Jurnal Linguistik Komputasional
Indonesian Association for Computational Linguistics

JLK
Vol 3, No 2
September
2020

Pemanfaatan *Epistemic Network Analysis* sebagai Pendukung Analisis Sentimen dalam *Collaborative Learning*

Roy Parsaoran, Jonathan Bernad, Tifani Astadini, Hapnes Toba, Maresha C. Wijanto, Mewati Ayub
Fakultas Teknologi Informasi Universitas Kristen Maranatha
Jalan Surya Sumantri no 65 Bandung 40164
1772044@maranatha.ac.id, 1772004@maranatha.ac.id, 1872060@maranatha.ac.id,
hapnestoba@it.maranatha.edu, maresha.cw@it.maranatha.edu, mewati.ayub@it.maranatha.edu

***Abstrak*—Banyak metode *blended learning* telah diterapkan pada sistem pembelajaran modern. Salah satu metode pembelajaran yang paling banyak digunakan adalah *collaborative learning* yang menekankan pembelajaran melalui diskusi kelompok. Data yang direkam selama sesi *collaborative learning* dapat bermanfaat untuk meningkatkan interaksi antar anggota kelas, termasuk dosen. Dengan menggunakan analisis sentimen, pembahasan dapat dikategorikan apakah pembahasan berjalan dengan baik atau tidak, juga dapat diketahui anggota kelompok mana yang paling aktif dan berdampak positif terhadap pekerjaan yang diberikan kepada kelompok. Pada penelitian ini, pendekatan analisis sentimen akan dipadukan dengan *Epistemic Network Analysis* (ENA) sehingga dapat melihat gambaran grafis kontribusi masing-masing anggota dalam diskusi kelompok. Hasil percobaan kami menunjukkan bahwa ENA menampilkan wawasan yang lebih baik tentang aktivitas siswa daripada hanya menggunakan analisis sentimen.**

Kata kunci—analisis jaringan epistemik, analisis sentimen, pembelajaran kolaboratif, klasifikasi *random forest*, penambangan teks

Abstract— A lot of blended learning methods have been applied to modern learning system. One of the most practiced learning methods is collaborative learning, which combines and extends group discussion. The recorded data during a collaborative learning session could be useful to enhanced the interaction among the class members, including the lecturer. Using sentiment analysis, the discussion can be categorized whether the discussion goes well or not, it can also be seen which group members are most active and have a positive impact on the work assigned to the group. In this preliminary research, sentiment analysis approach will be combined with Epistemic Network Analysis (ENA) so that it can see a graphical depiction of each member's contribution in a group discussion. Our experimental results show that ENA displays better insights of the students activities than only using the sentiment analysis.

Keywords—collaborative learning; epistemic network analysis; random forest classifier; sentiment analysis; text mining

I. PENDAHULUAN

Proses pembelajaran merupakan salah satu unsur penting untuk mencapai keberhasilan dalam kegiatan belajar mengajar, khususnya di dunia pendidikan tinggi. Dalam proses pembelajaran itulah terjadi proses transformasi ilmu pengetahuan serta nilai-nilai. Ketika proses pembelajaran berlangsung, terjadi interaksi antara dosen dengan mahasiswa yang memungkinkan bagi dosen untuk dapat mengenali karakteristik serta potensi yang dimiliki mahasiswa. Demikian pula sebaliknya, pada saat pembelajaran mahasiswa memiliki kesempatan untuk mengembangkan potensi yang dimilikinya sehingga potensi tersebut dapat dioptimalkan. oleh karena itu, pendidikan bukan lagi memberikan stimulus akan tetapi usaha mengembangkan potensi yang dimiliki.

Pengetahuan itu tidak sekedar diberikan, akan tetapi dibangun oleh mahasiswa. Untuk dapat mengenali dan mengembangkan potensi mahasiswa tentunya dalam proses pembelajaran perlu pembelajaran yang bersifat aktif. Pembelajaran tidak lagi berpusat pada dosen tetapi berpusat pada mahasiswa dan dosen hanya sebagai fasilitator serta pembimbing. Dengan demikian, mahasiswa memiliki kesempatan yang luas untuk mengembangkan kemampuannya seperti mengemukakan pendapat,berpikir kritis, menyampaikan ide atau gagasan dan sebagainya. Belajar aktif sangat diperlukan oleh siswa untuk mendapatkan hasil yang maksimal. Ketika mahasiswa pasif, atau hanya menerima dari pengajar ada

Submitted 31-08-2020; Revised 23-09-2020; Accepted 27-09-2020

kecenderungan untuk melupakan apa yang telah diberikan pengajar. Ada beberapa faktor yang dapat meningkatkan tercapainya tujuan pembelajaran, salah satunya yaitu keaktifan mahasiswa dalam tugas-tugas kelompok atau sering juga diistilahkan dengan *collaborative learning.*

Setiap anggota kelompok diharapkan dapat saling mendukung satu dengan lainnya sehingga mencapai solusi atau produk bersama yang ditugaskan oleh dosen. Namun tidak jarang, situasi di dalam kelompok tidak seimbang. Hanya satu atau dua orang yang mengerjakan tugas kelompok sendirian. Atau bahkan tidak jarang pula tugas kelompok tersebut menjadi terbengkalai karena satu dengan lainnya saling mengandalkan tanpa adanya komunikasi yang terarah. Sarana komunikasi dan teknologi dalam menjalankan tugas kelompok sebenarnya sudah sangat banyak. Namun demikian tidak selalu hsailnya dapat secara transparan diperlihatkan. Pada riset yang dipaparkan dalam makalah ini, dilakukan analisis sentimen untuk *collaborative learning* untuk dapat kemudian divisualisasikan dalam bentuk *Epistemic Network Analysis* (ENA).

II. KAJIAN LITERATUR

A. *Collaborative Learning*

Model pembelajaran kolaboratif adalah pembelajaran yang dilaksanakan dalam kelompok, namun tujuannya bukan untuk mencapai kesatuan yang didapat melalui kegiatan kelompok. Setiap mahasiswa dalam kelompok didorong untuk menemukan beragam pendapat atau pemikiran. Pembelajaran tidak terjadi dalam kesatuan, namun pembelajaran merupakan hasil dari keragaman atau perbedaan. Pada kegiatan diskusi, mahasiswa dapat melakukan aktivitas seperti menginventarisasi berbagai informasi yang diperlukan, mengkomunikasikan pendapat, menimbang/menerima gagasan orang lain, atau mengambil suatu simpulan.

Mahasiswa yang mengalami kesulitan dapat bertanya baik kepada mahasiswa lain yang lebih pandai maupun kepada dosen. Tujuan model pembelajaran kolaboratif ini adalah untuk meningkatkan kemampuan mahasiswa yang kurang mengerti atau belum memahami suatu materi secara sempurna dan untuk pertukaran maupun interaksi dari sisi pikiran, pendapat dan penafsiran yang berbeda terhadap materi pembelajaran dan tugas yang diberikan. Tujuan tersebut diharapkan dicapai melalui pemberian tugas dan *sharing* antar mahasiswa, khususnya di dalam kelompok. Pemberian tugas dan aktivitas *sharing* menekankan pada pemahaman konsep dan dapat berpengaruh pada materi-materi selanjutnya [1].

B. *Analisis Sentimen*

Analisis sentimen disebut juga *opinion mining*, adalah bidang ilmu yang menganalisa pendapat, sentimen, evaluasi, penilaian, sikap dan emosi publik terhadap entitas seperti produk, jasa, organisasi,individu, masalah, peristiwa, topik, dan atribut mereka [2]. Analisis sentimen berfokus pada opini-opini yang mengekspresikan atau mengungkapkan sentimen positif, netral atau negatif.

Melalui analisis sentimen dapat diketahui bagaimana nilai tanggapan seseorang terkait dengan isu yang sedang dibahas. Dalam situasi pengerjaan tugas secara berkelompok, nilai sentimen juga memungkinkan digunakan untuk melihat performa seorang mahasiswa. Hal ini penting bagi dosen untuk bisa melihat sejauh mana seorang mahasiswa memahami tugas dan perannya dalam penyelesaian tugas kelompok.

C. *Text Mining*

Text mining (penambangan teks) merupakan aktivitas analisis suatu kumpulan teks dengan sumber data yang didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisis keterhubungan, keterkaitan dan kelas antar dokumen Definisi lain, *text mining* melingkupi sebuah proses ekstraksi informasi yang terpola yang berasal dari sejumlah besar sumber data teks, seperti dokumen Word, PDF, kutipan teks, atau bahkan sms (*tweet*).

Pemrosesan *text mining* secara umum dibagi menjadi dua tahap. Pertama diawali dengan merubah data teks tidak terstruktur ke data semi atau terstruktur, dan kedua dilanjutkan dengan melakukan ekstraksi informasi yang diteliti dari data teks terstruktur [2].

Text mining dapat dilakukan dengan membentuk model klasifikasi (*classifier*) ataupun hanya dengan memperhatikan frekuensi (*word count*) dan dilanjutkan dengan melakukan analisis. Penelitian yang dibahas dalam makalah ini menggunakan analisis *text mining* pada data *unstructured* dengan bantuan *open-source R*. Dalam melakukan analisis teks, R merangkai data yang acak sehingga menjadi data semi-struktur atau terstruktur dan data siap dilakukan transformasi [3].

Transformasi teks mengubah teks asli ke dalam bentuk teks yang lebih terstruktur untuk memudahkan proses analisis. *Text mining* untuk data-data dari laman *web* perlu memanfaatkan *Application Programming Interface* (API) untuk mengambil teks yang akan ditambang. Beberapa kendala dalam melakukan *text mining* adalah dalam pemakaian bahasa.

Perangkat-perangkat analisis teks sejauh ini hanya dapat digunakan secara optimal dalam bahasa-bahasa yang sudah mapan, seperti bahasa Inggris. Hal itu juga berpengaruh terhadap penelitian-penelitian mengenai *text mining* berbahasa Indonesia yang sangat jauh lebih sedikit sumber pengolahannya dibandingkan dengan penelitian dengan bahasa yang lain.

D. *Epistemic Network Analysis*

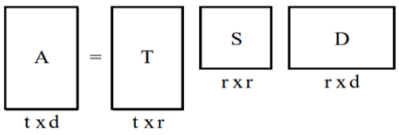
Epistemic Network Analysis adalah metode untuk mengidentifikasi dan mengukur koneksi antar elemen dalam data berkode – misalnya, melalui analisis sentimen – dan merepresentasikannya dalam model visualisasi

jaringan dinamis. Fitur utama dari ENA memungkinkan para peneliti untuk membandingkan jaringan yang berbeda, baik secara statistik maupun ringkasannya yang mencerminkan struktur koneksi yang berbobot [4]. Hasil visualisasi juga dapat memungkinkan pengguna untuk melihat data asli yang berkontribusi pada setiap koneksi dalam representasi jaringan / graf. Dengan demikian ENA dapat dimanfaatkan pula untuk menjawab berbagai pertanyaan penelitian, baik secara kualitatif maupun kuantitatif.

Para peneliti telah menggunakan ENA untuk menganalisis dan memvisualisasikan berbagai fenomena, termasuk: koneksi kognitif yang dibuat siswa sambil memecahkan masalah yang kompleks; interaksi antara berbagai daerah otak dalam data MRI; koordinasi tatapan sosial; integrasi keterampilan operasi selama prosedur bedah; dan berbagai hal lainnya [5].

E. *Singular Value Decomposition* (SVD)

Singular Value Decomposition adalah metode aljabar linier yang memecah matriks *A (terms-documents)* berdimensi *t* x *d* menjadi tiga matriks *TSD*. T adalah matriks kata (*terms*) berukuran *t* x *r*, S adalah matriks diagonal berisi nilai skalar (*eigen values*) berdimensi r x r, dan r ditentukan sebelumnya, dan D adalah matriks dokumen berukuran r x d. Dekomposisi nilai singular dari matriks A dinyatakan sebagai $A = TSD^T$, seperti yang diilustrasikan pada Gambar 1 berikut ini..



Gambar 1 Dekomposisi matriks A dengan SVD menjadi tiga matriks TSD^T

SVD dapat mereduksi dimensi dari matriks A dengan cara mengurangi ukuran r dari matriks diagonal S. Pengurangan dimensi dari matriks S dilakukan dengan cara mengubah semua nilai diagonal matriks S menjadi nol, kecuali untuk nilai diagonal dari dimensi yang tersisa. Pengalian ketiga matriks TSD^T akan membentuk matriks A awal dengan nilai setiap elemennya mendekati nilai sebenarnya [5].

F. *Random Forest Classifier*

Random Forest Classifier merupakan sebuah pendekatan dalam pembelajaran mesin. Teknik ini diperkenalkan oleh Leo Breiman dengan penelitiannya yang terbit pada tahun 2001 [6]. *Random forest* adalah sebuah metode pembelajaran *ensemble* yang menghasilkan beberapa pohon keputusan setiap kali dilakuan pelatihan. Hutan acak bekerja dengan menumbuhkan banyak pohon klasifikasi, sehingga pohon-pohon ini dapat diibaratkan sebagai hutan, lalu untuk objek baru diklasifikasikan dari masukan berupa sebuah vektor.

Metode ini akan menempatkan masukan vektor ke masing-masing pohon yang ada di dalam hutan. Setiap pohon akan memberikan klasifikasi dan penilaian untuk kelas yang dilatih. Hutan akan memilih klasifikasi yang memiliki penilaian paling banyak dari keseluruhan pohon yang ada (melalui penerapan konsep voting). Setiap pohon akan tumbuh seperti sebagai berikut :

1) Jika jumlah dalam kasus pada set latihan adalah N, maka sampel kasus N secara acak dengan pergantian dari data asli. Contoh ini akan menjadi himpunan pelatihan untuk pertumbungan jumlah pohon.
2) Jika ada variabel masukan M, sejumlah m<<M ditentukan sehingga pada setiap node, variabel m yang dipilih secara acak dari M dan perpecahan terbaik untuk m ini digunakan untuk berbagi *node*. Nilai m tetap konstan selama pertumbuhan hutan.
3) Setiap pohon akan tumbuh sebesar dan sejauh mungkin.

Tingkat kesalahan yang terjadi pada teknik *random forest* tergantung pada dua hal, sebagai berikut:

1) Korelasi antara dua pohon yang berada di hutan. Meningkatnya korelasi dapat juga meningkatkan tingkat kesalahan atau *error* pada hutan.
2) Kekuatan individu pada masing-masing pohon dalam hutan. Sebuah pohon dengan tingkat kesalahan yang rendah adalah penggolong (*classifier*) yang kuat. Meningkatnya kekuatan individu pohon dapat mengurangi tingkat kesalahan pada hutan [6].

G. *Confusion Matriks*

Pada dasarnya *confusion matrix* memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi sebenarnya. *Confusion matrix* berbentuk tabel matriks yang menggambarkan kinerja model klasifikasi pada serangkaian data uji yang nilai sebenarnya diketahui. Tabel I merupakan *confusion matrix* dengan 4 kombinasi nilai prediksi dan nilai aktual yang berbeda [7], [8].

TABEL I
CONFUSION MATRIKS

	0	1
0	*True Negative*	*False Negative*
1	*False Negative*	*True Positif*

True Positive (TP) dan *True Negative* (TN) adalah keadaan pada saat hasil sesuai dengan kondisi sebenarnya yang terjadi. *False Positive* (FP) dan *False Negative* (FN) adalah keadaan dimana hasil *outcome* tidak sesuai dengan kondisi yang sebenarnya terjadi. Berdasarkan *confusion matrix* kemudian dilakukan penghitungan *accuracy, precision, recall,* dan *f-measure* [9], [10], [11].

III. METODOLOGI PENELITIAN

Dalam bagian ini dijelaskan perencanaan yang lebih terperinci terhadap kerangka pemikiran yang telah dituliskan dalam bagian pendahuluan.

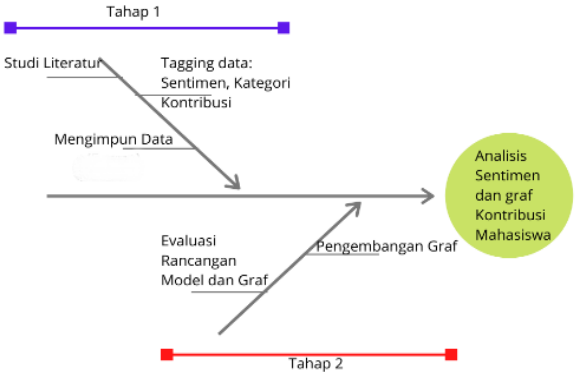
A. *Metodologi*

Secara garis besar penelitian yang direncanakan terbagi ke dalam dua tahapan besar. Gambar 2 memberikan ilustrasi pentahapan penelitian tersebut.

1) Tahap Pertama

Tahap pertama adalah tahap penghimpunan dan *tagging* data, meliputi kegiatan sebagai berikut:

- Melakukan Pengimpunan data melalui riwayat diskusi kelompok mata kuliah ERP dan Manajemen Proyek
- Melakukan ekstraksi kalimat dan kata-kata dari riwayat percakapan
- Melakukan *tagging* data seperti dicontohkan pada Tabel III.



Gambar 2 Metode Penelitian

Dalam Tabel III, secara umum dapat dilihat bahwa mahasiswa A dan B memiliki niat dan kontribusi yang lebih baik dibandingkan C. Ada terlihat porsi kontribusi yang menyatakan keunggulan maupun kelemahan (berdasarkan sentimen: positif atau negatif) dari tiap mahasiswa tersebut, sesuai 'kategori kontribusi' yang ditetapkan oleh dosen dalam rubrik penilaian, misalnya untuk contoh skenario ini adalah: teknologi, manajemen proyek dan sikap. Cara pengelompokan dapat dilakukan secara fleksibel tergantung pada kesepakatan penerapannya. Konsekuensinya adalah diperlukannya model yang mewakili pengelompokan tersebut.

Sampai tahap ini yang didapat adalah nilai sentimen dan kategori dari tiap kalimat diskusi. Tiap kalimat dapat dikategorikan tingkah laku ataupun proyek dll, lalu dapat juga ditentukan apakah kalimat tersebut termasuk dalam kelompok positif manajemen proyek, positif teknologi, positif sikap, negatif manajemen proyek, negatif teknologi, atau negatif sikap.

TABEL III
CONTOH *TAGGING* DATA

Mahasiswa	Isi Percakapan	Klasifikasi
A	*Lu buka drive new folder, folder nya lu share ke email" kita terus bikin google word di folder itu*	Positif (dari kata "*terus bikin*") Isu 'Teknologi (dari kata "*google word*")
B	*Gw masih ga ngerti ini teh proyek nya bener" kita bikin sampe jadi?*	Negatif (dari kata "*ga ngerti*") Isu Manajemen proyek (dari kata "*proyek*")
C	*Yu semangat teman-teman!*	Positif (dari kata "*yu*") Isu Sikap (dari kata "semangat teman-teman"

2) Tahap Kedua

Tahap Kedua adalah Evaluasi model dan pembuatan graf, meliputi kegiatan sebagai berikut:

- Melakukan pembuatan model dan graf dari *dataset*.
- Evaluasi rancangan model dan graf kemudian melakukan pengembangan graf tersebut.

B. *Data Penelitian*

Untuk tahap pertama penelitian akan difokuskan pada data riwayat percakapan kelompok pada mata kuliah ERP dan Manajemen Proyek pada sesi perkuliahan semester genap tahun akademik 2019/20. Data diambil dari berbagai media komunikasi dalam grup WhatsApp yang dihadiri oleh mahasiswa (anggota kelompok) dan asisten dosen mata kuliah terkait.

C. *Evaluasi dan Analisis*

Evaluasi keberhasilan dilakukan dengan menggunakan metrik pengukuran akurasi dari hasil analisis sentimen dan kategori kontribusi untuk setiap kalimat percakapan. Pertanyaan-pertanyaan pendukung riset dan menjadi bagian evaluasi mencakup pada:

- Berapa banyaknya hasil yang akurat tiap kalimatnya?
- Berapa hasil akhir dari analisis sentimen dalam satu diskusi?
- Berapa akuratkah grafik ENA dari pembentukan analisis sentimen dan kategori kontribusi?

IV. ULASAN PENELITIAN

A. *Pengolahan Data*

Data yang digunakan pada penelitian ini adalah data yang diambil dari riwayat pembicaraan kelompok pada tugas besar mata kuliah Manajemen Proyek. Terdapat sejumlah 476 teks dari riwayat diskusi dalam grup WhatsApp. Data tersebut diambil selama periode 1 semester ganjil 2019/2020. Setelah mengambil data percakapan lalu data tersebut dilakukan proses *tagging data* menjadi positif, negatif atau netral yang memiliki jumlah seperti pada Tabel IV.

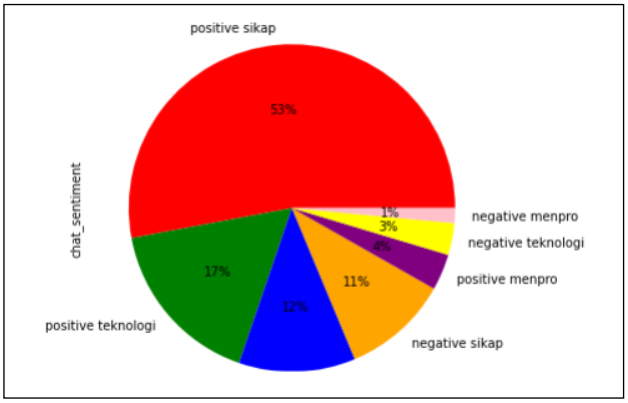
TABEL IV
JUMLAH DATA PENELITIAN

	Manajemen Proyek	Teknologi	Sikap
Positif	171	80	251
Negatif	7	15	50
Netral	55		

Kelompok teknologi merupakan riwayat pembicaraan yang membahas mengenai teknologi terbaru ataupun teknologi yang akan digunakan dalam diskusi kelompok, kelompok manajemen proyek atau manajemen proyek merupakan riwayat pembicaraan yang membahas mengenai pengelolaan proyek dalam diskusi kelompok tersebut dan kelompok sikap merupakan riwayat pembicaraan yang membahas mengenai sikap seseorang dalam merespon diskusi lainnya. Terdapat beberapa teks yang sulit dan tidak bisa diklasifikasikan yaitu kata-kata yang tidak memiliki makna dan tujuan.

B. *Pembentukan Model Klasifikasi*

Analisis statistik awal, dari hasil *tagging* manual dalam Gambar 3, menunjukkan bahwa nilai sentimen dalam kategori 'positif sikap' merupakan jumlah terbanyak (53%), diikuti dengan 'positif teknologi' (17%) dan gabungan lainnya (30%). Hal ini menunjukkan bahwa dalam berkomunikasi, para mahasiswa telah memperhatikan tata krama dalam interaksi kelompok.



Gambar 3. Grafik Perbandingan klasifikasi

Hasil prediksi model *random forest* untuk *dataset*

```
> datanew = Chat_History
> units = datanew[,c("Condition","Nama")]
> conversation =
datanew[,c("Condition","GroupName")]
> head(conversation)
> codeCols = c(
'Data','Positive.Menpro','Positive.Teknologi','
Positive.Sikap','Negative.Menpro','Negative.Tek
nologi','Negative.Sikap')
> codes = datanew[,codeCols]
> head(codes)
```

pengujian dapat dilihat pada Gambar 4. Proporsi data pelatihan dan pengujian adalah 80%:20% yang dilakukan secara acak dari 476 data. Akurasi umum yang dicapai dalam model adalah 57,29%. Kesulitan terbesar dalam pembentukan model adalah proses pembersihan data untuk kata-kata informal, seperti: *luh, loe, gue, gw, okay, okei,* dan sejenisnya.

	precision	recall	f1-score	support
negative sikap	0.50	0.08	0.14	12
negative teknologi	0.00	0.00	0.00	1
neutral	0.00	0.00	0.00	8
positive menpro	0.00	0.00	0.00	4
positive sikap	0.57	0.89	0.70	53
positive teknologi	0.78	0.39	0.52	18
accuracy			0.57	96
macro avg	0.31	0.23	0.23	96
weighted avg	0.52	0.57	0.50	96

0.5729166666666666

Gambar 4. Confusion matrix hasil klasifikasi *random forest* pada *dataset* pengujian

Selain kata-kata informal, seringkali muncul pula penggunaan singkatan-singkatan untuk kata-kata yang terbilang penting, misalnya yang terkait teknologi: *gdrive, drive, lib (library), pack (package)*, dll. Ketidakseragaman penulisan menjadi isu penting untuk dapat melakukan analisis sentimen secara baik. Model yang dikembangkan saat ini masih perlu dilengkapi dengan proses pembersihan teks dan normalisasi, khususnya untuk kata-kata pendukung kegiatan kerja kelompok dalam pembelajaran kolaboratif.

C. *Penerapan ENA*

Salah satu kelebihan ENA dalam melakukan analisis adalah adanya visualisasi graf berupa arah vektor, yang menggambarkan kecenderungan (berupa rata-rata nilai vektor) pada *dataset* sesuai kategori yang dianalisis. Dengan menggunakan ENA, analisis bukan dilakukan untuk memprediksi kategorinya, namun untuk menggambarkan 'kekuatan' kecenderungannya. Misalnya untuk seorang mahasiswa akan dapat dilihat 'kekuatan' yang dimilikinya, apakah dalam bidang teknologi, sikap ataupun kemampuan pengelolaan proyeknya. Untuk melakukan ENA, terdapat beberapa tahap hingga menjadi sebuah visualisasi [12], yaitu:

1. Indentify Columns to Accumulate

Sebelum menjalankan fungsi *ena.accumulate.data*, pertama-tama perlu mengindetifikasi setiap kolom dari data untuk digunakan sebagai unit, *conversation*, dan *codes*. Proses akumulasi sebenarnya membutuhkan kerangka data individual untuk masing-masing kategori (baris kelima Kode Program 1). Peneliti perlu mengelompokkan *dataset* menggunakan kolom yang diidentifikasi, dengan menggunakan Kode Program 1.

```
> datanew = Chat_History
> units = datanew[,c("Condition","Nama")]
> conversation =
datanew[,c("Condition","GroupName")]
> head(conversation)
> codeCols = c(
'Data','Positive.Menpro','Positive.Teknologi',
'Positive.Sikap','Negative.Menpro','Negative.T
eknologi','Negative.Sikap')
> codes = datanew[,codeCols]
> head(codes)
```

Kode Program 1. Identifikasi Kolom Sentimen Hasil *Tagging*

2. *Run the ENA Accumulation*

Dengan *dataset* yang telah teridentifikasi dan diberikan *tagging*, perlu ditunjukkan ukuran jumlah baris percakapan (stanza.window) yang akan dianalisis kecenderungannya. Pada percobaan ini digunakan ukuran *window* = 4, yang menunjukkan kemunculan suatu percakapan dan 3 baris sebelumnya, dengan menggunakan Kode Program 2.

```
> accum = ena.accumulate.data(units = units,
conversation = conversation,codes =
codes,window.size.back = 4)
```

Kode Program 2. *Run ENA Accumulation*

3. Generate the ENA Set

Menghasilkan model ENA dengan membangun pengurangan dimensi *adjacency vector* dalam objek data ENA , dengan menggunakan Kode Program 3.

```
set = ena.make.set(enadata = accum)
```

Kode Program 3. *Generate ENA Set*

4. Plot Units in Each Group

Melakukan *plotting unit* dari setiap kondisi (kategori kontribusi) dengan warna yang berbeda. Hal ini dilakukan dengan menggunakan Kode Program 4.

```
> first =
as.matrix(set$points$Condition$first)
> plot = ena.plot(set, scale.to =
"network", title = "Groups of Units")
> plot = ena.plot.points(plot, points =
first, confidence.interval = "box",
colors = c("red"))
```

Kode Program 4. *Plotting Unit*

5. Plotting Means

Plotting means dari grup unit dapat diselesaikan dengan fungsi *ena.plot.group* dan dengan melewati set point yang sama, fungsi tersebut akan menghitung rata-rata untuk *plot* , dengan menggunakan Kode Program 5.

```
> plot = ena.plot(set, scale.to = list(x=-
1:1, y=-1:1), title = "Groups and Means")
> plot = ena.plot.points(plot, points =
first, confidence.interval = "box", colors =
c("red"))
> plot = ena.plot.points(plot, points =
second, confidence.interval = "box", colors =
c("blue"))
> plot = ena.plot.group(plot, point = first,
colors =c("red"), confidence.interval =
"box")
> plot = ena.plot.group(plot, point = second,
colors =c("blue"), confidence.interval =
"box")
> plot$plot
```

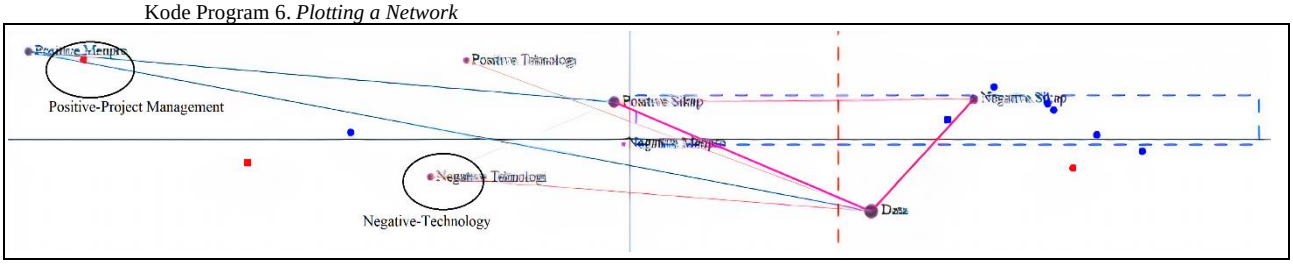
Kode Program 5. *Plotting Means*

6. *Plotting a Network*

Untuk menampilkan sebuah *network* yang mengaitkan antar kategori kontribusi dan mahasiswa dapat digunakan fungsi *ena.plot.network*. Fungsi ini memerlukan satu parameter, selain parameter *standard* awal plot, yang merupakan parameter *'network'* (keterhubungan antar kategori kontribusi) pada visualisasi, dengan menggunakan Kode Program 6.

```
> first.lineweights =
as.matrix(set$line.weights$Condition$first)
> second.lineweights =
as.matrix(set$line.weights$Condition$second)
> first.mean =
as.vector(colMeans(first.lineweights))
> second.mean =
as.vector(colMeans(second.lineweights))
> subtracted.mean = first.mean - second.mean
> head(first.mean, 5)
> head(second.mean, 5)
> head(subtracted.mean, 5)
> plot.first = ena.plot(set, title = "First")
> plot.first = ena.plot.network(plot.first,
network = first.mean)
> plot.first$plot
```

Kode Program 6. *Plotting a Network*



Gambar 5. Hasil ENA yang menggambarkan keterkaitan antara kategori kontribusi dalam bentuk graf beserta kecenderungan analisis sentimennya dalam *dataset* penelitian

7. *Plot Everything Together*

Keseluruhan bagian network digabungkan dengan menggunakan *library magrittr*. *Library magrittr* meneruskan hasil dari satu fungsi dengan menggunakan *forward-pipe* operator `%>%` (dalam bahasa R), sebagai parameter pertama ke fungsi berikutnya. Ini menghilangkan kebutuhan untuk menyediakan parameter *plot* untuk setiap pemanggilan, dengan menggunakan Kode Program 7.

```
> library(magrittr)
> library(scales)
> point.max = max(first, second)
> first.scaled = scales::rescale(first,
+ c(0,max(as.matrix(set$rotation$nodes))),
c(0,point.max))
> second.scaled = scales::rescale(second,
c(0,max(as.matrix(set$rotation$nodes))),
c(0,point.max))
> plot = ena.plot(set, title = "Plot with
Units and Network", font.family = "Times")
%>% ena.plot.points(points = first.scaled,
colors = c("red"))
%>% ena.plot.points(points
= second.scaled, colors = c("blue"))
%>% ena.plot.group(point = first.scaled,
colors = c("red"), confidence.interval =
"box")
%>% ena.plot.group(point = second.scaled,
colors = c("blue"), confidence.interval =
"box")
%>% ena.plot.network(network =
subtracted.mean)
> plot$plot
```

Kode Program 7. *Plot Everything Together*

V. Diskusi Hasil Penelitian

Setelah melalui tahap-tahap pengembangan yang disampaikan dalam bagian sebelumnya, metode ENA akan menghasilkan visualisasi sebagaimana diperlihatkan melalui Gambar 5. Setiap kategori kontribusi dalam *dataset* lihat contoh dalam Tabel III), yaitu: positif / negatif-manajemen proyek, positif / negatif-teknologi, dan positif / negatif-sikap akan menempati ruang vektor yang telah diproyeksikan ke dalam ruang dua dimensi, sebagai *node*, melalui hasil perhitungan SVD [5], [13], [14]. Hal ini sebenarnya menunjukkan bagaimana kemunculan kata-kata (*word co-occurrences*) dihitung nilai relasinya hingga membentuk topik dalam proses SVD tersebut, dan digambarkan sebagai sebuah *edge* dalam jaringan / graf kategori kontribusi.

Graf keterhubungan hasil visualisasi ENA untuk setiap kategori kontribusi direalisasikan dalam bentuk vektor dengan masing-masing kecenderungan arahnya. Dalam Gambar 5, kotak biru dan merah yang bergaris putus-putus adalah ruang 'keyakinan' dari kumpulan teks yang mengarah pada sebuah kategori kontribusi. Sebagai contoh, ruang keyakinan kalimat-kalimat bernuansa positif-manajemen proyek mengambil area yang cukup besar, dengan melingkupi kuadran 2 dan 3 (kiri atas, dan kiri bawah). Angka koordinat pada sumbu x dan y menunjukkan bobot (atau 'ketebalan') dari sebuah topik. Sumbu x dan y dapat diatur untuk memperlihatkan kecenderungan terhadap suatu kategori. Dalam Gambar 5, sumbu x menunjukkan arah sentimen (`first.scaled` pada Kode Program 7), dan sumbu y menunjukkan arah kategori kontribusi (`second.scaled` pada Kode Program 7).

Jika dibandingkan dengan hasil evaluasi model analisis sentimen pada Gambar 4, pembentukan visualisasi ENA memperlihatkan hubungan antara kategori kontribusi, yang lebih mendalam, daripada hanya sekedar memberikan klasifikasi sebuah kalimat sebagai positif atau negatif. Misalkan diketahui hasil prediksi terhadap kalimat *'Lu buka drive new folder, folder nya lu share ke email" kita terus bikin google word di folder itu'*, adalah positif-teknologi (lihat *tagging* dalam Tabel III), maka jika membandingkan hasil tersebut dengan Gambar 5, secara implisit dapat diharapkan pula bahwa mahasiswa yang menuliskan komentar tersebut memiliki kecenderungan untuk bersikap positif dan memiliki pemahaman yang baik untuk pengelolaan tugas (manajemen proyek). Ketiga kategori kontribusi tersebut memiliki arah vektor yang sama pada kuadran 2, di area kiri atas. Sebagai catatan arah vektor pada sumbu x dan y dapat diatur melalui pengubahan urutan *plotting area* dalam Kode Program 7.

Pada Gambar 5, warna merah menunjukan hasil ENA pada *dataset* percakapan kelompok matakuliah ERP sedangkan warna biru menunjukan hasil ENA pada *dataset* percakapan kelompok matakuliah Manajemen proyek. Hasil ENA percakapan matakuliah Manajemen proyek

memperlihatkan bahwa hasil diskusi tersebut lebih dominan pada positif-manajemen proyek dan negatif-sikap. Percakapan matakuliah ERP yang telah diproses oleh ENA, dominan dalam positif-sikap dan negatif-manajemen proyek. Diskusi kelas ERP pun memiliki area yang lebih besar pada grafik ENA menandakan ketersebaran kategori klasifikasi percakapan dengan lingkup percakapan yang lebih luas dibandingkan dengan diskusi kelas Manajemen proyek.

VI. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilaksanakan, metode ENA dapat menunjukkan kecenderungan kategori kontribusi yang terdapat dalam sebuah kumpulan teks. Pendekatan ENA mengubah hasil riwayat percakapan ke dalam bentuk reduksi matriks kata, dengan menggunakan SVD. Hasil reduksi tersebut lalu diproses menjadi grafik sebagai visualisasinya. Hasil visualisasi ENA dapat memperlihatkan keterhubungan antara komponen-komponen kategori kontribusi yang ditelaah dalam sebuah proses analisis. Hal ini akan sangat bermanfaat untuk melengkapi hasil dalam model klasifikasi biasa, seperti halnya dalam analisis sentimen, yang langsung memberikan prediksi tanpa memberikan kecenderungan yang terbentuk antar kategorinya dalam sebuah *dataset*.

Sebagai rencana pengembangan, hasil visualisasi ENA akan dibentuk untuk dapat memberikan graf keterhubungan antar mahasiswa. Dengan demikian karakteristik mahasiswa dapat divisualisasikan melalui kata-kata yang telah dituliskannya dalam sebuah diskusi, dan juga kecenderungan kontribusinya dalam kelompok kerja.

UCAPAN TERIMA KASIH

Penelitian ini didukung oleh hibah Kemenristek/BRIN nomor SK pelaksanaan 8/E1/KPT/2020 tertanggal 24 Januari 2020 melalui nomor kontrak 018/SP2H/AMD/LT-AMAND/LL4/2020.

DAFTAR PUSTAKA

[1] Dewi, M.R., Mudakir, I. and Murdiyah, S. Pengaruh Model Pembelajaran Kolaboratif berbasis Lesson Study terhadap Kemampuan Berpikir Kritis Siswa. *Jurnal Edukasi*, 3 (2), pp.29-33, 2016.

[2] B. Liu. *Sentiment Analysis and Opinion Mining*. USA: Morgan & Claypool Publishers, 2012.

[3] Hartanto, H. Text Mining dan Sentimen Analisis Twitter pada Gerakan LGBT. *Intuisi: Jurnal Psikologi Ilmiah*, 9 (1), pp.18-25, 2017.

[4] Mello, R.F. and Gašević, D., 2019, October. What is the effect of a dominant code in an epistemic network analysis?. In *International Conference on Quantitative Ethnography* (pp. 66-76). Springer, Cham.

[5] Shaffer, D.W., Collier, W. and Ruis, A.R. A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3 (3), pp.9-45, 2016.

[6] Breiman, L.. Random forests. *Machine learning*, 45 (1), pp.5-32, 2001.

[7] Arifin, O. and Sasongko, T.B. Analisa Perbandingan Tingkat Performansi Metode Support Vector Machine dan Naïve Bayes Classifier Untuk Klasifikasi Jalur Minat SMA. *SEMNASTEKNOMEDIA ONLINE*, 6 (1), pp.1-2, 2018.

[8] Sari, A.P., Saptono, R. and Suryani, E. The Implementation of Jaro-Winkler Distance and Naive Bayes Classifier for Identification System of Pests and Diseases on Paddy. *ITSMART: Jurnal Teknologi dan Informasi*, 7 (1), pp.1-7, 2018.

[9] Lestari, A.R.T., Perdana, R.S. and Fauzi, M.A. Analisis Sentimen Tentang Opini Pilkada DKI 2017 pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes dan Pembobotan Emoji. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN, 2548, p. 964X. 2017.

[10] Ling, J., Kencana, I.P.E.N. and Oka, T.B. Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, 3(3), pp.92-99, 2014.

[11] Sari, R. and Hayuningtyas, R.Y. Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website. *Indonesian Journal on Software Engineering (IJSE)*, 5(2), pp.51-60, 2019.

[12] Marquart, C.L. (2019). rENA: Epistemic Network Analysis. [ONLINE] Available at: https://cran.r-project.org/web/packages/rENA/index.html.

[13] Leydesdorff, L. and Welbers, K. The semantic mapping of words and co-words in contexts. *Journal of Informetrics*, 5(3), pp.469-475, 2011.

[14] Swiecki, Z. and Shaffer, D.W. iSENS: an integrated approach to combining epistemic and social network analyses. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 305-313), 2020.