

Analisis Ulasan Aplikasi MyPertamina Menggunakan *Topic Modeling* dengan *Latent Dirichlet Allocation*

Muhammad Adrinta Abdurrazzaq

Informatika, Fakultas Ilmu Komputer dan Desain, Institut Teknologi dan Bisnis Kalbis
Jalan Pulomas Selatan Kav. 22, Jakarta 13210

Email:muhammad.abdurrazzaq@kalbis.ac.id

Abstract: *The policy for using the MyPertamina application in purchasing subsidized fuel oil begins on July 1, 2022. Until August 17, 2022, the enormous demand for fuel in the community made this policy concern. Application developers can update applications according to user complaints and reviews. This study aims to map the problems in the MyPertamina application that many users complain about using topic modeling with the LDA algorithm. The study's results divided the comments into 7 with a perplexity value of -8.08844 and a topic coherence of 0.49860. These topics are related to user anxiety about the MyPertamina application usage policy, problems in the registration process, problems in the authentication process, and problems in the smooth use of applications related to bugs and signals. There are 5341 comments considered spam that does not correlate with the MyPertamina application.*

Keywords: *lda, latent Dirichlet allocation, perplexity, text clustering, topic modeling, topic coherence,*

Abstrak: *Kebijakan untuk menggunakan aplikasi MyPertamina dalam pembelian Bahan Bakar Minyak (BBM) bersubsidi dimulai pada tanggal 1 Juli 2022. Hingga tanggal 17 Agustus 2022, kebutuhan yang besar terhadap BBM di masyarakat membuat kebijakan ini menjadi perhatian. Pengembang aplikasi dapat memperbaiki serta memperbarui aplikasi sesuai dengan keluhan dan ulasan pengguna. Penelitian ini bertujuan untuk memetakan permasalahan yang ada pada aplikasi MyPertamina yang banyak dikeluhkan penggunanya menggunakan topic modeling dengan algoritma LDA. Hasil penelitian membagi komentar menjadi 7 dengan nilai perplexity -8.08844 dan topic coherence 0.49860. Topik – topik tersebut terkait tentang kecemasan pengguna terhadap kebijakan penggunaan aplikasi MyPertamina, permasalahan pada proses pendaftaran, permasalahan pada proses autentikasi, dan permasalahan pada kelancaran penggunaan aplikasi terkait bug dan sinyal. Terdapat 5341 komentar dianggap sebagai spam yang tidak berkorelasi dengan aplikasi MyPertamina.*

Kata kunci: *lda, klasterisasi text, koherensi topik, latent dirichlet allocation, pemodelan topik, perplexity*

I. PENDAHULUAN

Kebijakan untuk menggunakan aplikasi MyPertamina dalam pembelian Bahan Bakar Minyak (BBM) bersubsidi dimulai pada tanggal 1 Juli 2022. Kebijakan ini awalnya diterapkan pada 11 kabupaten/kota yang sampai saat ini telah diperluas menjadi 50 kabupaten/kota. Hingga tanggal 17 Agustus 2022, aplikasi MyPertamina tercatat telah mendapatkan 289 ribu komentar dari pengguna dengan *rating* 1.5 dari 5 di Google Playstore [1, 2]. Kebutuhan yang besar terhadap BBM di masyarakat membuat kebijakan penggunaan aplikasi MyPertamina dalam pembelian BBM menimbulkan berbagai macam komentar dan sentimen. Banyak masyarakat yang bergantung pada BBM bersubsidi, maka dari itu kemudahan dalam penggunaan aplikasi MyPertamina adalah hal utama yang harus

diperhatikan, agar kebijakan ini tidak dipandang sebagai pembatasan subsidi pada masyarakat yang membutuhkan. Hal ini sesuai dengan klaim Pertamina sendiri [3].

Dalam mengembangkan suatu aplikasi tentunya terdapat beberapa tahapan. Pada metode *waterfall* terdapat tahapan *maintenance* dimana pengembang memperbaiki serta memperbarui aplikasi sesuai dengan keluhan dan ulasan pengguna [4].

Ulasan pengguna untuk aplikasi MyPertamina pada Google Playstore yang berbentuk *rating* dan komentar dapat dimanfaatkan untuk memetakan topik – topik kekurangan aplikasi untuk kemudian menentukan fokus perbaikan dan peningkatan kualitas aplikasi.

Penelitian ini bertujuan untuk memetakan permasalahan yang ada pada aplikasi

MyPertamina yang banyak dikeluhkan penggunanya.

II. METODE PENELITIAN

A. Dasar Teori

1. Topic Modeling

Topic modeling adalah algoritma yang memiliki tujuan untuk memetakan suatu tema atau topik yang berada pada kumpulan dokumen yang tidak terstruktur. *Topic modeling* juga dapat diadaptasi penggunaannya ke banyak tipe data. *Topic modeling* juga digunakan untuk mencari pola pada data genetik, gambar, dan jaringan sosial [5].

Dokumen dari suatu sumber dapat mengandung informasi berguna yang tersembunyi, maka dari itu *topic modeling* berusaha untuk mencari makna tersembunyi yang ada pada dokumen [6].

2. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) adalah sebuah model generatif probablistik untuk sebuah korpus. LDA memandang sebuah dokumen tersusun atas data acak yang mengandung topik tersembunyi, dimana setiap topik dikarakterisasi melalui distribusi kata – kata [7].

LDA merupakan salah satu teknik paling banyak digunakan untuk *topic modeling*. LDA juga telah banyak dimodifikasi dan dikembangkan pada penelitian lain untuk meningkatkan kinerjanya [8].

B. Tahapan Penelitian

Penelitian ini dilakukan dengan 5 tahapan utama, yaitu koleksi data, praproses data, pemilihan jumlah topik, *hyperparameter tuning*, analisis hasil.

1. Koleksi Data

Koleksi data menggunakan Google Play Scraper dengan bahasa pemrograman Python. Ulasan yang dikoleksi hanya dari pengguna yang berasal dari wilayah Indonesia dengan jangka waktu dari 1 Juli 2022 hingga 17 Agustus 2022. Setelah proses *scraping* untuk mengkoleksi data, terdapat 93760 ulasan.

Contoh data yang dikoleksi dapat dilihat pada Tabel 1.

Tabel 1 Contoh ulasan yang dikoleksi

No.	Komentar	Rating	Tanggal
1	Ribet	1	2022-07-01
2	Aplikasi nya tidak jelas	1	2022-07-01
3	Mantap	5	2022-08-16

2. Praproses Data

Praproses data dilakukan untuk membersihkan teks komentar dari ulasan agar pemetaan topik lebih optimal. Tahapan praproses yang dilakukan, yaitu mengganti karakter *non-alphanumeric* menjadi spasi, mengubah multi spasi menjadi spasi tunggal, mengubah setiap huruf menjadi huruf kecil, menghapus kata – kata yang tidak memiliki pengaruh signifikan dan berada pada daftar kata *stopwords* seperti kata penghubung, mengubah setiap kata menjadi kata dasar, menghapus karakter yang berurutan dan berulang pada suatu kata sehingga kata tersebut hanya menggunakan satu dari beberapa karakter. Contoh perubahan teks setelah dilakukan praproses data dapat dilihat pada Tabel 2.

Tabel 2 Contoh data sebelum dan sesudah praproses

No.	Sesudah	Sebelum
1	Ribet	ribet
2	Aplikasi nya tidak jelas	aplikasi jelas
3	Mantaap	mantap

3. Pemilihan Jumlah Topik

Data yang telah melalui praproses kemudian dibagi menjadi dua bagian yaitu 80% data digunakan sebagai data latih model LDA dan 20% sebagai data uji, untuk menguji seberapa baik model yang dibentuk.

Pada penelitian ini digunakan nilai *perplexity* sebagai salah satu faktor untuk menentukan jumlah topik yang akan dibentuk. Nilai *perplexity* umum digunakan sebagai tolak ukur dalam mengevaluasi seberapa baik suatu model dalam mendeskripsikan data. Nilai *perplexity* yang rendah menunjukkan bahwa model tersebut lebih baik dibandingkan dengan model dengan nilai *perplexity* yang lebih tinggi. Nilai *perplexity* ini didapatkan dengan mengecek data uji ke model [9].

Nilai *perplexity* tidak selalu searah dengan persepsi manusia [10], maka dari itu selain nilai *perplexity*, penelitian ini akan menggunakan nilai *topic coherence*. *Topic coherence*

mengukur nilai sebuah topik dengan menghitung derajat kesamaan semantik antara kata – kata yang bernilai tinggi dalam topik [11].

4. Hyperparameter tuning

Pada LDA terdapat dua *hyperparameter* penting, yaitu alpha dan beta. Alpha adalah parameter yang mengatur bobot distribusi topik pada setiap dokumen, sementara beta adalah parameter yang mengatur bobot dari kata - kata untuk setiap topik.

Jika nilai alpha tinggi, maka sebuah dokumen dapat memiliki termasuk dalam beberapa topik. Sementara jika nilai alpha rendah, maka sebuah dokumen akan terkonsentrasi ke beberapa topik atau bahkan hanya dikategorikan ke dalam satu topik saja.

Jika nilai beta tinggi, maka ketergantungan sebuah topik terhadap kata – kata tertentu akan lebih menyebar. Sementara jika nilai beta rendah, maka ketergantungan sebuah topik akan lebih terkonsentrasi kepada beberapa atau bahkan satu kata saja [12].

5. Analisis Hasil

Analisis yang dilakukan pada penelitian ini yaitu menganalisis kata – kata yang memiliki bobot paling tinggi untuk setiap topik agar dapat diinterpretasikan ke dalam suatu topik yang spesifik, melihat distribusi topik dengan *rating*, dan menggunakan LDAvis [13] untuk memvisualisasikan hasil pemodelan dengan LDA.

III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan Gensim [14] untuk membangun model *topic modeling* dengan LDA. Tabel 3 menunjukkan nilai *perplexity* dan *topic coherence* dengan berbagai jumlah topik yang diterapkan pada data uji. Nilai *topic coherence* maksimal didapatkan saat jumlah topik bernilai tujuh, maka untuk selanjutnya komentar akan dikategorikan ke dalam tujuh topik.

Untuk mencari nilai alpha dan beta terbaik maka digunakan seluruh dokumen sebagai data latih model LDA. Nilai alpha yang digunakan adalah “symmetric”, 0.1, 0.01, dan 0.001. Sementara nilai beta yang digunakan adalah “symmetric”. 0.1, 0.01, dan 0.001. Kombinasi nilai alpha dan beta terbaik adalah nilai alpha 0.001 dan beta 0.1 yang memiliki nilai

perplexity -8.08844 dan *topic coherence* 0.49860.

Tabel 3 Nilai *perplexity* dan *topic coherence* berdasarkan jumlah topik pada data uji

Jumlah Topik	<i>Perplexity</i>	<i>Topic Coherence</i>
3	-7.40621	0.38136
4	-7.48723	0.45826
5	-7.57189	0.51449
6	-7.6226	0.49414
7	-7.6614	0.52495
8	-7.72659	0.49114
9	-7.78233	0.47571
10	-7.85874	0.47637

Pada Tabel 4 dapat dilihat rincian topik beserta kata – kata yang berpengaruh untuk masing – masing topik. Jika diperhatikan terdapat kata – kata yang tidak berkorelasi untuk aplikasi MyPertamina dan lebih berkorelasi untuk aplikasi permainan, yaitu pada topik 0 terdapat kata “war”, pada Topik 1 terdapat kata “gamenya”, pada Topik 5 terdapat kata “hero”, dan Topik 6 terdapat kata “game”, maka dapat dicurigai bahwa komentar – komentar tersebut merupakan komentar *spam*.

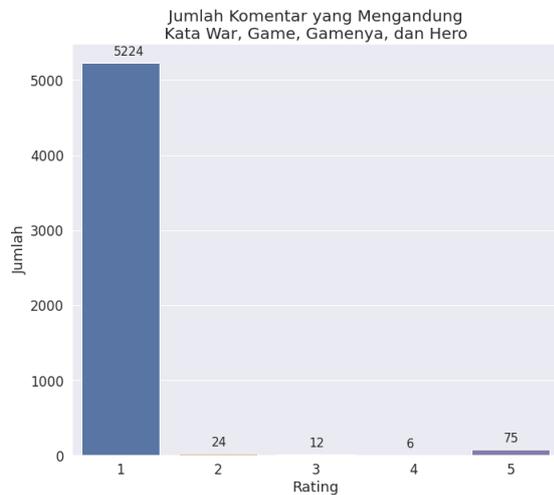
Terdapat 5341 komentar yang mengandung kata – kata tersebut, pada Gambar 1 dapat dilihat grafik sebaran jumlahnya berdasarkan *rating*, terlihat bahwa 98% komentar *spam* tersebut ber-*rating* 1. Contoh komentar *spam* dapat dilihat pada Gambar 2.

Tabel 4 10 kata paing berpengaruh pada setiap topik

Topik	10 Kata Paling Berpengaruh
0	kurang, aplikasi, moga, pertamina, kendaran, war, transaksi, top, bagus, tambah
1	ribet, bikin, hp, gamenya, repot, aplikasi, spbu, atur, main, pakai
2	susah, aplikasi, rakyat, bikin, nyusahin, jelek, masyarakat, orang, sampah, apk
3	daftar, aplikasi, eror, nya, grafik, masuk, lambat, data, burik, coba
4	pakai, beli, aplikasi, isi, bensin, hp, bm, bayar, pertamina, ok
5	sulit, bintang, kasih, aplikasi, mudah, hero, bagus, mantap, bantu, ruwet
6	nya, game, bug, bagus, tolong, baik, main, suka, jaring, lag

Setelah menghapus komentar *spam*, pada Gambar 3 dapat dilihat grafik distribusi jumlah komentar berdasarkan topik dan sentimennya. Keterangan sentimen didapatkan dengan cara mengkategorikan komentar yang memiliki

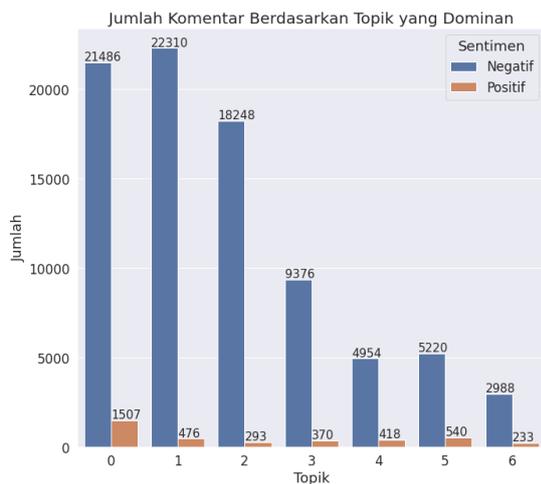
rating dibawah 3 sebagai komentar negatif, sementara sisanya adalah positif. Pada Gambar 3 juga setiap komentar hanya dikategorikan pada satu topik yang paling dominan. Topik 1 menjadi topik yang paling banyak menjadi topik dominan di seluruh komentar



Gambar 1 Jumlah komentar spam

content	score
Sebenarnya KEPEPET aja instal game ini... Karena pemerintah... icak icak ajalah, asal jgn nggak...	1
Anjay alok, kaya game sebelah ga ada pintunya tolong dong kasihin skin biar lebih semangat mainy...	1
Setelah Maintenance kemaren game ini jadi bnyak bug nya ,fps sama koneksi jaringan nya kurang st...	1
Game yang gak pelit skin, mantap saya dapat skin epic.	5
Tolong developer map nya bagusin, grafiknya burik banget, sering ngebug keluar tembok, senjatany...	1
Kirain game	1
pilihan hero ada banyak dan skill nya unik-unik, tapi tolong perbaiki matchmaking nya yang terla...	1
Untuk sekelas game karya anak bangsa ini game nya cukup bagus dan gamenya seru cuma ya itu masih...	3
Gamenya lumayan bagus, tapi ff lebih bagus anjay mabar	1
Game moba terbaik , saya yakin kedepanya bisa ngalahin mobile legend 🙌	1

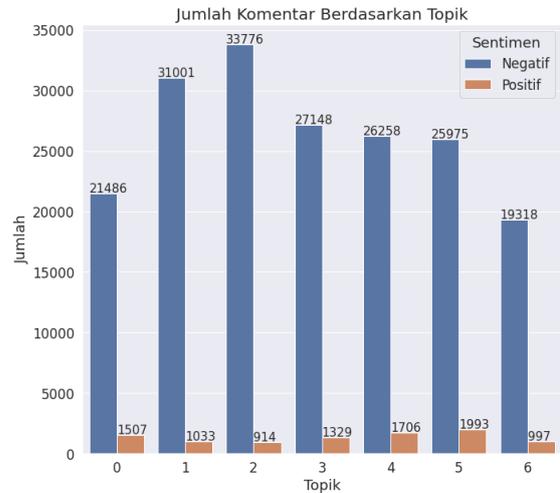
Gambar 2 Contoh komentar spam



Gambar 3 Grafik distribusi jumlah komentar berdasarkan topik yang dominan

Sementara pada Gambar 4, grafik menunjukkan distribusi jumlah komentar berdasarkan topik, ini berarti setiap komentar dapat dikategorikan ke lebih dari satu topik.

Berbeda dengan Gambar 3, Gambar 4 menunjukkan bahwa Topik 2 menjadi topik paling banyak terdapat pada seluruh komentar.



Gambar 4 Grafik distribusi jumlah komentar berdasarkan topik

Penelitian ini mengambil 10 sampel secara acak pada setiap topik untuk menganalisis topik komentar sebenarnya. Sampel yang diambil untuk Topik 0 yang dapat dilihat pada Gambar 5 cenderung berkomentar tentang kebijakan penggunaan aplikasi MyPertamina dan pengalaman penggunaan aplikasi MyPertamina secara umum. Pengguna mengeluhkan aplikasi MyPertamina membuat mereka menjadi sulit untuk membeli bensin bersubsidi.

content
Toh mendouksena, orewa hanya ingin beli bensin untuk anata si roda 2, why gitu loh ngab orewa harus.... arrgghhh baka
bikin ribet kasian orang yg gk bisa main hp atau yg gk punya paketan dan menurut saya ini bikin perusahaan linkaja makin kaya kayak ada konspirasi antara proyek burnn dengan swasta yaitu linkaja saya rasa ada permainan
Ruwet... Ruwet... Ruwet... Bukannya mempermudah, malah mempersulit rakyat. Segitu ruginya kah PERTAMINA dibawah kuasa AHOK, smpa2Z aplikasipun di paksa download buat dapetin cuan ?? 🙄 #Bintang1mypertamina
Menyulitkan masyarakat yg gaptek
G guna cck
Payah
galau cuma pengen rating 1 aja bantu org yang menderita
lebih baik dijadikan opsi, jangan kewajiban.
🙄
Tolong diperbaiki sistemnya suka nge lag kl lg maen. Sinyal tdk stabil.

Gambar 5 10 sampel Topik 0

Pada Gambar 6, terlihat 10 sampel komentar untuk Topik 1. Terlihat pengguna berkomentar tentang aplikasi MyPertamina yang dirasa akan merepotkan pengguna, sama halnya seperti pada Topik 0.

content
Ribet cuy
Laknatullah
Aplikasi gx terlalu berguna menurut saya, malah merepotkan
Meresahkan kehidupan dan sangat ribet
wkwkwk katanya di sptu dilarang memakai handphone tapi disini malah harus make aplikasi
Aplikasinya sering eror padahal saya suka menyapa pengguna sekitar. Tapi ga ada jawaban cewe2nya pada sombong
🙄
Ah Ribet
Masih banyak bug dan ribet
Wkwkwkwk... ajor
Sangat merepotkan dan membuat sulit rakyat kecil

Gambar 6 10 sampel Topik 1

kelancaran penggunaan aplikasi yang ditandai dengan kata *bug*, *lag*, dan sinyal.

content
Apaansi
Bad
Parah sekali
Seringleg
Bug
Bug
Sering Bug aplikasinya...tolong diperbaiki
Tidak bisa digunakan ketika tidak ada sinyal
Gabut doang
memperbualit

Gambar 11 sampel Topik 6

Gambar 12 merupakan hasil visualisasi model LDA dalam memetakan seluruh komentar dengan LDavis, Dapat dilihat 30 kata yang paling berpengaruh dalam menentukan topik pada keseluruhan dokumen dan elemen lingkaran yang menyatakan masing – masing topik.

IV. SIMPULAN

Penelitian ini mengkoleksi ulasan aplikasi MyPertamina dari tanggal 1 Juli 2022 hingga 17 Agustus 2022. Penelitian ini menggunakan LDA untuk *topic modeling* dengan nilai alpha 0.001 dan beta 0.1 yang membagi komentar pada ulasan menjadi 7 topik dengan nilai *perplexity* -8.08844 dan *topic coherence* 0.49860.

Beberapa kata yang berpengaruh dalam pembentukan topik tidak berkorelasi dengan aplikasi MyPertamina. Kata – kata tersebut yaitu war, game, gamenya, dan hero. Terdapat 5341 komentar yang memiliki kata – kata tersebut dan dianggap sebagai *spam*. 98% dari 5341 ulasan tersebut ber-*rating* 1.

Topik 1 merupakan topik yang paling dominan untuk keseluruhan komentar sedangkan Topik 2 merupakan topik yang paling banyak terkait pada keseluruhan komentar.

Dari hasil pengambilan sampel komentar untuk tiap topik, seluruh topik membahas tentang kecemasan dan keluhan pengguna terhadap kebijakan pemerintah terhadap penggunaan aplikasi MyPertamina dalam pembelian bensin bersubsidi. Topik 3 membahas tentang pendaftaran dan autentikasi, sementara pada Topik 4 pengguna berkomentar tentang pembayaran yang ditandai dengan kata “bayar” sebagai kata berpengaruh. Topik 5 lebih terkonsentrasi tentang pembahasan kelancaran penggunaan aplikasi yaitu terkait dengan kata *bug*, *lag*, dan sinyal.

Pada pengambilan sampel di Topik 4, terdapat komentar yang tidak berkorelasi

dengan aplikasi MyPertamina yaitu terkait dengan penggunaan kata “skin” yang lebih terkait dengan aplikasi permainan. Hal ini berarti masih terdapat komentar *spam* yaitu komentar yang tidak berkorelasi dengan aplikasi MyPertamina, maka dari itu dibutuhkan cara penyaringan yang lebih baik untuk menghasilkan data bersih.

DAFTAR RUJUKAN

- [1] Kurniawan D. & Sugiharto J. “Mulai 1 Juli Beli Peralite dan Solar Wajib Daftar Aplikasi MyPertamina,” 28 Juni 2022. [Online]. Available: <https://otomotif.tempo.co/>. [Diakses 18 Agustus 2022]
- [2] Fauzan R. “Heboh Beli BBM Wajib Paki MyPertamina per 1 Agustus, Pertamina Masih Menunggu Perpres,” 1 Agustus 2022. [Online]. Available: <https://ekbis.sindonews.com/>. [Diakses 18 Agustus 2022].
- [3] Riani S.P. & Rosana F.C. “Pertamina Sebut Kebijakan Beli Peralite Pakai MyPertamina Bukan untuk Pembatasan,” 30 Juni 2022. [Online]. Available: <https://bisnis.tempo.co/>. [Diakses 18 Agustus 2022]
- [4] Pressman, Roger S., and Rekayasa Perangkat Lunak. "Edisi 7." *Yogyakarta: Andi* (2012).
- [5] Blei, David M. "Probabilistic topic models." *Communications of the ACM* 55.4 (2012): 77-84.
- [6] Smatana, Miroslav, and Peter Butka. "TopicAE: a topic modeling autoencoder." *Acta Polytechnica Hungarica* 16.4 (2019): 67-86.
- [7] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3.Jan (2003): 993-1022.
- [8] Jelodar, Hamed, et al. "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey." *Multimedia Tools and Applications* 78.11 (2019): 15169-15211.
- [9] Zhao, Weizhong, et al. "A heuristic approach to determine an appropriate number of topics in topic modeling." *BMC bioinformatics*. Vol. 16. No. 13. BioMed Central, 2015.
- [10] Kuribayashi, Tatsuki, et al. "Lower perplexity is not always human-like." *arXiv preprint arXiv:2106.01229* (2021).
- [11] Stevens, Keith, et al. "Exploring topic coherence over many models and many topics." *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*. 2012.
- [12] Gupta, Rahul Kumar, et al. "Prediction of Research Trends using LDA based Topic Modeling." *Global Transitions Proceedings*(2022).
- [13] Sievert, Carson, and Kenneth Shirley. "LDavis: A method for visualizing and interpreting topics." *Proceedings of the workshop on interactive language learning, visualization, and interfaces*. 2014.
- [14] Rehurek, Radim, and Petr Sojka. "Gensim–python framework for vector space modelling." *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic* 3.2 (2011): 2.