

Recommendation System for Prospective Bride and Groom Using Cosine Similarity Algorithm

Apriliya Kurnianti*, Pascal Pahlevi, Inayah Mufidah

Universitas Muhammadiyah Yogyakarta, Jl. Brawijaya, , Tamantirto, Kasihan, Bantul,
Daerah Istimewa Yogyakarta 55183

*Corresponding author: apriliana@ft.umy.ac.id

Abstract

Ta'aruf is the beginning of a relationship between two individuals, and it is anticipated to endure until the pair is married. A group called Forum Ta'aruf Padi Melati, run by the regional leadership of Muhammadiyah Bantul, strives to make it easier for cadres Nasyi'atul Aisyiyah and Muhammadiyah Youth to find a life spouse. The current Forum ta'aruf Padi Melati continues to utilize WhatsApp to communicate with forum leaders and allows applicants to share their bio and partner-search criteria. The goal of this project was to design a site model that would assist the Ta'aruf Padi Melati Gathering leaders in performing coordination Likely Matching with the use of the cosine proximity calculation. To compare how well each candidate fits together, the Cosine Similarity method is used. The exploration process is completed in stages, starting with a meeting with the leaders of the jasmine rice ta'aruf discussion. Next, the Cosine comparability strategy is used to characterize the classification matches into bunches according to the degree of rule similarity. The result is the extent to which the pair criteria each participant predicted match.

Keywords: *Cosine similarity; Ta'aruf; Website*

1. Introduction

The presence of the internet provides an alternative for every individual to find a life partner (mate) without having to meet face to face. Ta'aruf is a form of interpersonal communication between men and women who want to get to know each other before heading to marriage [1].

The presence of various platforms unwittingly has a major influence on social changes in society, one of which is in the ta'aruf process. In this modern era, technology is increasingly developing and making it easier for humans to work and meet their needs. As is the case in the matter of finding a soul mate or partner which is usually implemented in digital form [2].

A recommendation System is a system designed to predict an item that matches the user's interests, where the item will be recommended to the user. The application of recommendations in a system usually predicts an item, such as recommendations for movies, music, books, news, and so on that attracts users. This system runs by collecting data from users either directly or indirectly.

In archival, the database will be full of large amounts of archived document data records. The amount of data will cause the data in the database to have similarities between documents from one. This can be solved by applying data mining, the concept of data mining in searching for documents using cosine similarity [3].

Cosine Similarity is a method for measuring the degree of similarity between two vectors. The calculation in this method is carried out by calculating the cosine value between two vectors [4]. Each vector represents each word in each document (text) that is compared and forms a triangle so that cosine law can be applied stating that.

$$\text{Cos: } \cos \alpha = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

The description of equation [1] :

- A : Vector A, which will be compared the similarity
- B : Vector B, to which the similarity will be compared
- A.B : Dot product between vector A and vector B
- |A| : Vector length A
- |B| : Vector length B

The classification method used in this system is by comparing the similarity or similarity between the criteria of prospective pairs with the first keyword, then how to compare the similarity or similarity between the criteria of prospective pairs with the second keyword, and so on until the eighth keyword. Then researchers the highest number of similarities between the eight keywords. (Wade et al., n.d.).

The similarity value of the two vectors is said to be similar when the value of the Cosine Similarity is 1. Cosine similarity is used in positive spaces, where the result is limited between the values of 0 and 1. If the value is 0 then the document is said to be similar if the result is 1 then the value is said to be not similar [5].

In carrying out a design, it must be done in accordance with the stages of the application to be built. With the creation of weighting values, it will be easier to do applications that have been designed with predetermined methods [7].

2. Method

Cosine similarity is a calculation that expresses how similar two or more vectors are. It is the cosine of the angle between two vectors, usually non-zero and located inside a product space [8]. A cosine similarity is a value confined to the range between 0 and 1. The two vectors are orthogonal or perpendicular to one another if the value is near to zero. The angle is less and the pictures are more comparable when the value is nearer to one.

3. Results and Discussion

This research will discuss the application of the Cosine Similarity method to recommend potential companions. For this reason, a calculation of the value of who is a worthy companion will be recommended or chosen to become a potential life partner later.

In carrying out a design, it must be done in accordance with the stages of the application to be built. With the creation of weighting values, it will make it easier to create applications that have been designed with predetermined methods [9].

3.1. Application of the Cosine Similarity calculation method

The process of determining potential pairs is carried out by matching the queries of potential pairs (desired criteria). Before calculating the similarity value between the

queries of potential pairs, the preprocessing stage is carried out first. Preprocessing stages include:

1. Case folding: It is a step to convert all letters in the document to lowercase letters and characteristics other than letters will be eliminated
2. Tokenizing/parsing: This is the stage of breaking sentences into single words or phrases (Parsing)
3. Stop word removal: It is the stage of taking important words from the results of the tokenizing or parsing stage. The stop word removal process is done by removing the stop list or stop word. Stopword can be a contact word, a pronoun, a preposition, or something else.
4. Stemming: This is the stage of transformation of a word into its root word. Here are some examples of calculations using the cosine similarity method. For example, the query inputted by men is:
 - a. Expected Age of Prospective Spouse: Younger
 - b. Criteria for prospective spouses: Faithful, Faithful, Diligent, Loyal, Patient, Responsible, Honest, Friendly, Outgoing, Considerate, Romantic, Educated, Earning, and Sincere.

The query data above will be compared with the data of potential companions. In this case, examples of data on prospective partners who will be sought for matches are as many as 5 women, namely:

1. First woman
 - a. Expected Age of Prospective Spouse: Older
 - b. Criteria for prospective spouses: Faith, Faith, Diligent, Loyal, Patient, Responsible, Honest, Simple, Friendly, Considerate, Educated, Earning, Sincere
2. Second lady
 - a. Expected Age of Prospective Spouse: Older
 - b. Criteria for prospective spouses: Faith, Faith, Diligent, Loyal, Patient, Responsible, Honest, Simple, Friendly, Considerate, Romantic, Educated, Earning, Obedient, and Sincere, can demand the right path.
3. Third woman
 - a. Expected Age of Prospective Spouse: Older
 - b. Criteria for prospective spouses: Faithful, Faithful, Loyal, Patient, Responsible, Honest, Considerate, Educated, Earning
4. Fourth woman
 - a. Expected Age of Prospective Spouse: Older
 - b. Criteria for prospective spouses: Faithful, Faithful, Diligent, Loyal, Patient, Responsible, Honest, Simple, Considerate, Indifferent, Educated, Earning, Submissive
5. Fifth woman
 - a. Expected Age of Prospective Spouse: Older

- b. Criteria for prospective spouses: Faithful, Faithful, Diligent, Loyal, Patient, Responsible, Honest, Simple, Considerate, Indifferent, Educated, Earning, Submissive
6. Fifth woman
- a. Expected Age of Prospective Spouse: Regardless of Age
 - b. Criteria for prospective spouses: Faithful, Faithful, Diligent, Loyal, Patient, Responsible, Outgoing, Considerate, Romantic, and Submissive.

The next calculation step is to make calculations by comparing the age data of prospective partners expected by the male party with the prospective spouses of the female party, and the criteria for potential spouses from the parties men will be compared to the criteria for potential partners on the Women's side.

For example:

The male side data is A

Women's side data is B 1,B2,B3,B4,B5,B6

Cosine similarity (CS) = (A.B)/(|| A|| || B||)

Sample data 1 :

- Example of how to calculate the dot product between A and B1
 - = 1.1 + 1.1 + 1.1 + 1.1 + 1.1 + 1.1 + 1.1 + 0.1 + 1.1 + 0.0 + 1.0 + 1.1 + 1.0 + 0.0 + 1.1 1.1 + 0.0 + 1.1 = 12

The result of the comparison calculation between A and B1, B2, B3, B4, B5, and B6 is seen in table 1.

Table 1 Calculation results A and B

No	Kata kunci	B1	B2	B3	B4	B5	B6
1	Beriman	1	1	1	1	1	1
2	seiman	1	1	1	1	1	1
3	rajin	1	1	0	0	1	1
4	setia	1	1	1	1	1	1
5	sabar	1	1	1	0	1	1
6	bertanggungjawab	1	1	1	1	1	1
7	jujur	1	1	1	1	1	0
8	seederhana	0	0	0	0	0	0
9	ramah	1	1	0	0	0	0
10	tertutup	0	0	0	0	0	0
11	supel	0	0	0	0	0	1
12	perhatian	1	1	1	1	1	1
13	romantis	0	1	0	0	0	1
14	cuek	0	0	0	0	0	0
15	berpendidikan	1	1	1	1	1	0
16	berpenghasilan	1	1	1	1	1	0
17	penurut	0	0	0	0	0	0
18	ikhlas	1	1	0	0	0	0
	Total	12	13	9	8	10	9

- Calculation of the magnitude of vector A :
 - = $\sqrt{12 + 12 + 12 + 12 + 12 + 12 + 12 + 02 + 12 + 02 + 12 + 12 + 12 + 02 + 12 + 12 + 02 + 12} = 3.741657$
- Calculation of the magnitude of the vector B1 :

$$= \sqrt{12 + 12 + 12 + 12 + 12 + 12 + 12 + 12 + 12 + 02 + 02 + 12 + 02 + 02 + 12 + 12 + 02 + 12} = 3.605551$$

The overall magnitude result of vectors B1 – B6 is shown in Table 2.

Table 2 Total magnitude calculation results of vector B

No	Kata kunci	B1	B2	B3	B4	B5	B6
1	Beriman	1	1	1	1	1	1
2	seiman	1	1	1	1	1	1
3	rajin	1	1	0	0	1	1
4	setia	1	1	1	1	1	1
5	sabar	1	1	1	0	1	1
6	bertanggungjawab	1	1	1	1	1	1
7	jujur	1	1	1	1	1	1
8	seederhana	1	1	0	0	1	1
9	ramah	1	1	0	0	0	1
10	tertutup	0	0	0	0	0	0
11	supel	0	0	0	0	0	0
12	perhatian	1	1	1	1	1	1
13	romantis	0	1	0	0	0	1
14	cuek	0	0	0	0	1	0
15	berpendidikan	1	1	1	1	1	1
16	berpenghasilan	1	1	1	1	1	1
17	penurut	0	1	0	0	1	1
18	ikhlas	1	1	0	0	0	0
TOTAL		3.605551	3.872983	3	2.828427	3.605551	3.741657

- Results of cosine similarity calculation

$$= (12)/(3.741657 * 3.605551) = 0.889499 \text{ (89\%)}$$

The result of the recommendation of a potential partner is shown in Table 1.

Table 3 Similarity level calculation results

Hasil Similarity (Calon Pasangan A)	B1	B2	B3	B4	B5	B6
	0.889499	0.897085	0.801784	0.755928946	0.741249	0.642857
	89%	90%	80%	76%	74%	64%

Based on the results of testing the five women to be matched, the similarity of the criteria for potential partners is male A and Female B2. Results should be clear and concise. Show only the most significant or main findings of the research. Discussion must explore the significance of the results of the work. Adequate discussion or comparison of the current results to the previously published articles should be provided to show the positioning of the present research (if available)[10].

3.2. Implementation

The implementation of the recommendation system for prospective life partners is designed using the Vs Code application with PHP programming language using the MySQL database. The start page display will show a glimpse of the jasmine rice ta'aruf forum as in figure 1.



Figure 1 of the application start page

The main menu in Figure 1 displays main menu options such as Home, Registration, About Us, Download, and Profile. Next is the login page as shown in Figure 2.



Figure 2 Login Page

Application users are required to sign in first before they can use the application. Login is divided into two users, namely as a manager and a user as in Figure 2. Next is a page to display the match rate of pairs that only admins see as shown in Figure 3.

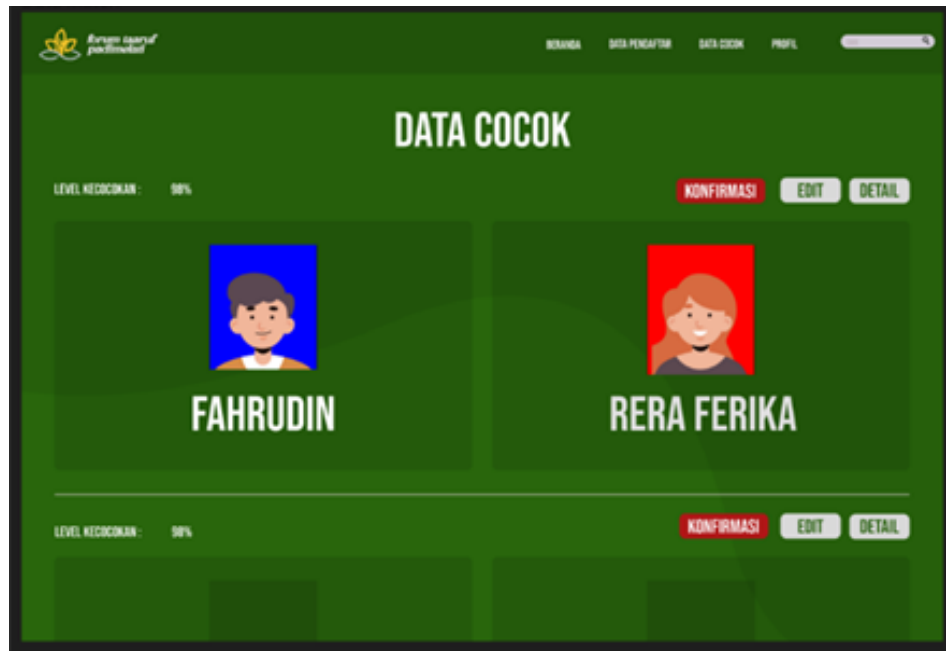


Figure 3 Data matching page

The matched data page is visible to admins only. Because the concept of this application is Islamic matchmaking, participants of this ta'aruf application can only register. Where after the application makes a matching classification, then the participant will be contacted by the admin to proceed to the next ta'aruf process.

4. Conclusion

The study's conclusion is that Islamic matching may be done by the brothers and sisters with the use of a search suggestion system called Syari'ah Mate. There is ta'aruf with using a web-based application using the cosine algorithm similarity, which will help to obtain results recommendations for the suitability of potential spouses from sisters and Ikhwan more precisely, and based functional testing using a black box to the matchmaking recommendation system syari'ah, valid results are obtained.

References

- [1] Elysia, E., Chatra, E., & Arif, E. (n.d.). "Transformation Of The Meaning Of Ta'aruf In The Digital Era". (In Indonesian) *Jurnal Komunikasi Global*, 10(1), 2021.
- [2] Ummareng, C., Nurhaeni, dan F., *Studi Sistem Informasi STMIK Dipanegara Makassar Jl Perintis Kemerdekaan Km, P., & Studi Manajemen Informatika STMIK Dipanegara Makassar Jl Perintis kemerdekaan Km, P.* (2018). Recommendation system to find a sharia doctor using android-based cosine similarity algorithm (Vol. 7, Issue 2).
- [3] Kurniadi, D., Farisa, S., Haviana, C., & Novianto, A. (2020). "Implementation of the Cosine Similarity Algorithm in the document archive system at Sultan Agung Islamic University". (In Indonesian) *TRANSFORMATIKA*, 17(2), 124–132
- [4] Saptono, R., Prasetyo, H., & Irawan, A. (n.d.). "Combination of Cosine Similarity Method and Conditional Probability for Plagiarism Detection in the Thesis Documents Vector Space Model". (In Indonesian) <http://digilib.uns.ac.id>.
- [5] Vendyansyah, N., & Pranoto, Y. A. (n.d.). "Designing and Making Applications to Detect Similarity of Answers Using the Cosine Similarity Method". (In Indonesian)

- [6] Wahyuni, R. T., Prastiyanto, D., & Suprpto, D. E. (n.d.). “Application of Cosine Similarity Algorithm and TF-IDF Weighting in Thesis Document Classification System”. (In Indonesian)
- [7] Azhari, A., Buulolo, E., & Silalahi, N. (2022). “Text Rank-Based Thesis Lecturer Recommendation System using the Cosine Similarity Method”. (In Indonesian) *Pelita Informatika: Informasi Dan Informatika*, 10(3).
- [8] R. Saptono, H. Prasetyo, and A. Irawan, “Combination of Cosine Similarity Method and Conditional Probability for Plagiarism Detection in the Thesis Documents Vector Space Model,” *J. Telecommun. Electron. Comput. Eng. JTEC*, vol. 10, no. 2–4, pp. 139- 143–143, Jul. 2018.
- [9] S. D. Priyani, E. Ripmiatin, and S. Arifin, “Implementation of Cosine Similarity and Time Interval Entropy Method to Identify Bot Spammer Account on Twitter,” *ITSMART J. Teknol. Dan Inf.*, vol. 6, no. 2, pp. 51–57, Feb. 2018.
- [10] A. Oktavianus and S. Hansun, “Job Vacancies Recommendation System With Glassdoor Api And Simple Additive Weighting Method,” *Tek. Dan Ilmu Komput.*, vol. 7, no. 26, Mar. 2018.