

PEMODELAN PERSEPSI PEMBELAJARAN ONLINE MENGUNAKAN LATENT DIRICHLET ALLOCATION

Jerhi Wahyu Fernanda¹

¹Institut Agama Islam Negeri (IAIN) Kediri

Email : fernanda.jerhi@iainkediri.ac.id

ABSTRAK

Latent Dirichlet Allocation (LDA) merupakan metode untuk pemodelan topik adalah yang didasarkan kepada konsep probabilitas untuk mencari kemiripan suatu dokumen dan mengelompokkan dokumen-dokumen menjadi beberapa topik atau kelompok. Metode ini masuk dalam *unsupervised learning* karena tidak ada label atau target pada data yang dianalisis. Penelitian ini bertujuan untuk mengelompokkan persepsi tentang pembelajaran online ke dalam beberapa topik menggunakan metode LDA. Data penelitian ini adalah data primer yang dikumpulkan melalui formulir online. Hasil analisis menunjukkan bahwa pemodelan LDA menggunakan 6 topik memiliki *coherence score* paling besar. Hasil visualisasi data text menggunakan wordcloud didapatkan kata tidak memiliki frekuensi kemunculan terbesar. Penentuan jumlah topik yang optimal berdasarkan *coherence score*, didapatkan pemodelan LDA dengan 6 topik adalah yang paling optimal. secara garis besar terdapat beberapa kata yang saling beririsan dengan topik yang lain. Hasil pemodelan memberikan gambaran bahwa persepsi/pandangan mahasiswa terdapat pembelajaran online terkait pemahaman materi yang diberikan dosen, sinyal atau jaringan internet, kuota, dan tugas. Pada kata-kata terkait pemahaman materi, mahasiswa memberikan pandangan bahwa mereka tidak dapat memahami dengan baik materi yang diberikan oleh dosen.

Kata kunci : *Latent Dirichlet Allocation* (LDA), *unsupervised learning*, pembelajaran online

PENDAHULUAN

Pembelajaran secara *online* merupakan solusi dalam dalam melakukan pembelajaran di masa pandemi *Covid-19*. Meskipun saat ini sudah dilaksanakan Pembelajaran Tatap Muka (PTM) secara terbatas, pembelajaran secara *online* tetap dilakukan untuk institusi pendidikan supaya pembelajaran tetap berjalan dengan baik. Platform *Learning Management System* (LMS) seperti seperti *google classroom*, *schoolology*, *kahoot*, maupun *elearning* yang disediakan institusi pendidikan merupakan aplikasi yang mendukung

pembelajaran secara *online* atau daring.

Dalam pelaksanaan pembelajar-an secara *online*, pasti akan ditemui beberapa kendala. Jaringan internet dan gawai, pengelolaan pembelajaran, penilaian, dan pengawasan merupakan kendala yang dihadapi dalam pembelajarn tersebut [1]. Adanya kendala-kendala tersebut, memaksa institusi pendidikan harus melakukan evaluasi terhadap pembelajaran *online* yang telah dilakukan. Salah satu evaluasi yang dapat dilakukan dengan menganalisis tingkat penerimaan siswa maupun mahasiswa terhadap penggunaan teknologi yang digunakan dalam pembelajaran secara *online*.

Hidayah 2021 melakukan penelitian tentang Penerimaan Aplikasi Pembelajaran Online Menggunakan *Technology Acceptance Model 3 Dan Partial Least Square Structural Equation Model*. Kesimpulan yang didapatkan dari penelitian tersebut adalah variabel *Perceived Ease of Use* berpengaruh signifikan terhadap *Perceived Usefulness* dan *Behavioral Intention to Use*. Variabel *Perceived Usefulness* memiliki pengaruh yang signifikan terhadap *Behavioral Intention to Use* dan *Behavioral Intention to Use* berpengaruh signifikan terhadap *Actual System Use*.

Evaluasi dari pembelajaran online juga dapat dilakukan dengan melakukan analisis terhadap persepsi peserta didik terhadap kendala yang dihadapi ketika pembelajaran *online*. Analisis ini akan memberikan gambaran tentang kendala, keunggulan, bahkan saran dari sistem tersebut. Analisis terhadap persepsi ini menggunakan metode *Latent Dirichlet Allocation* (LDA).

LDA merupakan metode untuk pemodelan topik adalah yang didasarkan kepada konsep *probabilitas* untuk mencari kemiripan suatu dokumen dan mengelompokkan dokumen-dokumen menjadi beberapa topik atau kelompok [2]. Metode ini masuk dalam *unsupervised learning* karena tidak ada label atau target pada data yang dianalisis.

LDA juga telah digunakan dalam pemodelan topik oleh peneliti-peneliti lain. LDA digunakan untuk pemodelan Topik Pengguna Twitter Mengenai Aplikasi “Ruangguru”. Penelitian tersebut memberikan hasil model LDA mengelompokkan *twett* dalam aplikasi ruang guru kedalam 28 buah topik

dengan topik yang sering diperbincangkan adalah diskon ruangguru [3]. LDA juga digunakan dalam analisis Sentimen dan Pemodelan Topik dalam Optimalisasi Pemasaran Destinasi Pariwisata Prioritas di Indonesia [4]. Berdasarkan uraian di atas, penelitian ini bertujuan untuk mengelompokkan persepsi tentang pembelajaran *online* ke dalam beberapa topik menggunakan metode LDA.

METODELOGI PENELITIAN

Sumber Data dan Variabel Penelitian

Penelitian ini menggunakan data primer yang dikumpulkan peneliti melalui survei menggunakan formulir online. Responden adalah mahasiswa dari satu Fakultas pendidikan di salah satu Perguruan Tinggi Negeri di Kota Kediri dengan total 581. Variabel dalam penelitian ini adalah persepsi mahasiswa terhadap pembelajaran online.

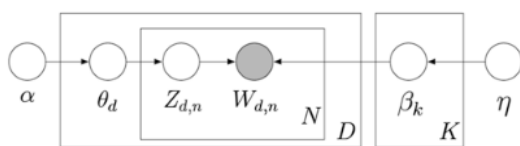
Metode Analisis

Analisis data dalam penelitian ini menggunakan software R dan terdiri dari 3 tahap. Tahapan tersebut adalah sebagai berikut.

1. *Preprocessing* data. Pada tahap ini dilakukan pemrosesan terhadap data text yang didapatkan sebelum dilakukan pemodelan menggunakan LDA. *Preprocessing* data sangat bermanfaat untuk mengolah data awal yang masih tidak terstruktur dengan berbagai macam bentuk dan komponen (*unstructured*) menjadi sebuah data yang terstruktur sehingga dapat dilakukan pemodelan menggunakan metode LDA [5]. *Preprocessing* dalam penelitian ini meliputi 7 tahap dengan langkah-langkah sebagai berikut.

- a. Mengubah semua huruf menjadi huruf kecil
 - b. Menghilangkan angka, maupun tanda baca
 - c. Menghilangkan stopwords
 - d. Menghapus spasi berlebihan
 - e. Mengubah kata menjadi kata dasar (*stemming*)
 - f. Membuat *corpus*
 - g. Membuat *Dokumen Term Matrix* (DTM).
2. Visualisasi data. Visualisasi data menggunakan *wordcloud* dan *barplot*. Visualisasi ini akan bermanfaat untuk memberikan gambaran tentang kata-kata yang memiliki frekuensi kemunculan paling besar.
 3. Pemodelan topik menggunakan metode LDA. Pemodelan LDA merupakan metode statistika yang menerapkan dasar-dasar dalam pencarian topik dalam kumpulan dokumen-dokumen.

LDA dapat mengatasi permasalahan *overfitting* pada *Probabilistic Latent Semantic Analysis* (PLSA). Visualisasi model LDA disajikan pada gambar 1 [6].



Gambar 1 *plate notation* LDA

Penjelasan symbol-simbol pada gambar di atas adalah sebagai berikut.

- α = parameter dirichlet untuk proporsi topik per dokumen
- θ_d = Distribusi topik dokumen
- $Z_{d,n}$ = Penetapan topik kata ke n dalam dokumen d
- $W_{d,n}$ = Kata yang diamati dari kata ke n dalam dokumen d
- η adalah dirichlet parameter atas

distribusi kata terhadap topik

β_k = Distribusi kata pada topik K

K = Kumpulan topik

N = Kumpulan kata dalam dokumen

D = Kumpulan dokumen dalam korpus

Algoritma yang digunakan dalam pemodelan LDA adalah sebagai berikut [7].

- a. Lakukan proses pemberian label topik setiap kata (1 sampai i) pada setiap dokumen 1 sampai j mengikuti pola distribusi multinomial (Dirichlet distribution). Sebagai ilustrasi misal dokumen 1 terdiri dari 25 persent topik 1, 25 persen topik 2, dan 50% topik 3.
- b. Secara probabilitas, untuk setiap kata (i), ada 1 sampai (i) kata untuk suatu topik (k), sebagai ilustrasi kata tersebut memiliki rata-rata probabilitas sebesar 0,25 untuk suatu topik secara statistik
- c. Untuk setiap kata (i) dalam dokumen (j) dan topik (k), hitung proporsi kata-kata dalam dokumen yang dimasukkan pada topik tersebut. probabilitas topik (k) dengan dokumen dilambangkan dengan (j), $p(k|j)$, dan proporsi kata (i) dalam topik (k) dari semua dokumen (probabilitas kata (i) dengan topik (k) dinotasikan dengan $p(i|k)$).
- d. Lakukan proses resampling dengan memberi topik baru pada kata berdasarkan hasil perkalian $p(k|j)$ dengan $p(i|k)$.
- e. Lakukan proses pemberian topik baru ke kata sampai iterasi yang telah ditentukan, sampai pada kondisi tidak ada perubahan topik pada kata (*steady state*).
 Pada pemodelan LDA, harus dilakukan pemilihan topik yang optimal yang dapat merepresentasikan data yang ada. Penentuan topik yang optimal, ditinjau dari nilai *coherence score* setiap topik. Topik yang memiliki *coherence score* tertinggi, merupakan topik yang

optimal dalam pemodelan LDA [8]. yang optimal. Pada penelitian ini, dilakukan perhitungan *coherence score* dengan menggunakan topik. Berdasarkan 20 topik tersebut, dilakukan perhitungan *coherence score* untuk menentukan jumlah topik yang optimal.

4. Interpretasi hasil analisis pemodelan LDA. Pada tahap ini, dilakukan interpretasi terhadap setiap topik yang didapatkan dari pemodelan dengan LDA. Interpretasi hasil juga dilakukan secara keseluruhan untuk mendapatkan topik yang akan dapat merepresentasikan kendala atau permasalahan dalam pembelajaran online.

HASIL DAN PEMBAHASAN

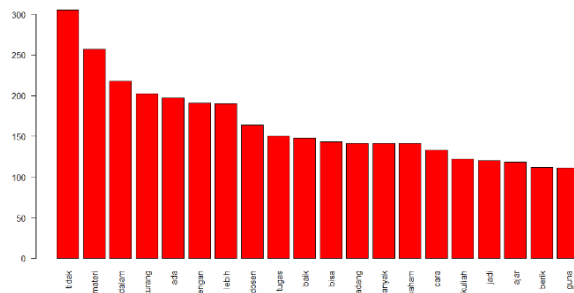
Analisis data text pada penelitian ini diawali dengan membuat visualisasi dari persepsi/pandangan mahasiswa terhadap pembelajaran online yang telah dilakukan. Visualisasi data menggunakan *wordcloud* dengan hasil disajikan pada gambar di bawah ini.



Gambar 2. *Wordcloud* persepsi

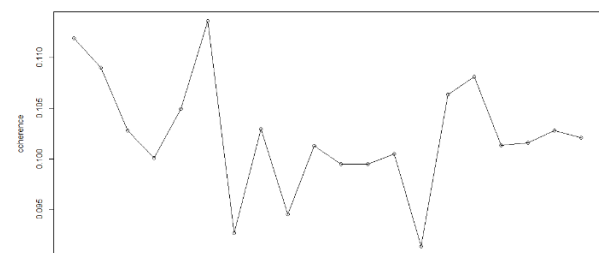
Gambar 2 memberikan informasi tentang 50 kata yang memiliki frekuensi kemunculan paling besar. Kata “tidak” merupakan kata yang memiliki frekuensi paling besar. Berdasarkan gambar tersebut juga dapat dilihat bahwa kata-kata lain yang memiliki frekuensi kemunculan lebih besar dibandingkan dengan kata-kata yang

lainnya adalah materi, kurang, sinyal.



Gambar 3. *Barplot* kata dengan frekuensi terbesar

Barplot merupakan alternative penyajian data text selain menggunakan *wordcloud*. Gambar 2 merupakan *barplot* yang juga memberikan informasi kata-kata yang memiliki frekuensi kemunculan paling besar dan disajikan dalam bentuk diagram batang yang telah diurukan dari yang terbesar sampai yang memiliki frekuensi paling kecil. Gambar 3 merupakan *barplot* dari persepsi/pandangan dari mahasiswa, dan dapat dilihat kata-kata seperti tidak, materi, dosen, tugas merupakan kata unik yang terkait dengan pembelajaran secara online.

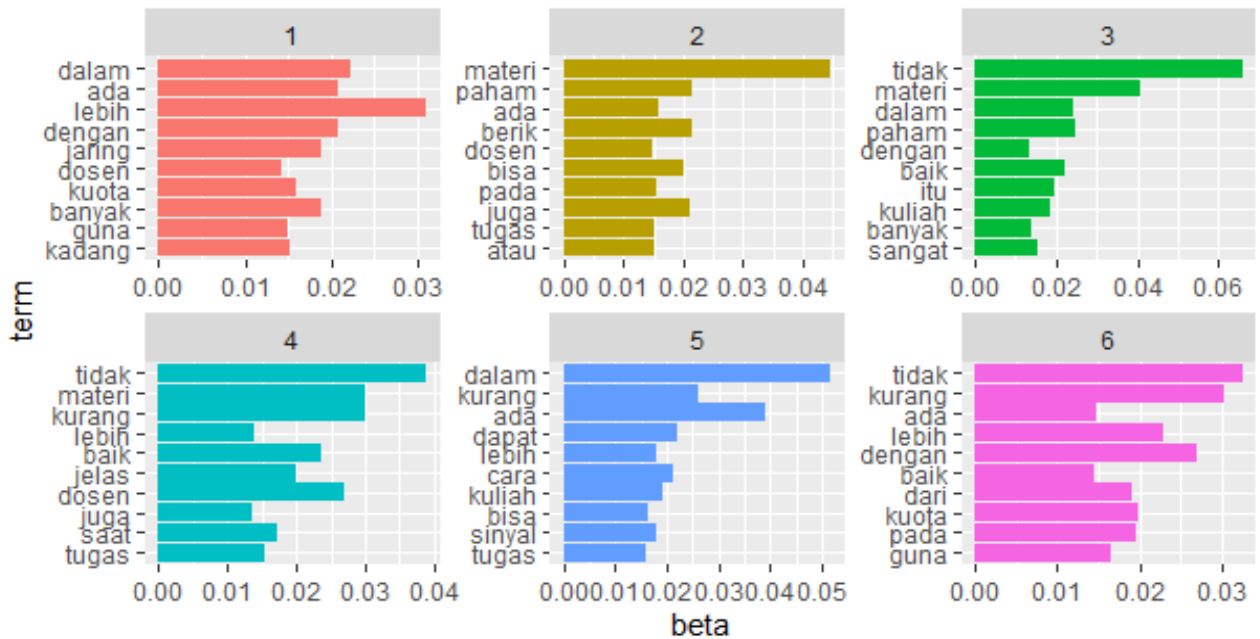


Gambar 4. Plot *coherence score* setiap topik

Analisis selanjutnya adalah pemodelan data menggunakan *Latent Dirichlet Allocation* (LDA). Pada pemodelan LDA, penentuan banyak topik dilakukan menggunakan komponen penting dalam proses analisis. Penentuan topik yang akan digunakan dalam pemodelan LDA, didasarkan pada *coherence score*.

Tahap awal pemodelan topik dengan LDA, adalah dengan melakukan pemodelan menggunakan 20 topik. *Coherence score* setiap topik disajikan pada gambar 4. Berdasarkan gambar 4, diketahui bahwa banyak topik yang memiliki *coherence score* paling besar adalah pada nilai 6.

Berdasarkan hasil tersebut, maka pemodelan LDA dilakukan dengan menggunakan 6 topik digunakan. Penelitian lain juga menjelaskan bahwa topik yang memiliki nilai *coherence score* akan memiliki peluang paling besar untuk muncul dalam suatu kasus yang teliti [9].



Gambar 5. Kata-kata yang paling umum di setiap topik

Hasil pemodelan LDA menggunakan 6 topik disajikan pada gambar 5. Gambar tersebut memberikan informasi tentang kata-kata yang umum atau terkait dengan topik tertentu. Pada topik 1 ini, kata “jaring” memiliki kata asli “jaringan”. Perubahan tersebut dikarenakan proses *steaming*. Penelitian lain juga memberikan informasi bahwa pada proses *steaming* juga ditemukan kata yang tidak dapat dirubah kedalam kata dasarnya (*steaming*) [10]. Topik 1 ini dapat mewaliki bahwa kendala yang dihadapi mahasiswa dalam perkuliahan online adalah terkait jaringan internet yang tidak optimal atau kurang bagus. Selain itu terdapat kata kuota yang merepresentasikan bahwa dalam pembelajaran online, yang menggunakan beberapa platform, akan boros di kuota internet. Selain itu kata kuota juga dapat

merepresentasikan bantuan kuota dari pemerintah. Pada topik 1 juga terdapat kata dosen. Kata ini terkait dengan materi yang diberikan oleh dosen.

Pada topik 2, terdapat kata materi, paham, beri, juga, dan dosen. Topik 2 ini merepresentasikan suatu masalah yaitu tentang pemahaman materi yang diberikan oleh dosen. Pada topik 2 juga muncul kata tugas. Kata ini merepresentasikan bahwa pemberian tugas sangat banyak dan tidak disertai dengan penjelasan. Topik 3 terdiri dair beberapa kata seperti tidak, materi, paham, kuliah, dan banyak. Hasil topik 3 ini juga mempertegas dari topik 2, mahasiswa tidak dapat memahami materi yang diberikan. Hasil ini juga mempertegas hasil pada topik 2. Pada topik 3 ini juga terdapat kata kuliah dan banyak. Dua kata ini merepresentasikan bahwa selama

pembelajaran online kuliah menjadi sangat banyak.

Topik 4 memiliki terdiri dari kata tidak, materi, kurang, dosen, dan tugas. Pada topik ini dapat dilihat bahwa materi yang diberikan oleh dosen kurang dapat diterima dengan baik oleh mahasiswa. Kata tugas juga muncul lagi pada topik 4. Topik 5 terdiri dari kata-kata yang dimana kata-kata tersebut sudah muncul di topik yang lain seperti kurang, tugas, lebih. Pada topik 5 ini terdapat kata sinyal. Kata sinyal ini memberikan gambaran bahwa dalam pembelajaran online, mahasiswa terkendala dengan sinyal internet. Kondisi ini dimungkinkan terkait dengan lokasi geografi tempat tinggal mahasiswa tersebut. Topik 6 juga terdiri dari kata-kata yang beririsan dengan topik yang lain. Pada topik 6 terdapat kata kuota dimana. Kata ini juga muncul pada topik 1 dan kondisi yang terkait dengan kuota adalah platform yang digunakan untuk pembelajaran online boros kuota, dan terkait dengan bantuan kuota dari pemerintah.

Berdasarkan hasil pemodelan LDA dengan 6 topik, secara garis besar terdapat beberapa kata yang saling beririsan dengan topik yang lain. Hasil pemodelan memberikan gambaran bahwa persepsi/pandangan mahasiswa terdapat pembelajaran online terkait pemahaman materi yang diberikan dosen, sinyal atau jaringan internet, kuota, dan tugas. Pada kata-kata terkait pemahaman materi, mahasiswa memberikan pandangan bahwa mereka tidak dapat memahami dengan baik materi yang diberikan oleh dosen. Hasil ini didukung oleh *wordcloud* pada gambar 1 yang memberikan informasi bahwa kata yang memiliki frekuensi kemunculan paling besar adalah kata tidak

KESIMPULAN

Hasil pemodelan LDA, terdapat 6 topik yang dibentuk berdasarkan coherence score. Hasil analisis terhadap 6 topik tersebut, terdapat kata-kata yang saling beririsan dengan topik lain. Berdasarkan pemodelan menggunakan LDA, terdapat beberapa

persepsi/pandangan terhadap pembelajaran online yang telah dilakukan seperti mahasiswa tidak dapat memahami dengan baik materi yang diberikan oleh dosen, jaringan internet yang tidak baik yang dapat mengganggu proses pembelajaran online, kuota internet yang dimiliki mahasiswa cepat habis karena menggunakan beberapa platform LMS, terkait tugas yang diberikan kadang tanpa disertai penjelasan.

DAFTAR PUSTAKA

- [1] H. A. Rigianti, "Kendala Pembelajaran Daring Guru Sekolah Dasar Di Kabupaten Banjarnegara," *Elem. Sch.*, vol. 7, no. Juli, pp. 297–302, 2020.
- [2] T. Kwartler, *Text mining in practice with R*, vol. 90, no. 13. 2017.
- [3] B. W. Arianto and G. Anuraga, "Topic Modeling for Twitter Users Regarding the 'Ruangguru' Application," *J. ILMU DASAR*, vol. 21, no. 2, p. 149, 2020, doi: 10.19184/jid.v21i2.17112.
- [4] Y. A. Singgalen, "Analisis Sentimen dan Pemodelan Topik dalam Optimalisasi Pemasaran Destinasi Pariwisata Prioritas di Indonesia," *J. Inf. Syst. Informatics*, vol. 4, no. 1, pp. 459–470, 2021, [Online]. Available: <http://journal-isi.org/index.php/isi/article/view/171>
- [5] H. Jaka, "Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining," *Inform. UPGRIS*, vol. 1, pp. 1–9, 2015.
- [6] D. Blei, L. Carin, and D. Dunson, "Probabilistic topic models," *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 55–65, 2010, doi: 10.1109/MSP.2010.938079.
- [7] R. Bali, D. Sarkar, B. Lantz, and C. Lesmeister, *R: Unleash Machine Learning Techniques*. 2016.
- [8] F. Alattar and K. Shaalan, "Emerging Research Topic Detection Using Filtered-LDA," *AI*,

- vol. 2, no. 4, pp. 578–599, 2021, doi: 10.3390/ai2040035.
- [9] Y. Sahria and D. H. Fudholi, “Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode Topic Modeling LDA,” *J. Rekayasa Sist. dan Teknol. Inf.*, vol. 4, no. 2, pp. 336–344, 2020.
- [10] 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.