

Analisis Review Siswa Selama Pembelajaran pada Masa Pandemi Menggunakan Metode Topic Modelling LDA

Ari Eka Prasetyanto¹, Kusri², Anggit Dwi Hartanto³

^{1,2,3}Teknik Informatika, Fakultas Teknik, Universitas Amikom Yogyakarta

E-mail: ¹arieka@students.amikom.ac.id, ²kusri@amikom.ac.id, ³anggit@amikom.ac.id

Abstrak – Dalam meningkatkan kualitas pendidikan yang semakin baik sekolah dapat mendapatkan masukan dari berbagai pihak baik itu orang tua murid, guru dan siswa. Sekolah dapat memanfaatkan e-learning dalam mengumpulkan pendapat siswa di sekolah. Namun dari hasil pendapat siswa sekolah masih kesulitan dalam menyimpulkan topik apa yang paling banyak di bahas oleh siswa. Dalam penelitian ini penulis mencoba untuk melakukan pemodelan topik dengan menggunakan LDA untuk studi kasus di SMK Negeri 1 Ngasem dalam proses pembelajaran selama pandemic covid-19. Dari hasil pengambilan dataset review pendapat siswa terdapat 1619 review yang kemudian dilakukan pemodelan topik dengan hasil bahwa topik “pembelajaran” yang banyak di review oleh siswa.

Kata Kunci — Topic Modelling, Latent Dirichlet Allocation, LDA, Text Mining, e-learning

1. PENDAHULUAN

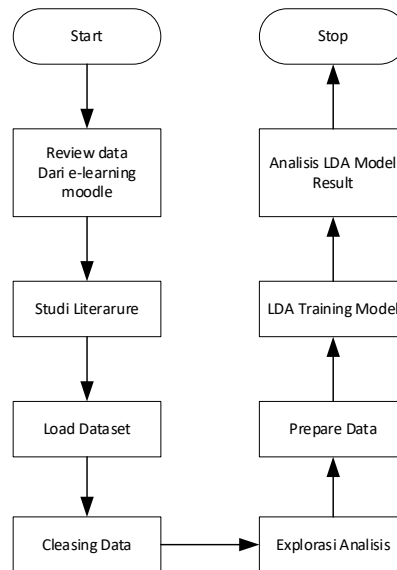
Pandemi yang dialami dunia akibat covid 19 telah membawa peradaban baru bagi kehidupan seluruh umat manusia. Banyak kebiasaan masyarakat yang mulai bergeser menjadi serba digital. Wajah dunia pendidikan juga mengalami banyak perubahan. Disaat kegiatan belajar mengajar secara luring harus dihentikan secara total semenjak pandemi melanda, maka secara otomatis proses pembelajaran dilakukan secara *full* daring. Seperti yang kita ketahui bersama, hal tersebut sudah berjalan selama 2 tahun terakhir dengan berbagai cerita yang mewarnainya, baik suka maupun duka.

Berdasarkan hal di atas, maka pada kesempatan kali ini penulis ingin melakukan *topic modeling* pada hasil *review* pendapat siswa di elearning yang digunakan sekolah, terkait pembelajaran pada masa pandemi. Menurut Blei et al. (2003) dalam Xie dan Xing (2013), topic modeling adalah model yang dibentuk untuk memodelkan teks dan mendeteksi topik yang belum diketahui dari sebuah dokumen [1]. Dengan adanya analisis ini, diharapkan akan bisa memberikan bahan untuk pihak manajemen sekolah dalam mengambil keputusan terkait inovasi pembelajaran yang akan dilakukan kedepannya.

LDA (*latent Dirichlet allocation*) adalah model yang dibentuk untuk mencari topik di dalam dokumen dan peluang kemunculan kata dalam sebuah topik. Model ini menghasilkan keluaran berupa campuran topik di dalam sebuah dokumen yang menyebar Dirichlet. Memodelkan LDA dengan data Twitter untuk mengidentifikasi respon masyarakat New Delhi terhadap polusi udara [2] sedangkan memodelkan LDA dengan data Weibo untuk mengidentifikasi respon masyarakat China terhadap Covid-19 [3].

2. METODE PENELITIAN

Data yang digunakan dalam penelitian ini berupa review siswa dalam pembelajaran selama masa pandemi covid-19 yang dilakukan di e-learning moodle. Review siswa dilakukan dengan memberikan komentar atau tanggapan terkait pembelajaran selama pandemi di SMKN 1 Ngasem kabupaten Kediri. Kemudian dalam hasil review siswa dilakukan studi literatur, proses pengumpulan data, preprocessing data, melakukan pemodelan LDA, melakukan analisis dan evaluasi melalui beberapa tahapan seperti gambar1 berikut: [4]



Gambar 1. Tahap Penelitian

Langkah awal yang harus dilakukan pada penelitian ini yaitu melakukan sebuah studi literatur, yang dimana melakukan studi dari jurnal, internet dan beberapa media lainnya yang membahas mengenai topik modelling LDA. Semua informasi yang telah didapatkan digunakan sebagai acuan teori serta pembahasan dalam penelitian ini.

2.1 Pengumpulan Data

Proses pengumpulan data, peneliti menggunakan data primer dalam proses pengumpulan data. Data diperoleh langsung dari hasil review siswa di SMKN 1 Ngasem Kabupaten Kediri. Untuk mendapatkan dataset yaitu dengan pengambilan data langsung di aplikasi e-learning. Dataset yang diperoleh yakni dari tanggal 21 Januari 2022 hingga 24 Januari 2022. Jumlah dataset yang didapatkan yaitu sebanyak 1619 data review siswa (table 1). Variable data yang digunakan sebanyak 5 variabel diantaranya yakni *Surname*, *First name*, *Email address*, *Started on*, *Response*.

Tabel 1. Data Review Siswa

| No | Kompetensi Keahlian | Kelas | Jumlah review |
|--------|--|-------|---------------|
| 1 | TKRO (Teknik Kendaraan Ringan Otomotif) | X | 84 |
| | | XI | 72 |
| | | XII | 72 |
| 2 | Tata Boga | X | 90 |
| | | XI | 96 |
| 3 | TKJ (Teknik Komputer dan Jaringan) | XII | 79 |
| | | X | 103 |
| | | XI | 95 |
| 4 | Multimedia | XII | 61 |
| | | X | 99 |
| | | XI | 95 |
| 5 | DPIB (Desain Pemodelan & Informasi Bangunan) | XII | 61 |
| | | X | 93 |
| | | XI | 84 |
| 6 | TPM (Teknik Pemesinan) | XII | 42 |
| | | X | 91 |
| | | XI | 65 |
| 7 | Perhotelan | XII | 59 |
| | | X | 64 |
| | | XI | 59 |
| Jumlah | | XII | 55 |
| | | | 1619 |

2.2 Pre Pemrosesan Data

Data *preprocessing* merupakan tahap awal dalam *text mining* yang bertujuan untuk mengubah teks dari bahasa manusia menjadi format yang dikelola mesin, menyusun teks tidak terstruktur, dan memelihara kata kunci yang berguna untuk mewakili topik [5]

Pada pemrosesan data dilakukan *cleaning dataset* yang bertujuan untuk menghapus data yang tidak diperlukan. Pada langkah yang selanjutnya yakni dengan melakukan *exploratory data* dengan tujuan untuk dapat memverifikasi data yang akan diperlukan sehingga sudah siap pada pemodelan dengan *wordcloud*

2.3 Latent Dirichlet Allocation (LDA)

LDA adalah dalam model probabilistik dalam topic modelling pada data teks untuk mendapatkan informasi berupa model topik [6]. Pemodelan topik digambarkan pada dokumen yang secara semantic dan tersembunyi dalam sebuah teks yang jumlahnya besar serta menemukan sebuah informasi dari data teks yang tidak terstruktur. Pada LDA sendiri merupakan sebuah dokumen yang dapat direpresentasikan untuk model campuran pada berbagai topik yang disebut sebagai laten, yang merupakan topik yang dapat dikarakteristik oleh kata. Proses LDA diawali dengan konversi dokumen ke dalam bentuk *dictionary* dan mengkonversi *dictionary* ke dalam bentuk matriks dokumen atau corpus, kemudian membentuk model topik dengan menggunakan algoritma LDA.

Dalam Memodelkan data dengan LDA dapat dilakukan dengan dua pemodelan dengan penentuan jumlah topik yang berbeda [1]. Langkah-langkah pemodelan LDA dijelaskan sebagai berikut:

- a. Membentuk Bag of Words (BoW) [7] per kata dari data *review*. Kata diasumsikan sebagai satu kosakata atau dua kosakata (frasa).
- b. Memodelkan LDA dengan Gibbs *sampling* pada setiap kata dalam data. Algoritma Gibbs *sampling* sebagai berikut:
 - i. Labelkan setiap kata di setiap *review* dengan sebuah topik secara acak.
 - ii. Pada kata ke-*i* dalam *review* ke-*k*, keluarkan kata tersebut dari model BoW lalu hitung:

$$P(z_i = j, x_i = k | w_i = m, z_{-i}, x_{-i}) = \frac{C_{mj}^{WT} + \beta}{\sum_{n=1}^V C_{nj}^{WT} + V\beta} \frac{C_{kj}^{DT} + \alpha}{\sum_{n=1}^T C_{kn}^{DT} + T\alpha} \dots\dots\dots(1)$$

Dengan:

- $(z_i = j, x_i = k)$: penempatan kata ke-*i* pada topik ke-*j* dan *review* ke-*k*
- $w_i = m$: observasi dimana kata ke-*i* adalah kata ke-*m* di dalam daftar kata
- z_{-i}, x_{-i} : seluruh penempatan topik dan *review* tanpa termasuk kata ke-*i*
- C_{mj}^{WT} : jumlah kemunculan kata ke-*m* ditempatkan pada topik ke-*j* tanpa termasuk kata ke-*m*.
- C_{kj}^{DT} : jumlah kemunculan topik ke-*j* pada *review* ke-*k* tanpa termasuk kata ke-*m*
- α : parameter sebaran *Dirichlet* distribusi topik terhadap *review*
- β : parameter sebaran *Dirichlet* distribusi kata terhadap topik
- V : jumlah kata
- T : jumlah topik

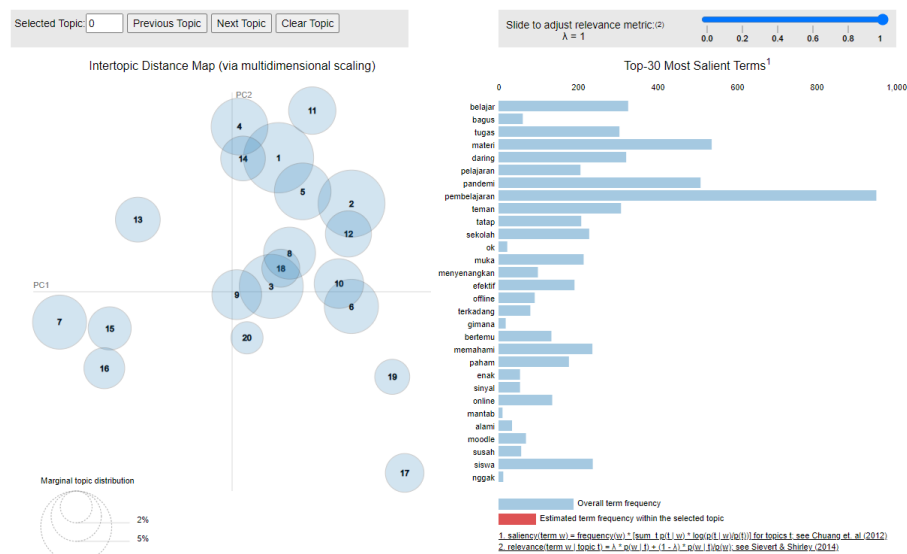
- iii. Tempatkan kata ke-*i* ke topik yang memiliki nilai $P(z_i = j, x_i = k | w_i = m, z_{-i}, x_{-i})$ terbesar.
- iv. Ulangi langkah ii dan iii pada setiap kata di seluruh *review*.
- v. Ulangi langkah iv sebanyak iterasi yang diinginkan

- c. Menentukan matriks kata-topik dan dokumen-topik
- d. Menghitung jumlah total dari distribusi kata-topik dan dokumen-topik, dan menyimpan hasil matriks.

Analisis dan evaluasi dilakukan dengan analisis kualitatif serta diperhatikan *term-term* dari hasil pengelompokkan dari topik yang telah terbentuk. Dari hasil pengelompokkan ini yaitu topik yang telah didapatkan, langkah selanjutnya yaitu melakukan tahapan analisis untuk menjelaskan data yang telah diperoleh. Untuk tahapan evaluasi dilakukan dengan memperhatikan hasil dari penilaian terhadap manfaat, nilai ataupun dari keseimbangannya dari hasil penelitian ini. Evaluasi yang dilakukan pada penelitian ini yaitu pada evaluasi topik terhadap hasil *review* siswa.

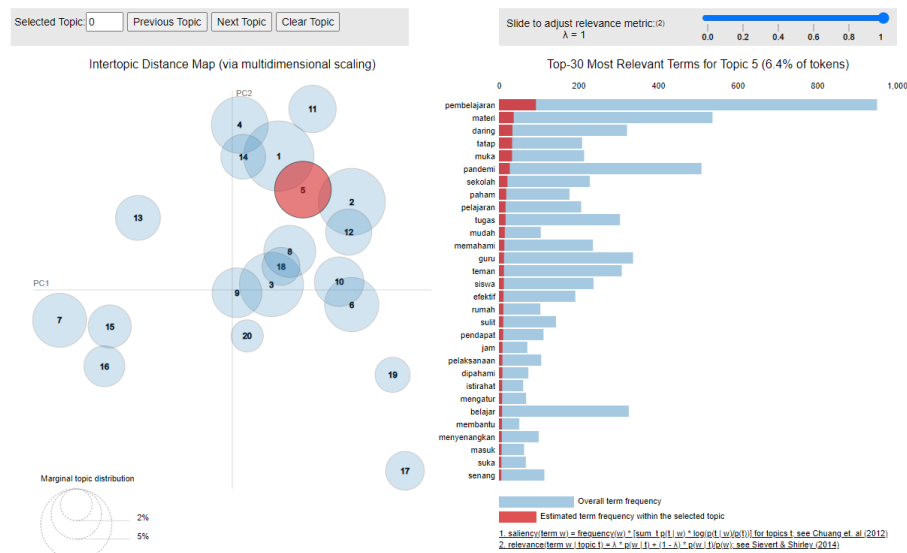
| | |
|-----|--|
| | '0.019*"nggak" + 0.019*"pembelajaran" + 0.019*"guru" + 0.016*"anak" + '0.014*"smk" + 0.014*"langsung"' |
| T14 | '0.059*"pembelajaran" + 0.038*"teman" + 0.036*"pandemi" + 0.035*"materi" + '0.030*"siswa" + 0.027*"guru" + 0.022*"tugas" + 0.022*"belajar" + '0.017*"memahami" + 0.014*"bertemu"' |
| T15 | '0.050*"pembelajaran" + 0.048*"belajar" + 0.023*"daring" + 0.019*"sulit" + '0.019*"tugas" + 0.017*"pandemi" + 0.017*"sinyal" + 0.017*"online" + '0.015*"teman" + 0.013*"alami"' |
| T16 | '0.053*"pembelajaran" + 0.031*"pelajaran" + 0.027*"tatap" + 0.025*"teman" + '0.025*"muka" + 0.025*"tugas" + 0.023*"daring" + 0.023*"pandemi" + '0.020*"guru" + 0.015*"langsung"' |
| T17 | '0.079*"pembelajaran" + 0.060*"pandemi" + 0.036*"materi" + 0.028*"efektif" + '0.017*"siswa" + 0.017*"offline" + 0.016*"guru" + 0.014*"semoga" + '0.013*"online" + 0.013*"memahami"' |
| T18 | '0.084*"pembelajaran" + 0.033*"materi" + 0.031*"daring" + 0.030*"tatap" + '0.030*"muka" + 0.025*"pandemi" + 0.019*"sekolah" + 0.017*"paham" + '0.015*"pelajaran" + 0.015*"tugas"' |
| T19 | '0.085*"materi" + 0.050*"pembelajaran" + 0.044*"daring" + 0.033*"paham" + '0.031*"pandemi" + 0.025*"tugas" + 0.023*"guru" + 0.020*"sulit" + '0.016*"memahami" + 0.015*"belajar"' |
| T20 | '0.059*"pembelajaran" + 0.034*"pelajaran" + 0.024*"memahami" + 0.023*"siswa" + '0.018*"pandemi" + 0.017*"efektif" + 0.017*"teman" + 0.016*"belajar" + '0.015*"muka" + 0.014*"tugas"' |

Untuk memudahkan visualisasi dari hasil penelitian menggunakan algoritma LDA, penulis menggunakan LDAvis dari *python*. Pada bagian kiri terdapat simbol lingkaran (*circle*) yang mewakili dari topik, terdapat 20 topik dari dataset yang telah digunakan.



Gambar 3. Hasil Visualisasi dengan LDAvis review siswa

Jika memilih salah satu topik (contoh : topik 5), maka kata-kata yang terletak pada sisi kanan adalah kata-kata yang terdapat pada topik tersebut.



Gambar 4. Hasil Visualisasi dengan LDAvis review siswa untuk topik 5

Dengan adanya visualisasi data dari model LDA, maka akan mempermudah dalam menganalisa data, serta kata-kata yang terdapat pada sebuah topik

4. SIMPULAN

Pemodelan topik dengan menggunakan LDA untuk data review siswa untuk proses pembelajaran pada masa mandemi dapat menghasilkan 20 topik dari dataset sebanyak 1619 data, komentar yang diambil merupakan komentar dalam bahasa Indonesia. Kemudian teknik Topik Modeling menggunakan metode LDA sangat berguna untuk mendapatkan informasi dari sekumpulan data teks yang sangat besar yang tidak terstruktur.

5. SARAN

Untuk penelitian selanjutnya dapat menggunakan teknik steeming dengan menggunakan sastrawi. Dapat memodelkan topik dengan kasus yang berbeda.

DAFTAR PUSTAKA

- [1] K. F. Hakim, P. Silvianti, and A. M. Soleh, "Latent Dirichlet Allocation Dalam Identifikasi Respon Masyarakat Indonesia Terhadap Covid-19 Tahun 2020-2021," *Xplore J. Stat.*, vol. 10, no. 3, pp. 248–257, 2021, doi: 10.29244/xplore.v10i3.836.
- [2] S. Gurajala, S. Dhaniyala, and J. N. Matthews, "Understanding Public Response to Air Quality Using Tweet Analysis," *Soc. Media Soc.*, vol. 5, no. 3, 2019, doi: 10.1177/2056305119867656.
- [3] X. Han, J. Wang, M. Zhang, and X. Wang, "Using social media to mine and analyze public opinion related to COVID-19 in China," *Int. J. Environ. Res. Public Health*, vol. 17, no. 8, 2020, doi: 10.3390/ijerph17082788.
- [4] Wahyudin, "Aplikasi Topic Modeling Pada Pemberitaan," *Semin. Nas. Off. Stat. Pengemb. Off. Stat. dalam mendukung Implementasi SDG's*, pp. 309–318, 2019.
- [5] M. Y. Hendrawan, N. W. K. Projo, M. Y. Hendrawan, and N. W. K. Projo, "Topic Modelling in Knowledge Management Documents BPS Statistics Indonesia," pp. 119–130.
- [6] A. Syaifuddin, R. A. Harianto, and J. Santoso, "Analisis Trending Topik untuk Percakapan Media Sosial dengan Menggunakan Topic Modelling Berbasis Algoritme LDA," *J. Intell. Syst. Comput.*, vol. 2, no. 1, pp. 12–19, 2021, doi: 10.52985/insyst.v2i1.150.
- [7] S. Deepu, R. Pethuru, and S. Rajaraajeswari, "A Framework for Text Analytics using the Bag of Words (BoW) Model for Prediction," *Int. Conf. Innov. Comput. Netw.*, pp. 320–323, 2016.