International Journal of Advanced Trends in Computer Science and Engineering

Fuzzy Sugeno Algorithm for Clustering Document Management

Layli Chusna Nikmatul Kamila[1], Indra Ranggadara[2]

Faculty of Computer Science, Mercu Buana University, Jakarta, Indonesia, 41815120173@student.mercubuana.ac.id

Faculty of Computer Science, Mercu Buana University, Jakarta, Indonesia, indra.ranggadara@mercubuana.ac.id

ABSTRACT

The administration process exists in every department of the company, and it has caused a large quantity of the documents. Before the documents are processed, the administration staff need to classify which document should be processed first. Because the clustering process has not integrated into a system, it takes a long time, and it caused the administration process got delayed. With this problem, it is expected the document management system using
Fuzzy Sugeno will help to reduce the problems. Fuzzy Sugeno works on clustering the document by the value of the category and the acceptance input by the user. The Fuzzy Sugeno helps administration staff to know which documents are in high priority and low priority to prevent the delay of the administration process. The document management system using Fuzzy Sugeno is help the administration staff to process the documents efficiently and it is improving the administration process in the company.

**Key words** : Administration, Clustering, Document, Fuzzy Sugeno, Management

1. INTRODUCTIONS

Administration process is everywhere putting its hands to every new opportunity in the company[1]. The administration process is quite high due to the presence of several offices spread across several regions. The administrative process, such as delivery, approval, and billing. With so many processes, many documents are circulated, and archives are stored in different sections according to their needs.

Manage information from documents in a large number of texts is not an easy job. Therefore we need a method that can organize and classify documents automatically to facilitate the search for information relevant to needs[2]. Also, information is relating to the explanation, interpretation, and relationships with other material about specific processes[3].

It is hard to manage the documents when document management not integrated into a system. These documents are circulating and also require approval from several parts required, such as budgeting, human resources, legal, and compliance, so that later these documents can be continued to the next process.

The high delivery and approval process raises some problems on the administration process, such as delay administration process, missing documents, and the large numbers of outstanding document causes the administration process to be longer and late. The obstruction of the administration process will have a negative impact on the company as disclosed much time is wasted in searching for the desired documents.
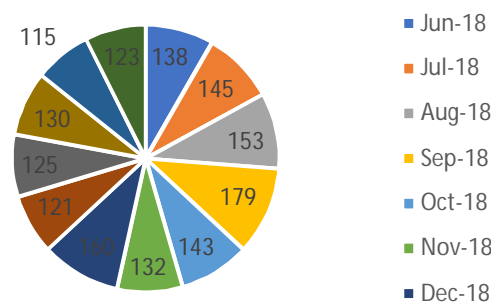
Figure 1: Quantity of the documents (monthly)

The table above shows the amount of the document processed by one branch office. From the amount above the documents need to be classify by the category before the documents are processed. The classification process takes time and it is caused the administration process much longer.

These conditions require improvements and changes in the administration process. This is done to help the company to work optimally and reduce the problems on the administration process of the company. The form of improvement that will be carried out is by creating a document management system for the company using Fuzzy Sugeno. A method used in document management is clustering or classifying documents[4].

In this research, the Fuzzy Sugeno positioning method on the clustering of the documents. Currently, the clustering of the document is based on the document title and the recipients, but in this case, the company needs to classified the document by the priority to know which documents should be process in a high priority. So, by the document management system administration staff will work efficiently because the document already classified with the priority. By this research, the system could analyze the document type based on the category and the acceptance date to classify which document is in high or low priority to make the administration process more efficient.

2. RELATED WORKS

Fuzzy logic is created because boolean logic does not have high accuracy, it only has 0 and 1 logic numbers[5]. That's why the Fuzzy logic is closer in spirit to human thinking and natural language than conventional logical systems[6]. The essence of fuzzy control is to build a model of the human expert who is capable of controlling the plant without thinking in terms of mathematical model[7]. Sugeno has better processing time since the weighted average replace the time-consuming defuzzification process[8].

By the research of Comparision of Mamdani and Sugeno Fuzzy Inference System for Deciding the Set Point for a Hydro Power Plant Dam Reservoir Based on Power Generation Requirement, said that Mamdani-type FIS and Sugeno-type FIS perform quite similar, but Sugeno-type FIS allows the evaluation of gate opening work at its full capacity with smooth operational performance. Although the designing of both systems is same but the output membership functions of Sugeno-type can only be either constant or linear and also the crisp output is generated in different ways for both FISs. Sugeno-type FIS has also an advantage that it can integrated with neural networks and genetic algorithm or other optimization techniques so that the system can adapt to system characteristic efficiently[9].

By the research of Application of Sugeno Fuzzy Inference System to Determine Amount of Drug Purchases, the author use 3 variables like stocks, purchases, and sales to determine the amount of drugs purchases. Based on the test result, with the Sugeno method the accuracy is 88,88% by counting with MAPE (Mean Absolute Percentage Error) formula[10].

Based on the research of Fuzzy Sugeno Logic Based on Medical Record Data in Diseases the authors use fuzzy inference system sugeno with 4 variable inputs. In this research they got error less than 1% and it is mean the result works for fast and valid information based on the input to the system[11].

By the research of Fuzzy Logic Comparison Analysis of Tsukamoto, Sugeno, and Mamdani Methods the authors give 2 inputs to predict the amount of registration students in the future. Based on the comparison of 3 fuzzy logic, Sugeno method has the higher error with on 86,41%[12].

In this research, to determine the priority of the documents use 2 inputs from the document, that are category and the acceptance date. Each of the input has score and criteria that will be count to determine the priority of the documents.

3. RESEARCH METHOD

Fuzzy inference system with Sugeno method has a characteristic that the output (consequent) is not a fuzzy set, but a linear equation with variable in accordance with the variables input[13]. In this research, the output is a priority of the document. The form of Sugeno method in general is :

IF (x_1 is A_1) (x_2 is A_2) (X_3 is A_3) … (X_n is A_n) THEN y is B

The following steps that will carried out in this research by using fuzzy inference system with Sugeno method are :

A. Collecting Data

In this step, we collect the document received by department in a month. From the data, we can specify the document based on the title, acceptance date, and category. After that, it will be classified by the rules on the next steps. The total of the data is 1664 from June 2018 – May 2019, on the table below, show the sample of data taken from the database.

Table 1: Sample Of Data

No	Document Title	Incoming Date	Category
1	Procurement Contract of Server	01-May-19	Contract
2	Invoice of Outing events in Sydney	03-May-19	Invoice
3	Monthly Operational Report Project Online Charging System	10-May-19	Report
4	Training Quotation of ABC Company	15-May-19	Quotation
5	Invoice of meeting events in Hotel A	15-May-19	Invoice
6	Contract of Managed Services Campaign Project	17-May-19	Contract

B. Fuzzy Inference System Process

Fuzzy inference systems, also known as fuzzy-rule-based systems, fuzzy associative memories, fuzzy models, or fuzzy controllers if it used as controllers[14]. Fuzzy inference system is composed by 5 functional blocks :

1. Rule base contains a number of fuzzy if-then rules.
   In this research based on the data it has 9 rules :

   a. *If* the category of the document is low *and* the acceptance date is near *then* the priority is low
   b. *If* the category of the document is medium *and* the acceptance date is near *then* the priority is low
   c. *If* the category of the document is high *and* the acceptance date is near *then* the priority is medium
   d. *If* the category of the document is low *and* the acceptance date is medium *then* the priority is medium
   e. *If* the category of the document is medium *and* the acceptance date is medium *then* the priority is medium
   f. *If* the category of the document is high *and* the acceptance date is medium *then* the priority is high
   g. *If* the category of the document is low *and* the acceptance date is medium *far* the priority is high
   h. *If* the category of the document is medium *and* the acceptance date is far *then* the priority is high
   i. *If* the category of the document is high *and* the acceptance date is far *then* the priority is high

27

Table 2 : Fuzzy if-then rules

Category / Acceptance Date	Near	Medium	Far
Low	Low	Low	High
Medium	Low	Medium	High
High	Medium	High	High

From the table above, it showed the fuzzy if-then rules that have 9 rules that already mention above.

2. The database defines the membership functions of the fuzzy sets used in the fuzzy rules. The membership function is a graph that represents the magnitude of the degree of membership of each input variable that is in an interval between 0 and 1[15]. Below is the membership functions of fuzzy sets :

Table 3: Type of the Documents

Document Types	Category	Score
Report and Quotation	Low	5
Contract	Medium	7,5
Invoice	Tinggi	10

The documents are classified into some types such as Report, Contract and Invoice. Each of the the type has category and score, from this classification the system determine the priority of the document and calculate with the score of acceptance date as shown at the table below. From the table above it is also define on a formula :

$$\mu Low = \{x \le 5$$

$$\mu Medium = \begin{cases} x \le 7,5 \\ \dfrac{x - 5}{7,5 - 5} \\ \dfrac{10 - x}{10 - 7,5} \end{cases}$$

$$\mu High = \begin{cases} x \le 10 \\ \dfrac{x - 7,5}{10 - 7,5} \end{cases}$$

Table 4:Category of Acceptance Date

Acceptance Date	Category	Score
<= 3 days	Near	5
> 3 days	Medium	7
>= 7 days	Far	10

From the table above it shown that acceptance date has their own category and score based on the distance between today's date and the acceptance date. Below is the formula to define the membership functions :

$$\mu Near = 4$$

$$\mu Medium = \begin{cases} x \le 7 \\ \dfrac{x - 4}{7 - 4} \\ \dfrac{10 - x}{10 - 7} \end{cases}$$

$$\mu Far = \begin{cases} x \le 10 \\ \dfrac{x - 7}{10 - 7} \end{cases}$$

Table 5 Category of the Priority

Priority	Score
Low	100
Medium	200
High	300

The table above shown the category of priority, it is classified as low, medium, and score.

3. Decision-making unit to performs the inference operation on the rules. From the fuzzy rules above each of the membership functions combine using logic $\cap$ to take the minimum value.

4. The fuzzification interface transforms the crisp inputs into degrees of the match with linguistic values.

5. Defuzzification interface to transform the result of interference onto a crisp output.

4. RESULT AND ANALYSIS

The membership functions of the documents type and the acceptance date is define with the category and score. Based on the fuzzy rules the system should determine the priority of the documents to help the administration staff process the documents. Below is the data of the document :

Table 6: Data of a document

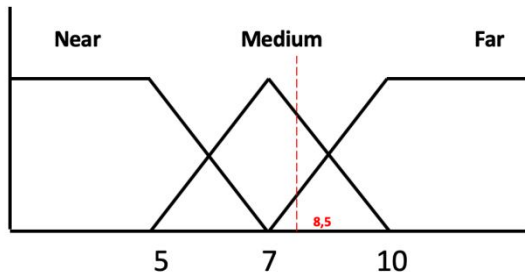| Document Title	Acceptance Date	Category
Invoice of Meeting Events in Hotel A	05-Dec-19	Invoice
Score	**8,5**	**7**

From the data above it is also can transform to a graphic of membership functions below :

Acceptance Date: $4 < x < 10$

28
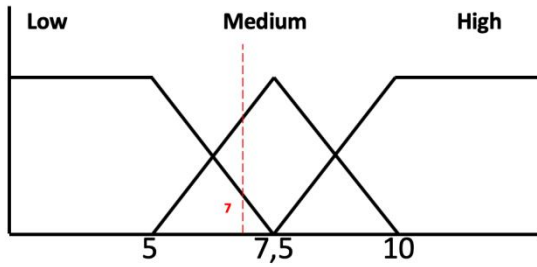
Figure 2: The membership functions of document types

Type of the document : $5 < x < 10$

Figure 3: The membership functions of acceptance date of

the document

29

From the membership functions above to determine the priority it should count by the formula below :

$$\alpha \text{ medium} = \mu \text{ medium } [7] = \frac{7-5}{7,5-5} = 0,8$$

$$\alpha \text{ low} = \mu \text{ low } [7] = \frac{7,5-7}{7,5-5} = 0,2$$

$$\alpha \text{ medium} = \mu \text{ medium } [8,5] = \frac{10-8,5}{10-7} = 0,5$$

$$\alpha \text{ far} = \mu \text{ far } [8,5] = \frac{8,5-7}{10-7} = 0,5$$

Then, decision-making unit to performs the inference operation on the rules. From the fuzzy rules above each of the membership functions combine using logic to take the minimum score and multiplied with the score of the priority :

α medium $\cap$ α medium $= 0,5$
α medium $\cap$ α far $= 0,5$
α low $\cap$ α medium $= 0,2$
α low $\cap$ α far $= 0,2$

Then defuzzification interface to transform the result of interference onto a crisp output.

From the calculation above, the next step is defuzzification process using average formula :

$$z = \frac{0,5(200)+0,5\,(300)+0,2\,(200)+0,2\,(300)}{0,5+0,5+0,2+0,2} = 250$$

So, the priority score of the document is 250, which means the priority of the document is high, and it is a document that needs to be processed as soon as possible.

Below is the implementation of Fuzzy Sugeno to the document management system.

Figure 4: Input a document to the system

The user needs to input the document title, incoming date, sender, recipient, and the category of the document. After that system will classify the document into priority such as high, medium, and low based on the score of category and incoming date.


Figure 5: Grouping of the document based on the priority

From the picture above user can easily make the priority of which document need to processed first. Based on the priority classification the administration process could improve the administration process on the company and simplify the job of administration staff. The testing of the Fuzzy Sugeno is using Visual Basic programming language.
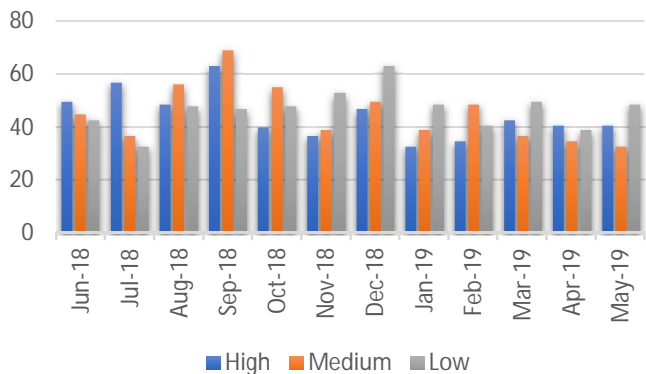
Figure 6: Classification of the documents based on the priority

From the testing using Fuzzy Sugeno the 1664 data can be classify the priority. So for the next incoming document the administration staff can easily make the priority to each documents to be proceed.
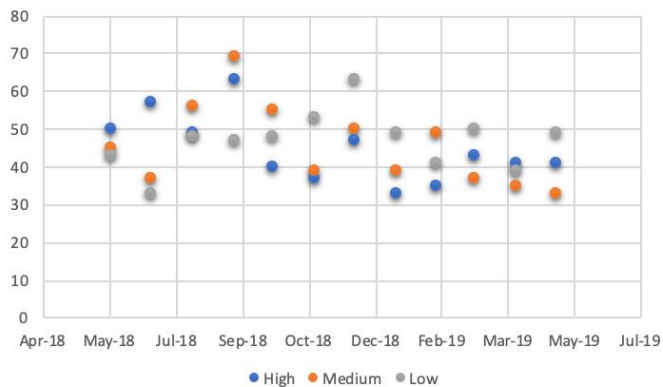
Figure 7: Classification of the documents based on the priority

(scatterplott)

From the picture above, it is shows the classification of the documents in a scatterplot.

4. CONCLUSION

Based on the research above can be concluded that this research helps the company to improve the administration process by the classification of the document on the document management system. By using the Fuzzy Sugeno method, the administration staff can process the document based on the priority and the administration process done on time.

Based on the 2 variables input of the documents, the Fuzzy Sugeno worked to determine the classification of the priority of the documents. This research still has some limitations, because the data on this research only take from 1 branch office and some of the company has a lot of branch office. So, by this research the author suggest the next author will work for the document management system for a company that connected some branch office.

REFERENCES

[1] "The Study of Administration Author (s): Woodrow Wilson Source : Political Science Quarterly , Vol . 2 , No . 2 (Jun ., 1887), pp . 197-222 Published by : The Academy of Political Science Stable URL : http://www.jstor.org/stable/2139277," vol. 2, no. 2, pp. 197–222, 2010. https://doi.org/10.2307/2139277

[2] J. Samodra, S. Sumpeno, and M. Hariadi, "Klasifikasi Dokumen Teks Berbahasa Indonesia dengan Menggunakan Naïve Bayes," pp. 1–4, 2009.

[3] Y. S. Sari, "PROTOTYPE OF KNOWLEDGE MANAGEMENT SYSTEM (KMS) E-PROCUREMENT WEB-BASED : CASE STUDY AT PT . SIGMA PRO 77," vol. 5, no. 06, 2018.

[4] R. T. Wahyuni, D. Prastiyanto, and E. Supraptono, "Jurnal Teknik Elektro," *J. Tek. Elektro*, vol. 9, no. 1, pp. 18–23, 2017.

[5] A. Hermawan, I. Ranggadara, A. Suharso, I. Prihandi, and Suhendra, "Fuzzy tsukamoto for determinen linen priority," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 2749–2755, 2019. https://doi.org/10.35940/ijeat.A9764.109119

[6] M. C. Pradhan, "An Intelligent Fuzzy Based Technique Of Making Food Using Rice Cooker," vol. 5, no. May, pp. 1–7, 2015.

[7] A. Kaur and A. Kaur, "Comparison of Mamdani-Type and Sugeno-Type Fuzzy Inference Systems for Air Conditioning System," *Int. J. Soft Comput. Eng.*, 2012.

[8] K. Jain and A. Soni, "Comparision of Mamdani and Sugeno Fuzzy Inference System for Deciding the Set Point for a Hydro Power Plant Dam Reservoir Based on Power Generation Requirement," no. 2, pp. 1–4, 2015.

[9] O. A. Egaji, A. Griffiths, M. S. Hasan, and H. N. Yu, "A comparison of Mamdani and Sugeno fuzzy based packet scheduler for MANET with a realistic wireless propagation model," *Int. J. Autom. Comput.*, vol. 12, no. 1, pp. 1–13, 2015. https://doi.org/10.1007/s11633-014-0861-y

[10] S. L. M. Sitio, "Penerapan Fuzzy Inference System Sugeno untuk Menentukan Jumlah Pembelian Obat (Studi Kasus: Garuda Sentra Medika)," *J. Inform. Univ. Pamulang*, vol. 3, no. 2, p. 104, 2018. https://doi.org/10.32493/informatika.v3i2.1522

[11] N. T. Pradipta *et al.*, "Logika Fuzzy Sugeno Berbasis Data Rekam Medik Pada Penyakit," *J. Ilm. Inform.*, vol. 2, no. 1, pp. 59–67, 2017.

[12] M. Irfan, L. P. Ayuningtias, and J. Jumadi, "Analisa Perbandingan Logic Fuzzy Metode Tsukamoto, Sugeno, Dan Mamdani (Studi Kasus : Prediksi Jumlah Pendaftar Mahasiswa Baru Fakultas Sains Dan Teknologi Uin Sunan Gunung Djati Bandung)," *J. Tek. Inform.*, vol. 10, no. 1, pp. 9–16, 2018. https://doi.org/10.15408/jti.v10i1.6810

[13] A. Bahroini, A. Farmadi, and R. A. Nugroho, "Prediksi Permintaan Produk Mie Instan Dengan Metode Fuzzy Takagi-Sugeno," *Klik - Kumpul. J. Ilmu Komput.*, vol. 3, no. 2, p. 220, 2016. https://doi.org/10.20527/klik.v3i2.62

[14] J. R. Jang, "ANFIS : Adap tive-Ne twork-Based Fuzzy Inference System," vol. 23, no. 3, 1993. https://doi.org/10.1109/21.256541

[15] Lodewyik Rahakbauw Dorteus, "Penerapan Logika Fuzzy Metode Sugeno Berdasarkan Data Persediaan Dan Jumlah Permintaan (Studi Kasus : Pabrik Roti Sarinda Ambon) Application of Fuzzy Logic Method Sugeno To Determine the Total Production of Bread ," *J. Ilmu Mat. dan Terap.*, vol. 9, pp. 121–134, 2015.