

6. 264-IMEIS-SIMILARITY

By Adi Raharjo

Using Supervised Machine Learning to Predict Sales in Marketplaces: Case study Predicting Sales of Padimas Bread in Marketplaces in Indonesia

Adi Raharjo¹, Nur Ichsan Utama,², Muharman Lubis³

Abstract

This project intends to apply supervised machine learning to anticipate sales of Padimas bread in marketplaces in Indonesia, with an emphasis on evaluating sales data to gain more profits and predict future income. Data from the Shopee, Tokopedia, and TikTok markets in 2023 was analyzed, employing techniques like exploratory sales data analysis and machine learning. The analysis findings encompass the top-selling products, the highest sales figures, regions with the most substantial sales, overall market sales, sales patterns, and revenue forecasts. The primary discoveries encompass the widespread appeal of chocolate toast, the most substantial sales of banana chocolate toast in West Java, and Shopee as the marketplace with the most significant sales and revenue. Sales trends exhibit a pattern of oscillation around an average value, but income from sales demonstrates a downward tendency until the 30th day. The strategic ramifications of this analysis encompass augmenting the production of sought-after goods, amplifying sales in specific regions, and delving into prospective marketplaces.

Keywords : Supervision, Machine Learning, Padimas Bread, Marketplace

INTRODUCTION

Human behavior and habits changed the COVID-19 pandemic. Some changes were favorable, like an increase in the consumption of fresh vegetables and fruits and a shift towards cooking at home. Nevertheless, specific behavioral changes might also have negative implications, such as an elevated intake of foods rich in fat, salt, and sugar and frequent snacking outside of regular meals.

The COVID-19 epidemic has significantly influenced the buying patterns of internet consumers. Online purchases experienced a substantial surge during the lockdown period, leading to greater familiarity and comfort among individuals in purchasing products and services through online platforms. Consumers have increased their inclination to purchase essential items online and circumvent brick-and-mortar establishments. Furthermore, there has been a surge in the utilization of food and beverage delivery applications and websites. Consumer

preferences for companies and products underwent alterations during the lockdown period. Nevertheless, even after the shutdown, specific patterns in online transactions persist, indicating enduring shifts in consumer conduct. The online marketplace is highly appealing to consumers due to its minimal seller interaction and various conveniences such as efficient product selection, competitive pricing, and the flexibility of shopping from anywhere without the need to visit physical markets. (Bazrkar & Hosseini, 2023; Chou, Chuang, Chou & Oliva, 2023)

Padimas, a bread product that has experienced growth since the 1980s and primarily relied on traditional markets for marketing, is now compelled to enter the realm of digital marketing in order to showcase and sell its products in online marketplaces. Therefore, with the outbreak of COVID-19, Padimas began expanding its presence in online marketplaces such as Shopee, Tokopedia, and TikTok.

It is essential to examine sales data from different products in the marketplace to gain extra benefits, such as identifying sales patterns and determining if they are increasing or decreasing. Regional growth or decline trends: Which product trends are currently in high demand among customers? It is also crucial to determine which products are likely to become top-selling or prioritized in future sales based on the available sales data. (Cong, Luo, Pei, Zhu & Zhang, 2022; Efat, Hajek, Abedin, Azad, Jaber, Aditya & Hassan, 2022)

They are engaging in e-commerce through online marketplaces while utilizing consumer profiling. This publication elucidates the use of consumer profiles in aiding sellers in online marketplaces to make astute selections about price, advertising, and other sales methods. (Dhanaraj, Rajkumar & Hariharan, 2020) Product transport costs can be minimized, and production efficiency can be enhanced by utilizing external vendors or adopting blended distribution methods. Notwithstanding apprehensions regarding consumer privacy and seller competition, consumer profiling continues to be a crucial determinant for sellers in online marketplaces to enhance their sales and earnings. (Kitanaka, Kwiatek & Panagopoulos, 2021; Li, 2021)

An advanced machine learning framework is developed to predict the purchase behavior of online clients using dynamic pricing. This study examines the utilization of machine learning in forecasting online customer purchasing patterns. (Ramu & Yeruva, 2023; Sarkar, Ayon, Mia, Ray, Chowdhury, Ghosh & Puja, 2023) It employs various machine learning techniques and statistical algorithms to analyze data from diverse sources, such as visitor data, purchase

history data, website behavior data, and contextual comprehension. (Martínez, Schmuck, Pereverzyev Jr, Pirker & Haltmeier, 2020) The journal discusses a versatile machine-learning framework that can be implemented with suitable machine-learning algorithms on various datasets. This framework assists enterprises in boosting their profits by accurately forecasting purchasing behavior. Furthermore, the magazine explores the pivotal role of machine learning in the customization of dynamic pricing and other innovations in e-commerce.

METHOD

Prior to the implementation of big data and machine learning principles in the business, there was an instance where a company saved a substantial amount of data on its storage and incurred significant expenses to maintain it. The firm leaders likely carefully considered their plans for the future use of the data.

It appears to be a forecast or prophecy. In 2017, economists widely proclaimed that data has surpassed oil as the world's most valuable resource. Data has evolved into a valuable resource for organizations to ascertain their course of action. Hence, possessing the capability to excavate and investigate data is vital.

Exploratory Data Analysis is the process of examining a particular dataset in order to identify and understand patterns and trends. The capacity to meticulously examine data, including identifying flaws in data collection and processing, detecting violations of statistical assumptions, and formulating intriguing hypotheses. Exploratory Data Analysis (EDA) is crucial in a data analyst's endeavor to extract

valuable insights from a given dataset. During the settlement procedure, it is necessary to establish the following steps:

- a. Identification of the problem.
Discover and ascertain sales issues for enhancement.
- b. Data acquisition.
Gather sales data from the sales department.
- c. Procedure Identification.
Identify the tools and techniques used for analyzing data and the considerations that must be considered.
- d. Determining the Software Package.
Enumerate examples of data analysis applications, such as Python.
- e. Implementation of Analysis
Recommendations for enhancing sales approach.

Machine Learning is employed to forecast the purchasing patterns of online customers. This method involves utilizing diverse machine learning techniques and statistical algorithms to analyze data from multiple sources, such as visitor data, purchase history data, website behavior data, and contextual comprehension. The process incorporates a versatile machine-learning framework that can be applied to each dataset using suitable machine-learning techniques. This capability enables organizations to enhance their profits by accurately predicting purchase behavior.

Machine Learning is categorized into two main branches: Supervised machine learning and Unsupervised machine learning. Supervised machine learning can be categorized into two main methods: Classification and Regression.

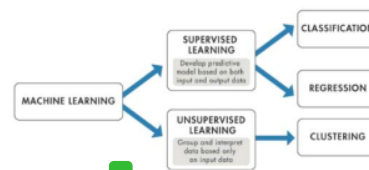


Figure 1. Types of Machine Learning

Supervised learning is a machine learning technique that relies on labeled data or a pre-existing dataset provided by the designer. These pre-configured datasets are intended to train "supervised" algorithms to classify or forecast a scenario precisely. Supervised learning is commonly employed for two primary purposes:

1. Categorization
2. Estimation

Unsupervised learning is a machine learning methodology that uses algorithms to examine and detect data patterns without human assistance or involvement. Under this mode, we cannot furnish information regarding the anticipated output resulting from an input to the algorithm. Instead, the algorithm autonomously discerns potential patterns from a given dataset. Unsupervised learning is typically employed for three purposes:

- a. Clustering refers to grouping similar data points based on their characteristics.
- b. Association involves finding relationships or patterns among different data items.
- c. Dimensionality reduction aims to reduce the number of variables or features in a dataset while preserving essential information.

Supervised machine learning is the approach we shall employ to examine padimas bread items' sales data.

RESULTS AND DISCUSSION

The dataset we will utilize comprises sales data for Padimas bread on three Indonesian marketplaces: Shopee, Tokopedia, and TikTok, specifically for 2023. To clarify, Padimas bread was first offered in the 1980s; however, its sales have recently commenced on a newly established marketplace in 2023.

The initial stage of initiating the process of gathering data involves importing sales data in Excel files into the data frame. Here is the application:



Figure 2. Load the Padimas bread sales data on the Shopee, Tokopedia and Tiktok Marketplaces

Subsequently, we will proceed to scrutinize the data for any instances of missing information. If such cases are identified, they will be eliminated from the dataset we intend to utilize. Please find the application below:



Figure 3. Deletion of Nan

To identify the padimas bread items that customers are interested in, we can utilize the word cloud library, a regularly employed tool for gauging sentiment and assessing the level of demand for a product among buyers in the marketplace. Here is how it may be applied:



Figure 4. Code and Result of Most-in Demand products

Figure 4 presents a comprehensive overview of the top-selling products across all data, providing insights into the bread products that are most in-demand among customers in the marketplace. Based on the statistics above, it is evident that the Chocolate Toast product is the most coveted among consumers in the marketplace, with Chocolate Nut Mooncake being the second most popular choice. This data can be a benchmark for determining which bread production quantities should be augmented.

Next, we will analyze the data to determine the top-selling products. A successful sale is indicated by an order status of "Processed." Therefore, we will first filter the data to include only orders with the status "Processed." The analysis is expected to reveal the products with the highest sales value. Here is the application:

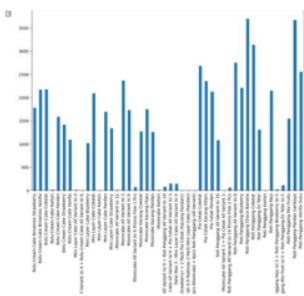


Figure 5. Sales by Product category

The choco banana toast is the most popular item among purchasers, followed by the crimson pandan bread in sales. The acquired findings diverge from the most prevalent products, indicating that the high-demand products may

In order to determine the most significant sales, the initial procedure involves filtering and categorizing the data according to sales per province. The data utilized for this purpose is the sales data that has been effectively processed. Here is the application:

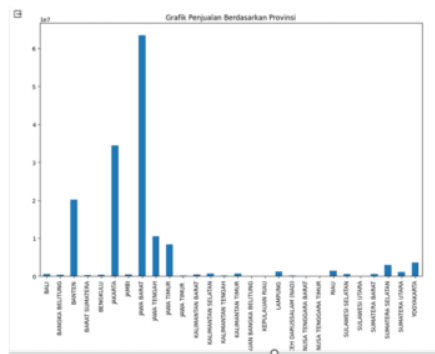


Figure 6 illustrates the graph visualization, indicating that the West Java province has the highest nominal income in sales.

[illegible]

Upon examining the graph depicted in Figure 7, it becomes evident that the Shopee marketplace exhibits the most substantial sales figures. This observation indicates that the purchasing capacity at Shopee surpasses other markets.

[illegible]

Based on the graph depicting the daily income from sales in the marketplace exhibits lateral fluctuations. However, it consistently maintains a very stable income level daily, with occasional growth.

To forecast the income for the next 30 days based on daily income data, we will employ linear regression. This task requires the utilization of various packages and models, which are listed below:

```

Day Order  Month Order  Year Order  Nominal
0      10      2      2023      2000
1      1      2      2023      95200
2      7      3      2023      62000
3      27     3      2023      38310
4      4      4      2023      17500
...
152    22     10     2023      154050
153    23     10     2023      206000
154    24     10     2023      120075
155    25     10     2023      245515
156    26     10     2023      73100
[157 rows x 4 columns]

```

Figure 9. Load daily sales revenue data using linear regression package

Before making predictions, it is necessary to partition the imported data into training and testing datasets. Additionally, a model must be constructed. We utilize the hardware capabilities of the tensorflow package to do sequential modeling using the deep learning method. Here are the code and results:

```

def split_sequences(sequence, n_steps):
    sequences = list()
    for i in range(1, len(sequence)):
        seq_in = sequence[i - n_steps:i]
        seq_out = sequence[i]
        sequences.append([seq_in, seq_out])
    return sequences

seq_in, seq_out = sequences[seq_in_sequences[seq_in_idx], seq_out_idx]
X.append(seq_in)
y.append(seq_out)

def rnn_model(x, y):
    model = Sequential()
    model.add(LSTM(64, input_shape=(n_steps, 1)))
    return model

rnn_model.compile(optimizer='adam')
train_data_loader = DataGenerator(X, y)
rnn_model.fit(train_data_loader.get_data(), train_data_loader.get_labels(),
              validation_data=train_data_loader.get_data(),
              validation_labels=train_data_loader.get_labels())

model = keras.Sequential()
model.add(LSTM(64, input_shape=(n_steps, 1), activation='tanh', return_sequences=True))
model.add(LSTM(64, input_shape=(n_steps, 1), activation='tanh', return_sequences=True))
model.add(LSTM(64, input_shape=(n_steps, 1), activation='tanh', return_sequences=True))
model.compile(optimizer='adam', loss='mse')
model.summary()

```

Model: "sequential_8"

Layer (type)	Output Shape	Param #
lstm_12 (LSTM)	(None, 1, 64)	16880
lstm_13 (LSTM)	(None, 64)	16824
lstm_14 (LSTM)	(None, 1)	85

Total params: 49980 (190.12 KB)
 Trainable params: 49980 (190.12 KB)
 Non-trainable params: 0 (0.00 KB)

Figure 10. Sequential Modeling

[illegible]

```
plt.plot(history.history['loss'], marker='x')
plt.title('Grafik perkembangan Model')
plt.xlabel('Epochs')
plt.ylabel('Error (loss)')
plt.grid()
plt.show()
```

Figure 11. Code and result of Model Training

The training yielded test results using 75% of the training data and 100 epochs. The loss initially started high at epoch one but gradually decreased. As the loss decreased, the accuracy of the analysis and learning improved. To visually represent the progression of the model above, we execute the following code:

```

get_forecast(figure=(8,4))
plt.plot(actual, label='Actual')
plt.plot(preds, label='Predicted')
plt.grid()
plt.title('Data Predicted vs Data Actual')
plt.xlabel('Time')
plt.ylabel('Number of Sales')
plt.legend()
plt.show()

```

Figure 12. Actual vs Predictive

There is a negligible disparity between the projected data and the factual data. Both data patterns are nearly identical, indicating that the model can effectively forecast future sales.

```

def predict_proba(self, X):
    """
    Predict the probability of each class for each instance in X.
    """
    X = self._preprocess(X)
    X = self._normalize(X)
    X = self._embed(X)
    X = self._project(X)
    X = self._predict(X)
    return X

def predict(self, X):
    """
    Predict the class for each instance in X.
    """
    X = self._preprocess(X)
    X = self._normalize(X)
    X = self._embed(X)
    X = self._project(X)
    X = self._predict(X)
    return X

def _preprocess(self, X):
    """
    Preprocess the input data X.
    """
    X = self._check_X(X)
    X = self._check_y(X)
    X = self._check_X_y(X)
    return X

def _normalize(self, X):
    """
    Normalize the input data X.
    """
    X = self._check_X(X)
    X = self._check_y(X)
    X = self._check_X_y(X)
    return X

def _embed(self, X):
    """
    Embed the input data X.
    """
    X = self._check_X(X)
    X = self._check_y(X)
    X = self._check_X_y(X)
    return X

def _project(self, X):
    """
    Project the input data X.
    """
    X = self._check_X(X)
    X = self._check_y(X)
    X = self._check_X_y(X)
    return X

def _predict(self, X):
    """
    Predict the class for each instance in X.
    """
    X = self._check_X(X)
    X = self._check_y(X)
    X = self._check_X_y(X)
    return X

```

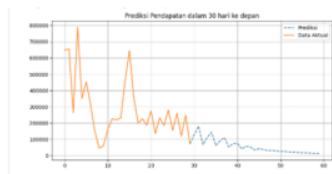


Figure 13. Sales revenue predictions in the next 30 days

The blue line represents the sales forecast for the following 30 days. According to the predicted data, the sales revenue exhibits a consistent downward trend each day, reaching its lowest point on the 30th and final day.

CONCLUSION

Based on the sales data analysis, it is evident that Padimas' chocolate toast is their best-selling bread product. The choco banana toast is the top-selling product, with the red pandanus toast in second place. The province of West Java has the highest sales and generates the highest income. Shopee is the leading marketplace in sales and revenue compared to other marketplaces. The sales trend exhibits a lateral fluctuation, alternating between upward and downward movements while maintaining a flat pattern. The revenue from sales exhibited a consistent downward trajectory, reaching its lowest point on the final day, the 30th day.

Based on the findings above, the corporation can implement strategic measures such as prioritizing the production of high-demand products, expanding sales in specific regions, exploring new marketplaces, and boosting overall sales. Upon observing unfavorable patterns and forecasts, the corporation can promptly devise strategies to enhance sales and income objectives.

REFERENCES

6. 264-IMEIS-SIMILARITY

ORIGINALITY REPORT

3%

SIMILARITY INDEX

PRIMARY SOURCES

1	apps.dtic.mil Internet	15 words — 1%
2	staging.globalforestwatch.org Internet	12 words — 1%
3	medium.com Internet	11 words — 1%
4	Hossein Taheri, Maria Gonzalez Bocanegra, Mohammad Taheri. "Artificial Intelligence, Machine Learning and Smart Technologies for Nondestructive Evaluation", Sensors, 2022 Crossref	9 words — < 1%
5	www.ipswitch.com Internet	8 words — < 1%

EXCLUDE QUOTES ON
EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES OFF
EXCLUDE MATCHES OFF