Research Supervisor Recommendation System Based on Topic Conformity

Ridwan Rismanto
Departemen of Information Technologi, State Polytechnic of Malang , Malang, Indonesia
Email: rismanto@polinema.ac.id

Arie Rachmad Syulistyo
Departemen of Information Technologi, State Polytechnic of Malang , Malang, Indonesia
Email: arie.rachmad.s@gmail.com

Bebby Pramudya Citra Agusta
Departemen of Information Technologi, State Polytechnic of Malang , Malang, Indonesia
Email: pramudyabebby1@gmail.com

Abstract—In a higher education such as universities, final project are under supervision of one or more supervisors with a similar research interest or topic. The determination of the final project supervisor is an important factor in the work of the student's final project. However, the lack of information about the supervisor can hamper students in making the determination of the supervisor. Thus, a system is needed that can facilitate students in determining the final project or thesis advisors in accordance with the research topic. This problem is the basis of this research. The study is conducted by developing a web-based system and applying the TF-IDF word weighting and cosine similarity method. TF-IDF method is a way to give the weight of the relationship of a word to the document. The cosine similarity is a method for calculating the similarity between two objects expressed in two vectors by using keywords from a document as a measure. The results of the advisor recommendation system can provide recommendations to students regarding the final assignment advisor who has conducted research in accordance with the topic of the student's final assignment written in Indonesian. In 20 testing, the accuracy of the comparison of the results of the system recommendations with the actual data obtained an average of 75% by comparing system recommendation with actual assigned supervisor.

Index Terms—Higher Education, Recommendation System, TF-IDF Weighting, Cosine Similarity, Research Supervisor

I. INTRODUCTION

In selecting research topic, students need some literature and also references as research material to be made. Unlike previous years, 2018/2019 academic year, the Department of Information Technology of the State Polytechnic of Malang has a research group to determine which student research groups will be taken. The research groups include Intelligent System; Computer Network, Architecture and Data Security; Computer Vision; Information System; Multimedia and Games. Each research group has members consisting of several lecturers who can be used as supervisors for the final project. The difference in the focus of the supervisor's research is what often becomes the consideration of students in the selection of supervisors in accordance with the research topic.

Students will conduct guidance with the final project supervisor during the final assignment process. The final project supervisor has an important role because he has the responsibility to ensure that students are able to arrange the final project well until the final project is ready to be tested and qualified. Therefore, the appropriateness of student research topics with the focus of lecturer research is an important factor because it will influence the guidance that will be carried out by students during the final assignment process.

Students need to choose supervisor that is linear with their research topic. To find the right supervisor is sometimes a problem for the student because of the different research area and interest for each prospective supervisor.

The purpose of this study is to create a system that can provide supervisory lecturer recommendations in accordance with student research topics. Student research topics will be sought in common with the corresponding lecturer research journals automatically. The application of text mining can be used to analyze data obtained from lecturer research journals and queries from students. So that it is expected that the cosine similarity method can facilitate students in determining supervisors in accordance with the research topic.

The research supervisor recommendation system will extract required features which are the title and abstract

of recent researches of prospective supervisors. To increase the accuracy of the recommendation system, the list of prospective supervisors will be filtered first based on the research topic.

II. LITERATURE REVIEW

In [1,2,3] cosine similarity was used to perform semantic, granularity-based assesment of similarity between short text string and hybrid recommendation system based on content and collaborative filtering. Considering the research body which cointains title and abstract, it can be used as a base to perform recommendation for research supervisor to the student based on their title and abstract similarity. The cosine similarity algorithm was also used to provide recommendations for movie reviews and sightseeing location reviews written in Japan [4]. It shows that by using this methodology, candidates can be recommended without having to manually apply natural language processing, such as preparing stopwords for each category. A content-based recommender system also incorporate cosine similarity methodology to provide journal recommendation system and gained good result for subjects with higher density of technical vocabulary and shorter corpus documents [5]. However, in that research, only the abstract of the document is considered. The algorithm also used in Hirroe Wijaya Ani Kesuma and Feddy Setio Pribadi in 2016 to conduct a research entitled Application of "Cosine Similarity in the Application of the Commercial Law". The research carried out resulted in a system that could search for indexes in the Criminal Code [6]. The cosine similarity algorithm was also used in a study by Ade Riyani, Muhammad Zidni Naf'an, and Auliya Burhanudin in 2019 entitled "Application of Cosine Similarity and Weighting of TF-IDF to detect document similarities". This research produces a system that can detect the similarity of text and its plagiarism value [7].

A recommender engine has developed to match final-year project student with supervisor [8], this system incorporates questionairre to construct the corpus data. This is not feasible for a condition where the corpus data is dynamic. In [9], a multi-criteria decision support system has developed using TF-IDF and Okapi BM52. Since this is a multi-criteria decision support system, a weighting of each crieria is needed and performed using quissionaire, resulting in too many steps needed to execute before actual recommendation system can be performed. TF-IDF also used for several recommender system [10,11,12,13,14]. One research also propose a recommendation system for academic collaboration in undergraduate research [15] based on the undergraduates' profiles and incorporates rule-based filtering techniques. A decision support system by applying the simple additive weighting method proposed by Pristiwanto [16] for the determination of the final assignment supervisor also provides similar aim. However, this research has a weakness because it only covers 2 areas of expertise,

namely computers and non-computers, so the simple additive weighting method is not appropriate for the research to be conducted which can handle many areas of expertise.

III. TEXT MINING

Text mining is one of the special fields of data mining. Text mining can be defined as a process of extracting information where a user interacts with a group of documents using an analysis tool which is a component of data mining [17]. Text mining is used to process documents before the similarity process is carried out. Text mining is the process of extracting patterns (useful information and knowledge) from a large number of unstructured data sources. Text mining has a purpose and uses the same process as data mining, but has different inputs. Input for text mining is data that is not (or lacking) structured, such as Word documents, PDFs, quoted text, etc., while input for data mining is structured data

Cosine similarity is used in positive space, where results are bounded between values 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. Note that this limit applies to a number of dimensions, and Cosine similarity is most often used in high dimensional positive space [18] Text mining can provide solutions to problems such as processing, organizing / grouping and analyzing large amounts of unstructured text [19,20].

A. Prepocessing Data

Before the dataset is entered into the proposed model, preprocessing data is first performed. This process is needed to prepare the data to be able to further processed by the algorithm and to increase accuracy by minimizing bias and noise caused by non basic words, unimportant terms and unmatching font cases. At this stage several things were done, including case folding, tokenizing, filtering and stemming [17].

1) Case Folding

The stage of changing all letters in a document into lowercase letters and characters other than letters will be eliminated[21]. Table 1 shows the case folding example.

Table 1. Case Folding Example

Before	After
Implementasi Metode	implementasi metode
PENERAPAN	Penerapan

2) Tokenizing

Tokenizing is a process carried out by someone to make a sentence more meaningful or in a way by breaking the sentence into words or phrases ("Parsing") [22] as can be seen in Table 2.

Table 2. Tokenizing Example

Before:	Penentuan Dosen Pembimbing		
After:	penentuan	dosen	pembimbing

3) Filtering

At this stage, taking important words from the tokenizing results by removing stopwords (words that are less important). Stopwords can be either subject or conjunctions. Filtering is done by determining which terms will be used to represent the document so that it can describe the contents of the document and distinguish the document. The example of filtering as shown in Table 3.

Table 3. Filtering Example

Before:	hari yang indah
After:	hari indah

4) Stemming

Is the process of changing the form of words into basic words. This method is the process of changing the form of words into basic words that adjust to the structure used in the stemming process. Table 4 shows the example of stemming process.

Table 4. Stemming Example

Before	After
penentuan	tentu
menggambarkan	gambar
pembimbing	bimbing

B. Weighting tf-idf

The TF-IDF method is a method for calculating the weight of each word that is most commonly used in information retrieval. We utilized TF-IDF to find the most influental words in the student research title and abstract text, which is later will be matched with prospective research title and abstracts database by utilizing cosine similarity method.

This method is also known to be efficient, easy and has accurate results [23]. Word weighting is very influential in determining the similarity between document and query. This method combines two concepts for weight calculation, namely the frequency of occurrence of a word in a particular document and the inverse frequency of the document containing the word. Factors that play an important role in word weighting are [4]:

C. Term Frequency

TF (term frequency) which is the value or number of words / terms in a collection of documents [25]. There are some kinds of frequency term like :

1) Raw tf

The Tf value of a term is calculated based on the appearance of the term in the document.

$$tf = 1 + \log (tf) \qquad (1)$$

2) Logarithmic tf

In obtaining Tf values, this method uses logarithmic functions in mathematics.

3) Binnary tf

This method produces a Boolean value based on the appearance of the term in the document. It will be worth 0 if the term is not in a document, and value 1 if the term is in the document. So that the number of terms appearing on the document has no effect.

4) Augmented tf

TF value is the number of occurrences of a term in a document. The max value (Tf) is the most number of occurrences of the term in the same document. Tf calculation that will be used in the implementation of the supervisor lecturer recommendation system in the system that the author builds is Raw Tf

$$tf = 0.5 + 0.5 \text{ x } tf\text{max } (tf) \qquad (2)$$

D. Invers Document Frequency

Inverse Document Frequency (IDF) values for each token (word) in each document in the corpus.

Idf is calculated using the following formula [24]:

$$idf = \log\frac{D}{df} \qquad (3)$$

Where :

idf = Inverse document frequency
D = Total Documents
df = Frequency of documents from term
log = To minimize the effect relative to tf

$$W = tf \text{ } x \text{ } idf \qquad (4)$$

The term weight is calculated using the formula:
Where :

W = Weight of document
tf = Frequency term
idf = Inverse document frequency

E. Cosine Similarity Method

Cosine similarity method is a method used to calculate similarity (level of similarity) between two objects. In the vector space model algorithm the formula is used to find the weight of the document and the weight of the keywords. Calculation of the cosine value between the two vectors is known as the Cosine Similarity method. The formula used in cosine similarity is as follows [1].

$$im\,(q, d_j) = \frac{q, d_j}{|q|x|d_j|} = \frac{\sum_{i=1}^{t} w_{iq} x\, w_{ij}}{\sqrt{\sum_{i=1}^{t}(w_{iq})^{2}} x \sqrt{\sum_{i=1}^{t}(w_{ij})^{2}}} \qquad (5)$$

Where :
q = Vector query, which will be compared similarity
d = Vector document j, which will be compared similarity
| q | = Length of the query vector
| d | = Document vector length j
Wiq = Weight of the word i in the query
Wij = Weight of the word i in document j

IV. RESEARCH METHODOLOGY

In conducting a research, it takes a reference to the implementation called the research methodology. The research methodology consists of several stages of a structured / systematic research framework starting from the initial stages of research to obtaining the desired results. Fig. 1 describes the research methodology.

1. Idntification Of Problem

This stage is the first step to determine the formulation of the problem from research. In this case observing problems related to the factors that cause difficulty in determining the concentration of the student's thesis topic. The existing problems, then analyzed to find out how to solve these problems and determine the scope of the problem to be examined.

2. Study Literature

This stage is the stage to support the writing and development of the system to be made. This is a reference that is relevant to the same or relevant case. The author looks for references from journals, the internet, and books.

3. Data Set Collection

The method of taking data that the author uses is the document study method. Document studies are data collection methods that are not addressed directly to research subjects. Document study is a type of data collection that examines various kinds of documents that are useful for analysis material. The data used are Polinema informatics journal data from the website http://jip.polinema.ac.id and data from the Information Technology Polynema lecturer.

4. System Requairement Analysis

a. System Requirements
• The system can display lecturer recommendations
• The system can display lecturer data
• The system can display lecturer research data
• The system can run the cosine similarity method to calculate the similarity between documents.
b. Hardware Requirements
The minimum specifications required to run the software above are as follows:
• Processor: Intel (R) Core (TM) i3-6200 CPU @

2.30 GHz - 2.40 Ghz
• RAM: 4GB DDR4
• Minimum HDD: 200GB
c. Software Requirements
The software requirements for designing and running applications include the following:
• Operating System: Windows 10
• System companion software: Mozilla Firefox browser
• Software for local server: XAMPP
• Database: MySql

5. System Design

In this stage, a number of designs are needed in the Development of Information Systems Suitability for Research Topics with Lecturer Research Focus including:

• Design an overview of the features contained in the application using Use Case as can be seen in Fig. 2 and flowchart in Fig. 3.

There are 2 actors in the supervisor's recommendation system, namely admin and students. The first actor is the admin, explained in the picture that the admin can manage lecturer data such as lecturer data input, delete lecturer data, or edit lecturer data. Admin can also manage lecturer research data such as lecturer research data input, delete, or edit lecturer research data. Only the name of the lecturer in the lecturer data can be inputted into the lecturer research data. But the admin can only manage lecturer data and lecturer research data if it has logged in.
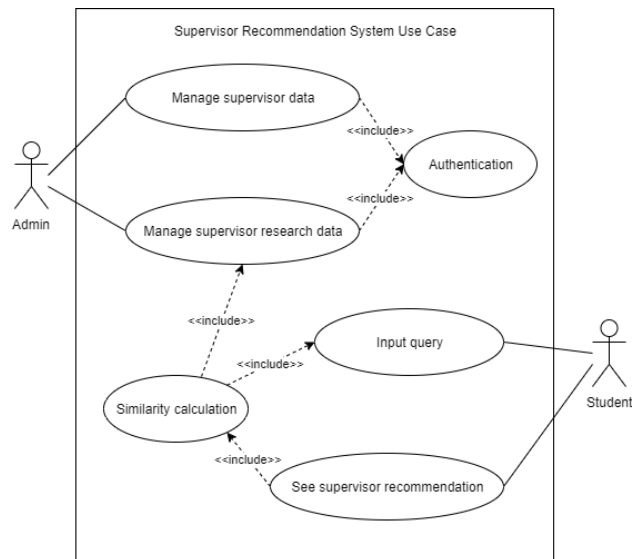

Fig. 2. Use Case

The second actor is a student. Students do not log in to enter the system. Students must enter the title, abstract, and topic that are the topic of their research. After the student enters the title, the student can see and know the lecturer who matches the research topic. The data that is displayed as the final result is the name of the lecturer and the focus of the lecturer's research.

- Plan how the process will work using the Activity Diagram.
- Design an interface description of an application using mockup.
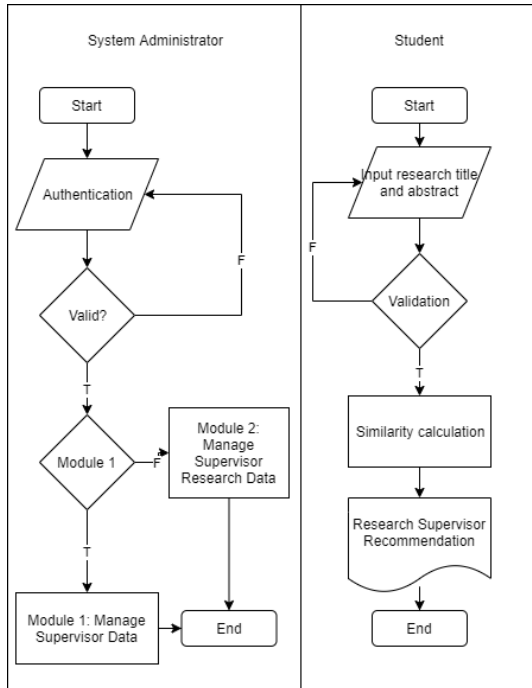
6. Implementation

Fig. 3. Flowchart

System implementation phase refers to the application design. Implementation of the system is done using the PHP language. The database used is MySql. Implementation of the system includes:

• Making the application interface in the form of a website.
• Database creation and data entry - required data such as lecturer data and lecturer research data.
• Application of the cosine similarity method for calculating document similarity.

7. Testing

Stages of testing on the system that has been made on the problems found, Testing used on the website that is using functional testing that is testing blackbox and testing accuracy

8. Conclussion

The conclusion stage will be based on the results of the system testing and also includes constructive suggestions for future system development.

V. SIMULATION AND EXAMPLE

The following is an example of a calculation using the Cosine Similarity method. The query entered by the

student is as seen in Fig. 4, and the supervisor data in Fig. 5, 6 and 7:

Query

Title: Sistem Penyiraman Tanaman Sayur Secara Aeroponik Berdasarkan Suhu Dan Kelembapan Berbasis Iot Menggunakan Metode Fuzzy

Abstract: Dalam meningkatkan produksi sayuran para petani menggunakan teknik hidroponik dalam mengatasi penggunaan pestisida yang berlebihan, akan tetapi teknik pengontrolan penanaman hidroponik yang digunakan masih menggunakan konvensional. Salah satu teknik penanaman hidroponik yang membutuhkan pengontrolan penyiraman secara tepat yaitu teknik aeroponik, dimana air nutrisi disemprotkan langsung ke akar secara berkala disesuaikan dengan kondisi lingkungan tanaman. Sistem pengontrolan penyiraman konvensional ini dapat digantikan dengan sebuah sistem yang otomatis, dimana dalam pengontrolan penyiramannya dapat digantikan dengan Raspberry Pi yang dihubungkan dengan sensor suhu dan kelembapan untuk melihat kondisi lingkungan tanaman dan pompa hidroponik untuk mengalirkan air ke akar tanaman. Pengambilan keputusan kontrol durasi penyiraman berdasarkan kondisi lingkungan tanaman yaitu menggunakan Metode Fuzzy Mamdani. Dalam penelitian ini diperoleh beberapa faktor yang mempengaruhi durasi penyiraman yaitu, pada suhu tertinggi 33°C dengan kelembapan 44% mendapati durasi 11 detik, pada suhu terendah 25°C dengan kelembapan 69%-70% mendapati durasi 8 detik. Penggunaan Metode Fuzzy Mamdani telah dapat menentukan durasi penyiraman berdasarkan kondisi lingkungan tanaman.
Topics: Computer Network, Architecture, and Data Security

Fig. 4. Query

Supervisior Data 1

Title: Sistem Informasi Logistik Stock Opname Menggunakan Quick Response Code Berbasis Android

Abstract: Sistem Informasi Logistik Stock Opname memberikan informasi mengenai barang yang terdapat di gudang antara lain jumlah tiap barang dan detail barang berupa merk dan nomor seri barang. Informasi tersebut ditampilkan secara mobile melalui aplikasi yang berbasis sistem operasi Android. Bagian utama proses logistik yang berpengaruh besar pada perusahaan yang menerapkannya adalah stock opname. Stock opname adalah proses menghitung jumlah barang yang ada di gudang dan mencocokkannya dengan catatan pembukuan persediaan. Proses tersebut dipermudah dengan menggunakan QR (Quick Response) Code yang diaplikasikan kedalam sistem operasi Android sehingga proses stock opname tidak lagi dilakukan secara manual. Dalam system ini juga terdapat tampilan web untuk admin yang berguna untuk menampilkan dan mengolah semua data barang termasuk hasil stock opname yang sudah dilakukan. Selain itu, juga disediakan fitur search dan filter data untuk memudahkan pencarian barang dan fitur untuk mencetak laporan dari data tersebut. Terdapat fitur transaksi masuk dan transaksi keluar sebagai proses untuk melakukan pembelian barang dari supplier dan penjualan barang kepada bengkel yang membutuhkan.
Topics: Information Systems, Computer Networks, Architecture, and Data Security

Fig. 5. Supervisor Data 1

Supervisior Data 2

Title: Implementasi Iot Pada Sistem Monitoring Dan Pengendali Sirkulasi Air Tanaman Hidroponik

Abstract: Hidroponik merupakan budidaya menanam dengan memanfaatkan air tanpa menggunakan tanah dengan menekankan pada pemenuhan kebutuhan nutrisi bagi tanaman. Penerapan hidroponik lebih efisien di daerah yang memiliki ruang hijau terbatas. Salah satu jenis hidroponik adalah Deep Flow Technic (DFT) merupakan jenis hidroponik yang menerapkan aliran nutrisi secara kontinyu dan terdapat genangan setangah dari diameter pipa yang menggenangi akar tanaman. Namun, seringkali para penggiat hidroponik sistem DFT ini mengalami kegagalan selama proses pertumbuhan tanaman, dikarenakan kurangnya penjagaan terhadap unsur tumbuh tanaman seperti sirkulasi air, intensitas cahaya, suhu, kelembaban dan pH air yang menyebabkan tanaman tersebut tidak tumbuh optimal. Maka diperlukan sistem monitoring dan pengendali sirkulasi air pada hidroponik DFT berbasis IoT untuk mengantisipasi terjadinya perubahan pada unsur tumbuh tanaman. Tanaman yang digunakan adalah sawi daging (pakchoy). Data unsur tumbuh tanaman diakusisi oleh sensor yang terintegrasi dengan Raspberry Pi. Pada proses monitoring menggunakan website akan menampikan data unsur tumbuh tanaman berupa pH, suhu, kelembaban dan ketinggian air pada tandon hidroponik. Suhu dan kelembaban digunakan sebagai parameter pengendali sirkulasi air yang diproses menggunakan Metode Fuzzy Sugeno untuk menyalakan atau mematikan pompa pada hidroponik DFT. Dari hasil pengujian yang telah dilakukan, sistem dapat melakukan monitoring unsur tumbuh tanaman yang ditampilkan pada website secara realtime serta pengendalian sirkulasi air secara otomatis. Sistem yang diterapkan dalam hidroponik tanaman sawi daging ini pun menghasilkan pertumbuhan jumlah daun dan tinggi tanaman yang signifikan.

Topics: Computer Network, Architecture, and Data Security

Fig. 6. Supervisor Data 2

Supervisior Data 3

Title: Pengembangan Game Marine Ekosistem Hero Menggunakan Fuzzy Logic

Abstract: Game merupakan hiburan yang biasanya diangkat dari masalah dalam kehidupan sehari-hari, salah satunya adalah masalah ekosistem laut. Pada saat ini kebanyakan kondisi ekosistem laut sangat memprihatinkan dikarenakan kurangnya kepedulian masyarakat akan kebersihan ekosistem laut. Untuk menyampaikan pesan agar masyarakat peduli pada lingkungan laut diperlukan media yang interaktif, maka dari itu dibuat sebuah Game yang bernamakan Marine Ecosystem Hero. Game ini memberikan misi agar player dapat mengumpulkan sampah sebanyak mungkin untuk mendapatkan poin sebelum waktu habis. Poin dan sisa waktu digunakan untuk menghitung total skor. Oleh karena itu, dibutuhkan sebuah metode yaitu Fuzzy Logic untuk menghitung skor akhir player sehingga dapat ditentukan apakah player dapat lolos ke level selanjutnya atau tidak. Hasil dari penelitian ini menunjukkan bahwa implementasi metode Fuzzy Logic berjalan dengan baik dalam Game Marine Ecosystem Hero. Game ini dinilai untuk memberikan nilai score yang sesuai pada setiap levelnya. Sesuai hasil dari 30 responden game ini memenuhi kepuasan user sebanyak 63%.

Topics: Multimedia and Games, Computer Vision

Fig. 7. Supervisor Data 3

First step is to filter the topics of the query with relevant topics of supervisor data. Therefore, the result of the filtering is as follow:

a. Q = The query entered

b. D1 = Supervisior 1
c. D2 = Supervisior 2

Next step is to calculate the similarity between the title text and abstract text of query data and supervisor data using TF-IDF and cosine similarity. For the sake of simplicity, the calculation showed in Table 5-10 has been cropped. Step 1 and 2 below are based on equation 1-5.

1. Similarity calculation of the title text between query and supervisor data.

Table 5. TF-IDF of Title Text

Term	tf			df	idf	Wdt=tf.idf		
	Q	D1	D2		log(n/df)	Q	D1	D2
sistem	1	1	1	3	0	0	0	0
siram	1	0	0	1	0.47712	0.47712	0	0
tanam	1	0	1	2	0.17609	0.17609	0	0.17609
sayur	1	0	0	1	0.47712	0.47712	0	0
aeroponik	1	0	0	1	0.47712	0.47712	0	0

Table 6. Inner Product

Term	Q-D1	Q-D2
sistem	0	0
siram	0	0
tanam	0	0.03101
sayur	0	0
aeroponik	0	0

Table 7. Term Square

Term	Q	D1	D2
sistem	0	0	0
siram	0.22764469	0	0
tanam	0.03100813	0	0.03101
sayur	0.22764469	0	0
aeroponik	0.22764469	0	0

2. Similarity calculation of the abstract text between query and supervisor data.

Table 8. TF-IDF of Abstract Text

Term	tf			df	idf	Wdt=tf.idf		
	Q	D1	D2		log(n/df)	Q	D1	D2
tingkat	1	0	0	1	0.47712	0.47712	0	0
produk	1	0	0	1	0.47712	0.47712	0	0
sayur	1	0	0	1	0.47712	0.47712	0	0
tani	1	0	0	1	0.47712	0.47712	0	0
guna	1	1	1	3	0	0	0	0

Table 9. Inner Product

Term	Q-D1	Q-D2
tingkat	0	0
produk	0	0
sayur	0	0
petani	0	0
guna	0	0

Table 10. Term Square

Term	Q	D1	D2
tingkat	0.22764	0	0
produk	0.22764	0	0
sayur	0.22764	0	0
petani	0.22764	0	0
guna	0	0	0

The final results of the final assignment supervisor's recommendation are obtained from the sum of the final assignment weights, the final assignment topic weight, and the final assignment abstract. The final results are shown in Table 11.

Table 11. The result of Cosine Similarity

Document	Target	Cosine Similarity Calculation
Q-D1	Title	0.0324026
	Abstract	0.0029901
Q-D2	Title	0.0372116
	Abstract	0.0636921

Based on the data in table 1 the composition of the document for the recommendation of the final project supervisor is calculated by the formula:

composition = (percentage of title x result of cosine title) + (percentage of abstract x result of cosine abstract).

For example the percentage given is:
a. Title composition: 40% or 0.4
b. Abstract composition: 60% or 0.6

So that it can be seen the total score of the supervisor who recommended the system is as follows:

a. D1 = (0.4 x 0.0155684) + (0.6 x 0.1255684) = 0.08157
b. D2 = (0.4 x0.0354623) + (0.6 x 0.7996822) = 0.49399

The conclusion based on the above calculation is known that document D2 (Supervisior 2) has the highest level of similarity with the query (Q) that is 0.493994. The system will rank each document after calculating the similarity between the query and data entry from the composition results. So the system will display the results of the appropriate lecturer recommendations based on rank.

VI. RESULT

1. Web Application Implementation

This research is implemented in a web application to provide user interface and database management system. The database table structure can be seen at Table 6.

Table 12. Database Table Structure

Table	Field	Type
Auth	id_login	Int
	user_login	Varchar
	pass_login	Varchar
Supervisor	id_supervisor	Int
	name_supervisor	Varchar
	research_topic	Varchar
spv_research_list	id_research_list	Int
	id_supervisor	Int
	research_title	Varchar
	research_abstract	Varchar
results	id_results	Int

id_supervisor	Int
research_title	Varchar
research_topic	Varchar
score	Float

The *auth* table is used to store user credential data. The *supervisor* table contains list of prospective supervisor including their research topic. Table *spv_research_list* are recent research list of the prospective supervisor. In this table, previous research title and abstract are stored. The *results* table store the recommendation result calculated by the system.

To provide user interface for the user, form-based web application are developed which can be seen at Fig 8, 9, 10, 11 and 12.

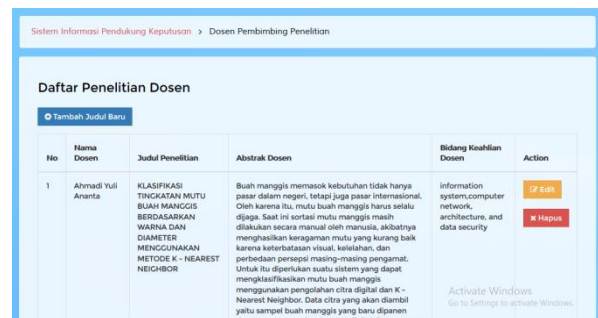Fig. 8. Prospective supervisor list

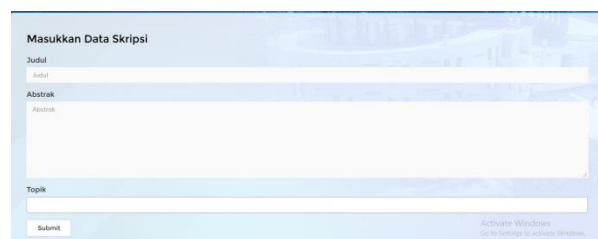Fig. 9. Previous research of prospective supervisor list

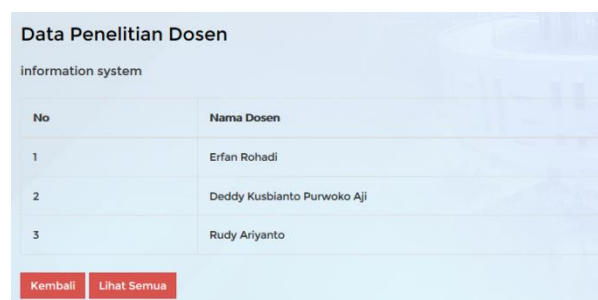Fig. 10. Student title, abstract and topic research input

Fig. 11. Supervisor recommendation result

Fig. 12. Detailed recommendation result

Prospective supervisor list page provide the list of prospective supervisor along with their research topic or interest. *Previous research of prospective supervisor list* page shows the title and abstract of previous researches performed by the prospective supervisor. *Student research input* page is where student input their title, abstract and topic to be processed by the recommendation system. *Supervisor recommendation result* shows the recommended supervisor. *Detailed recommendation result* page shows further recommendation with the score of each supervisor.

2. Accuration Testing

To test the accuracy of the system, the accuracy testing is done by comparing the results recommended by the system with actual data. In testing the accuracy of 20 experiments conducted where each test is tested with a different query. It aims to see the value of accuracy from the comparison of the system recommendation data with the data. The result of accuration testing can be seen in Table 7.

Table 13. Accuration Result

Testing	System	Actual	Result
Q1	S43, S21, S39	S39, S7	True
Q2	S19, S20, S43	S7	False
Q3	S21, S17, S35	S21, S34	True
Q4	S43, S19, S24	S43	True
Q5	S43, S11, S17	S8, S52	False
Q6	S19, S24, S3	S24, S39	True
Q7	S48, S30, S15	S19, S33	False
Q8	S43, S12, S11	S12, S23	True
Q9	S30, S48, S9	S48, S22	True
Q10	S41, S30, S35	S37	False
Q11	S30, S9, S12	S9, S39	True
Q12	S24, S3, S52	S52	True
Q13	S9, S35, S41	S9	True
Q14	S48, S30, S12	S48, S22	True
Q15	S24, S8, S3	S52, S8	True
Q16	S30, S41, S37	S30, S41	True
Q17	S43, S11, S22	S22, S45	True
Q18	S24, S11, S22	S17, S2	False
Q19	S41, S30, S9	S35	True
Q20	S48, S30, S12	S48, S22	True

Where :
Sn = supervisor who has id n

$$Accuration = \frac{15}{5} * 100\% = 75\%$$

VII. CONCLUSION

Simulation in manual calculations using cosine similarity as a method of determining the supervisor shows that the method can be used and can provide the results of the supervisor's recommendations in accordance with the research topic of the query. In its application the system can make it easier for students to find supervisors in accordance with the students' research topics. Of the 20 queries used as test material by comparing the results of recommendations from the system with actual data shows an accuracy rate of 75%

Future work for this research is to automate the classification of the research list topics of prospective research supervisor. One possible method is to incorporate machine learning approach, such as Naïve Bayes classification or Neural Network classification algorithm. The usage of automatic classification can minimize the steps needed in term of adding new prospective supervisor research list data.

REFERENCES

[1] H. Kaur, R. Maini, "Granularity-Based Assessment of Similarity Between Short Text Strings," *Proceedings of the Third International Conference on Microelectronics, Computing and Communication Systems*, 2019.

[2] A. Gupta, P. Kumar, "Efficient Hybrid Recommendation Model Based on Content and Collaborative Filtering Approach," *Emerging Trends in Expert Applications and Security*, Advances in Intelligent Systems and Computing 841, Springer, 2019.

[3] W. Wang, Q. Feng, F. Zhou, Y. Cao, H. LV, "Personalized employment recommendation method based on semantic matching of requirements," *Journal of Physics*: Conf. Series 1213, 2019.

[4] Takano Y. et al. "Improving Document Similarity Calculation Using Cosine-Similarity Graphs," *Advanced Information Networking and Applications*. AINA 2019. Advances in Intelligent Systems and Computing, vol 926. Springer, Cham, 2019.

[5] M. Wijewickrema, V. Petras, and N. Dias, "Selecting a text similarity measure for a content-based recommender system,", The Electronic Library, Vol. 37 No. 3, pp. 506-527. 2019.

[6] Hirroe Wijaya Ani Kesuma, dan Feddy Setio Pribadi, 2016, Penerapan Cosine Similarity dalam Aplikasi Kitab Undang-Undang Hukum Dagang, Vol.8 No.1, Universitas Negeri Semarang. 2016.

[7] Ade Riyani, Muhammad Zidni Naf'an, dan Auliya Burhanudin, 2019, *Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk mendeteksi kemiripan dokumen,*Vol.2 No.1. Maret 2019, Intitut Teknologi Telkom Purwokerto

[8] Ismail, M. H., Razak, T. R., Hashim, M. A., & Ibrahim, A. F. (2019). A Simple Recommender Engine for Matching Final-Year Project Student with Supervisor, 2019.

[9] M.A. Hasan, D.G Schwartz, "A Multi-criteria Decision Support System for Ph.D. Supervisor Selection: A Hybrid Approach," Proceedings of the 52nd Hawaii International Conference on System Sciences, 2019.

[10] Wang, Z., Hahn, K., Kim, Y. et al. "A news-topic recommender system based on keywords extraction," Multimeda Tools Application 77: 4339. 2018.

[11] Sripadh T., Ramesh G. "Personalized Research Paper Recommender System," Hemanth D., Smys S. (eds) Computational Vision and Bio Inspired Computing. Lecture Notes in Computational Vision and Biomechanics, vol 28. Springer, Cham, 2018.

[12] S. Mohtaj and F. Tavakkoli, "Maglet: a Persian Journal Recommender System," 9th International Symposium on Telecommunications (IST), Tehran, Iran, 2018, pp. 348-352. 2018.

[13] Nakayama, K., Yamada, M., Shimada, A., Minematsu, T. & Taniguchi, R.i. "Learning Support System for Providing Page-wise Recommendation in e-Textbooks," K. Graziano (Ed.), Proceedings of Society for Information Technology & Teacher Education International Conference (pp. 1078-1085). Las Vegas, NV, United States: Association for the Advancement of Computing in Education (AACE), 2019

[14] Ayala-Gómez, F., Daróczy, B., Benczúr, A., Mathioudakis, M., Gionis, A. "Intelligent and Fuzzy Systems applied to Language & Knowledge Engineering," Journal of Intelligent & Fuzzy Systems, vol. 34, no. 5, pp. 3089-3100, 2018.

[15] B. Hawashin, M. Lafi, T. Kanan, A Mansour, "An efficient hybrid similarity measure based on user interests for recommender systems," Expert Systems, e12471, (2019).

[16] Irinanda SYahroni Wahyu, Dachlan Harie Soekotjo, Musim Muhammad Aziz "Identifikasi Kemiripan Teks Menggunakan Class Indexing Based Dan Cosine Similarity Untuk Klasifikasi Dokumen Pengaduan", Research Gate, 2018

[17] Ade Riyani, Muhammad Zidni Naf'an, dan Auliya Burhanudin, 2019, *Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk mendeteksi kemiripan dokumen,*Vol.2 No.1. Maret 2019, Intitut Teknologi Telkom Purwokerto

[18] Salam Abu, Wicaksana Verdian Putra, Khafiiz Hasturi "Sistem Rekomendasi Penentuan Dosen Pembimbing Tugas Akhir Dengan Menggunakan Algoritma Rabin-Karp", Techno.com, Vol. 14, 2015

[19] Soeparno. Dasar-dasar Linguistik. Yogyakarta: Mitra Gama Widya. 1993.

[20] Y. Wang, M. Wang, W. Xu, "A Sentiment-Enhanced Hybrid Recommender System for Movie Recommendation: A Big Data Analytics Framework," Wireless Communications and Mobile Computing Volume 2018, Article ID 8263704, 2018.

[21] Rizki Tri, Dhidik Prastiyanto, dan Eko Supraptono, 2017, *Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi*, Vol.9 No.1, Juni 2017, Universitas Negeri Semarang

[22] V. Imbar, M. Ayub, and A. Rehatta, "Implementasi Cosine Similarity dan Algoritma Smith-Waterman untuk Mendeteksi Kemiripan Teks," J. Inform., vol. 10, hal. 31–42, 2014.

[23] Dewa Ayu Rai Ariantini, Arie S. M. Lumenta, Agustinus Jacobus. Teknik Informatika Universitas Sam Ratulangi Manado, Indonesia. 110216022,ayurai15@gmail.com, al@unsrat.ac.id, a.jacobus@unsrat.ac.id "Pengukuran Kemiripan Dokumen Teks Bahasa Indonesia Menggunakan Metode Cosine Similarity"

[24] Gunawan, Indah Lestari, dan Muhammad Ihsan, 2015, *Sistem Pendukung Keputusan Pemilihan Pembimbing dan Penguji Proyek Akhir di Politeknik Caltex Riau*, September 2015, Politeknik Negeri Batam

[25] Nurdiana Ogie, Jumadi, Nursantika Dian"Perbandingan Metode Cosine Similarity Dengan Metode Jaccard Similarity Pada Aplikasi Pencarian Terjemah Al-Qur'an Dalam Bahasa Indonesia", JOIN, vol.1 No.1, 2016

Authors' Profiles

Ridwan Rismanto was born on March 18, 1986 He received bachelor degree at Electronic Engineering Polytechnic Institute of Surabaya in 2008 and master degree at Sepuluh Nopember Institute of Technology at 2011. His main research interests are machine learning, learning technology and object oriented programming.

Arie Rachmad Syulistyo was born on August 24, 1987 He received the S.T. and M.Kom. degrees from Brawijaya University and Indonesia University, in 2011 and 2015, respectively. His research interests include mobile programming, object based programming, enterprise application, and applied multimedia.

Bebby Pramudya Citra Agusta was born on September 11, 1996. She got his bachelor's degree from State Polytechnic of Malang, Department of Information Technology in 2019.