Jurnal Teknik Informatika (JUTIF)
Vol. 4, No. 3, June 2023, pp. 503-509

DOI: https://doi.org/10.52436/1.jutif.2023.4.3.751
p-ISSN: 2723-3863
e-ISSN: 2723-3871

IMPROVING ONLINE MEETING EFFICIENCY USING LATENT DIRICHLET ALLOCATION (LDA) AND SOCIAL NETWORK ANALYSIS (SNA) METHODS

Megananda Hervita Permata Sari[*1], Utomo Budiyanto[2]

[1,2]Master of Computer Science, Faculty of Information Technology, Universitas Budi Luhur, Indonesia
Email: [1]2111602153@student.budiluhur.ac.id, [2]utomo.budiyanto@budiluhur.ac.id

(Article received: December 13, 2022; Revision: January 30, 2023; published: June 26, 2023)

Abstract

The pandemic period can change the habits of a person and organization, where all meetings are not held face-to-face/offline but virtually, so it is not uncommon for meetings to be attended by employees who are not Persons in Charge (PIC) on certain meeting topics. This study aims to identify trends in time, day, and duration of meetings within the Secretariat General of the Ministry of Finance and to cluster meeting matters into several themes so that further identification can be carried out to provide recommendations for units having duties related to the meeting using networking analysis. This study uses the Natural Language Processing (NLP) method with Latent Dirichlet Allocation (LDA) which can conclude the factors that represent topics to produce topic clustering and Social Network Analysis (SNA) modeling using the Degree Centrality method to find out the closest relationship between topics and names. unit based on the highest centrality value and the possibility of a unit attending a meeting that discusses a particular topic. Data used in this reseacrh are meetings held during April 2020 up to April 2022 with 59,891 data records. The modeling results shows clustering result dashboard based on meeting topics and to produce an analysis of which meeting topics are often discussed and become a concern. The results of the research are expected to be used to provide recommendations to unit leaders to assign meeting dispositions for each PIC to attend the meeting.

Keywords: *degree centrality, latent dirichlet allocation, natural language processing, online meeting, social network analysis.*

1. INTRODUCTION

In today's new normal era, virtual meeting has become a new habit. Virtual meeting provides convenience in terms of flexibility [1]. However, that flexibility comes with consequences [2]. First, virtual meeting has reduced travel time from one meeting to another. This leads to a situation where meetings can be conducted in sequence in a day with merely quick pause in between. It is common that employees attended more than one meetings at the same time. Second, flexibility in terms of time and space allows meetings to be conducted after office hours and on weekend. This definitely gives impacts to employees' life, such as work-life balance. Third, meeting duration should also be a concern. Virtual meeting is different from the physical meeting in terms of the situation, which forces the employees to see the screen for as long as the meeting conducted. A research also shows that camera usage on virtual meetings give impacts to exhaustion. [3] Another problem happened often was that the meeting was not attended by the related *Person in Charge* (PIC). Meanwhile, the related PIC did not receive the meeting invitation. These issues can be addressed by creating a better meeting schedule and a more efficient meeting in terms of duration and the related meeting participants. Due to these issues, meeting analysis is required. For this research, data were acquired from the Office Automation application, specifically the meeting modul, owned by the Ministry of Finance.

The meeting data used comes from 13 Echelon 2 Units at the Secretariat General of the Ministry of Finance during the period April 2020 to April 2022 with 59,891 data records.

Data mining is a process to mine information and find insights without explicit assumption derived from large data, in which the information gained should have 3 characteristics, namely novel, effective, and practical. Data mining can be categorized into two main categories, which are predictive and descriptive [4].

The method widely applied in the data mining is Cross-Industry Standard Process Model for Data Mining (CRISP-DM) [5]. The data mining process in the CRISP-DM consists of six stages, namely (1) Business Understanding; (2) Data Understanding; (3) Data Preparation; (4) Modeling; (5) Evaluation; dan (6) Deployment.

This study uses the LDA method to map meeting topics held by all echelon 2 units of the Secretariat General so that interrelationships between topics can be obtained. With this model it is possible to produce an analysis of which meeting topics are often discussed and become a concern.

This research applies *Latent Dirichlet Allocation* (*LDA*) method [6]. LDA is an unsupervised machine learning model that can cluster topics automatically. LDA identifies and formulates the patterns from the words in the document as well as recommend some topics automatically. The more documents used as an input, the more accurate the model would be produced since LDA applies an unsupervised learning.

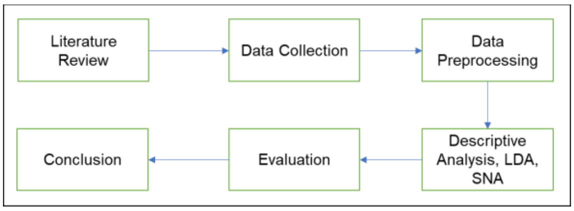
In addition to that, this research also implements *Social Network Analysis* (SNA) method. SNA is a method used to identify relation between units or employees using graphic analysis technique. SNA is often used by calculating the centrality score, which identifies the central actor in a network [7].

The results of this research can then be included in the OA application to help leaders get PIC recommendations for meetings that are appropriate by providing highlight on the name of the unit concerned.

2. RESEARCH METHOD

2.1. Research Stages

The initial stage of this research is literature study, specifically journals with topics related to the research topic. Literature study aims to find the benchmark and support this research. The next stage is data collection for topic modelling process. Prior analysis, preprocessing data should be conducted, in this case data cleansing. The cleansed data can then be used for clustering the meeting topics using LDA method and SNA to identify the most important unit in every topic. After analysis, the next stage is evaluating the model by testing and calculating the accuracy score to define the model performance. The last stage is making conclusion of the research. The flow of the research is as shown by Picture 1.



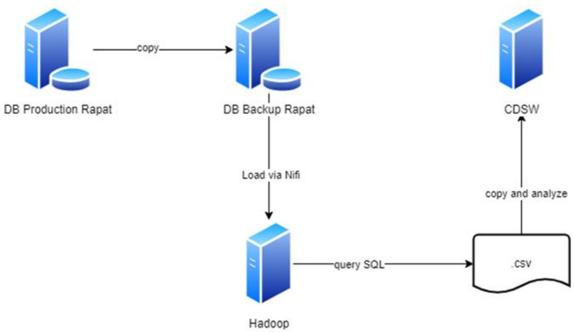
Picture 1. Research Stages.

2.2. Literature Review

Several research on topic modeling and network analysis have been conducted to comprehend the trending topics. A research of [8] combines the LDA dan *Network Centrality* method in order to identify the criminal expert, such as hackers, in hackers forum. Applying LDA, the forum members were eliminated by 79%. Then the data was further analyzed by applying *network centrality* method to identify the popular members that likely have good expertise. Another research recommends a method or tool that can automatically analyze an underground forum and identify the main hacker, namely HackerRank, in which the tool combined CA (*content analysis*) dan SNA (*social network analysis*). HackerRank was then tested to five different underground forums separately. Compare to using social network analysis and content analysis separately, HackerRank increases the coverage rate in the five underground forums by 3.14% and the average of 16,19% [9]. LDA and SNA can also help scientists and researchers to understand science and technology research development by analyzing research trend evolution and humanoid technology structure based on publication database and patent. For policy-makers, it can be used as a guidance for policy formulation in the future and technology investment strategy [10]. LDA, SNA, and sentiment analysis are also applied in *e-commerce* research to find *insights* on consumers behavior and wish lists [11]. The results are positive sentiment trend using sentiment analysis, namely e-commerce having games attracting more customers using topic modeling (LDA), identification of best-selling product using descriptive analysis, and maximization of social network using social media (SNA).

2.3. Data Collection

This is the initial stage of collecting the data from the source of database production (db_ekeu_rapat). The data is restored *to database staging server* applying *copy* between *servers mechanism*. The total number of data used for this research is 59.891 *records.* The data is sent to Cloudera machine using EL (*extract-load*) *tools* Nifi. The flow of data collection is shown on Picture 2.



Picture 2. Flow of Data Collection

2.4. Data *Preprocessing*

The next stage after data collection is preprocessing data, which aims to remove and revise some characters in order to produce a more accurate analysis of meeting topics. Some main steps in data cleansing are [12]:

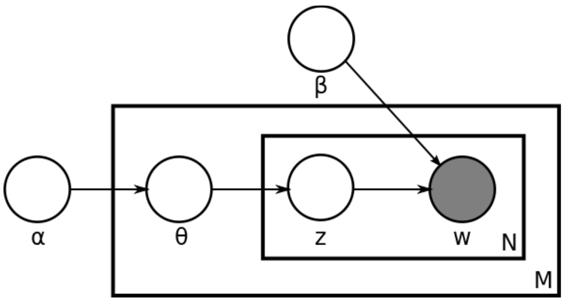
a. Lemmatization, which transforms word into its root based on the Bahasa dictionary.
b. Tokenization, which divides text into smaller parts (sentence, word, biagrams). This division covers text based on word and ignore punctuation.

c. *Stopword removal*, which filters common words and chooses important and unique words from the tokenization result.
d. *Pos Tagger*, which categorizes word class, such as nouns (NN) and pronouns (NNP).

2.5. Modeling Using LDA Method

After completing the preprocessing, the next stage is clustering the meeting topics by using LDA method.

LDA is topic modeling using Bayesian estimation technique. It is a technique that can conclude factors representing topics from every element into several topic groups. The functions of LDA are summarizing data, clustering, providing weight on each cluster. *Imaginary random process* in the model assumes the document coming from certain topics and every topic consisting of word distribution, showing that LDA is generative. In LDA, the inputs are the numbers of topics wished and words representing each topic [13][14]. An overview of the graphical model of LDA is presented in Picture 3.



Picture 3. LDA Model.

2.6. *Social Network Analysis*

The next analysis stage is identifying the most important or the central units and connecting units by applying the *Social Network Analysis* (SNA) method. The SNA aims to recognizing the network among units and groups in an organization [15]. In this case, *graph* was built based on units as the node.

3. HASIL DAN PEMBAHASAN

3.1. Data Collection Result

To analyze meeting data, the data required are meeting titles, meeting schedule (date, hour, duration), meeting participants and units for both host and attendees. The data are collected from two tables, namely meeting table and attendance table. The collected data are transformed through query, which are then loaded in the form of CSV file. Transformation process includes join between tables using SQL query. The result of the query is then exported to CSV file as shown on Picture 4:

	Id	NamaRapat	unit	tgl_mulai_rapat	jam_mulai_rapat	tgl_akhir_rapat	jam_akhir_rapat
0	80C00001-FB54-4414-AC8A-05F77CD400F4	Persiapan Sidang Redaksi Buletin Kinerja	Subbagian Pengelolaan Kinerja dan Risiko II, B...	2022-04-05T00:00:00	10:00	2022-04-05T00:00:00	12:00
1	828F2153-EFD2-4ED3-AFAE-B867553C89EF	Pembahasan rapat internal pokja I terkait dokm...	Biro Manajemen Barang Milik Negara dan Pengada...	2022-04-05T00:00:00	13:00	2022-04-05T00:00:00	17:00
2	9AE3A0A8-8560-4570-843B-897C98991CBC	Rapat Agenda Setting KSP	Subbagian Publikasi Elektronik, Bagian Manajem...	2022-04-05T00:00:00	14:00	2022-04-05T00:00:00	15:30
3	031904F6-9E37-4DE9-B2C5-4D0B6F3B8180	pembahas internal pokja 1 terkait Pengadaan J...	Biro Manajemen Barang Milik Negara dan Pengada...	2022-04-05T00:00:00	08:30	2022-04-05T00:00:00	12:00
4	D3CDA637-FF20-47E4-AD2E-EAFDCD18ED9F	Editorial konten smindrawati	Subbagian Publikasi Elektronik, Bagian Manajem...	2022-04-05T00:00:00	08:00	2022-04-05T00:00:00	09:00
...	...	...	...	...	...	...	...
59886	3E462504-47F9-40E5-8858-A6F3F32F2BF8	Rapat Evaluasi Implementasi e-Kemenkeu dan Per...	Subbagian Proses Bisnis II, Bagian Strategi Or...	2021-01-08T00:00:00	NaN	2021-01-08T00:00:00	NaN
59887	8FAE5848-8D33-45AF-9900-BAF6A6CF67A6	Pembahasan Hasil One on One Usulan Pengembanga...	Subbidang Perancangan dan Pengembangan Aplikas...	2021-01-07T00:00:00	NaN	2021-01-07T00:00:00	NaN
59888	1758890D-33FD-4862-B1F7-E60090669C74	Rapat Tim Fokus	Subbagian Umum, Bagian Tata Usaha, Pusat Pembi...	2021-01-06T00:00:00	NaN	2021-01-06T00:00:00	NaN
59889	BF3E25A7-DF01-4834-A04A-F0F3D7EB8A25	Rapat Uji Coba Modul Rapat	Subbidang Perancangan dan Pengembangan Aplikas...	2021-01-06T00:00:00	NaN	2021-01-06T00:00:00	NaN
59890	AE0168EC-2DCA-45F2-8857-EBE196175247	Rapat Rencana Kerja Bagian TU PPPK Tahun 2021	Subbagian Umum, Bagian Tata Usaha, Pusat Pembi...	2021-01-05T00:00:00	NaN	2021-01-05T00:00:00	NaN

Picture 4. Data Collection Result.

3.2. *Data Preparation*

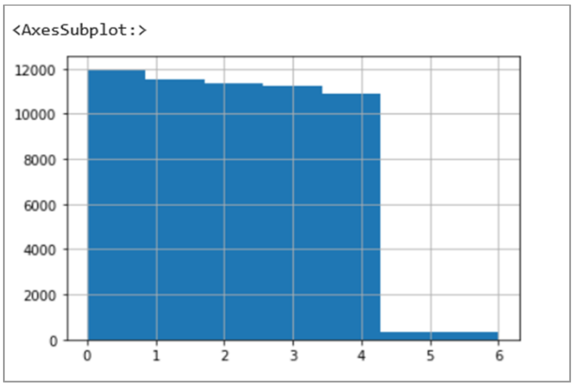
Data preparation activities covers choosing table and column to be used and transforming the data into the new database as the raw data that can be processed further using the data mining method.

Steps conducted in this process are loading the dataset from the file into the database, fixing the null values using imputer, cleansing the meeting title using lemmatization to change words into its root words, choose unique words from the tokenized words, and use stopword to remove common words (meeting, discuss, coordination, regarding, FGD, 2021 and 2022). The result of data preparation is used for the next stage.

3.3. Descriptive Analysis Results

After data preprocessing, the next stage is analysis. One of the initial analyses that can be conducted is descriptive analysis. In descriptive analysis, some statistical trends can be identified, such as time, day, and meeting duration.

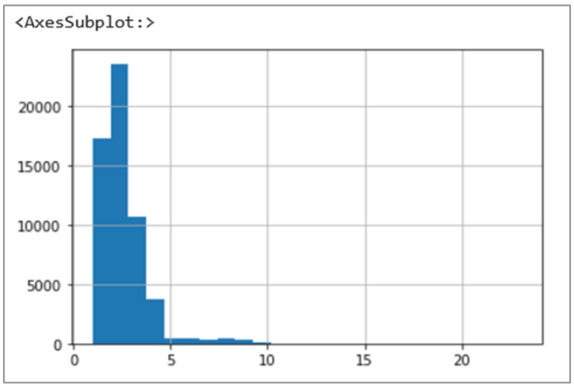
For the attendance table data, descriptive analysis covers the total number of meetings on each day, meeting start time, meeting finish time, and duration meeting trends. The descriptive analysis for the total number of meetings on each day can be seen on Picture 5.



Picture 5. Descriptive Analysis on Number of Meetings

The result shows that Monday has the highest number of meetings. Then the number decreases up to Sunday. Furthermore, the data shows that some meetings are held on weekends (Saturday and Sunday).

The next analysis is the meeting duration trend, which shows the longest meeting held for about 2-3 hours as seen on Picture 6.



Picture 6. Descriptive Analysis on Meeting Duration.

3.4. LDA Analysis Result

LDA method analysis is applied to cluster meeting titles into several clusters. LDA Topic Modeling can be seen on Picture 7.

```
def get_lda_topics(model, num_topics):
    word_dict = {};
    for i in range(num_topics):
        words = model.show_topic(i, topn = 20);
        word_dict['Topic # ' + '{:02d}'.format(i+1)] = [i[0] for i in words];
    return pd.DataFrame(word_dict);
```

```
from gensim.models import LdaMulticore
```

Picture 7. Topic Modeling Using LDA

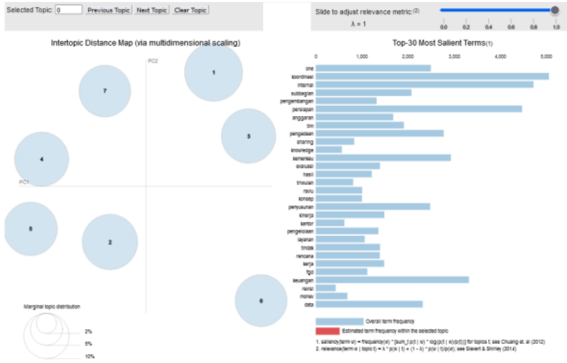
The next step is experimenting topic classification by changing the number of clusters into 7, 6, or 5 clusters to gain the best result.

In the first experiment, the cluster number is set at 7 topics or clusters. The result is as shown on Picture 8 and Picture 9.

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05	Topic # 06	Topic # 07
0	koordinasi	koordinasi	internal	one	internal	persiapan	koordinasi
1	subbagian	internal	one	tim	keuangan	keuangan	kemenkeu
2	persiapan	persiapan	pengadaan	internal	pengadaan	koordinasi	persiapan
3	penyusunan	subbagian	kemenkeu	data	anggaran	penyusunan	biro
4	bahan	pengembangan	rencana	pengadaan	koordinasi	aplikasi	keuangan
5	internal	keuangan	pengelolaan	evaluasi	data	one	penyusunan
6	tim	kemenkeu	penyusunan	lanjutan	persiapan	pengembangan	data
7	kemenkeu	rkmk	persiapan	paket	kementerian	kementerian	hasil
8	aplikasi	teknis	evaluasi	hasil	bmn	kinerja	konsep
9	tindak	pajak	kerja	koordinasi	kerja	layanan	fgd
10	keuangan	pengelolaan	tik	manajemen	paket	sharing	tim
11	teknis	biro	keuangan	biro	reviu	tindak	triwulan
12	knowledge	rencana	program	persiapan	subbagian	pelaksanaan	kerja
13	sdm	bahan	aplikasi	jabatan	lanjutan	bahan	lanjutan
14	pegawai	evaluasi	konsep	kajian	kemenkeu	biro	pengadaan
15	monev	tindak	hasil	jasa	jasa	kemenkeu	kementerian
16	kinerja	kementerian	anggaran	kinerja	dokumen	perangkat	kinerja
17	data	pelaksanaan	jf	keuangan	lingkungan	data	jf
18	biro	pengadaan	secondment	penyusunan	aplikasi	internal	anggaran
19	perencanaan	sekretariat	iku	is	kantor	lingkungan	aplikasi

Picture 8. LDA Modeling Result using 7 Topics (1)

In the picture above is the result of clustering of 7 clusters, where 7 clusters can be calculated for coherence values to determine the best model value.



Picture 9. LDA Modeling Result using 7 Topics (2)

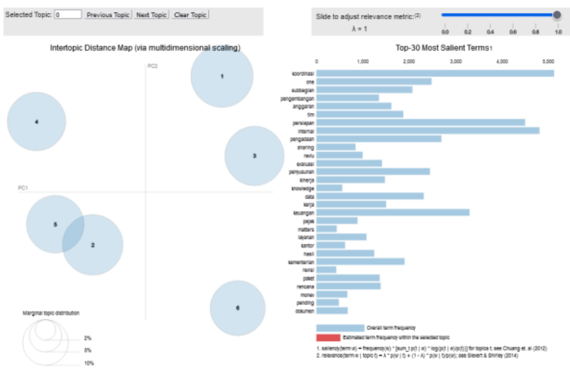
The result of 7 clusters shows that there is not overlapping topics with coherence score at -3.93. Then the second experiment applies 6 clusters. The analysis result is shown on Picture 10 and Picture 11.

[44]:

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05	Topic # 06
0	koordinasi	koordinasi	internal	one	internal	persiapan
1	persiapan	internal	one	tim	keuangan	keuangan
2	subbagian	persiapan	kemenkeu	data	pengadaan	koordinasi
3	penyusunan	kemenkeu	pengadaan	internal	anggaran	penyusunan
4	kemenkeu	pengembangan	rencana	pengadaan	koordinasi	aplikasi
5	tim	subbagian	penyusunan	evaluasi	data	kementerian
6	bahan	keuangan	pengelolaan	lanjutan	persiapan	pengembangan
7	internal	biro	persiapan	hasil	kerja	kinerja
8	teknis	rkmk	kerja	biro	kementerian	one
9	aplikasi	pajak	konsep	koordinasi	bmn	kemenkeu
10	tindak	teknis	evaluasi	paket	kemenkeu	layanan
11	keuangan	pengelolaan	tik	manajemen	lanjutan	biro
12	monev	bahan	aplikasi	jasa	reviu	pelaksanaan
13	kinerja	evaluasi	hasil	jabatan	paket	sharing
14	knowledge	sekretariat	keuangan	kinerja	subbagian	fgd
15	sdm	rencana	program	persiapan	aplikasi	tindak
16	pegawai	jf	jf	kemenkeu	dokumen	bahan
17	biro	tindak	secondment	kajian	iku	perangkat
18	subbag	tata	iku	sesi	perangkat	data
19	data	pelaksanaan	data	keuangan	kantor	triwulan

Picture 10. LDA Modeling Result using 6 Topics (1)

In the picture above is the result of clustering of 6 clusters, where 6 clusters can be calculated for coherence values to determine the best model value.



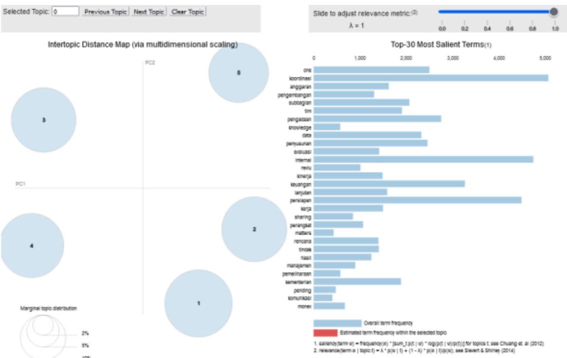
Picture 11. LDA Modeling Result using 6 Topics (2)

The result of 6 topics shows that there is overlapping topics and the coherence score is lower than the 7 topics, which is -4.11. The last experiment uses 5 topics with result as shown on Picture 12 and Picture 13.

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	koordinasi	koordinasi	internal	one	internal
1	persiapan	persiapan	one	tim	keuangan
2	penyusunan	internal	kemenkeu	data	pengadaan
3	subbagian	pengembangan	pengadaan	internal	koordinasi
4	aplikasi	keuangan	persiapan	lanjutan	persiapan
5	kemenkeu	kemenkeu	penyusunan	pengadaan	anggaran
6	bahan	biro	rencana	kinerja	data
7	tim	subbagian	aplikasi	evaluasi	kementerian
8	kinerja	rkmk	pengelolaan	biro	bmn
9	keuangan	teknis	keuangan	persiapan	kerja
10	tindak	tindak	kerja	hasil	lanjutan
11	internal	kementerian	evaluasi	koordinasi	kemenkeu
12	teknis	pajak	tik	manajemen	perangkat
13	knowledge	pelaksanaan	konsep	jabatan	aplikasi
14	fgd	bahan	hasil	paket	reviu
15	monev	pengelolaan	laporan	keuangan	paket
16	biro	evaluasi	layanan	layanan	lingkungan
17	pegawai	rencana	program	penyusunan	subbagian
18	data	jf	jf	kajian	rkmk
19	sharing	briefing	biro	jasa	fgd

Picture 12. LDA Modeling Result using 5 Topics (1)

In the picture above is the result of clustering of 5 clusters, where 5 clusters can be calculated for coherence values to determine the best model value.



Picture 13. LDA Modeling Result using 5 Topics (2)

The modeling result using 5 topics shows coherence score at -4.12 without overlapping topics. Comparing the coherence scores of 7, 6, and 5 topics, the model applying 7 topics is the best model since it has the highest coherence score.

3.5. SNA Analysis Result

This research also applies SNA modeling using the Degree Centrality method. Degree Centrality is a method to determine the possibility of a unit to attend the meeting discussing particular topics. The score recommends the closest relation between units (unit in charge) and certain topics based on the highest value of degree centrality. This method recommends particular units to attend the meeting.

The first step is joining the topics derived from the LDA modeling and the name of the units. The results for topic 0, 1, and 2 are as follows:

i. Topic 0

For Topic 0, the closest unit to attend the meeting is the Division of Organization Strategy and Business Process as shown on Picture 14.

```
Top 10 nodes by degree centrality for Topic 0.0:
Bagian Strategi Organisasi dan Proses Bisnis-Biro Organisasi dan Ketatalaksanaan | 137
Bagian Organisasi dan Jabatan Fungsional I-Biro Organisasi dan Ketatalaksanaan | 137
Bagian Tata Kelola dan Pelayanan Publik-Biro Organisasi dan Ketatalaksanaan | 137
Bagian Organisasi, Tata Laksana, dan Komunikasi-Biro Umum | 137
Bagian Umum-Sekretariat Pengadilan Pajak | 136
Sekretaris Pengganti I A-Sekretariat Pengadilan Pajak | 136
Sekretaris Pengganti XIX B-Sekretariat Pengadilan Pajak | 136
Bagian Teknologi Informasi dan Komunikasi-Sekretariat Pengadilan Pajak | 136
Sekretaris Pengganti XVII B-Sekretariat Pengadilan Pajak | 136
Sekretaris Pengganti III B-Sekretariat Pengadilan Pajak | 136
```

Picture 14. SNA Topic 0

ii. Topic 1

The closest unit to attend the meetings regarding topic 1 is the Internal Compliance Division as shown on Picture 15.

```
Top 10 nodes by degree centrality for Topic 1.0:
Bagian Kepatuhan Internal-Biro Umum | 133
Bagian Keuangan-Biro Umum | 132
Bagian Manajemen Barang Milik Negara-Biro Umum | 132
Bagian Pengembangan Strategi dan Kinerja-Biro Manajemen Barang Milik Negara dan Pengadaan | 132
Bagian Administrasi Kementerian-Biro Umum | 132
Bagian Manajemen Hubungan Media dan Kelembagaan Masyarakat-Biro Komunikasi dan Layanan Informasi | 132
Bagian Organisasi, Tata Laksana, dan Komunikasi-Biro Umum | 132
Bagian Manajemen Strategi Komunikasi-Biro Komunikasi dan Layanan Informasi | 132
Bagian Umum-Sekretariat Komite Pengawas Perpajakan | 132
Bagian Perbendaharaan-Biro Perencanaan dan Keuangan | 132
```

Picture 15. SNA Topic 1

iii. Topic 2

For topic 2, the closest unit to attend the meeting is the Administration and Data Management Division as shown on Picture 16.

```
Top 10 nodes by degree centrality for Topic 2.0:
Bagian Tata Usaha dan Pengelolaan Data-Pusat Analisis dan Harmonisasi Kebijakan | 130
Bagian Umum-Sekretariat Pengadilan Pajak | 130
Bagian Kepatuhan Internal-Biro Umum | 126
Bagian Perencanaan dan Pengadaan Sumber Daya Manusia-Biro Sumber Daya Manusia | 126
Bagian Sumber Daya Manusia-Biro Umum | 126
Bagian Manajemen Informasi Sumber Daya Manusia-Biro Sumber Daya Manusia | 125
Bagian Administrasi Putusan, Persidangan, dan Kepatuhan Internal-Sekretariat Pengadilan Pajak | 125
Bagian Akuntansi dan Pelaporan Keuangan-Biro Perencanaan dan Keuangan | 125
Bagian Advokasi IV-Biro Advokasi | 124
Bagian Pengembangan Sumber Daya Manusia-Biro Sumber Daya Manusia | 124
```

Picture 16. SNA Topic 2

The results above show that some units have the closest relation to certain topics in a meeting. This can help the executives to assign units to attend certain meetings in accordance with their main tasks.

3.6. Evaluation

The results of descriptive analysis and modeling show that Monday has the highest number of meetings. There are some meetings held on weekends (Saturday and Sunday) and some meetings are held at 00.00. In addition, the meeting is generally held for 1-2 hours. This slightly exceeds the recommended duration for virtual meeting (below 1 hour) [16]. For LDA analysis, the number of topics determined to produce good model without overlapping topics are 7 topics. The result of the *Degree Centrality* analysis shows that the meeting title can determine the *unit in charge* that has to attend the meeting, so the executives are able to assign the right PIC to attend the meeting with particular topics.

4. DISCUSSION

Pandemic leads meeting to be held virtually. Employees often attend the virtual meeting in quite a long duration, which leads to zoom fatigue [17]. In addition, in order to hold an effective meeting, it is required to identify the meeting trends in certain session (interval). This aims to better manage meeting schedule.

Research in the literature review section indicates LDA to be the method widely used for clustering data. Meanwhile, network analysis uses the Degree Centrality, Closeness Centrality, and Betweenness Centrality.

This research applies CRISP DM methodology with the initial stage of literature review up to evaluation using data from the meeting modul of Office Automation application. The modeling method applied is LDA for clustering data and SNA using Degree Centrality to determine the central unit in a network. The data acquired from the database production server is 59.891 records, which is then cleansed and analyzed. The cleansing and analysis process are conducted in python language. The modeling is evaluated using coherence score, which is currently not at the ideal value. Therefore, it requires further data cleansing to gain better score.

The results of this research can provide recommendations for leaders to determine the appropriate meeting PIC for each unit.

5. CONCLUSION

This research is expected to provide efficiency impacts. Clustering and Network Analysis are expected to provide prediction or recommendation regarding related unit to receive meeting invitation on a particular topic. The result of this research can be an input to the executives in assigning the right PIC to attend certain meetings.

Based on the modeling and testing, it can be concluded that LDA method can be used to identify and cluster topics of the meeting. The number of clustered topics determined in this research are seven topics with coherence score at -3.93. In terms of SNA, the model shows some units with the closest relation of topics. This helps the executives in determining certain unit to be the Unit in Charge in a certain meeting.

Nevertheless, the coherence score shows that the model can be improved in order to cluster the topics better, both in terms of the ideal cluster number and better topics clusters. Furthermore, additional data could be added, such as meeting schedule and minute of meeting, so that there will be more information gained, more accurate topic clusters and better model.

REFERENCES

[1] H. Pratama, M. N. A. Azman, G. K. Kassymova, and S. S. Duisenbayeva, "The Trend in Using Online Meeting Applications for Learning During the Period of Pandemic COVID-19 : A Literature Review," *J. Innov. Educ. Cult. Res.*, vol. 1, no. 2, pp. 58–68, 2020, doi: 10.46843/jiecr.v1i2.15.

[2] K. A. Karl, J. V Peluchette, and N. Aghakhani, "Virtual Work Meetings During the COVID-19 Pandemic : The Good , Bad , and Ugly," *Small Gr. Res.*, vol. 53, no. 3, pp. 343–365, 2022, doi: 10.1177/10464964211015286.

[3] K. R. Khurya and D. Prayoga, "Eye fatigue during pandemic covid-19: literatire review," *J. Ilm. Permas J. Ilm. STIKES Kendal*, vol. 11, no. 4, pp. 515–524, 2021, [Online]. Available: http://journal.stikeskendal.ac.id/index.php/PSKM

[4] J. Yang *et al.*, "Brief introduction of medical database and data mining technology in big data era," *J. Evid. Based. Med.*, vol. 13, no. 1, pp. 57–69, 2020, doi: 10.1111/jebm.12373.

[5] F. Martinez-Plumed *et al.*, "CRISP-DM Twenty Years Later : From Data Mining Processes to Data Science Trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 8, pp. 3048–3061, 2021, doi: 10.1109/TKDE.2019.2962680.

[6] A. Rahman, R. B. Waskitho, M. F. A. U. Nuha, and N. A. Rakhmawati, "Klasterisasi Topik Konten Channel Youtube Gaming Indonesia Menggunakan Latent Dirichlet Allocation," *J. Inf. Eng. Educ. Technol.*, vol. 05, no. 01, pp. 78–83, 2021.

[7] D. Inayah and F. L. Purba, "Implementasi Social Network Analysis Dalam Penyebaran Informasi Virus Corona (Covid-19) di Twitter," *Semin. Nas. Off. Stat.*, vol. 2020, no. 1, pp. 292–299, 2020, doi: 10.34123/semnasoffstat.v2020i1.573.

[8] J. W. Johnsen and K. Franke, "Identifying Proficient Cybercriminals Through Text and Network Analysis," *2020 IEEE Int. Conf. Intell. Secur. Informatics*, pp. 1–7, 2020, doi: 10.1109/ISI49825.2020.9280523.

[9] C. Huang, Y. Guo, W. Guo, and Y. Li, "HackerRank : Identifying key hackers in underground forums," *Int. J. Distrib. Sens. Networks*, vol. 17, no. 5, pp. 1–12, 2021, doi: 10.1177/15501477211015145.

[10] R. Kumari, J. Y. Jeong, and B. Lee, "Topic modelling and social network analysis of publications and patents in humanoid robot technology," *J. Inf. Sci.*, vol. 47, no. 5, pp. 658–676, 2021, doi:

10.1177/0165551519887878.

[11] Afolabi, I. T, and E. E. Israel, "SNA , LDA And Sentiment Analysis For Competitive Intelligence In Agro E-Commerce," *Int. J. Comput. Sci. Eng.*, vol. 11, no. 1, pp. 43–60, 2022.

[12] J. Joung and H. M. Kim, "Automated Keyword Filtering in Latent Dirichlet Allocation for Identifying Product Attributes From Online Reviews," *J. Mech. Des.*, vol. 143, no. 8, pp. 1–6, 2021, doi: 10.1115/1.4048960.

[13] C. Natalia, F. Suprata, F. P. S. Surbakti, and S. Clarence, "Penentuan Standar Spesifikasi Kerja di Café Berdasarkan Big Data dengan Metode LDA dan AHP," *J. Rekayasa Sist. Ind.*, vol. 10, no. 2, pp. 211–225, 2021, doi: 10.26593/jrsi.v10i2.5228.211-226 Jurnal.

[14] N. S. ZD, D. Saepudin, and A. A. Rohmawati, "Prediksi Pergerakan Volatilitas Pasar Saham Berdasarkan Financial News," *e-Proceeding Eng.*, vol. 7, no. 1, pp. 2870–2880, 2020.

[15] H. N. R. Achaqie, "Pemanfaatan Social Network Analysis (SNA) Untuk Analisis Pemetaan Komunikasi Pada Pegawai Bank Syariah di Jawa Tengah," *J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 2, no. 1, pp. 42–49, 2022.

[16] T. Annisa, "10 Rapat yang Efektif dan Tips Mewujudkannya," *EKRUT Media*, 2021. https://www.ekrut.com/media/6-tips-menciptakan-rapat-yang-efektif (accessed Nov. 08, 2022).

[17] G. Fauville, M. Luo, A. C. M. Queiroz, J. N. Bailenson, and J. Hancock, "Zoom Exhaustion & Fatigue Scale," *Comput. Hum. Behav. Reports*, vol. 4, 2021, doi: 10.1016/j.chbr.2021.100119.