**INTERNATIONAL JOURNAL of DYNAMICS in ENGINEERING and SCIENCES (IJDES)**
**LLDIKTI WILAYAH X**
Vol. 8, No. 2, Oct 2023
e-ISSN: 2502-0692

# OPTIMIZING BUSINESS MANAGEMENT BY PREDICTING SALES USING THE C4.5 ALGORITHM

**Chintya Ones Charli [1,]*, Sarjon Defit [2,], Ariyanto M[3]**
[123] Univesitas Putra Indonesia YPTK Padang
Jl. Lubuk Begalung Padang

DOI : 10.22216/jod.v8i2.2444
*Correspondence should be addressed to chintiaonescharli@upiyptk.ac.id


**Article Information**

Submitted :
02 Agust 2023

Accepted :
20 Sept 2023
Published :
01 Oct 2023

**Abstract**

With the growing development of the business sector, it causes intense competition, namely how to keep the business stable and maintain business continuity by implementing good business management. Padang Vape Shop Seventeen Vapehouse 2 is one of the vape shops in Padang. The problem faced by the shop is that the shop often buys or orders vape products that are less attractive to consumers from suppliers. By ordering vape products that are less desirable to consumers, it can cause piles of vape in stores and of course this will increase inventory costs and can be detrimental to the store. And also with these conditions causes a lack of optimization in business management. One way to overcome this problem is by implementing business management by predicting sales by using data mining techniques, namely the c4.5 algorithm. By using sales data at the Padang Seventeen Vapehouse 2 Vape Shop, 40 product data were obtained. Of the 40 data, 22 are in demand, and 18 are not in demand. From this data, the shop knows which products are in demand and which are not in demand, so the shop can order vape products for the future according to the products consumers are interested in.

**Keywords**: Business Management, Data Mining, C4.5 Algorithm, Sales Prediction

## 1. Introduction

Information technology is a technological development that is used to process information, collect, compile, store and process information in various ways to produce quality information, namely information that is relevant, accurate and timely and is information that can support decision making. This technology basically uses computers for data processing and connects computer network systems with other computers as desired, and communications technology so that data can be distributed and accessed widely [1] . Data mining is a process that connects various statistical, artificial intelligence, mathematics, and machine learning techniques to generate and extract useful information and related knowledge from large amounts of data [2] . Algorithm C4.5 is an algorithm used to build decision trees. Decision trees are very powerful and well-known methods for classification and prediction. The decision tree method can turn large problems into decision trees that represent rules [3] . Prediction or can also be called forecasting is a method used to predict something that has not yet happened. Prediction is also an activity that predicts in a structured way about something that might happen in the future, which is sourced from past and present information[ 4].

## 2. Methods

In taking or collecting data in this study, namely by way of interviews. Then collect or collect sales transaction data with the permission of the Padang Seventeen Vapehouse 2 Vape Shop, in order to obtain transaction data for the continuation of this research.

The general way of Algorithm C4.5 for building decision trees is as follows:

a. Select attribute as root
b. Create a branch for each value
c. Divide cases into branches
d. Repeat the process for each branch until all cases in the branch have the same class.

The first thing to do is to select the attribute as the root is the highest value for the existing attribute, and the formula used for the basic entropy is as follows [5] :

$$Entropy\ (S) = -\sum pi\ log\ 2\ (pi)$$

Information :

S : Set of cases
N : Number of partition S
pi : The proportion of Si to S

After finding the entropy value for an attribute, then looking for the gain value to determine which attribute to use as a node for the decision tree[5].

Data analysis is a process carried out to collect data that has been completed. In this research, researchers have obtained data so that it can be directly used for research data. The predictors used to predict vape sales at Padang Seventeen Vapehouse 2 vape shops: (1) price: cheap, affordable and expensive, (2) aroma: soft and strong (3) nicotone content: high and low, can be seen in table 1 following:

| No | Product name | Price | Aroma | Nicotine Content | Decision |
|---|---|---|---|---|---|
| 1 | Slank Vi Saltnic | Cheap | Strong | Tall | Less popular |
| 2 | Saltnic Paradeva | Affordable | Strong | Tall | Less popular |
| 3 | CN Mango | Affordable | Strong | Low | Less popular |
| 4 | Mango Breath | Cheap | Strong | Low | Less popular |
| 5 | Exo Mango | Cheap | Soft | Low | Bestseller |
| 6 | Million ice watermelon | Cheap | Soft | Low | Bestseller |
| 7 | Lcu Lychee | Cheap | Soft | Low | Bestseller |
| 8 | Lunar Hexom | Expensive | Soft | Low | Bestseller |
| 9 | Ice Cream Mall | Expensive | Soft | Low | Bestseller |
| 10 | Laklai | Expensive | Strong | Low | Less popular |
| 11 | The one | Expensive | Soft | Low | Bestseller |
| 12 | Brazilian Mango | Affordable | Soft | Low | Bestseller |
| 13 | Upod Saltnic | Cheap | Soft | Low | Bestseller |
| 14 | Mango Killer | Cheap | Soft | Low | Bestseller |
| 15 | Million One | Cheap | Soft | Low | Bestseller |
| 16 | Lcu Tiramisu | Cheap | Soft | Low | Bestseller |
| 17 | Slank V2 | Expensive | Strong | Low | Less popular |
| 18 | Ice Lond V2 | Affordable | Soft | Tall | Less popular |
| 19 | Miru Mango | Cheap | Strong | Low | Less popular |
| 20 | Miru Strawberry | Cheap | Soft | Low | Bestseller |
| 21 | Una Strawberry Chess Cake | Cheap | Strong | Low | Less popular |
| 22 | Recall savage Bluberry | Expensive | Strong | Low | Less popular |
| 23 | Joice Mango | Cheap | Soft | Low | Bestseller |
| 24 | Lcu Wice Apple Saltnic | Cheap | Soft | Low | Bestseller |
| 25 | Whale Dream Hokado | Expensive | Strong | Low | Less popular |
| 26 | Million Freezes | Cheap | Soft | Low | Bestseller |
| 27 | Opium V1 | Cheap | Strong | Low | Less popular |
| 28 | Opium V2 | Cheap | Strong | Low | Less popular |
| 29 | Mix Max Mango | Affordable | Strong | Low | Less popular |
| 30 | Aloe vera grap | Cheap | Soft | Low | Bestseller |
| 31 | Vanilla Liceous | Cheap | Soft | Tall | Less popular |
| 32 | Kaze Salt | Cheap | Strong | Tall | Less popular |
| 33 | Banana Rolla | Cheap | Soft | Low | Bestseller |
| 34 | Waffles | Cheap | Soft | Low | Bestseller |
| 35 | Shaks | Affordable | Soft | Low | Bestseller |
| 36 | Bandana Bread | Expensive | Soft | Low | Bestseller |
| 37 | Bestie | Expensive | Soft | Low | Bestselle |

| No | Product name | Price | Aroma | Nicotine Content | Decision |
|---|---|---|---|---|---|
| 38 | Snacks | Affordable | Strong | Low | Less popular |
| 39 | Avoris | Cheap | Soft | Low | Bestseller |
| 40 | Exo Melon | Cheap | Strong | Low | Less popular |

Table 1. Sales Data at the Padang Seventeen Vapehouse 2 Vape Shop

## 3. Results and Discussion

Calculation of Algorithm C4.5 starts by selecting the root attribute which first looks for the total number of cases.

| Node 1 | | Amount | Bestseller | Less popular | Entropy | Gains |
|---|---|---|---|---|---|---|
| Total | | 40 | 22 | 18 | 0.992774454 | |
| Price | | | | | | 0.046075839 |
| | Expensive | 9 | 5 | 4 | 0.991076060 | |
| | Affordable | 7 | 2 | 5 | 0.863120569 | |
| | Cheap | 24 | 15 | 9 | 0.954434003 | |
| Aroma | | | | | | 0.744484344 |
| | Strong | 16 | 0 | 16 | 0 | |
| | Soft | 24 | 22 | 2 | 0.413816850 | |
| Nicotine Content | | | | | | 0.159982113 |
| | Low | 35 | 22 | 13 | 0.951762676 | |
| | Tall | 5 | 0 | 5 | 0 | |

Table 2. Calculation of node 1

Node 1 calculation results in table 2 Total entropy row is calculated by the following equation:

***Total Entropy***

*Entropyl ( Total) )* = $(-\frac{22}{40} * \log_2\left(\frac{22}{40}\right)+(-\frac{18}{40} * \log_2\left(\frac{18}{40}\right)$

*Total Entropy* = 0.992774454

***Entropy Price***

*Entropy* ( Price , Expensive ) = $(-\frac{5}{9} * \log_2\left(\frac{5}{9}\right)) + (-\frac{4}{9} * \log_2\left(\frac{4}{9}\right)) = 0.99107606$

*Entropy* ( Price , Affordable )) = $(-\frac{2}{7} * \log_2\left(\frac{2}{7}\right)) + (-\frac{5}{7} * \log_2\left(\frac{5}{7}\right))= 0.863120569$

*Entropy* ( Price , Cheap )) = $(-\frac{15}{24} * \log_2\left(\frac{15}{24}\right)) + (-\frac{9}{24} * \log_2\left(\frac{9}{24}\right))= 0.954434003$

***Entropy Scent***

*Entropy* ( Scent , Strong) = $(-\frac{0}{16} * \log_2(\frac{0}{16})+(-\frac{16}{16} * \log_2(\frac{16}{16})) = 0$

*Entropy* ( Scent , Soft) = $(-\frac{22}{24} * \log_2(\frac{22}{24})+ (-\frac{2}{24} * \log_2(\frac{2}{24}))= 0.41381685$

***Entropy Nicotine Content***

*Entropy* ( Nicotine Content , High ) = $(-\frac{0}{5} * \log_2(\frac{0}{5}))+ (-\frac{5}{5} * \log_2(\frac{5}{5}))=0$

*Entropy* ( Nicotine Content, Low ) = $(-\frac{22}{35} * \log_2(\frac{22}{35})) + (-\frac{13}{35} * \log_2(\frac{13}{35})) = 0.951762676$

**Next look for Gain**

Gain (Total, Price ) = $0.992774454 - \left(\frac{9}{40} * 0{,}99107606\right) + \left(\frac{7}{40} * 0{,}863120569\right) + \left(\frac{24}{40} * 0{,}954434003\right)$

= 0.046075839

Gain (Total, Scent ) = $0.992774454 - \left(\frac{16}{40} * 0\right)+\left(\frac{24}{40} * 0{,}41381685\right)$

= 0.744484344

Gain (Total, Nicotine Content ) = $0.992774454 - \left(\frac{5}{40} * 0\right) + \left(\frac{35}{40} * 0{,}951762676\right)$= 0.159982113

From the results of table 2 it can be seen that the highest gain attribute value is Scent, which is equal to 0.744484344. Thus aroma can be the first node.

| Node 1.1 | | Amount | Bestseller | Less popular | Entropy | Gains |
|---|---|---|---|---|---|---|
| Total-Aroma(Soft) | | 24 | 22 | 2 | 0.413816850 | |
| Price | | | | | | 0.074169827 |
| | Expensive | 5 | 5 | 0 | 0 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Affordable | 3 | 2 | 1 | 0.91829583 4 | |
| | Cheap | 16 | 15 | 1 | 0.33729006 7 | |
| Nicotine Content | | | | | | 0.41381685 |
| | Low | 22 | 22 | 0 | 0 | |
| | Tall | 2 | 0 | 2 | 0 | |

Table 3 Calculation of nodes 1.1

***Total Entropy***

*Entropy(Total)* = $(-\frac{22}{24} * \log_2\left(\frac{22}{24}\right) + (-\frac{2}{24} * \log_2\left(\frac{2}{24}\right)$

*Total Entropy* = 0.41381685

***Entropy Price***

*Entropy* ( Price , Expensive ) = $(-\frac{5}{5} * \log_2\left(\frac{5}{5}\right)) + (-\frac{0}{5} * \log_2\left(\frac{0}{5}\right)) = 0$

*Entropy* ( Price , Affordable )) = $(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)) + (-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)) = 0.918295834$

*Entropy* ( Price , Cheap )) = $(-\frac{15}{16} * \log_2\left(\frac{15}{16}\right)) + (-\frac{1}{16} * \log_2\left(\frac{1}{16}\right)) = 0.337290067$

***Entropy Nicotine Content***

*Entropy* ( Nicotine Content , High ) = $(-\frac{0}{2} * \log_2(\frac{0}{2})) + (-\frac{2}{2} * \log_2(\frac{2}{2})) = 0$

*Entropy* ( Nicotine Content , Low ) = $(-\frac{22}{22} * \log_2(\frac{22}{22})) + (-\frac{0}{22} * \log_2(\frac{0}{22})) = 0$

**Next look for *Gain***

Gain (Total, Price ) = $0.41381685 - \left(\frac{5}{24} * 0\right) + \left(\frac{3}{24} * 0{,}918295834\right) + \left(\frac{16}{24} * 0{,}337290067\right)$

= 0.074169827

Gain (Total, Ni cotine Content ) = $0.41381685 - \left(\frac{22}{24} * 0\right) + \left(\frac{2}{24} * 0\right)$

**= 0.41381685**

From the results of table 3 it can be seen that the highest gain attribute value is Ni cotine Content which is equal to 0.41381685. Thus the Ni cotine Content can be the final search node. There are two attribute values of Ni cotine Content . From the two attribute values, the final decision is obtained for the Low attribute "Laris" and for the High attribute "Less Laris".

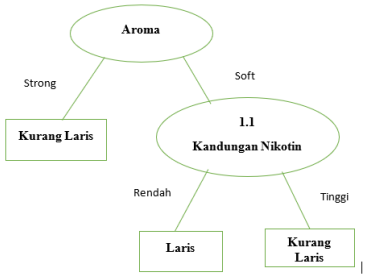


Figure 1. Decision tree results using manual calculations

## 4. Conclusion

Based on the results of research that has been carried out on the Application of Data Mining with Algorithm C4.5 in Predicting vape sales at Padang Seventeen Vapehouse 2 vape shops can be summed up as follows. If the scent is strong then the decision is "Kurang Laris", and if the scent is soft and the nicotine content is low then the decision is "Laris", and if the scent is soft and the nicotine content is high then the decision is " Less Selling".

From the research conducted, the results of this study can help the Padang Seventeen Vapehouse 2 Vape Shop in predicting sales of best-selling and less-selling products. This is proven by the processing of product data as much as 40 data. Of the 40 data, 22 were in demand, and 18 were less in demand. With the implementation of Business Management using the C.45 Algorithm, it can help the Padang Seventeen Vapehouse 2 Vape Shop in optimizing business management where the shop can order products according to the products that are in great demand by consumers, so the shop can reduce losses to the shop. With the implementation of Business Management using the C4.5 Algorithm, it can help the Padang Seventeen Vapehouse 2 Vape Shop in optimizing inventory so that there is no buildup at the Padang Seventeen Vapehouse 2 Vape Shop.

.

[1] RS Naibaho, "The Role and Planning of Information Technology in Companies," *J. War.* , no. April, p. 4, 2017, [Online]. Available: https://media.neliti.com/media/publications/290731-peranan-dan-peplanan-technology-inform-ad00d595.pdf

[2] Enda Esyudha Pratama, H. Sastypratiwi, and Yulianti, "Analysis of Information Trends Related to Covid-10 Based on

Social Media Big Data Using Data Mining Methods," *J. Inform. Polynema* , vol. 7, no. 2, pp. 1–6, 2021, doi: 10.33795/jip.v7i2.453.

[3] ED Sikumbang, F. Ariani, T. Handayani, K. Ramanda, SH Sukmana, and A. Supriyatna, "Application of Algorithm C. 45 To Determine the Level of Customer Satisfaction with Telkomsel Prepaid Cards," vol. 6, no. September, pp. 811–820, 2022.

[4] SP Dewi, N. Nurwati, and E. Rahayu, "Application of Data Mining to Predict Sales of Best Selling Products Using the K-Nearest Neighbor Method," *Build. Informatics, Technol. Sci.* , vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.

[5] Yusuf Maulana, Riki Winanjaya, and Fitri Rizki, "Application of Data Mining with the C.45 Algorithm in Predicting Tempe Sales," *Bull. Comput. Sci. Res.* , vol. 2, no. 2, pp. 53–58, 2022, doi: 10.47065/bulletincsr.v2i2.163.