Eduvest – Journal of Universal Studies
Volume 5 Number 9, September, 2025
p- ISSN 2775-3735- e-ISSN 2775-3727

Analysis of the Classification of Data on the Launch of Apple Mobile Phone Prices in China and Pakistan Using the Decision Tree Algorithm in Python Programming

Kevin Vincentio Benedict, Novi Rukhviyanti
STMIK Indonesia Mandiri, Indonesia
Email: kevinbenedict27@gmail.com, novi.rukhviyanti@stmik-im.ac.id

ABSTRACT

This study aims to explain the analysis of data by employing data classification and applying the concept of the Decision Tree Algorithm and programming tools developed using Python programming. This approach serves as a flexible and objective method for data analysis and visualization. In today's world of machine learning, data requires high performance and accurate results when addressing data cases. The data used consists of NoSQL data or datasets in CSV format. These CSV documents contain tables, and the resulting output will be the latest datasets. The process involves data analysis, classification, and categorization of Apple mobile phone products in China and Pakistan to determine whether the quality is high or low based on the launch price of Apple mobile phones. Subsequently, decision tree modeling is built to explain the factors affecting classification and to aid analysis of the launch prices of Apple mobile phone products in China and Pakistan. The results are executed in Python programming, a language distinguished by excellent characteristics for data processing and visualization. The datasets will undergo testing alongside supporting factors in Python programming data modeling, which provides flexibility in modeling, execution, and analysis, thereby enabling problems to be solved effectively and objectively.

KEYWORDS *Data classification, NoSql, Data modeling, Decision tree, Python.*

INTRODUCTION

Modern Technology has reached a very good development stage, it has been applied in several industrial fields in the health sector to education the application of various fields has different patterns so that it can be used efficiently and complexly until now, Technology has opened up many opportunities to improve human welfare, including in terms of health, education, and access to information (Agustina et al., 2021; Herlina, 2022; Nuansari & Ratri, 2022; Santoso et al., 2023).

Technology in this era aims to provide information so that all work and management of design systems can be applied easily in the work process, technology in the field of informatics is in great demand in the current era and even

work that is needed such as data management work and visualized through its hardware and supported by software that works together to facilitate data processing (Novita, 2022) and for data modeling that It needs to be considered in building data, namely by conducting data analysis.

In the field of informatics, there are several sources of work that have different language levels that have different characteristics and the programming language has different levels, therefore encouragement and efforts are needed to support these human resources must continue to be honed both in knowledge and in training (Aryansuka Mautama Putra & Mujiati, 2022; Fauzi & Siregar, 2019; Mintawati et al., 2023; Muis Wirasujatma, 2022; Rukhvianti et al., 2021).

Over time, the company in the field of information technology has the goal of making it easier to complete projects of developers and software houses, many things that companies do, one of which is to evaluate the goals that need to be achieved to advance the company in the future. The quality of human resources that play a role in it can be measured from employee performance (Rukhviyanti & Ambarwati, 2023) in the time scale that it takes will provide an improvement in program skills in program planning and program implementation.

The backend programming language will be used Python, In the backend programming language that has been explained, the backend also needs to master database management because one of the most important tasks of the backend developer is how to process, create, read, compress and delete data stored in the database, therefore data management is one of the important backend developer skills and the database is divided into two categories, namely relational databases (SQL) and Non-relational databases (NoSQL) These two categories have a very significant difference. SQL databases are relational databases which means that the data is organized into tables and each table has a specific structure, so that the tables are connected to each other through relationships. NoSQL databases are non-relational databases. This means that the data is stored in a document set. There is no specific structure for this document, and this data is not connected to each other through relationships. So this data is better suited for storing data that doesn't need to be accessed in a certain way.

In this study, the Data Source used NoSQl-based Datasets, in NoSQl-based Datasets has several tables, tables, columns regarding mobile phone products, specifications, launch prices in several countries, and launch year. Because the purpose of this research article focuses on the selected product and the work uses the predetermined ones, so this study only targets apple mobile phone products and launch prices in China and Pakistan and the table will be processed in python and meotode and data analysis is needed as a basis for management and to manage data to develop a structure so that data classification can be used in this study, one of

which is Find labels and averages as the key so that the results can be extracted data by selecting as a new label.

Several previous studies have emphasized the importance of technology-based data management in supporting effective analysis and decision-making. For example, Moniruzzaman & Hossain (2013) highlighted that NoSQL databases provide significant advantages in handling unstructured and large-scale data, especially in big data contexts that require high flexibility. However, their study did not focus on specific applications within certain industries. Similarly, Stonebraker (2018) compared SQL and NoSQL databases in terms of performance and efficiency, but his analysis mainly centered on technical system aspects without linking them to practical applications such as technology product classification. This research fills that gap by applying NoSQL-based data management to Apple product datasets, specifically launch prices in China and Pakistan, processed through Python to achieve more efficient data classification.

The objective of this study is to analyze Apple product data using a non-relational database approach to produce relevant classification structures, while the contribution lies in providing practical insights for developers, software houses, and academics on how Python and NoSQL can enhance data management effectiveness and support strategic decision-making in the field of information technology.

RESEARCH METHOD

Input Data

In the implementation of the Data Design System, a quantitative method is used based on the results of data analysis obtained from the KAAGLE website in the implementation and developed by determining the launch of Apple product prices in China and Pakistan and identifying the pattern of Apple mobile product price data so that the data displays the results of the table and the statistical results of the calculation.

Data Understanding

Data Understanding is a component for data grouping that will later be executed to carry out the data classification analysis process so that it requires the main label as the basis and the label will be processed in the Data Classification.

Table Data Filter

The Data Filter in the table is one of focusing on the problems in this journal so that it will target several table columns that will be processed later.

Table 1 Filter data table

Company Name	Model Name	Launched Price (Pakistan)	Launched Price (China)
1. Apple	Product Iphone	Launched Price Apple Iphone	Launched Price Apple
2. Apple	.....	.....	Iphone
…….			…..

Cleaning Data Table

This data cleaning is the deletion of data on which the data value has a variable value that is not known by the system, so this deletion is directed to the column of the Launched Price Pakistan and Launched Price China tables.

Table 2. Cleaning Data Table

Launched Price (Pakistan)	Launched Price (China)
PKR 224,999	CNY 5,799
PKR 234,999	CNY 6,099
…….	…….

Mean or Average

It is a formula in mathematics as the calculation of the average search in a data sample.

$$\frac{Data\ Amount}{Total\ Data} = (Average) \quad (1)$$

The formula is used in the formation of the Column Avarage Price table which will later be used as the calculation of the Average in each row of launched prices of China and Pakistan so that if implemented in the form of a formula it will look like this.

$$\frac{Launched\ Price\ China}{Launched\ Pakistan} = Avarage\ Price \quad (2)$$

Median

Median is a component used to find the middle value in the substitution of a data, the median mathematically uses a formula as a calculation in a lot of data values by targeting the middle value value. The application of the median has many processes that will be addressed so that the median can be implemented through several problems in the design of the data design system to the process of a data.

The design of this article measures how much data in the column of the Avarage price table as the basis for the calculation results will be given a new table column called Price Label, this label includes a condition where high and low values are determined.

$$\text{MEDIAN (High)} = x = \frac{(n+1)}{2} \quad (3)$$

$$\text{MEDIAN (Low)} = = \frac{x\left(\frac{n}{2}\right) + x\left(\frac{n}{2} + 1\right)}{2} \quad (4)$$

Data Classification Flowchart

The development of a data classification system provides easy data in data processing, data testing and sorting the results will provide a new data from the previous data and aim to average the quality of the mobile product launch price in each country based on certain features of the product. The Flowchart Decision Tree process as an implementation data flow in data visualization using a decision tree, for the construction of a system design this requires label data, this label is a label resulting from data classification, namely Price Label which will be processed into a new variable, namely Encode Label, so as to aim for label prediction in the data test

Flowchart Decision Tree

The Flowchart Decision Tree process as an implementation data flow in data visualization using a decision tree, for the construction of a system design this requires label data, this label is a label resulting from data classification, namely Price Label which will be processed into a new variable, namely Encode Label, so as to predict labels in the data test.

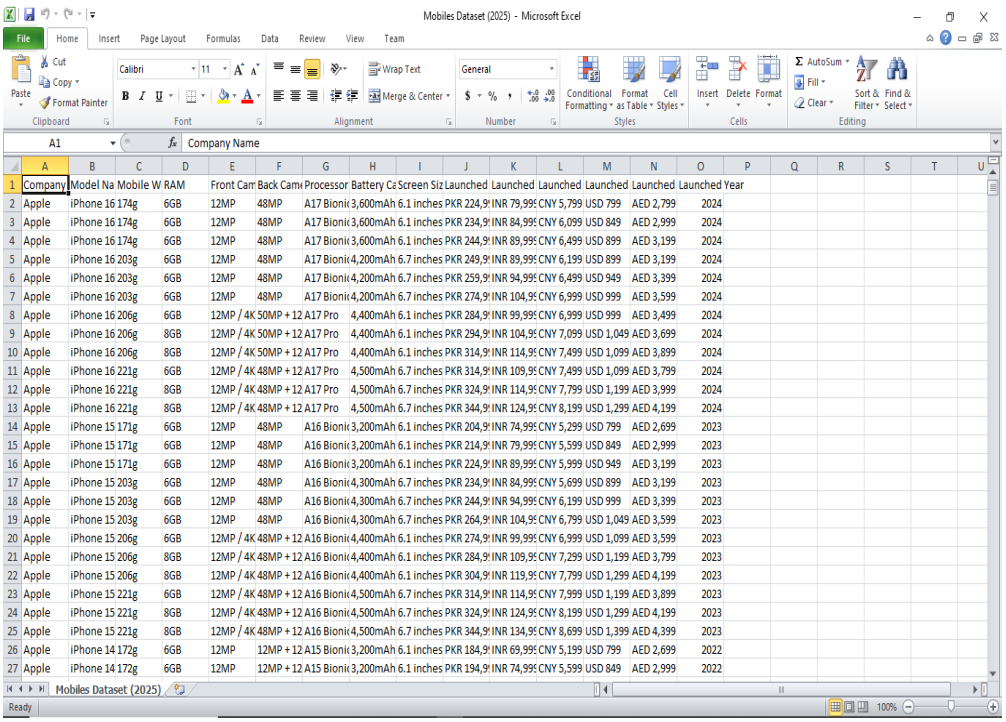
Python Programming

Data Classification Analysis has many ways in training or data modeling so that it requires the processing equipment so that it can be completed properly, the equipment used is Python, Python is very flexible and Objective in the application of data processing.

RESULT AND DISCUSSION

Data Normalization

Data Classification Analysis has many ways in training or data modeling so it requires the equipment to work on it so that it can be completed well, the equipment used by Python is very Flexible and Objective. which is sourced from Kaagle a Data-based website, a lot of data has been processed on the website so that the data is published so that the data can be used properly.



Mobiles Dataset (2025) - Microsoft Excel

Company	Model Na	Mobile W	RAM	Front Cam	Back Cam	Processor	Battery Ca	Screen Siz	Launched	Launched	Launched	Launched	Launched	Launched Year
Apple	iPhone 16	174g	6GB	12MP	48MP	A17 Bionic	3,600mAh	6.1 inches	PKR 224,9	INR 79,99	CNY 5,799	USD 799	AED 2,799	2024
Apple	iPhone 16	174g	6GB	12MP	48MP	A17 Bionic	3,600mAh	6.1 inches	PKR 234,9	INR 84,99	CNY 6,099	USD 849	AED 2,999	2024
Apple	iPhone 16	174g	6GB	12MP	48MP	A17 Bionic	3,600mAh	6.1 inches	PKR 244,9	INR 89,99	CNY 6,499	USD 899	AED 3,199	2024
Apple	iPhone 16	203g	6GB	12MP	48MP	A17 Bionic	4,200mAh	6.7 inches	PKR 249,9	INR 89,99	CNY 6,199	USD 899	AED 3,199	2024
Apple	iPhone 16	203g	6GB	12MP	48MP	A17 Bionic	4,200mAh	6.7 inches	PKR 259,9	INR 94,99	CNY 6,499	USD 949	AED 3,399	2024
Apple	iPhone 16	203g	6GB	12MP	48MP	A17 Bionic	4,200mAh	6.7 inches	PKR 274,9	INR 104,9	CNY 6,999	USD 999	AED 3,599	2024
Apple	iPhone 16	206g	6GB	12MP / 4K	50MP + 12	A17 Pro	4,400mAh	6.1 inches	PKR 284,9	INR 99,99	CNY 6,999	USD 999	AED 3,499	2024
Apple	iPhone 16	206g	8GB	12MP / 4K	50MP + 12	A17 Pro	4,400mAh	6.1 inches	PKR 294,9	INR 104,9	CNY 7,099	USD 1,049	AED 3,699	2024
Apple	iPhone 16	206g	8GB	12MP / 4K	50MP + 12	A17 Pro	4,400mAh	6.1 inches	PKR 314,9	INR 114,9	CNY 7,499	USD 1,099	AED 3,899	2024
Apple	iPhone 16	221g	6GB	12MP / 4K	48MP + 12	A17 Pro	4,500mAh	6.7 inches	PKR 314,9	INR 109,9	CNY 7,499	USD 1,099	AED 3,799	2024
Apple	iPhone 16	221g	8GB	12MP / 4K	48MP + 12	A17 Pro	4,500mAh	6.7 inches	PKR 324,9	INR 114,9	CNY 7,799	USD 1,199	AED 3,999	2024
Apple	iPhone 16	221g	8GB	12MP / 4K	48MP + 12	A17 Pro	4,500mAh	6.7 inches	PKR 344,9	INR 124,9	CNY 8,199	USD 1,299	AED 4,199	2024
Apple	iPhone 15	171g	6GB	12MP	48MP	A16 Bionic	3,200mAh	6.1 inches	PKR 204,9	INR 74,99	CNY 5,299	USD 799	AED 2,699	2023
Apple	iPhone 15	171g	6GB	12MP	48MP	A16 Bionic	3,200mAh	6.1 inches	PKR 214,9	INR 79,99	CNY 5,599	USD 849	AED 2,999	2023
Apple	iPhone 15	171g	6GB	12MP	48MP	A16 Bionic	3,200mAh	6.1 inches	PKR 224,9	INR 89,99	CNY 5,999	USD 949	AED 3,199	2023
Apple	iPhone 15	203g	6GB	12MP	48MP	A16 Bionic	4,300mAh	6.7 inches	PKR 234,9	INR 84,99	CNY 5,699	USD 899	AED 3,199	2023
Apple	iPhone 15	203g	6GB	12MP	48MP	A16 Bionic	4,300mAh	6.7 inches	PKR 244,9	INR 94,99	CNY 6,199	USD 999	AED 3,399	2023
Apple	iPhone 15	203g	6GB	12MP	48MP	A16 Bionic	4,300mAh	6.7 inches	PKR 264,9	INR 104,9	CNY 6,799	USD 1,049	AED 3,599	2023
Apple	iPhone 15	206g	6GB	12MP / 4K	48MP + 12	A16 Bionic	4,400mAh	6.1 inches	PKR 274,9	INR 99,99	CNY 6,999	USD 1,099	AED 3,599	2023
Apple	iPhone 15	206g	8GB	12MP / 4K	48MP + 12	A16 Bionic	4,400mAh	6.1 inches	PKR 284,9	INR 109,9	CNY 7,299	USD 1,199	AED 3,799	2023
Apple	iPhone 15	206g	8GB	12MP / 4K	48MP + 12	A16 Bionic	4,400mAh	6.1 inches	PKR 304,9	INR 119,9	CNY 7,799	USD 1,299	AED 4,199	2023
Apple	iPhone 15	221g	6GB	12MP / 4K	48MP + 12	A16 Bionic	4,500mAh	6.7 inches	PKR 314,9	INR 114,9	CNY 7,999	USD 1,199	AED 3,899	2023
Apple	iPhone 15	221g	8GB	12MP / 4K	48MP + 12	A16 Bionic	4,500mAh	6.7 inches	PKR 324,9	INR 124,9	CNY 8,199	USD 1,299	AED 4,199	2023
Apple	iPhone 15	221g	8GB	12MP / 4K	48MP + 12	A16 Bionic	4,500mAh	6.7 inches	PKR 344,9	INR 134,9	CNY 8,699	USD 1,399	AED 4,399	2023
Apple	iPhone 14	172g	6GB	12MP	12MP + 12	A15 Bionic	3,200mAh	6.1 inches	PKR 184,9	INR 69,99	CNY 5,199	USD 799	AED 2,699	2022
Apple	iPhone 14	172g	6GB	12MP	12MP + 12	A15 Bionic	3,200mAh	6.1 inches	PKR 194,9	INR 74,99	CNY 5,599	USD 849	AED 2,999	2022

Figure 1. Datasets Mobile (2025). CSV

Load Datasets

In the table above, the datasets that will later be processed and processed to require that the datasets can be processed using python require calling the datasets into the program.

```python
import pandas as pd

# Step 1: Load Dataset
data = pd.read_csv('Mobiles Dataset (2025).csv', encoding='latin1')
```

Figure 2. Load Datasets 1

Subsetting Data

```python
# Step 2: Fungsi untuk mengambil produk dari Apple
def filter_apple_products(df):
    df_apple = df[df['Company Name'].str.lower() == 'apple']
    return df_apple.copy()

# Step 3: Panggil fungsi
apple_data = filter_apple_products(data)
```

Figure 3.2 Subsetting Data 1

Data subsetting calls several table columns in the data selection so that the data used is only Apple products and filters the dataframe containing product information from various companies. The purpose of filtering the data will later be as new data after processing.

Cleaning Data

For data cleansing on research as cleaning up the launch price column for Pakistan and China by removing non-digit characters and converting the results to a float data type, this is useful to ensure that the price data can be used for further analysis in the absence of unwanted characters.

```
# Step 4: Bersihkan harga (hapus titik dan karakter non-digit)
apple_data['Launched Price (Pakistan)'] = (
    apple_data['Launched Price (Pakistan)']
    .astype(str)
    .str.replace(r'[^0-9]', '', regex=True)
    .astype(float)
)

apple_data['Launched Price (China)'] = (
    apple_data['Launched Price (China)']
    .astype(str)
    .str.replace(r'[^0-9]', '', regex=True)
    .astype(float)
)
```

Figure 4. Cleaning Data 1

Classification of Average Data

To determine the features in the classification requires labels and labels are based on two columns containing the launch price of apple products in Pakistan and China. The .mean(axis=1) function functions to calculate the average of Apple products horizontally where later the results will calculate the average of Apple products in both China and Pakistan for each product. So the result will create a new column table with the table name Avarage Price.

```
# Step 5: Hitung rata-rata harga
apple_data['Average Price'] = apple_data[['Launched Price (Pakistan)', 'Launched Price (China)']].mean(axis=1)
```

Figure 5. Calculate Average 1

After getting the Output Avarage Price, Calculate the median value from the Avarage Price column and will store it in the variable 'median_price'.using the Median to give a good idea of the average price without being influenced by the extreme value.

```
# Step 6: Labelkan harga berdasarkan rata-rata
median_price = apple_data['Average Price'].median()
```

Figure 6. Median

After performing the Median process, the process creates a new table column with the name Price Label which aims to provide the Avarage Price result by determining certain feature features.

```
apple_data['Price Label'] = apple_data['Average Price']
.apply(lambda x: 'Tinggi' if x >= median_price else 'Rendah')x
```

Figure 7. Determining the Price Label

The process on the price label has 2 variables, the price label variable and the Avarage price variable and the lambda calculation x:'High' if x >= median price else 'low')x shows the lambda function which checks whether the average price (x) is greater than or equal to the median. If so, the label is high otherwise the label is low.

```
# Step 7: Tampilkan hasil
print("Data produk Apple dengan label harga:")
print(apple_data[['Company Name', 'Model Name', 'Launched Price (Pakistan)', 'Launched Price (China)', 'Average
```

Figure 8. Calling Program Results 1

A pre-processed program at the end will call a result such as Print Data of Apple products with a price tag that is a description of the new dataset. The next process calls the apple_data data which in the process is the table columns "Company name", "Model Name", "Launched Price (Pakistan)", "Launched Price (China)", "Average Price", "Price Label".

```
Data produk Apple dengan label harga:
  Company Name           Model Name  Launched Price (Pakistan)  \
0        Apple       iPhone 16 128GB                   224999.0
1        Apple       iPhone 16 256GB                   234999.0
2        Apple       iPhone 16 512GB                   244999.0
3        Apple  iPhone 16 Plus 128GB                   249999.0
4        Apple  iPhone 16 Plus 256GB                   259999.0

   Launched Price (China)  Average Price Price Label
0                  5799.0       115399.0      Rendah
1                  6099.0       120549.0      Rendah
2                  6499.0       125749.0      Tinggi
3                  6199.0       128099.0      Tinggi
4                  6499.0       133249.0      Tinggi
```

Figure 9. Program Results

The print process will display the data as it is called and the results of the calculations that have been made.

```
# Step 8: Simpan seluruh hasil ke CSV
output_filename = 'Hasil_Program.csv'
apple_data[['Company Name', 'Model Name', 'Launched Price (Pakistan)',
            'Launched Price (China)', 'Average Price', 'Price Label']].to_csv(output_filename, index=False)

# Konfirmasi
print(f"\n✅ Seluruh data hasil berhasil disimpan ke file: {os.path.abspath(output_filename)}")
```

Figure 10 Export CSV File Results

The data that has been calculated will be exported to a CSV file so that the results can be structured and provide the latest dataset.

	A	B	C	D	E	F
1	Company	Model Name	Launched Price (Pakistan)	Launched Price (China)	Average Price	Price Label
2	Apple	iPhone 16 128GB	224999	5799	115399	Rendah
3	Apple	iPhone 16 256GB	234999	6099	120549	Rendah
4	Apple	iPhone 16 512GB	244999	6499	125749	Tinggi
5	Apple	iPhone 16 Plus 128GB	249999	6199	128099	Tinggi
6	Apple	iPhone 16 Plus 256GB	259999	6499	133249	Tinggi
7	Apple	iPhone 16 Plus 512GB	274999	6999	140999	Tinggi
8	Apple	iPhone 16 Pro 128GB	284999	6999	145999	Tinggi
9	Apple	iPhone 16 Pro 256GB	294999	7099	151049	Tinggi
10	Apple	iPhone 16 Pro 512GB	314999	7499	161249	Tinggi
11	Apple	iPhone 16 Pro Max 128GB	314999	7499	161249	Tinggi
12	Apple	iPhone 16 Pro Max 256GB	324999	7799	166399	Tinggi
13	Apple	iPhone 16 Pro Max 512GB	344999	8199	176599	Tinggi
14	Apple	iPhone 15 128GB	204999	5299	105149	Rendah
15	Apple	iPhone 15 256GB	214999	5599	110299	Rendah
16	Apple	iPhone 15 512GB	224999	5999	115499	Rendah
17	Apple	iPhone 15 Plus 128GB	234999	5699	120349	Rendah
18	Apple	iPhone 15 Plus 256GB	244999	6199	125599	Rendah
19	Apple	iPhone 15 Plus 512GB	264999	6799	135899	Tinggi
20	Apple	iPhone 15 Pro 128GB	274999	6999	140999	Tinggi
21	Apple	iPhone 15 Pro 256GB	284999	7299	146149	Tinggi
22	Apple	iPhone 15 Pro 512GB	304999	7799	156399	Tinggi
23	Apple	iPhone 15 Pro Max 128GB	314999	7999	161499	Tinggi
24	Apple	iPhone 15 Pro Max 256GB	324999	8199	166599	Tinggi
25	Apple	iPhone 15 Pro Max 512GB	344999	8699	176849	Tinggi
26	Apple	iPhone 14 128GB	184999	5199	95099	Rendah
27	Apple	iPhone 14 256GB	194999	5599	100299	Rendah

Figure 11 CSV Program Results

Decision Tree

The results of the Datasets will be used to provide a visualization of the decision tree and the application of the concept of a complex data structural Decision tree in the arrangement of nodes that are interconnected with the data that has been implemented, the Decision tree requires labels resulting from data classification so that the label data will be tested and in the application of the Decision tree concept requires Encode the label to be Numerical because, Decision tree can only read numerically Decision tree is a machine that reads a number so it requires adjustment, The Price Label is adjusted to the variable Price_Label_Code and then adjusted Low will be filled with the number '0' and High '1'.

```
apple_data['Price_Label_Code'] = apple_data['Price Label'].map({'Rendah': 0, 'Tinggi': 1})
```

Figure 12. Encode Decision Tree

```
X = apple_data[['Launched Price (Pakistan)', 'Launched Price (China)']]
y = apple_data['Price_Label_Code']
```

Figure 13. Data Preparation and Training

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

clf = DecisionTreeClassifier(max_depth=3, random_state=42)
clf.fit(X_train, y_train)
```

Figure 14. Data Modeling and Data Split

Dividing the dataset into 80% percent for training and 20% for testing and random_state = 42 is the result that the split runs consistently every time it is run. Then the development of DecisionTreeClassifier forms a decision tree model and max_depth=3 is a limit to the depth level of a decision tree of up to 3 levels to avoid overflitting. And Clf.fit (x_train, y_train) to train models based on x_train and y_train training data.

```
y_pred = clf.predict(X_test)
print("\n📊 Classification Report:")
print(classification_report(y_test, y_pred, target_names=['Rendah', 'Tinggi']))
```

Figure 15. Model Evaluation

Y_pred = clf.predict(x_test) Use the trained model (clf) to predict the labels on the test data (x_test) so that the results of the prediction are stored in (y_test).

```
plt.figure(figsize=(10, 6))
plot_tree(clf, feature_names=X.columns, class_names=['Rendah', 'Tinggi'], filled=True)
plt.title("Decision Tree: Prediksi Price Label")
plt.show()
```

Figure 16. Decision Tree Visualization

Decision tree visualization is a result based on calculations from Encode label data, data training data preparation, data modeling, and data evaluation. In the decision tree script, feature_names label the feature name so that the graph is easy

to read in the pre-designed build scheme, class_names displays the results of the low and high target labels and filled=true gives a color based on the impurity value of the node that is an informative visualization.

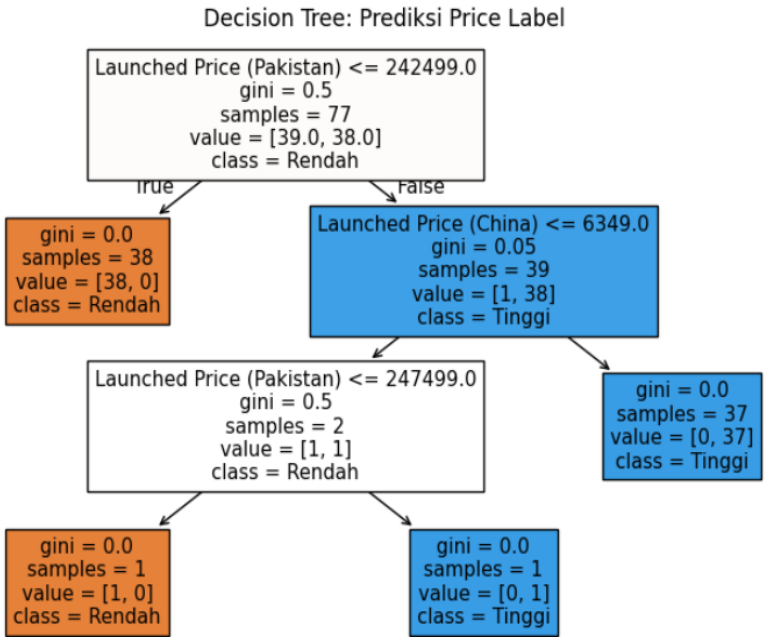


Figure 17. Program Visualization Results

The results of this study demonstrate that the application of the Decision Tree algorithm provides a clear visualization of classification outcomes, particularly in distinguishing price categories (Low and High) in the Apple smartphone dataset. The process involved encoding categorical labels into numerical form, preparing the dataset with an 80/20 split for training and testing, and optimizing the depth of the tree (max_depth=3) to prevent overfitting. These steps ensured that the decision tree model was interpretable while maintaining predictive consistency.

This finding aligns with the theory of supervised learning in machine learning, which emphasizes the transformation of categorical variables into numerical representations to enable algorithmic processing (Han, Kamber, & Pei, 2011). The importance of balancing model complexity and interpretability, as demonstrated by setting tree depth, is also supported by Breiman et al. (1984), who highlighted that decision trees are prone to overfitting if not constrained.

Previous studies further strengthen these results. Research by Kotsiantis (2013) found that decision trees offer strong classification performance while being easier to interpret compared to more complex algorithms, making them suitable for decision-making in practical contexts. Similarly, Patel and Upadhyay (2020) demonstrated that applying decision tree models with proper data preprocessing improved classification accuracy in e-commerce product datasets, particularly when categorical features were effectively encoded into numerical values.

By situating the findings within these theoretical and empirical frameworks, this study confirms that decision trees remain a powerful and efficient tool for

classification tasks in structured datasets. Moreover, this research contributes by applying decision tree modeling specifically to smartphone product data, offering a practical example of how price segmentation can be automated and visualized for business insights.

CONCLUSION

This study successfully demonstrated the application of data classification and decision tree modeling in analyzing Apple smartphone datasets, specifically focusing on launch prices in China and Pakistan. By implementing Python-based processing to calculate mean and median values and creating new variables such as Average Price and Price Label, the research achieved its objective of developing a structured classification system that distinguishes high and low price categories. The decision tree visualization further validated the effectiveness of this approach by producing an interpretable model that simplifies data labeling, training, testing, and evaluation. These findings confirm that decision tree methods can efficiently support data-driven decision-making in price segmentation and product analysis. Beyond addressing the research objectives, this study contributes to practical data management by highlighting the usability of Python in handling large-scale datasets with minimal complexity. For future research, expanding the scope to include additional smartphone brands, incorporating more diverse market variables, and comparing decision tree results with advanced machine learning algorithms such as Random Forest, Gradient Boosting, or Neural Networks would enrich the analysis and provide broader insights into global product pricing strategies.

REFERENCES

Agustina, D., Putri, M. A., & Ramadhan, M. G. (2021). Pemetaan riset strategi pemasaran bank syariah: Analisis bibliometrik. *MALIA (Terakreditasi)*, 12(2). https://doi.org/10.35891/ml.v12i2.2417

Aryansuka Mautama Putra, I. P. G., & Mujiati, N. W. (2022). Pengaruh motivasi, pelatihan dan kompetensi terhadap pengembangan karir karyawan bank. *Buletin Studi Ekonomi*, 27(2). https://doi.org/10.24843/bse.2022.v27.i02.p06

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC Press.

Fauzi, F., & Siregar, M. H. (2019). Pengaruh kompetensi dan kinerja karyawan terhadap pengembangan karir di perusahaan konstruksi. *Jurnal Manajemen Universitas Bung Hatta*, 2(1).

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Herlina, V. (2022). Analisis bibliometrik terhadap pemetaan riset social media marketing. *Jurnal Pekommas*, 7(2), 157–168. https://doi.org/10.56873/jpkm.v7i2.4920

Kotsiantis, S. B. (2013). Decision trees: A recent overview. *Artificial Intelligence Review*, 39(4), 261–283. https://doi.org/10.1007/s10462-011-9272-4

Mintawati, H., Rahayu, A. S., Puspita, D., Fauzan, F., & Safhira, F. (2023). Pengaruh pelatihan dan pengembangan terhadap kinerja karyawan di perusahaan pelayanan jasa. *Jurnal Ilmiah IPS dan Humaniora (JIIH)*, 1(1), 41–53. https://doi.org/10.61116/jiih.v1i1.38

Novita, N. (2022). Manajemen proyek sistem informasi pengolahan data apotek berbasis database. *Methosisfo Journal of Information Systems*, 2(1), 9–17.

Nuansari, S. D., & Ratri, I. N. (2022). Pemetaan riset teori agensi: Bibliometrik analisis berbasis data Scopus. *Implementasi Manajemen & Kewirausahaan*, 2(1), 24–37. https://doi.org/10.38156/imka.v2i1.105

Patel, H., & Upadhyay, D. (2020). Decision tree algorithm for e-commerce dataset classification. *International Journal of Computer Applications*, 176(39), 17–22. https://doi.org/10.5120/ijca2020920142

Rukhvianti, N., Rosida, & Ramdhani, M. A. (2021). Pengaruh pelatihan dan pengembangan melalui e-learning terhadap kompetensi karyawan di perusahaan X. *Seminar Nasional: Inovasi & Adopsi Teknologi 2021*, 10(9), 85–93.

Rukhviyanti, N., & Ambarwati, A. (2023). Pengaruh beban kerja dan motivasi kerja terhadap produktivitas kerja PT Toyota-Astra Motor NVDC Karawang. *Techno-Socio Economy*, 16(2), 197–205. https://doi.org/10.32897/techno.2023.16.2.2820

Santoso, R. A., Rukhviyanti, N., & Hayati, N. (2023). Pemetaan lanskap riset human development index dan technology menggunakan data Scopus dengan analisis bibliometrik. *Media Jurnal Informatika*, 15(2), 153–164. https://doi.org/10.35194/mji.v15i2.3480

Wirasujatma, M. (2022). Pengaruh komitmen, pelatihan dan kompetensi terhadap kinerja karyawan di PT X. *ATRABIS: Jurnal Administrasi Bisnis (e-Journal)*, 8(2), 90–98. https://doi.org/10.38204/atrabis.v8i2.1051