Recommender System for Tourist Destinations in Indonesia Using Matrix Factorization Method

Danny Matthew Saputra[#], Nadya Angelia[#], Novi Yusliani[#]

[#] Faculty of Computer Science, Sriwijaya University, Indonesia
E-mail: danny[at]ilkom.unsri.ac.id, nadya.angel126[at]gmail.com, novi_yusliani[at]unsri.ac.id

A B S T R A C T S

Indonesia has various tourist destinations. The large number of tourist destinations makes people confused about choosing a suitable tourist destination. The recommendation system is an appropriate way to help Indonesians choose tourist destinations that suit their preferences. One recommendation system method is matrix factorization. This research uses a matrix factorization algorithm, Alternating Least Square (ALS). The dataset used is Indonesia Tourism Destination from Kaggle. Based on research that has been carried out, this algorithm is successful in predicting tourist attractions that suit users. The evaluation results are an MAE value of 1.27203388032266, while the RMSE value is 1.475271987.

Manuscript received Jul 09, 2024; revised Aug 02, 2024 accepted Aug 06,2024 Date of publication Sep 30, 2024 International Journal,

Keywords / Kata Kunci — ***Recommendation System, Matrix Factorization, Alternating Least Square, Tourist Destinations, Collaborative Filtering***

CORRESPONDING AUTHOR

Nadya Angelia
Faculty of Computer Science, Sriwijaya University, Indonesia
Email: nadya.angel126[at]gmail.com

1. Introduction

A recommendation system is an application that can provide recommendations regarding an item so that it can be used to help users make decisions [1]. Recommendation systems can also help users determine product choices that suit the user's needs and interests. One type of recommendation system is collaborative filtering. This type has advantages such as being able to produce unexpected recommendations according to market trends. Collaborative filtering is also suitable for use when information about items is not available or difficult to obtain because it does not require detailed information about the recommended items.

One method of collaborative filtering is matrix factorization. Based on previous research regarding the matrix factorization method, the results showed that the model with the best performance for recommendation engines was matrix factorization compared to the K-Nearest Neighbor model with an MAE value of 0.6417 and an RMSE of 0.8305 [2]. Research on recommendation engines using matrix factorization shows that the accuracy of the recommendation system is 86.556% (RMSE) and 89.573% (MAE). This research states that the model created using the matrix factorization method succeeded in learning 80% of the existing data so that it could predict the tested data and results. machines are good enough to predict [3]. Furthermore, this method has the advantage that it can work efficiently and provide good recommendations even when faced with a large number of users and items. This is shown from the RMSE test results, namely 0.8068042 on the recommendation engine compared with the Restricted Boltzmann machine algorithm [4]. One algorithm of the matrix factorization method is Alternating Least Squared (ALS). This algorithm is an algorithm built for large-scale collaborative filtering problems. This algorithm works well in overcoming data scalability and sparseness problems.

Tourism is one sector that can increase the development of the country's economy. The tourism sector plays a significant role as a source of foreign exchange for the country and has the potential to boost the national economy, especially in reducing the unemployment rate and increasing a country's productivity [9]. However, the large number of tourist destinations in Indonesia makes people confused about choosing a suitable destination. Apart from that, many people have difficulty finding a variety of tourist attractions and a system is needed that helps people determine tourist destinations. So, the research was created.

2. RESEARCH METHODOLOGY

2.1. Types of Data and Data Sources

The type of dataset used in Matrix Factorization method research is secondary data. The dataset is Indonesia Tourism Destination (ITD). This dataset includes tourism interactions in five cities in Indonesia, namely Jakarta, Yogyakarta, Semarang, Bandung, and Surabaya. ITD consists of 4 packages consisting of 10,000 tourist interaction rating data, 437 tourist attractions, and 300 tourist data. This dataset was last updated in 2021. This dataset can be accessed on the Kaggle page.

TABLE 1. Example of Data Used

User Id	Place Id	Place Ratings
1	179	**3**
1	21	2
6	116	5
12	429	3
8	202	3

2.2. Matrix Factorization

Matrix Factorization in recommendation systems is an approach used in collaborative filtering to represent user preferences and item properties in matrix form. This allows users and items to mutually produce information which shows that the two have a high relationship and correspondence so that they can influence each other. Matrix Factorization is also a part of the Collaborative Filtering method, the matrix factorization method models the interactions formed between users and items through ratings given by users. This method can break down the user-item rating matrix into two matrices, the user matrix and the item matrix, which can be multiplied together to predict the empty rating.

Alternating Least Square (ALS) is a matrix factorization algorithm built for large-scale Collaborative Filtering problems. The advantage of this algorithm is that ALS works very well to overcome data scalability and sparseness. ALS takes as input a matrix containing interactions between users and items. This algorithm aims to estimate the rating matrix. The ALS algorithm factors the user-item ranking matrix into two matrices, namely the matrix representing the user and the matrix representing the item so that the product of the two matrices approaches the original ranking matrix. This is done by updating the user-item matrix alternately while keeping the other matrices constant. The equation for updating the user matrix latent factors can be seen in:

$$X_u = (\textstyle\sum_{r_{ui} \in r_{*i} \,.\, y_i Y_i^T + \lambda I_k})^{-1} . \textstyle\sum_{r_{ui} \in r_{*i}} r_{ui} y_i \quad (1)$$

Information:

y_i = item latent factor matrix
y_i^T = transposed item latent factor matrix
λ = regulatory parameter
I_k = identity matrix
r_{ui} = rating matrix

Meanwhile, to update the item latent factor matrix through the equation:

$$Y_i = (\textstyle\sum_{r_{ui} \in r_{*i} \,.\, x_u x_u^T + \lambda I_k})^{-1} . \textstyle\sum_{r_{ui} \in r_{*i}} r_{ui} x_u \quad (2)$$

Information:

x_u = user latent factor matrix
x_u^T = transposed user latent factor matrix
λ = regulatory parameters
I_k = identity matrix
r_{ui} = rating matrix

In updating the rating matrix to obtain appropriate results, this is done using a formula:

$$R' = X_u \,.\, Y_i^T \quad (3)$$

Information:

R' = rank prediction matrix
X_u = user factor matrix that has converged
Y_i^T = item factor matrix that has converged

List Programming

```
X_train, X_test = data.randomSplit([0.80, 0.20])
als = ALS(rank = 20, maxIter=5, regParam=0.6)
model = als.fit(X_train.select(["user", "item", "rating"]))
```

2.3. Mean Absolute Error (MAE)

MAE is one of the evaluation methods used in recommendation systems to measure the accuracy of a given system. MAE calculates the absolute difference between the predicted value and the actual value and then calculates the average of the predictions.
The calculation to get the MAE value can be seen in this equation:

$$\text{MAE} = \frac{\left|X - X_i\right|}{n} \quad (4)$$

Information:
X = Predicted value
X_i = Actual value
n = Number of data

List Programming

```
predictions = model.transform(X_train.select(["user", "item"]))

ratesAndPreds = X_test.join(predictions, (X_test.user == Thank you for reaching out.
        predictions.user) & (X_test.item == predictions.item) ,\
        how='inner').select(X_test.user, X_test.item, \
        X_test.rating, predictions.prediction)
# ratesAndPreds.show()
ratesAndPreds = ratesAndPreds.select([col('rating').alias('label'),\
        col('prediction').alias('raw')])
ratesAndPreds = ratesAndPreds.withColumn('label', \
        ratesAndPreds['label'].cast(FloatType()))

# test = ratesAndPreds.select('raw').where(col('label')==3.0)
print(ratesAndPreds.count())
evaluator = RegressionEvaluator(predictionCol='raw')
mae = evaluator.evaluate(ratesAndPreds, {evaluator.metricName: "mae"})
```

2.4. Root Mean Squared Error (RMSE)

RMSE is a technique for evaluating linear regression models which is done by measuring the accuracy of the model prediction results. The RMSE calculation is done by squaring the error difference (prediction - observation) dividing the result by the amount of data and then rooting it.

$$\text{RMSE} = \frac{\sqrt{\sum(X - Y)^2}}{n} \quad (5)$$

Information:
X = Predicted value
Y = Actual value
n = Number of data

List Programming

```
predictions = model.transform(X_train.select(["user", "item"]))

ratesAndPreds = X_test.join(predictions, (X_test.user == \
        predictions.user) & (X_test.item == predictions.item) ,\
        how='inner').select(X_test.user, X_test.item, \
        X_test.rating, predictions.prediction)
# ratesAndPreds.show()
ratesAndPreds = ratesAndPreds.select([col('rating').alias('label'),\
        col('prediction').alias('raw')])
ratesAndPreds = ratesAndPreds.withColumn('label', \
```

```
        ratesAndPreds['label'].cast(FloatType()))

# test = ratesAndPreds.select('raw').where(col('label')==3.0)
print(ratesAndPreds.count())
evaluator = RegressionEvaluator(predictionCol='raw')
rmse = evaluator.evaluate(ratesAndPreds, {evaluator.metricName: 'rmse'})
```

2.5. Software Architecture

The initial data processing stage is processed first through data cleaning or deleting several data attributes that are not needed in the research. Next, dimensionality reduction is carried out which aims to avoid noisy patterns and memory errors due to the computing process on very large data. At this stage, tourist attractions are filtered based on the ratings obtained, namely more than equal to 20 ratings.

The data will be divided into training data and testing data. Training data will be processed for model training with Matrix Factorization. The Matrix Factorization algorithm used is Alternating Least Squared (ALS). After training the model, its performance will be tested using testing data. Then the prediction results will be compared with the actual data from the training data. This comparison will be measured by calculating accuracy using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

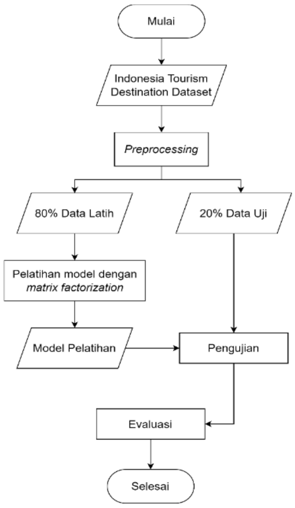

FIG 1. Flowchart of the Recommender System

3. RESULTS AND DISCUSSION

The research experiment was carried out using 300 user data, 320 tourist attraction data, and 8,162 rating interaction data that had been previously cleaned. This research will also be carried out in 5 scenarios. These scenarios have different configurations in the latent factor parameters. Of the five different latent factors ten different iterations will be carried out. Based on research from the journal Matrix Factorization Based Recommendation System using Hybrid Optimization Technique and Movie Recommender System Based on Collaborative Filtering Using Apache Spark, the best regularization parameters were taken at lambda 0.1 and 0.6.

Through the calculation results, the average RMSE value for predictions made on test data with a regularization parameter of 0.1 is 1.700842504. Meanwhile, the average MAE value in predictions made on test data at parameter 0.1 is 1.40940150. The average MAE value for predictions made on test data is 1.276121082. Meanwhile, the average RMSE value for predictions made on test data is 1.476800363. The RMSE value is higher than the MAE value because RMSE calculates the squared average of the absolute difference between the prediction and the actual value, while MAE only calculates the absolute average of the difference. Thus, RMSE gives greater weight to prediction errors. Fig.2 shows the comparison between training data and testing data with some iterations.

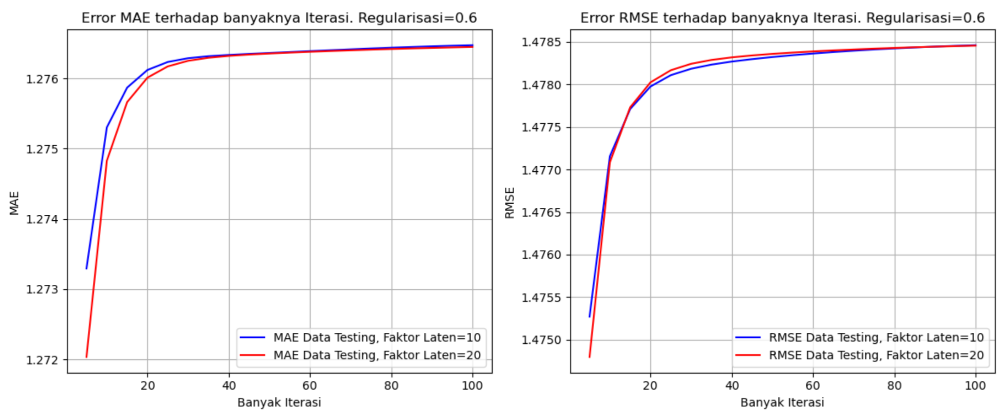

FIG 2. Comparisons of MAE dan RMSE with $\alpha = 0.6$

Based on research data, it can be concluded that the Alternative Least Square (ALS) model for the Indonesia Tourism Destination dataset is working well. This algorithm can work optimally in the 5th iteration. This is because, in the 5th iteration, the model has reached its optimal point, namely when the prediction error is minimal before entering the overfitting phase. In subsequent iterations, the model will become too specific to the training data and start picking up noise in the data. The best regularization parameter is 0.6 with a minimum MAE error value of 1.272 and a minimum RMSE error value of 1.475. This is due to the limited number of datasets resulting in a lack of variation in the data. As a comparison, in the Recommendation Engine journal using the Alternating Least Square (ALS) algorithm on Goodreads which uses 981,756 data, the minimum MAE error is 0.5213, while the RMSE error is 0.672. Therefore, the larger the number of datasets in the model training process, the smaller the prediction error will be.

4. CONCLUSIONS

The conclusions derived from the research and analysis presented in the previous chapter indicate the successful development of a tourist attraction recommendation system for Indonesia using the matrix factorization method and the Alternating Least Squares (ALS) algorithm. This system effectively provides recommendations for tourist attractions, achieving a minimum MAE (Mean Absolute Error) value of 1.2720 and a minimum RMSE (Root Mean Square Error) value of 1.475. These values indicate that the system is quite accurate and reliable in suggesting tourist destinations that match user preferences.

For future research on recommendation systems utilizing the matrix factorization method and the ALS algorithm, the following suggestions are proposed: Firstly, incorporate more varied datasets to enhance the model's performance and its ability to learn from the data more effectively. A more diverse dataset will enable the system to better understand different patterns and preferences, leading to more accurate recommendations. Secondly, consider using the cosine similarity technique to improve the accuracy of the dataset. Cosine similarity can help in measuring the similarity between different items more precisely, thus improving the overall recommendation quality. These recommendations aim to refine the effectiveness and precision of the matrix factorization approach in providing high-quality recommendations for tourist attractions in Indonesia, ultimately contributing to a better user experience and increased satisfaction among travelers

ACKNOWLEDGMENT

I would like to express my deepest gratitude to those who have provided invaluable support throughout the completion of this research. First and foremost, I would like to thank Mrs Novi Yusliani, M.T and Mr. Danny Matthew Saputra, M.Sc., my thesis advisors, for their continuous guidance, insightful feedback, and unwavering support. Their expertise and encouragement have been instrumental in the successful completion of this work.

I am also deeply grateful to my fellow researchers and colleagues for their collaboration, insightful discussions, and technical assistance. Their camaraderie and support have made this journey an enriching experience. A heartfelt thank you to my family, whose love, patience, and understanding have been my foundation. Their unwavering belief in me has been a constant source of motivation. Finally, I would like to acknowledge Sriwijaya University for providing the financial support and resources necessary for this research. Without their assistance, this project would not have been possible. Thank you all for your unwavering support and encouragement.

REFERENSI

[1] G.F. Fernanto, R. Intan, and S. Rostianingsih, "Sistem Rekomendasi Mata Kuliah Pilihan Menggunakan Metode User Based Collaborative Filtering Berbasis Algoritma Adjusted Cosine Similarity," Jurnal Infra, vol.1, no. 7, pp 1-2, 2019.

[2] J. E. Prayogo, et al., "Analisis Perbandingan Model Matrix Factorization dan K-Nearest Neighbor dalam Mesin Rekomendasi Collaborative Berbasis Prediksi Rating". Jurnal Informatika Universitas Pamulang 5(4):506-514.

[3] S. Martin, J. I. Sitohang, and B. Jonathan, "Mesin Rekomendasi Menggunakan Algoritma Alternating Least Square (ALS) pada Goodreads," Jurnal CoreIT, vol. 6, no. 2, pp. 79-84, 2020.

[4] Y. Pratama, et al., "Restricted Boltzmann Machine and Matrix Factorization-Alternating Square Algorithm for Development Tourist Recommendation System," Journal of Physics: Conference Series, 2022.

[5] S. Gosh, et al., "Recommendation System for E-commerce Using Alternating Least Squares (ALS) on Apache Spark," in Intelligent Computing and Optimization 2020 (ICO 2020), Bangladesh, Feb. 2021.

[6] K. B. P. Jati, A. T. Wibowo, and R. Rismala, "Implementasi dan Analisis Online – Updating Regularization Kernel Matrix Factorization Model pada Sistem Rekomendasi," in Proceedings of the Indonesia Symposium On Computing, Aug. 2015.

[7] D. Bokde, S. Girase, and Mukhopadhyay, "Matrix Factorization Model in Collaborative Filtering Algorithms: A Survey," Procedia Computer Science, vol. 49, pp. 136-146, 2015.

[8] Suryanto and Muqtadir, "Penerapan Metode Mean Absolute Error (MAE) dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi," SAINTEKBU: Jurnal Sains dan Teknologi, vol. 1, no. 11, pp. 78-83, 2019.

[9] A. Permata Yakup, "Pengaruh Sektor Pariwisata Terhadap Pertumbuhan Ekonomi di Indonesia," Skripsi Program Magister Ilmu Ekonomi, Fakultas Ekonomi dan Bisnis, Universitas Surabaya, 2019 (unpublished).

[10] Perwitasari, R., Afwani, R., & Anjarwani, S. E. (2020). Penerapan Metode Rational Unified Process (RUP) dalam Pengembangan Sistem Informasi Medical Check Up pada Citra Medical Centre. Jurnal Teknologi Informasi, Komputer, dan Aplikasinya (JTIKA), 1(2), 77.