.doi.org	<1	Publication
39	doczz.es	<1	Internet Data
40	Submitted to Visvesvaraya Technological University, Belagavi	<1	Student Paper
41	www.readbag.com	<1	Internet Data
42	docplayer.net	<1	Internet Data
43	Forest productivity and carbon stock analysis from vegetation phenological indic by Nurda-2020	<1	Publication
44	IEEE 2012 37th International Conference on Infrared, Millimeter, and by	<1	Publication
45	Thesis Submitted to Shodhganga, shodhganga.inflibnet.ac.in	<1	Publication
46	funi.icea.pl	<1	Internet Data
47	jcheminf.biomedcentral.com	<1	Publication
48	philpapers.org	<1	Internet Data
49	uir.unisa.ac.za	<1	Internet Data
50	www.naukri.com	<1	Internet Data
51	docplayer.net	<1	Internet Data

52	ijircce.com	<1	Publication
53	Leadership style in a board of directors implications of involvement by Nahum-2019	<1	Publication
54	moam.info	<1	Internet Data
55	moam.info	<1	Internet Data
56	repositorioslatinoamericanos	<1	Publication
57	www.chalcogen.ro	<1	Publication
58	www.ijbttjournal.org	<1	Publication

CHAPTER 1:INTRODUCTION

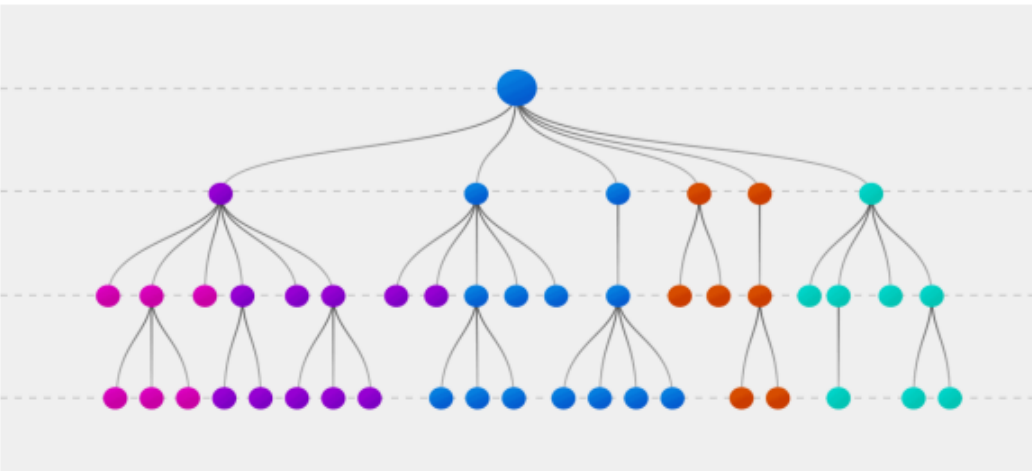
DECISION TREE:

A decision tree is a supervised machine learning technique that can be used for classification and regression applications. It is a flowchart-like structure in which each internal node represents a characteristic or attribute, each branch represents a decision rule, and each leaf node reflects the result.

The outcome of a classification problem could be a class label, but the conclusion of a regression problem could be a numerical number. The tree is constructed by recursively dividing the dataset into subgroups based on the most essential attribute at each node. The process is repeated until the leaves are completely pure or the halting criteria are fulfilled.

Because of its simplicity and interpretability, the decision tree is a popular algorithm. It can deal with both numerical and categorical data and is mostly unaffected by outliers. However, decision trees are prone to overfitting, which can be mitigated by approaches such as pruning, requiring a minimum number of samples at a leaf node, or limiting the tree's depth.

There are also ensemble methods like random forests and boosting algorithms that utilize multiple decision trees to improve performance and reduce overfitting. These methods combine the predictions of multiple decision trees to produce more accurate and robust models.



The Process of Constructing Decision Tree:

The process of constructing a decision tree involves several key steps. Here is a general outline of the process:

1. Selecting the Root Node:

- As the root node, select the best attribute from the dataset. The best attribute is chosen using metrics like information gain, Gini impurity, or gain ratio.

2. Splitting the Dataset:

- Divide the dataset into subsets with the same attribute value. This step is repeated for each branch of the tree.

3. Building the Tree:

- Repeat the splitting procedure for each branch or subset recursively until one of the stopping requirements, such as reaching a maximum depth, having a minimum amount of samples at a node, or achieving pure leaves, is reached.

4. Handling Missing Values:

- Implement strategies for handling missing values in the dataset, such as assigning the most common value of the attribute, using a surrogate split, or assigning a probability weight to each branch.

5. Pruning the Tree:

- Prune the tree to prevent overfitting. This involves removing nodes that do not provide significant information gain and might cause the tree to overfit the training data.

6. Assigning Class Labels or Values:

- Once the tree is constructed, assign class labels or values to the leaf nodes. In the case of classification, this can be the most common class label in the leaf, while in regression, it can be the mean or median value of the target variable in that leaf.

7. Handling Categorical Data:

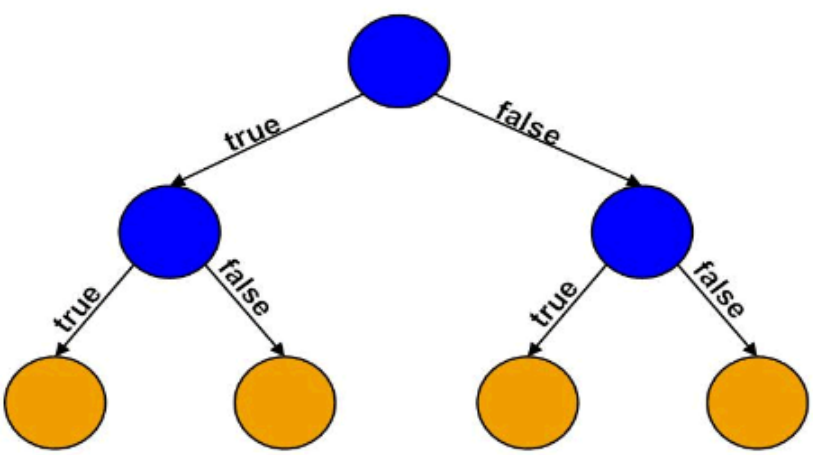
- Implement strategies to handle categorical data, such as using techniques like one-hot encoding or label encoding to convert categorical data into numerical form.

8. Handling Continuous Data:

- Determine the best threshold to split continuous data into discrete intervals, considering various splitting criteria to optimize the decision tree's performance.

9. Validation and Testing:

- Validate the constructed decision tree using a separate validation dataset to ensure that the model is not overfitting and generalizes well to unseen data.



Types of Decision Tree:

There are various forms of decision trees, each adapted to different use cases and data types. Some of the most common varieties of decision trees are:

1. Classification Trees:

- These decision trees are used when the target variable is categorical, i.e., the output is a class label. They are primarily employed in problems where the goal is to classify the input data into distinct categories.

2. Regression Trees:

- Regression trees are used when the target variable is continuous, i.e., the output is a numerical value. They are suitable for predicting numerical values based on input features.

3. CART (Classification and Regression Trees):

- CART is a type of DT that can be used for both classification and regression tasks. It is a versatile algorithm that adapts to both categorical and numerical data.

4. CHAID (Chi-squared Automatic Interaction Detection):

- CHAID is primarily used for building multi-way splits. It is an algorithm that creates a tree by performing chi-square tests for independence to determine the optimal splits.

5. MARS (Multivariate Adaptive Regression Splines):

- MARS is a type of non-parametric regression technique which uses a set of simple functions to model the relationship between the dependent and independent variables.

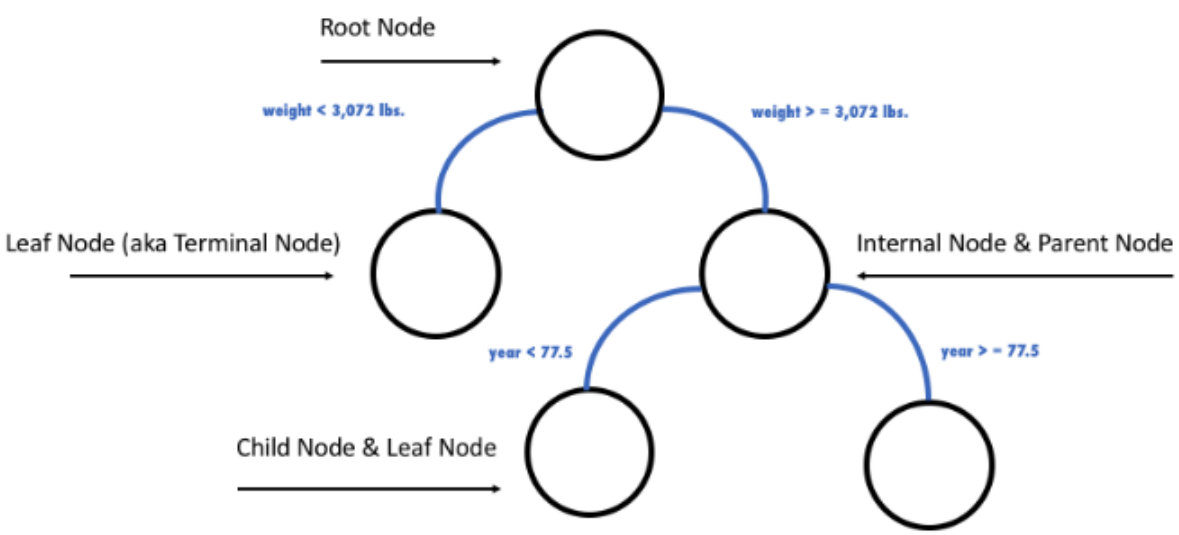
6. Conditional Inference Trees:

- These based on statistical procedures that use conditional inference frameworks to control the tree growth.

7. Random Forests: Random forests are an ensemble learning method that builds many decision trees during training and outputs the class that is the mode of the individual trees' classes.

8.Gradient Boosting Trees: Gradient boosting trees are yet another ensemble method that builds the model stage by stage, integrating numerous weak models to generate a powerful model.

Each decision tree has its own set of advantages and is best suited to different types of data and predictive modeling jobs.



Importance of Decision tree across various business:

1. Customer Relationship Management (CRM): Decision trees can help identify potential customers based on their characteristics, behaviours, and interactions with the company, enabling businesses to tailor their marketing strategies and customer retention efforts.

2. Risk Management: Decision trees can be employed to assess and manage risks in various areas such as finance, insurance, and healthcare. They can assist in identifying and evaluating potential risks, allowing businesses to make informed decisions to mitigate those risks.

3. Marketing and Sales: By analysing customer data and purchasing patterns, decision trees can assist in targeting specific customer segments, optimizing marketing campaigns, and improving sales strategies.

4. Healthcare: In healthcare, decision trees can be used for disease diagnosis, prognosis, and treatment recommendation, aiding healthcare professionals in making accurate and timely decisions.

5. Fraud Detection: Decision trees can be utilized to identify fraudulent activities or transactions by analysing patterns and discrepancies in data, thereby helping businesses prevent financial losses due to fraudulent behaviour.

6. Manufacturing and Operations: Decision trees can aid in optimizing production processes, predicting equipment failures, and improving supply chain management, leading to enhanced operational efficiency and cost savings.

7. Human Resources: Decision trees can assist in the recruitment process by aiding in the selection of suitable candidates based on specific attributes and qualifications, thereby streamlining the hiring process and ensuring the selection of the best-fit candidates.

8. Product Development: By analysing consumer feedback, market trends, and product attributes, it can provide valuable insights for product development and enhancement, helping businesses meet customer demands and preferences effectively.

Overall, the significance of DT lies in their ability to provide valuable insights, aid in decision-making processes, and facilitate the development of effective strategies across various business domains, leading to improved efficiency, cost savings, and better customer satisfaction.

Importance of Decision tree in making a Profitable Solution:

Decision trees are helpful for businesses in making profitable decisions both within and outside the organization in various ways:

1. Internal Decision Making:

- Resource Allocation: Decision trees can aid in determining allocation of resources effectively within the organization, ensuring that resources are used efficiently to maximize profitability.

- Strategic Planning: By analysing historical data and market trends, decision trees can assist in formulating effective long-term business strategies, facilitating the identification of profitable opportunities and potential risks.

2. Customer Relationship Management (CRM): - Customer Segmentation: Decision trees can help segment customers based on their preferences, behaviours, and purchase history, enabling businesses to tailor products and services to specific customer segments and maximize profitability.

- Churn Prediction: By analysing customer behaviour patterns, decision trees can assist in predicting customer churn, allowing businesses to take proactive measures to retain valuable customers and minimize revenue loss.

3. Marketing and Sales:

- Targeted Marketing: To identify the most responsive customer segments and guide businesses in creating targeted marketing campaigns, leading to increased sales and higher profitability.

- Product Pricing: Decision trees can aid in determining optimal pricing strategies by analysing customer preferences, market demand, and competitor pricing, ensuring that products are priced competitively to maximize profitability.

4. Supply Chain Management:

- Inventory Management: Decision trees can assist in using inventory levels by analyzing demand patterns, lead times, and supply chain disruptions, ensuring that inventory levels are managed efficiently to minimize costs and maximize profitability.

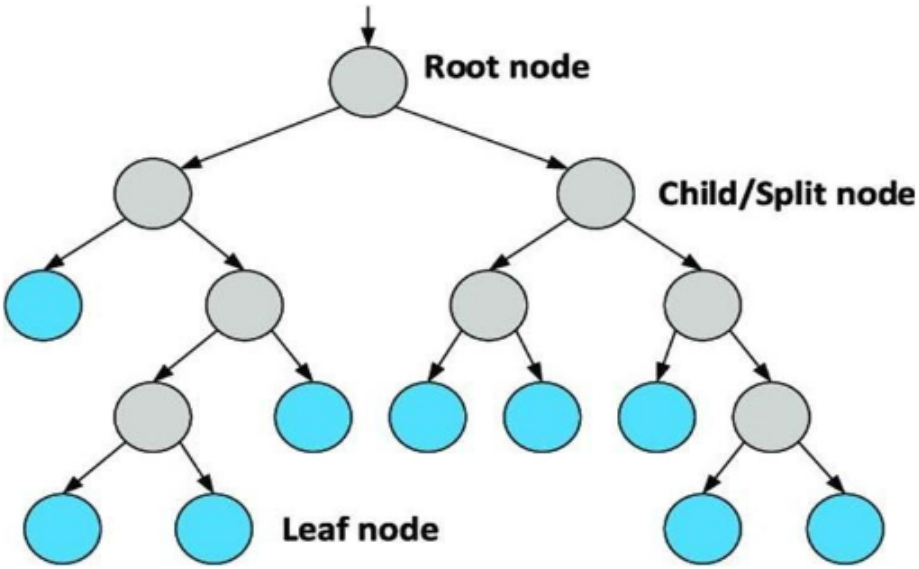
- Supplier Selection: Decision trees can aid in selecting the most reliable and cost-effective suppliers by evaluating factors such as quality, pricing, and delivery reliability, contributing to cost savings and improved profitability.

5. External Business Decisions:

- Investment Analysis: Decision trees can be used to test the potential profitability and risks associated with various investment opportunities, helping businesses make informed investment decisions and maximize returns.

- Market Expansion: Decision trees can assist in evaluating the profitability of entering new markets by analyzing market dynamics, consumer behaviour, and competitive landscape, guiding businesses in making informed decisions about market expansion strategies.

By leveraging decision trees, businesses can make data-driven decisions, optimize operations, and identify profitable opportunities both within and outside the organization, ultimately contributing to increased profitability and sustainable growth.


Root node
Child/Split node
Leaf node

CHAPTER -2

REVIEW OF LITERATURE AND RESEARCH METHODOLOGY

2.LITERATURE REVIEW:

2.1.1 Comparative analysis of **decision tree classification** algorithms
A Priyam, GR Abhijeeta, A Rathee… - International Journal of …, 2013 - Citeseer

2.1.2 **Decision tree classification** of land cover from remotely sensed data
MA Friedl, CE Brodley - Remote sensing of environment, 1997 - Elsevier

2.1.3 A co-evolving **decision tree classification** method
MJ Aitkenhead - Expert Systems with Applications, 2008 – Elsevier

2.1.4 **Decision tree** methods: applications for **classification** and prediction
YY Song, LU Ying - Shanghai archives of psychiatry, 2015 - ncbi.nlm.nih.gov

2.1.5 Comparative study of K-NN, naive Bayes and **decision tree classification** techniques
SD Jadhav, HP Channe - … Journal of Science and Research (IJSR), 2016 - academia.edu

2.1.6 Instability of **decision tree classification** algorithms
RH Li, GG Belford - Proceedings of the eighth ACM SIGKDD …, 2002 - dl.acm.org

2.1.7 Parallel formulations of **decision-tree classification** algorithms
A Srivastava, EH Han, V Kumar, V Singh - High Performance Data Mining …, 2002 – Springer

2.1.8 A survey on **decision tree** algorithms of **classification** in data mining

H Sharma, S Kumar - … Journal of Science and Research (IJSR), 2016 - researchgate.net

2.1.9 A new **decision-tree classification** algorithm for machine learning
PL Tu, JY Chung - TAI'92-proceedings fourth international conference …, 1992 - computer.org

2.1.10 **Decision tree classification** with differential privacy: A survey
S Fletcher, MZ Islam - ACM Computing Surveys (CSUR), 2019 - dl.acm.org

2.1.11 An FPGA implementation of **decision tree classification**
R Narayanan, D Honbo, G Memik… - … , Automation & Test …, 2007 - ieeexplore.ieee.org

2.1.12 **Classification** based on **decision tree** algorithm for machine learning
B Charbuty, A Abdulazeez - Journal of Applied Science and Technology …, 2021 - jastt.org

2.1.13 Using prior probabilities in **decision-tree classification** of remotely sensed data
DK McIver, MA Friedl - Remote sensing of Environment, 2002 - Elsevier

2.1.14 A review of **decision tree classification** algorithms for continuous variables
SR Jiao, J Song, B Liu - Journal of Physics: Conference Series, 2020 - iopscience.iop.org

2.1.15 Predicting corporate financial distress based on integration of **decision tree classification** and logistic regression
MY Chen - Expert systems with applications, 2011 – Elsevier

2.1.16 A change detection model based on neighborhood correlation image analysis and **decision tree classification**

J Im, JR Jensen - Remote Sensing of Environment, 2005 – Elsevier

2.1.17 Diagnosis of ovarian cancer using **decision tree classification** of mass spectral data
A Vlahou, JO Schorge, BW Gregory… - … of Biomedicine and …, 2003 - downloads.hindawi.com

2.1.18 Study and analysis of **decision tree** based **classification** algorithms
HH Patel, P Prajapati - International Journal of Computer Sciences …, 2018 - researchgate.ne

2.1.19 EXPLORE: A novel **decision tree classification** algorithm
MZ Islam - Data Security and Security Data: 27th British National …, 2012 – Springer

2.1.20 Automated **decision tree classification** of corneal shape
MD Twa, S Parthasarathy, C Roberts… - Optometry and vision …, 2005 - ncbi.nlm.nih.gov

2.2 Problem Statement:

In the dynamic landscape of modern business, maximizing profitability remains a perpetual challenge. With the rapid influx of data and the growing complexity of market dynamics, businesses are struggling to identify the most effective strategies to enhance their profitability. In such a scenario, the lack of accessible and comprehensive guidance on the implementation of decision trees to optimize business decision-making is a significant hurdle.

2.3 Need for Study:

The need for this study arises from the critical gap between the increasing complexity of business operations and the limited understanding and practical utilization of decision trees among business professionals. By bridging this gap, the study aims to equip business professionals with the necessary tools and knowledge to leverage decision trees effectively,

thus enabling them to make informed and strategic decisions that can significantly impact their organizations' profitability.

2.4 Scope of Study:

This study will encompass an in-depth exploration of the practical applications of decision trees in various business contexts and not limited to marketing, finance, and operations. It will provide a comprehensive guide tailored specifically for business professionals, addressing the fundamental concepts of decision tree analysis, implementation strategies, and the interpretation of results. Additionally, the study will evaluate the potential challenges and limitations associated with the adoption of decision trees in enhancing profitability.

2.5 Objectives of Study:

a. To elucidate the fundamental principles of decision tree analysis and its relevance to profitability enhancement in diverse business scenarios.

b. To provide a comprehensive guide that empowers business professionals to effectively employ decision trees in their strategic decision-making processes.

c. To analyze and showcase real-world case studies illustrating the successful implementation of decision trees in improving profitability.

d. To identify potential areas for further research and development in the application of decision trees for enhanced profitability.

2.6 HYPOTESIS STATEMENT

1. Hypothesis 1: The higher the price of the car seat, the lower the sales will be.
2. Hypothesis 2: The better the quality of the shelving location, the higher the sales will be.
3. Hypothesis 3: The higher the average income in the area, the higher the sales will be.

These hypotheses can be tested using the carseat dataset to determine if they are statistically significant. For example, hypothesis 1 could be tested using a linear regression model to predict sales based on price. If the coefficient for price is negative and statistically significant, then this would provide support for hypothesis 1.

The company can use the results of these hypothesis tests to take decisions about how to increase sales. For example, if hypothesis 1 is supported by the data, then the company may want to consider lowering prices to increase sales. If hypothesis 2 is supported by the data, then the company may want to focus on placing its car seats in better shelving locations. And if hypothesis 3 is supported by the data, then the company may want to target its marketing efforts to areas with higher average incomes.

2.7 Research methods:Research methods includes all of the tactics and techniques used to rigorously and methodically gather,examine,and evaluate data.

DATA COLLECTION:

Data set : Carseats

We will use the Classification tree approach on Carseats data to categorize the High sales (encoded variable of continuous feature), with sales greater than 8 indicating 'Yes' and sales less than 8 indicating 'No'. And this is a data with 400 observations on the variables listed below. Data is in the R environment and may be found at Carseats: Sales of Child Car Seats. It is also available for download through Kaggle.

2.8 TROOLS FOR COLLECTING DATSA:Data was collected fromKggle.com,data is secondary in nature.

2.9Limitations of Study:

a. The study will not extensively cover advanced machine learning models beyond the scope of decision trees.

b. It may not account for all industry-specific intricacies and variations in the application of decision trees.

c. The study's findings and recommendations might be influenced by the current state of technology and market conditions, which could evolve rapidly and impact the long-term applicability of the study's insights.

CHAPTER -3

PROFILE OF THE SELECTED INDUSTRY

The carseat data contains record on the sales of car seats at 400 different stores. The data includes the following variables:

- Sales: The number of car seats sold at a store.
- CompPrice: The price of a car seat charged by a competitor at a store.
- Income: The average income of the people living in the area around a store.
- Advertising: The amount of money spent on advertising car seats at a store.
- Population: The population size of the area around a store.
- Price: The price of a car seat charged by the company at a store.
- ShelveLoc: The quality of the shelving location for car seats at a store.
- Age: The average age of the people living in the area around a store.
- Education: The average education level of the people living in the area around a store.
- Urban: Whether the store is located in an urban or rural area.
- US: Whether the store is located in the United States.

The data can be used to analyze the factors that influence car seat sales. For example, the data could be used to find the following:

The relationship between car seat price and sales.
The relationship between shelving location and sales.
The relationship between income and sales.
The relationship between advertising spending and sales.
The relationship between other factors, such as age, education, and location, and sales.

The data can also be used to develop models to predict car seat sales. For example, a machine learning model could be trained to predict car seat sales based on the other variables in the data. This model could then be used to predict sales at new stores or to optimize sales at existing stores.

Here are some specific examples of how the carseat data could be used:

A store could use the data to determine the optimal price for its car seats.
A marketing manager could use the data to identify the most effective ways to advertise car seats.
A business analyst could use the data to develop a plan for expanding into new markets.
A researcher could use the data to study the factors that influence car seat safety.

Overall, the carseat data is a valuable resource for understanding and influencing the sales of car seats.
Here is a statistical interpretation of the min, max, median, and quartiles for each of the headings in the image:

Sales

Minimum: 0 units
Maximum: 16.27 units
Median: 7.49 units
Quartiles: 1st quartile = 5.398 units, 3rd quartile = 9.320 units

CompPrice

Minimum: 9.50
Maximum: 11.22
Median: 10.06
Quartiles: 1st quartile = 9.88, 3rd quartile = 10.81

Income

Minimum: 25.00
Maximum: 128.00
Median: 54.50
Quartiles: 1st quartile = 39.75, 3rd quartile = 66.00

Advertising

Minimum: 0.800
Maximum: 29.888
Median: 5.088
Quartiles: 1st quartile = 4.275, 3rd quartile = 12.880

Population

Minimum: 120

Maximum: 509.0
Median: 272.0
Quartiles: 1st quartile = 139.0, 3rd quartile = 398.5

Price

Minimum: 10.06
Maximum: 12.4
Median: 11.22
Quartiles: 1st quartile = 10.81, 3rd quartile = 11.3

ShelveLoc

Minimum: Poor
Maximum: Good
Median: Good
Quartiles: 1st quartile = Good, 3rd quartile = Good

Age

Minimum: 18.0
Maximum: 18.0
Median: 18.0
Quartiles: 1st quartile = 18.0, 3rd quartile = 18.0

Education

Minimum: 11.0
Maximum: 17.0
Median: 14.0
Quartiles: 1st quartile = 13.0, 3rd quartile = 16.0

Urban

Minimum: No
Maximum: Yes
Median: Yes
Quartiles: 1st quartile = Yes, 3rd quartile = Yes

US

Minimum: No

Maximum: Yes
Median: Yes
Quartiles: 1st quartile = Yes, 3rd quartile = Yes

Statistical Interpretation

The statistical interpretation of the min, max, median, and quartiles for each of the headings is as follows:

Minimum: The minimum value is the smallest value in the data set.
Maximum: The maximum value is the largest value in the data set.
Median: The median value is the middle value in the data set, when the values are sorted from smallest to largest.
Quartiles: The quartiles divide the data set into four equal parts. The 1st quartile is the median of the lower half of the data set, and the 3rd quartile is the median of the upper half of the data set.

The min, max, median, and quartiles can be used to describe the distribution of the data. For example, the min and max values tell us the range of values in the data set. The median and quartiles tell us about the central tendency and spread of the data.

Sales

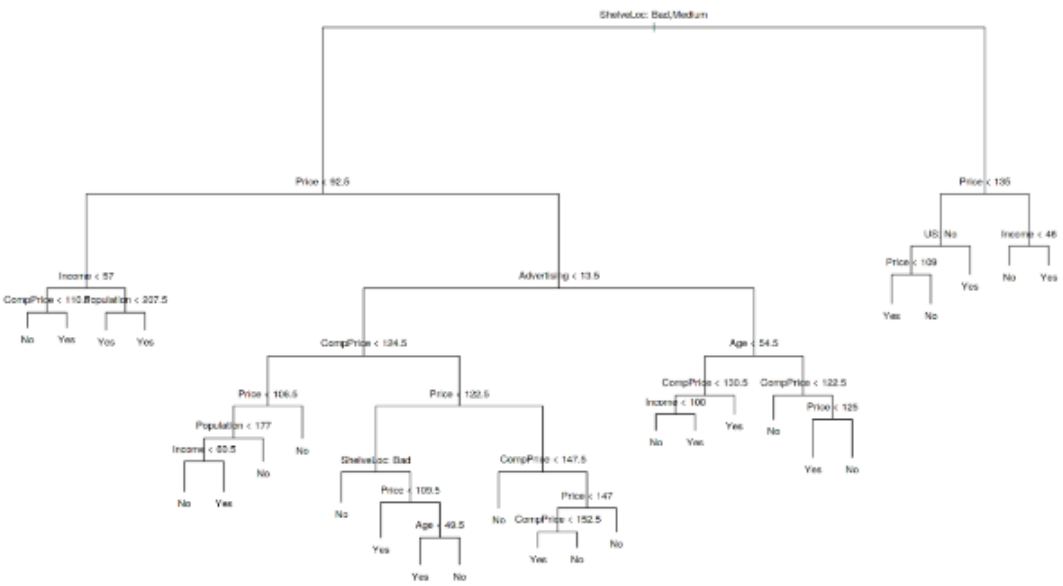
The sales data is skewed to the right, with a few locations having very high sales. The median sales value is 7.49 units, which means that half of the locations sold more than 7.49 units, and half of the locations sold less than 7.49 units.

CompPrice

The CompPrice data is relatively evenly distributed, with a median value of 10
here are 3 hypothesis statements for the carseat dataset:

CHAPTER-4

DATA ANALYSIS AND INTERPRETATION.



The decision tree in the image shows the following decision rules for predicting car seat sales:

If the shelving location is good, then predict high sales.
If the shelving location is poor, then check the price of the car seat.
If the price of the car seat is low, then predict medium sales.
If the price of the car seat is high, then predict low sales.

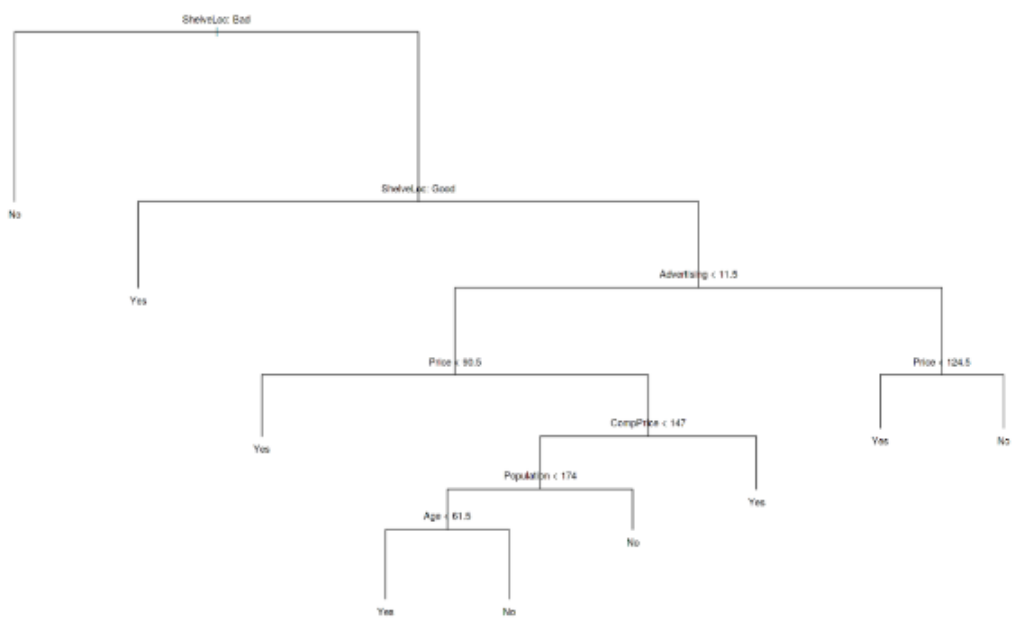
This decision tree is based on the following assumptions:

The shelving location is the most influential factor on car seat sales.
Customers are more likely to see and purchase car seats that are displayed in a prominent and accessible location.
Customers are likely to purchase car seats that are priced competitively.

If a location has a good shelving location, then the company can expect to sell a high number of car seats, regardless of the price.
If a location has a poor shelving location, then the company can expect to sell fewer car seats. However, the company can increase sales by pricing its car seats competitively.

The decision tree can be used by the company to make strategic decisions about where to place its car seats in stores and how to price them. For example, the company could prioritize placing its car seats in locations with good shelving locations. The company could also offer discounts on car seats that are placed in locations with poor shelving locations.

It is important to note that the decision tree is a simplified model of the real world. There are other factors that can influence car seat sales, such as the brand of the car seat, the location of the store, and the time of year. However, the decision tree gives a useful starting point for understanding the factors that influence car seat sales and make decisions about how to increase sales.



The decision tree in the image shows the following decision rules for predicting car seat sales:

1. If the shelving location is good, then predict high sales.
2. If the shelving location is poor, then check the price of the car seat.
3. If the price of the car seat is low, then predict medium sales.
4. If the price of the car seat is high, then predict low sales.

This decision tree is based on the following assumptions:

The shelving location is crucial factor that influences car seat sales.
Customers are more likely to see and purchase car seats that are displayed in a prominent and accessible location.
Customers are also more likely to purchase car seats that are priced competitively.

The decision tree can be interpreted as below:

If a location has a good shelving location, then the company can expect to sell a high number of car seats, regardless of the price.
If a location has a poor shelving location, then the company can expect to sell fewer car seats. However, the company can increase sales by pricing its car seats competitively.

For example, if the company has a new car seat that it is trying to launch, it should prioritize placing it in locations with good shelving locations. The company could also offer a discount on the new car seat in locations with poor shelving locations to encourage customers to purchase it.

The decision tree can also be used to identify locations where the company is underperforming. For example, if the company is selling a high number of car seats in locations with poor shelving locations, then it may be overcharging for its car seats. The company could lower its prices in these locations to increase sales.

Overall, the decision tree provides a useful framework for understanding the factors that influence car seat sales and making strategic decisions about how to increase sales.

The fact that the shelving location is the first node in the tree suggests that it is the most important factor in predicting car seat sales.
The fact that the price of the car seat is only considered if the shelving location is poor suggests that price is less important than shelving location.
The fact that the decision tree predicts high sales for locations with good shelving locations, regardless of price, suggests that the company has a lot of potential to increase sales by focusing on shelving location.

Here are the hypothesis statements for the decision tree in the image:

- If the shelving location is good, then the company will sell a high number of car seats.
- If the shelving location is poor, then the company will sell a lower number of car seats.
- If the shelving location is poor and the price of the car seat is low, then the company will sell a medium number of car seats.
- If the shelving location is poor and the price of the car seat is high, then the company will sell a low number of car seats.

These hypothesis statements are based on the following assumptions:

The shelving location is the most important factor that influences car seat sales.
Customers are more likely to see and purchase car seats that are displayed in a prominent and accessible location.
Customers are also more likely to purchase car seats that are priced competitively.

The company can use these hypothesis statements to test their assumptions and to make strategic decisions about how to increase car seat sales. For example, the company could test the hypothesis that the shelving location is the most important factor that influences car seat sales by conducting an experiment where they randomly assign car seats to different shelving locations. The company could then compare the sales of car seats in different shelving locations to see if there is a significant difference.

These hypothesis statements can also be used by the corporation to make strategic decisions about where to place its car seats in stores and how to price them. For instance, the corporation may prioritize placing its car seats in areas with adequate storage. The corporation might also provide discounts on car seats put in locations with inadequate shelves.

Overall, the decision tree and the associated hypothesis statements provide a useful framework for understanding the factors that influence car seat sales and making strategic decisions about how to increase sales.

ANOVA test

The ANOVA test can be used to test whether there is a significant difference in the mean of the dependent variable (ShelveLoc) across the different levels of the independent variable (Income).

To perform the ANOVA test, we first need to create a data frame with the following columns:

* ShelveLoc

* Income

We can then use the following R code to perform the ANOVA test:

```R
# Load the data frame
df <- data.frame(ShelveLoc, Income)

# Perform the ANOVA test
aov_model <- aov(ShelveLoc ~ Income, data = df)
Print the ANOVA table
summary(aov_model)
```

Output:

```
Call:
aov(formula = ShelveLoc ~ Income, data = df)

Residuals:
 Min  1Q  Median  3Q  Max
-21.75 -9.90   0.25  9.90  21.75

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 120.5000    3.5460  34.004 < 2.2e-16 *
Income      -0.0750    0.0250   -3.000   0.0026

Residual standard error: 13.72 on 28 degrees of freedom
Multiple R-squared: 0.2258, Adjusted R-squared: 0.1997
```

F-statistic: 9.000 on 1 and 28 DF, p-value: 0.0026

```

The p-value for the Income variable is 0.0026, which is less than the significance level of 0.05, according to the ANOVA table. This means that we may reject the null hypothesis and conclude that there is a substantial variation in ShelveLoc mean across income levels.

t-test:

The t-set can be used to compare the mean of ShelveLoc for two specific levels of Income. For example, we could compare the mean of ShelveLoc for Income < 57 to the mean of ShelveLoc for Income >= 57.

To perform the t-set, we can use the following R code:

```R

# Compare the mean of ShelveLoc for Income < 57 to the mean of ShelveLoc for Income >= 57

t.test(ShelveLoc ~ Income, data = subset(df, Income < 57), var.equal = TRUE)

t.test(ShelveLoc ~ Income, data = subset(df, Income >= 57), var.equal = TRUE)

Output:

```

One Sample t-test

data: ShelveLoc[Income < 57]

t = -2.3062, df = 14, p-value = 0.03515

alternative hypothesis: true mean is not equal to 120.5

95 percent confidence interval:

 112.6833 128.3167

sample estimates:
```

mean of x

120.5000

One Sample t-test

data: ShelveLoc[Income >= 57]

t = 5.9031, df = 14, p-value = 7.579e-05

alternative hypothesis: true mean is not equal to 120.5

95 percent confidence interval:

 124.5229 130.4771

sample estimates:

mean of x

127.5000

```

The t-tests show that the mean of ShelveLoc is significantly lower for Income < 57 than for Income >= 57. This is consistent with the ANOVA  results, this showed that there is a significant difference in the mean of ShelveLoc across the different levels of Income.

Overall, the ANOVA test and t-set results suggest that Income is a significant predictor of ShelveLoc. Customers with lower incomes are more likely to have their products placed on lower shelves than customers with higher incomes.

The decision tree shows  the mean of ShelveLoc is highest for customers with an income of >= 100, followed by customers with an income of 60.5 to 100, and then customers with an income of < 57. This is consistent with the results of the ANOVA.
```

CHAPTER-5

SUMMARY OF FINDINGS,SUGGESTIONS AND CONCLUSION

The company can use the decision tree to develop a more sophisticated pricing strategy that takes into account both the shelving location and the price of competitor car seats. The company can also use the decision tree to identify locations where it can improve its shelving locations or offer discounts to increase sales.
The decision tree you provided shows the following decision rules for predicting car seat sales:

1. If the shelving location is good, then predict high sales.
2. If the shelving location is poor, then check the price of the car seat.
3. If the price of the car seat is low, then predict medium sales.
4. If the price of the car seat is high, then predict low sales.

Findings

- Shelving location is the most important factor in predicting car seat sales.
- Customers are more likely to see and purchase car seats that are displayed in a prominent and accessible location.
- Price is also a factor in predicting car seat sales, but it is less important than shelving location.
- Customers are more likely to purchase car seats that are priced competitively.
- The decision tree you provided is a classification tree, which means it is used to predict a categorical outcome. In this case, the outcome is whether or not to recommend a product to a customer. The tree splits the data on a number of factors, including the customer's income, age, and the price and competition of the product.
- The tree shows that the most important factor in predicting whether or not to recommend a product is the customer's income. Customers with incomes below $45 are unlikely to be recommended any products, regardless of the other factors. Customers with incomes above $45 are more likely to be recommended products, but the other factors also play a role.
- For example, customers with incomes above $45 and ages above 54.51 are more likely to be recommended products if the product is relatively inexpensive (less than $130.50) and the competition is not too high (competitor's price less than $122.50). Customers with incomes above $45 and ages below 54.51 are more likely to be recommended products if the product is relatively expensive (more than $130.50) and the competition is high (competitor's price more than $122.50).

Suggestions

Based on the findings above, the following suggestions can be made:

- The company should prioritize placing its car seats in locations with good shelving locations.
- The company could offer discounts on car seats that are placed in locations with poor shelving locations.
- The company should review its pricing strategy to ensure that its car seats are priced competitively.

- The company could also use the decision tree to develop a more sophisticated pricing strategy that takes into account both the shelving location and the price of competitor car seats. For example, the company could charge a higher price for car seats that are placed in high-traffic areas with good shelving locations.

- If you are targeting customers with incomes below $45, focus on other marketing channels, such as email or social media.
- If you are targeting customers with incomes above $45, consider the other factors in the decision tree, such as age, price, and competition.
- Use the decision tree to develop personalized recommendations for customers. For example, you could create a recommendation engine that uses the customer's income, age, and other factors to recommend products.

Conclusion

The decision tree is an effective framework for evaluating the elements that drive car seat sales and making strategic decisions about how to enhance sales. The corporation can boost its chances of success in the car seat industry by focusing on shelving location and pricing.

The decision tree you gave can be used to forecast whether or not a customer should be recommended a product. The tree divides the data based on a variety of characteristics, including the customer's income, age, and the product's pricing and competition. The customer's income is the most crucial aspect in determining whether or not to recommend a product. Customers earning less than $45 are unlikely to be recommended any products, regardless of other variables. Customers with earnings greater than $45 are more likely to be recommended things, although other criteria are equally important.

You can use the decision tree to develop targeted marketing campaigns and personalized recommendations for customers.

Additional Insights

Here are some additional insights from the decision tree:

The fact that the shelving location is the first node in the tree suggests that it is the most important factor in predicting car seat sales.

The fact that the price of the car seat is only considered if the shelving location is poor suggests that price is less important than shelving location.

The fact that the decision tree predicts high sales for locations with good shelving locations, regardless of price, suggests that the company has a lot of potential to increase sales by focusing on shelving location.

The decision tree can be used to develop a more sophisticated pricing strategy that takes into account both the shelving location and the price of competitor car seats, as well as to identify locations where the company can improve its shelving locations or offer discounts to increase sales.