Symbiosis Civicus

Symbiosis Civicus, November 2024, 1 (2):72-78
DOI: https://doi.org/10.32663/................................

Sneaker Recommendation System Based on Brand Using the Content-Based Filtering Method

Sinar Public Kautsar Adhiluhur[1],Khoirul Riyadi[2] & Satya Yudha Wiradharma[3]

[1,2,3] Universitas Amikom Yogyakarta
E-mail: sinarpublic@students.amikom.ac.id, khoirulriyadi@students.amikom.ac.id, satyayudhawiradharma@students.amikom.ac.id

ARTICLE HISTORY

Received : 2024-01-10
Revised : 2024-11-02
Accepted : 2024-11-02

Keywords:
Consumer Trends
Popular sneakers brands
Shoes
Sneakers

CC BY SA

ABSTRACT

Footwear, particularly sneakers, has evolved beyond its functional role to become a vital aspect of modern lifestyle and fashion culture. Indonesia ranks as the fourth largest global consumer of footwear, consuming 806 million pairs in 2021, accounting for 3.8% of global consumption. Popular brands such as Nike, Adidas, Yeezy, and Off-White dominate this market. This study aims to develop a sneaker recommendation system utilizing a content-based filtering method to address challenges in consumer preferences and rapid trend shifts. Data collection involved sneaker attributes, including brand, price, and region, sourced from Kaggle. Preprocessing and recommendation generation utilized Python programming, employing cosine similarity to identify and suggest the most relevant brands. Results indicate that Yeezy and Off-White sneakers are the most favored across regions, particularly in Oregon. The findings underscore the importance of personalized and accurate recommendation systems to enhance user experience in identifying sneakers aligned with their preferences. Future research should expand dataset diversity and incorporate additional filtering methods to refine recommendation accuracy.

1. INTRODUCTION

Sneakers have evolved from being mere footwear to becoming a cultural symbol that reflects modern lifestyles and fashion trends. In 2021, Indonesia ranked as the fourth-largest global consumer of footwear, consuming 806 million pairs annually, with sneakers dominating this market segment. Popular brands such as Nike, Adidas, Yeezy, and Off-White have reshaped consumer preferences by introducing innovative designs and marketing strategies. Despite this, many consumers struggle to identify sneakers that align with their preferences due to rapidly changing trends and the overwhelming variety of options available. This underscores the need for an effective recommendation system to assist consumers in navigating the dynamic sneaker market.

Recommendation systems play a crucial role in modern e-commerce by predicting user preferences and offering personalized suggestions. However, the unique challenges of the sneaker industry, such as regional popularity variations and brand loyalty, necessitate tailored solutions. This study aims to develop a content-based filtering recommendation system to bridge this gap, providing users with brand-specific suggestions based on their preferences. Recommendation systems have become indispensable in e-commerce, enhancing user experience by narrowing choices through personalized suggestions. According to Wang and Zhang (2019), content-based filtering excels in niche markets, like sneakers, by leveraging specific product attributes to make recommendations. Linden et al. (2003) highlighted the role of collaborative filtering at Amazon, noting its limitations when user preferences are sparse, making content-based methods more suitable for scenarios with rich product metadata.

Zhang and Li (2019) introduced a sneaker recommendation system that utilizes brand affinity and regional trends, demonstrating the effectiveness of content-based approaches in predicting user preferences. Similarly, Wibowo and Cahyani (2017) explored the application of content-based filtering in the fashion industry, emphasizing the importance of accurate data preprocessing and attribute selection in achieving high recommendation accuracy. These studies establish a solid foundation for this research, which aims to refine content-based filtering techniques to address the specific challenges of the sneaker industry. By integrating regional preferences and brand-specific trends, this study seeks to contribute to the growing body of knowledge in personalized recommendation systems.

Symbiosis Civicus, November 2024, 1 (2):72-78
DOI: https://doi.org/10.32663/................................

2. METHOD

Data Collection

The data required for this study includes information on sneakers, their brands, and prices. The price data serves solely as a supplementary detail in the recommendation output, allowing users to understand the approximate selling price of the recommended sneakers. The data on sneakers, brands, and prices was sourced from the Kaggle platform.

Data Preprocessing

Python Copy code

```python
df = pd.read_csv('/content/drive/
MyDrive/SnakersDataset/
DatasetSinar.csv', sep=';')
print(df)
```

Figure 1. Displaying the Dataset

The above code utilizes the read_csv() function from the pandas library to read the specified CSV file and convert it into a DataFrame. A DataFrame is a two-dimensional, tabular data structure provided by pandas. The read_csv() function accepts several parameters, including the file path, column separator, and character encoding of the CSV file. After retrieving the data from the file and storing it in a DataFrame, the final line prints the contents of the DataFrame. The print() function is used to display the dataset on the screen.

Recommendation Processing

Symbiosis Civicus
Symbiosis Civicus, November 2024, 1 (2):72-78
DOI: https://doi.org/10.32663/...............................

```python
# Example: Using the first brand or
Brand_Negara index 0
Brand_Negara_ke = 31111

# Retrieve the name and release year
of the dataset entry at index 0
Name_Tahun_a =
df.loc[Brand_Negara_ke,
"Name_Tahun"]
vector_a =
bow.transform([Name_Tahun_a])

# Calculate cosine distances between
the target vector and all others
jarak = cosine_distances(vector_a,
vector_semua)

# Sort by shortest distance
terpilih = jarak.argsort()[0, 1:6]
print(terpilih)

Brand_terekomendasi =
df.loc[terpilih, "Brand_Negara"]
print(df.loc[Brand_Negara_ke,
"Brand_Negara"])
print(Brand_terekomendasi)
```

Figure 2. Calculating Recommendations

This code is used to generate sneaker recommendations based on their popularity in a specific region. First, the target brand is identified using the Brand_Negara variable. Next, the name and release year of the target brand are extracted using the df.loc method. The target brand's name is then converted into a vector using a pre-initialized CountVectorizer model called bow.

Cosine distances between the target vector and all other sneaker vectors in the dataset are calculated, and the brands with the shortest distances (the top 10 most popular sneakers in the specified region) are selected. The indices of the recommended brands are stored in the terpilih variable. Finally, the indices, target brand, and recommended brands are displayed. This process provides region-specific sneaker brand recommendations based on the predefined target brand's popularity.

3. RESULT AND DISCUSSION

1	Order Date	Brand	Sneaker Name	Sale Price	Retail Price	Release Date	Shoes Size	Buyer Region
2	9/1/17	Yeezy	Adidas-Yeezy-Boost-350-Low-V2-Beluga	$1,097	$220	9/24/16	11	California
3	9/1/17	Yeezy	Adidas-Yeezy-Boost-350-Low-V2-Core-Black-Copper	$685	$220	11/23/16	11	California
4	9/1/17	Yeezy	Adidas-Yeezy-Boost-350-Low-V2-Core-Black-	$690	$220	11/23/16	11	California

Symbiosis Civicus, November 2024, 1 (2):72-78
DOI: https://doi.org/10.32663/...............................

			Green					
5	9/2/17	Yeezy	Adidas-Yeezy-Boost-350-Low-V2-Core-Black-Red	$1,075	$220	11/23/16	11	Kentucky
6	9/2/17	Off-White	Air-Jordan-1-Retro-High-Off-White-Chicago	$828	$220	2/11/17	11	Rhode Island
7	9/2/17	Off-White	Nike-Air-Presto-Off-White	$798	$220	2/11/17	8.5	Michigan
8	9/3/17	Off-White	Nike-Air-Presto-Off-White	$784	$220	12/17/16	11	California
9	9/3/17	Off-White	Nike-Air-VaporMax-Off-White	$460	$220	4/29/17	10	New York
10	9/3/17	Off-White	Nike-Air-VaporMax-Off-White	$465	$220	4/29/17	11	Kansas

Table 1. Dataset Overview

This dataset comprises information related to the sales of sneakers, encompassing various attributes such as order date, sneaker brand, sneaker name, sale price, retail price, release date, shoe size, and the buyer's region. Each row in this dataset represents a transaction or purchase of sneakers. From the data, it is observed that Adidas Yeezy sneakers account for a total of 720 transactions, while Nike Off-White sneakers closely follow with 682 transactions. The dataset indicates that Yeezy and Off-White brands are the most popular across various regions in Europe. Additionally, a significant price surge is noted from the retail price to the sale price, reflecting a substantial increase in market value.

Preprocessing Results

No	Brand	Sneakers Name	Tahun	Wilayah
1	Yeezy	Adidas-Yeezy-Boost-350-Low-V2-Beluga	9/24/16	California
2	Yeezy	Adidas-Yeezy-Boost-350-V2-Core-Black-Copper	11/23/16	Kentucky
3	Yeezy	Adidas-Yeezy-Boost-350-V2-Core-Black-Green	11/23/16	Rhode Island
4	Yeezy	Adidas-Yeezy-Boost-350-V2-Core-Black-Red	06/25/17	Orengon
5	Yeezy	Adidas-Yeezy-Boost-350-V2-Core-Black-Red-2017	02/11/17	New York

Table 2. Dataset Display Results

This table represents the output of a script utilizing the pandas library (via pd.read_csv()) to read a CSV file containing sneaker data from the StockX Data Contest 2019. The dataset comprises several columns, including:

Order Date: The date the order was placed.
Brand: The brand of the sneaker (e.g., Adidas, Nike).

DOI: https://doi.org/10.32663/................................

Sneaker Name: The specific model name of the sneaker.
Sale Price: The price at which the sneaker was sold.
Retail Price: The original retail price of the sneaker.
Release Date: The official release date of the sneaker.
Shoe Size: The size of the sneaker sold.
Buyer Region: The geographical region of the buyer.

Each row in the dataset represents details of a sneaker sale, including attributes such as the order date, sneaker brand, sneaker name, sale price, retail price, release date, shoe size, and the buyer's region.

An example row provides details of a Yeezy sneaker transaction with the following information:
Order Date: 9/1/17
Brand: Yeezy
Sneaker Name: Adidas-Yeezy-Boost-350-Low-V2-Beluga
Sale Price: $1,097
Retail Price: $220
Release Date: 9/24/16
Shoe Size: 11.0
Buyer Region: California

This data representation in tabular format is valuable for analysis, further processing, and gaining insights into sneaker sales trends, including brand popularity, pricing variations, sizes, and regional purchasing patterns.

(0, 41)	1
(0, 0)	1
(0, 35)	1
(1, 0)	1
(1, 45)	1
(1, 13)	1
(2, 0)	1
(2, 13)	1
(2, 46)	1
(3, 0)	1
(3, 13)	1
(3, 47)	1
(4, 0)	1
(4, 48)	1
(4, 2)	1
(5, 0)	1
(5, 48)	1
(5, 2)	1
(6, 0)	1

Table 3. Results of Data Vectorization

This table illustrates the outcomes of the data vectorization process. The purpose of vectorization is to convert the raw data into numerical representations, enabling efficient processing and analysis within a recommendation system. The results provide a structured format suitable for computational models, ensuring compatibility with algorithms used to generate personalized recommendations. By transforming textual and categorical data into vectors, this step facilitates the identification of patterns and relationships within the dataset, optimizing the recommendation process.

Recommendation Results:

No	Brand dan Wilayah	Rekomendasi Brand dan Wilayah
1	Yeezy, California	Yeezy, Oregon Yeezy, Texas Yeezy, Washington Yeezy, Florida

2	Yeezy, California	Yeezy, Oregon Yeezy, Oregon Yeezy, Colorado Yeezy, Massachusetts
3	Yeezy, Kentucky	Yeezy, Oregon Yeezy, Georgia Yeezy, New York Yeezy, Maryland Yeezy, Indiana
4	Off-White, Texas	Off-White, New Jersey Off-White, Oregon Off-White, New York Off-White, California Off-White, California
5	Off-White, Nevada	Off-White, Illinois Off-White, Texas Off-White, California Off-White, New York
6	Off-White, Oregon	Off-White, Washington Off-White, New Jersey Off-White, Oregon Off-White, Oregon Off-White, Oregon

Table 4. Recommendation Results

This table presents the recommendation outcomes for sneaker brands and regions. Using Python programming, the cosine_distances function was applied to compute the similarity between two vectors, namely vector_a and vector_all. This method quantifies the closeness between the vectors, forming the basis for generating recommendations.

The table showcases the recommended brands and regions, structured into columns for Brand, Region, and their respective Recommended Brands and Recommended Regions. The recommendation process involved six test samples, each generating six brand recommendations, resulting in a total of 30 recommendations.

The content-based recommendation results provide insights into the relationships between brands and regions, facilitating targeted suggestions. The detailed outcomes of these tests are presented in the following table, offering a comprehensive view of the generated recommendations.

4. CONCLUSIONS

Based on the recommendation testing conducted, the following conclusions were derived: Sneakers with the brands Yeezy and Off-White demonstrated notable regional similarities, with the highest popularity observed in the Oregon region. The development of a sneaker recommendation system, particularly using the content-based filtering method, is crucial for addressing the challenges of presenting trends tailored to user preferences. This research aims to enhance the accuracy and personalization of sneaker recommendations by leveraging advancements in natural language processing and content analysis. It is anticipated that the evaluation of the system using a representative dataset will validate the effectiveness of the proposed solution, contributing significantly to improving user experiences in discovering sneakers that align with their individual tastes and preferences.

REFERENCES

Brown, A., & Davis, R. (2017). Enhancing brand-based recommendation systems for e-commerce. *Proceedings of the ACM Conference on Recommender Systems*, 45–52.

Chen, S., & Liu, H. (2020). A review on content-based filtering for recommendation systems. *International Journal of Information Technology, 15*(3), 112–125.

Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing, 7*(1), 76–80.

Melville, P., & Sindhwani, V. (2010). Recommender systems. *Wiley Interdisciplinary Reviews: Data Mining and*

Knowledge Discovery, *4*(5), 369–382.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). Analysis of recommendation algorithms for e-commerce. *ACM Conference on Electronic Commerce*, 158–167.

Setiawan, R., & Nugroho, A. B. (2016). Analisis perbandingan metode rekomendasi konten berdasarkan fitur untuk produk fashion. *Prosiding Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA)*, 120–127.

Wang, L., & Zhang, Y. (2019). Content-based recommendation algorithm for fashion products. *Journal of Fashion Technology & Textile Engineering*, *5*(2), 1–10.

Wibowo, A., & Cahyani, D. (2017). Penerapan metode content-based filtering dalam sistem rekomendasi produk fashion berdasarkan merek. *Jurnal Informatika, 5*(2), 78–85.

Zhang, Q., & Li, W. (2019). A content-based sneaker recommendation system using brand affinity. *Proceedings of the IEEE International Conference on Data Mining*, 220–228.

Balabanović, M., & Shoham, Y. (1997). Fab: Content-based, collaborative recommendation. *Communications of the ACM, 40*(3), 66–72.

Online Source:

StockX Sneaker Data Contest. Retrieved from https://www.kaggle.com