

**Analisis Sentimen UKSW Melalui Google Maps
Dengan Model IndoBERT**



DISUSUN OLEH :

Haleluya Noka

672022069

M. Rizadry Eka Syaputra

672022126

**UNIVERSITAS KRISTENSATYAWACANA
SALATIGA
2025**

DAFTAR ISI

Daftar Isi	1
Latar Belakang Masalah	2
Tinjauan Pustaka	3
Metode Penelitian.....	7
Hasil Pembahasan.....	8
Kesimpulan	21
Daftar Pustaka	21

Latar Belakang Masalah

Universitas Kristen Satya Wacana adalah salah satu universitas di Kota Salatiga, dimana universitas ini memiliki julukan sebagai “Kampus Indonesia Mini”. Tempat dimana mahasiswa dan mahasiswi yang ada berasal dari berbagai macam tempat dan berkuliah menjadi satu di Salatiga, tepatnya di Universitas Kristen Satya Wacana (UKSW). Namun, muncul suatu permasalahan umum yang sering muncul di berbagai kampus-kampus Indonesia yang ada. Mengenai fasilitas yang disediakan oleh pihak kampus bagi seluruh mahasiswa/mahasiswi yang sedang mengampu jenjang pendidikan terkhususnya di UKSW. Permasalahan ini, tentu sangatlah krusial bagi seluruh sivitas akademika di UKSW. Juga tentu permasalahan ini akan mempengaruhi calon mahasiswa/mahasiswi yang akan mendaftarkan dirinya di UKSW. Oleh karena itu, penelitian ini membawa permasalahan mengenai bagaimana sentimen dari tiap perseorangan yang sudah melihat atau merasakan dari fasilitas yang disediakan oleh pihak UKSW. Penelitian sentimen ini, dapat dilaksanakan dengan cara mengambil seluruh hasil review yang sudah dengan sengaja diberi oleh pengguna dari situs *Google Maps* agar pihak UKSW dapat mengerti mengenai sentimen mereka tentang fasilitas yang sudah diberikan. Alasan dibalik digunakannya situs *Google Maps* dan bukan situs lain seperti media sosial, ini dikarenakan fokus domain dari situs *Google Maps* hanya berfokuskan pada tempat yang dicari dan tidak melebar kedalam hal yang lain seperti contohnya pada situs lainnya. Selain mengambil seluruh hasil review pada situs *Google Maps*, pada penelitian ini perlu juga untuk ditentukan model mana yang cocok dengan dataset yang didapat. Dikarenakan dataset yang didapat adalah data bertipe teks, maka untuk melakukan hasil analisis sentimen diperlukan model yang dapat mengolah *Natural Language Processing* (NLP). Oleh karena itu, ditentukanlah model *deep learning* dari IndoBERT untuk dapat belajar dari dataset yang didapat. Ini dikarenakan model IndoBERT dapat mengolah data teks dengan fokus Bahasa Indonesia untuk menjadi suatu klasifikasi sentimen, selain itu model ini juga memiliki arsitektur dasar dari BERT (*Bidirectional Encoder Representations from Transformers*). Dimana, model *deep learning* ini jarang digunakan dalam penelitian dikarenakan model ini masih tergolong baru. Oleh karena itu dengan dilakukannya penelitian ini juga, diharapkan dapat membuat peneliti lain untuk dapat tertarik melakukan penelitian dengan menggunakan model dari IndoBERT juga. Jika ingin melakukan analisis sentimen dengan dataset berfokus pada Bahasa Indonesia. Namun, tetap tujuan utama dari dilakukannya penelitian ini adalah untuk mengetahui bagaimana keseluruhan sentimen yang diberikan untuk fasilitas yang sudah disediakan oleh pihak UKSW.

Tinjauan Pustaka

Judul Jurnal	Masalah	Metode Penelitian	Tujuan	Manfaat	Solusi	Saran
E. Yulianti and N. K. Nissa, "ABSA of Indonesian customer reviews using IndoBERT: single-sentence and sentence-pair classification approaches," <i>Bulletin of Electrical Engineering and Informatics</i>	Sulit untuk melakukan identifikasi aspek dan sentimen dalam ulasan konsumen	Menggunakan model <i>deep learning</i> IndoBERT untuk melakukan klasifikasi	Melakukan klasifikasi aspek dan sentimen secara akurat dan tepat	Meningkatkan pemahaman mengenai opini konsumen	Menggunakan pendekatan pasangan kalimat untuk konteks sentimen yang lebih baik	Perlu dilakukan eksplorasi lebih besar pada model <i>deep learning</i> IndoBERT
S. Riyadi, L. K. Salsabila, C. Damarjati, and R. A. Karim, "INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi 233 Sentiment Analysis of YouTube Users on Blackpink Kpop Group	Banyak opini pro-kontra terhadap grup <i>K-Pop Blackpink</i>	<i>Fine-tuning</i> dengan menggunakan model IndoBERT untuk analisis sentimen di <i>Youtube</i>	Melakukan analisis sentimen publik dari opini yang didapat dari <i>Youtube</i> terhadap <i>Blackpink</i>	Memberikan wawasan mengenai sosial media untuk fans dan manajemen dari <i>Blackpink</i>	Model IndoBERT lebih efektif pada teks pendek <i>Youtube</i>	Tambahkan data non- <i>Youtube</i> untuk hasil dapat lebih bersifat publik, tidak hanya berfokuskan pada data <i>Youtube</i> saja

Using IndoBERT						
S. Aras, M. Yusuf, R. Y. Ruimassa, E. A. B. Wambrauw, and E. B. Pala'ngan, "Sentiment Analysis on Shopee Product Reviews Using IndoBERT," <i>Journal of Information Systems and Informatics</i>	Review konsumen yang banyak dan bervariasi sulit diklasifikasi manual	<i>Fine-tuning</i> dengan menggunakan model IndoBERT untuk melakukan klasifikasi	Melakukan klasifikasi sentimen dari ulasan yang didapat dari ulasan <i>Shopee</i>	Dapat membantu penjual dalam memahami <i>feedback</i> produk yang dijual	Menggunakan model IndoBERT sebagai dasar model yang kuat	Kembangkan model IndoBERT untuk dapat melakukan klasifikasi aspek spesifik terhadap suatu produk saja
Fransiscus and A. S. Girsang, "Sentiment Analysis of COVID-19 Public Activity Restriction (PPKM) Impact using BERT Method," <i>International Journal of Engineering Trends and Technology</i>	Ketidakpastian opini masyarakat terkait kebijakan PPKM	Menggunakan model <i>deep learning</i> BERT untuk melakukan klasifikasi	Memahami dampak kebijakan PPKM berdasarkan opini masyarakat	Dapat sebagai informasi tambahan bagi pengambil kebijakan	Model BERT mampu menangkap konteks opini masyarakat dalam kebijakan sosial	Disarankan data dikombinasikan dengan media berita agar hasil yang didapat mempunyai skala yang lebih luas
Y. A. Singgalen,	Kebutuhan untuk	<i>Fine-tuning</i> dengan	Menyediakan	Membantu industri	Model IndoBERT	Tambahkan fitur lokasi

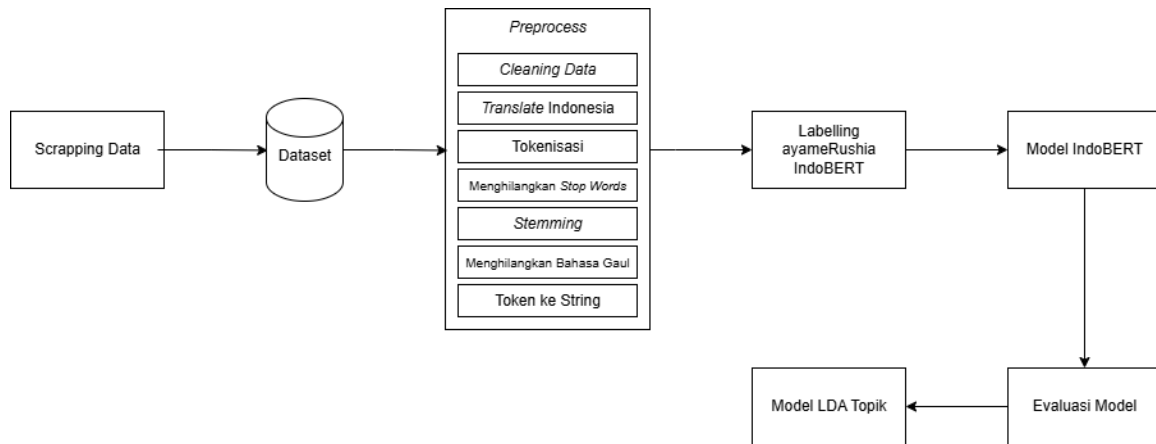
“Performance Analysis of IndoBERT for Sentiment Classification in Indonesian Hotel Review Data,” <i>Article in Journal of Information System Research</i>	memahami preferensi dari tamu hotel berdasarkan hasil review online	menggunakan model IndoBERT untuk melakukan klasifikasi sentimen	wawasan preferensi baru bagi tamu hotel	pariwisata dalam meningkatkan pelayanan	cocok untuk ulasan hotel dikarenakan hasil akurat	dan jenis hotel untuk konteks yang lebih luas
Y. A. Singgalen, “IndoBERT-Based Sentiment Analysis for Understanding Hotel Guests’ Preferences,” <i>Article in Journal of Computer System and Informatics</i>	Review hotel belum digunakan secara optimal untuk dilakukan pengambilan keputusan bisnis	Menggunakan model <i>deep learning</i> IndoBERT untuk memahami preferensi pengguna	Menyediakan wawasan mendalam terkait bagaimana kepuasan tamu terhadap suatu hotel	Membantu manajemen hotel dalam meningkatkan inovasi dan pelayanan	Model IndoBERT dapat membantu memahami ekspresi emosi dari pengguna	Perlu untuk menguji pada platform lain selain situs dari suatu hotel
H. Imaduddin, F. Yusfida A’la, and Y. S. Nugroho, “Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach”	Banyak review negatif / positif terkait aplikasi kesehatan Indonesia	Menggunakan model <i>deep learning</i> IndoBERT untuk melakukan sentimen analisis	Melakukan evaluasi performa sentimen aplikasi <i>e-health</i>	Dapat meningkatkan pengembangan aplikasi berbasis umpan balik / <i>feedback</i> terhadap pengguna	Model IndoBERT cocok untuk domain kesehatan dikarenakan hasilnya yang akurat	Perlu untuk menerapkan pada ulasan rumah sakit atau BPJS

P. Sayarizki and H. Nurrahmi, "Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates," <i>Journal on Computing</i>	Beragamnya opini masyarakat menjelang datangnya pemilu memicu perlunya analisis opini dalam politik	Menggunakan model <i>deep learning</i> IndoBERT pada media sosial kandidat presiden	Melakukan klasifikasi sentimen terhadap calon presiden yang mendaftar	Membantu untuk melakukan analisis opini dalam politik secara publik	Model IndoBERT efektif untuk teks pendek dalam opini politik	Perlu untuk dilakukan validasi dengan data <i>polling</i> atau survei data secara nyata agar data yang diolah dapat bersifat langsung
M. Z. Arifin, S. Y. Maulana, A. Noertjahyana, and A. Mohamed Asghaiyer, "Analyzing the Indonesian sentiment to rohingya refugees using IndoBERT model," <i>Bulletin of Social Informatics Theory and Application</i>	Banyak komentar sosial bernuansa negatif / positif tentang rohingya yang mengungsi di Indonesia	Menggunakan model <i>deep learning</i> IndoBERT pada komentar di media sosial dan berita	Mengetahui sentimen masyarakat Indonesia tentang pengungsi rohingya	Dapat menjadi Informasi penting bagi organisasi kemanusiaan	Model IndoBERT berhasil untuk melakukan deteksi opini sensitif terhadap pengungsi rohingya	Perlu studi lanjutan mengenai opini terkait etnis lainnya juga yang mengungsi di Indonesia
A. S. Rizky and E. Y. Hidayat, "Emotion Classification in Indonesian	Sulit untuk membedakan antara sentimen dan emosi dalam teks Bahasa Indonesia	Menggunakan model <i>deep learning</i> IndoBERT untuk melakukan	Melakukan klasifikasi emosi seperti marah, senang, dan sedih	Meningkatkan kemampuan teknologi dalam memahami emosi manusia	Model IndoBERT dapat bekerja secara akurat dalam memahami nuansa emosi	Tambahkan klasifikasi multi-label untuk emosi campuran agar hasil klasifikasi

Text Using IndoBERT”		klasifikasi emosi		terutama dalam teks Bahasa Indonesia	dengan teks Bahasa Indonesia	emosinya dapat lebih spesifik
-------------------------	--	----------------------	--	---	------------------------------------	-------------------------------------

Metode Penelitian

Metode yang dilakukan dalam penelitian ini meliputi beberapa tahap. Pertama, dataset yang didapat merupakan dataset yang diambil dengan cara melakukan *scrapping* melalui *Google Maps* dengan *key search* UKSW. Dimana diambil data dari kolom *reviews* pada *Google Maps*, untuk dapat melakukan penelitian mengenai sentimen dari fasilitas yang telah disediakan oleh Universitas Kristen Satya Wacana. Setelah dataset berhasil diambil, dataset perlu dilakukan tahap *preprocess* terlebih dahulu. Dimana, tahap *preprocess* ini ada beberapa tahapan didalamnya yang nantinya membuat hasil penelitian semakin akurat dan jelas. Tahapan dari *preprocess* ini diantaranya adalah *cleaning data*, penerjemahan dari Bahasa Inggris ke Bahasa Indonesia, tokenisasi, menghilangkan *stop words*, *stemming*, menghilangkan bahasa gaul, lalu mengubah kembali yang awal dipisah menjadi pertoken lalu digabungkan kembali menjadi kalimat utuh. Setelah melalui tahap *preprocess*, tahap selanjutnya adalah masuk dalam tahap *labelling* dimana dipakai model *pre-train* dari *ayameRushia* IndoBERT. Tahap *labelling* dengan menggunakan model *pretrain*, bertujuan untuk memberi label asli pada dataset agar model nantinya dapat belajar dari dataset yang sudah diberi label ini. Tahap selanjutnya, adalah melakukan implementasi model dari IndoBERT untuk dapat belajar dari dataset yang sudah diolah dari tahapan-tahapan sebelumnya. Setelah model dari IndoBERT berhasil belajar dari dataset, model IndoBERT dilakukan evaluasi dengan parameter *Accuracy*, *Recall*, *Precision*, *F1-score*, *ROC AUC*, *Classification Report*, dan terakhir *Confusion Matrix*. Tujuan dilakukan evaluasi dalam model dikarenakan model perlu diteliti apakah model IndoBERT sudah dapat belajar dengan baik dari dataset yang telah diberikan atau belum. Pada tahap terakhir, tiap komentar dari dataset yang sudah dilakukan proses sentimen analisis dikelompokkan menjadi pertopik. Semisal dengan topik jumlah kata yang sering muncul dalam tiap sentimen maupun dalam tiap fasilitas yang bersangkutan dengan UKSW, pengelompokan topik ini menggunakan model dari *Latent Dirichlet Allocation* (LDA). Untuk gambaran lebih jelas dari tahapan metode penelitian yang dilakukan, dapat dilihat pada Gambar 1 berikut.



Gambar 1. Tahapan metode penelitian

Hasil Pembahasan

- ***Business Understanding***

Tujuan utama dari dilakukannya penelitian ini adalah untuk melakukan analisis mengenai sentimen dari fasilitas yang telah disediakan oleh UKSW. Dengan fokus pada situs *Google Maps*, alasan dibalik digunakannya *Google Maps* dikarenakan situs ini hanya berfokus pada tempat yang dicari saja. Oleh karena itu, dengan digunakannya *Google Maps* diharapkan topik ataupun opini yang dikomentari oleh pengguna dari situs *Google Maps* tidak keluar dari pembahasan fasilitas UKSW yang dicari dalam *Google Maps*. Dengan hasil dari penelitian sentimen analisis ini, diharapkan selain dapat untuk menjadi dasar perubahan dan perbaikan fasilitas atau pelayanan dalam fasilitas yang disediakan oleh UKSW. Diharapkan juga hasil penelitian ini untuk dapat menjadi media pembelajaran hal baru, dimana pihak UKSW dapat mengetahui ulasan balik dari pengguna yang menggunakan fasilitas yang disediakan.

- ***Data Understanding***

Dataset yang didapat adalah dataset yang berawal dalam kolom *reviews* pada situs *Google Maps*, dengan pencarian pada seluruh fasilitas yang disediakan oleh UKSW. Dimana dataset didapatkan dengan cara mengumpulkan semua hasil dataset dengan teknik *scraping*, teknik dimana mengumpulkan data yang berasal dari situs website. Dataset yang berhasil dikumpulkan ini masih bersifat mentah dikarenakan masih belum dibersihkan dalam tahap *preprocess* dan juga masih belum ada *true label* dan harus untuk diberi pelabelan terlebih dahulu pada tahapan *labelling*, ini mengapa dataset ini masih tergolong mentah dan tidak bisa langsung digunakan didalam penelitian sentimen analisis. Jumlah dari dataset mentah yang

dikumpulkan ada sebanyak 2017, dengan hasil ulasan yang didapat dari seluruh fasilitas yang disediakan oleh UKSW pada situs *Google Maps*.

- ***Preprocess***

Tahap *preprocess* digunakan untuk membuat model nantinya saat ingin mengolah dataset yang digunakan sebagai dataset penelitian untuk dapat belajar dengan optimal setelah dataset tersebut disesuaikan dengan model yang digunakan. Untuk tahapan *preprocess*, terdapat beberapa tahapan didalamnya diantaranya adalah sebagai berikut.

- ***Cleaning Dataset***

Tahapan *cleaning* dataset, ini digunakan untuk menghilangkan beberapa karakter unik dalam suatu kalimat. Ini digunakan agar model nantinya dapat belajar dengan baik dari dataset yang diberikan. Karakter unik yang dihilangkan saat tahap *cleaning* dataset ini adalah seperti karakter kalimat yang memiliki ciri-ciri seperti *url* atau sebuah website, karakter unik seperti *mention* dalam sebuah kalimat, karakter unik seperti *hashtag* dalam sebuah kalimat, karakter unik seperti *emoticon* pada sebuah kalimat, karakter unik seperti tanda baca dalam sebuah kalimat, karakter unik seperti angka dalam sebuah kalimat, karakter unik seperti contohnya jumlah spasi yang berlebih dalam sebuah kalimat, dan terakhir karakter unik seperti emoji dalam sebuah kalimat. Namun, tidak hanya menghilangkan karakter-karakter unik saja. Dalam tahap *preprocess* juga, semua komentar hasil *review* dari situs *Google Maps* diubah menjadi huruf kecil semua atau biasa disebut sebagai *lower casing*.

- ***Translate Indonesia***

Tahapan untuk melakukan penerjemahan ke Bahasa Indonesia, ini biasanya tidak muncul pada tahap *preprocess* seperti jika langsung menggunakan dataset yang sesuai dengan Bahasa yang diinginkan. Namun, dikarenakan *default setting* dari situs *Google Maps* menggunakan Bahasa Inggris. Ini menyebabkan perlunya untuk melakukan penerjemahan ulang untuk hasil sentimen dari *Google Maps* dalam berbentuk Bahasa Indonesia. Fungsi yang digunakan dalam tahap penerjemahan dari Bahasa Inggris ke Bahasa Indonesia, ini menggunakan *GoogleTranslator*. Dimana, dataset setelah dilakukan pembersihan pada tahap sebelumnya. Dilanjutkan pada tahap penerjemahan ke Bahasa Indonesia memakai *GoogleTranslator*.

- **Tokenisasi**

Tahap tokenisasi ini adalah tahap dimana dalam suatu kalimat dipisah menjadi perkata, tahap ini krusial untuk dilakukan dikarenakan model nantinya dapat belajar dari potongan-potongan kata tersebut untuk menjadi suatu kesimpulan sentimen. Ini mengapa, setelah dataset berhasil untuk di terjemahkan ke Bahasa Indonesia. Penting sekali, untuk dilakukan tokenisasi pada dataset yang telah selesai diterjemahkan. Agar dapat membuat hasil analisis sentimen dari penelitian ini.

- **Menghilangkan *Stop Words***

Tahapan menghilangkan *Stop Words*, ini digunakan untuk menghilangkan kata-kata umum dalam kata-kata yang sudah dipisah dalam tahap tokenisasi sebelumnya. Tujuannya agar kata-kata umum yang sering muncul pada suatu kalimat namun tidak membawa makna ataupun suatu sentimen kepada suatu kalimat, ini agar tidak mengganggu proses model nantinya agar model dapat belajar dari dataset dengan baik.

- ***Stemming***

Tahap *stemming*, ini digunakan untuk mengembalikan kata-kata kerja yang memiliki imbuhan awal atau akhiran menjadi kata dasar saja. Proses *stemming* ini juga sangat krusial pada tahap *preprocess*, dikarenakan model cenderung dapat belajar untuk menyimpulkan kata-kata menjadi suatu sentimen dengan mengolah dari kata dasar. Jadi setelah kata-kata dipisah dari tahap tokenisasi dan untuk kata umum dihilangkan, maka tahap selanjutnya adalah mengubah kata-kata tersebut menjadi kata dasar.

- **Menghilangkan Bahasa Gaul**

Tahap menghilangkan bahasa *gaul* atau bahasa kekinian yang sering digunakan dalam berkomunikasi melalui internet, ini digunakan untuk membuat hasil kesimpulan analisis sentimen menjadi semakin optimal. Dikarenakan bahasa *gaul*, ini terkadang memiliki konotasi yang kurang jelas untuk disimpulkan menjadi suatu sentimen. Ini mengapa, penting untuk setelah diubah menjadi kata dasar. Dataset perlu dihilangkan kata-kata *gaul* agar dapat membuat hasil analisis sentimen yang optimal

- **Token ke String**

Tahap token ke *string*, ini adalah proses dimana menggabungkan kata-kata yang sudah terpisah dari tahap sebelumnya. Digabungkan menjadi satu kembali menjadi

Untuk hasil dataset setelah dilakukan tahap *preprocess*, telah dibuat visualisasi menggunakan *WordCloud* untuk dapat melihat gambaran umum dari hasil proses. Visualisasi *WordCloud* dapat dilihat dan dipahami pada Gambar 2 berikut.

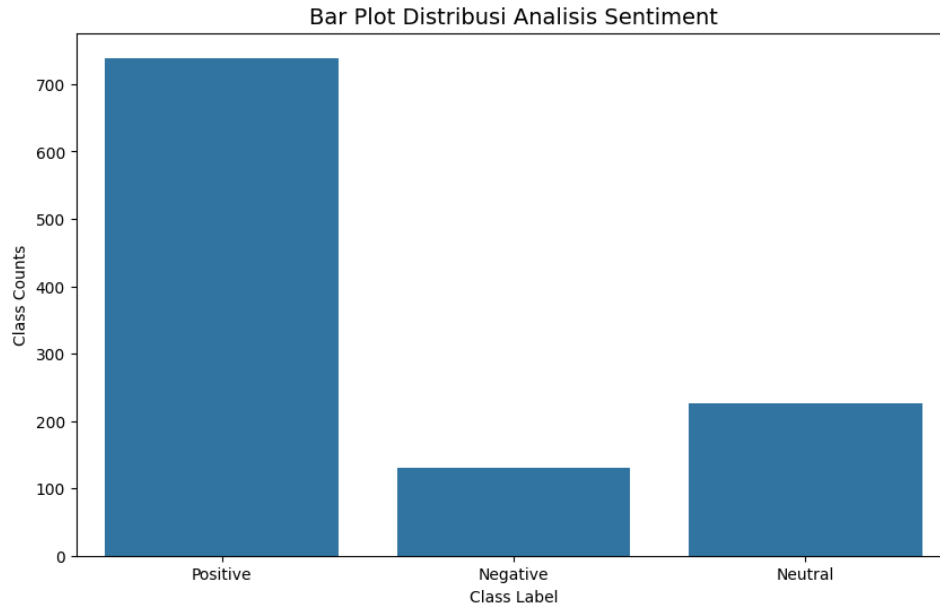


Gambar 2. Hasil *preprocessing*

- *Labelling*

Proses *labelling* digunakan dalam penelitian ini, untuk membuat *true label* atau label asli dalam pengelompokan sentimen seperti *Negative*, *Positive*, atau *Neutral*. Proses ini tergolong krusial, dikarenakan sebuah model klasifikasi penting untuk punya sebuah label kategori untuk ditentukan klasifikasinya. Oleh karena itu, sebelum model dilakukan implementasi pada penelitian ini untuk membuat kesimpulan analisis sentimen. Dilakukan terlebih dahulu proses *labelling*, proses ini memakai fungsi dari model *pre-train ayameRushia* IndoBERT. Dimana model *pre-train* ini, memakai dasar arsitektur dari IndoBERT yang sudah dilatih dengan fokus dataset teks Bahasa Indonesia untuk menghasilkan suatu sentimen. Alasan dibalik digunakannya model ini, dikarenakan model *ayameRushia* IndoBERT mempunyai hasil akurasi yang tinggi saat mengolah suatu data teks dengan Bahasa Indonesia menjadi suatu sentimen sebesar 0,93. Juga selain model ini mempunyai tingkat akurasi yang tinggi, model *ayameRushia* mudah untuk dilakukan implementasi contohnya dalam melakukan pelabelan dalam dataset. Dalam tahap *labelling*, dibuatlah hasil visualisasi *bar chart* untuk melihat hasil

pelabelan sentimen pada dataset yang dapat dilihat pada Gambar 3. Dari Gambar 3, menunjukkan bahwa hasil *labelling* terdapat sentimen *Positive* sebanyak 739, *Neutral* sebanyak 227, dan *Negative* sebanyak 131.



Gambar 3. Visualisasi *bar chart* hasil pelabelan

- **IndoBERT Model**

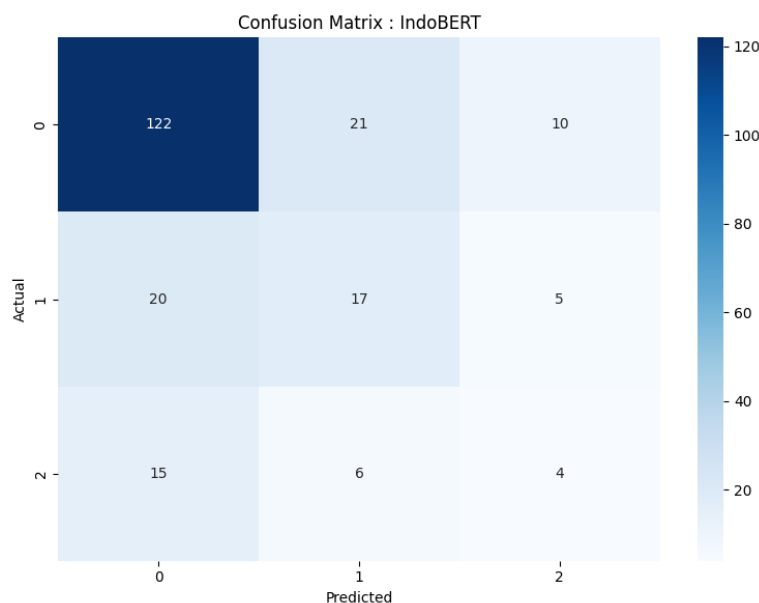
Pada tahap implementasi dengan model IndoBERT, dilakukan pengubahan pada hasil tahap *labelling*. Dengan tipe data awal bersifat kategori dirubah menjadi numerik, dimana untuk sentimen *Positive* dirubah menjadi 0, *Neutral* dirubah menjadi 1, dan terakhir *Negative* dirubah menjadi 2. Setelah tipe data dirubah menjadi numerik, dilakukanlah implementasi model IndoBERT untuk dapat melakukan pembobotan kata dengan teknik pembobotan CLF (*Confident Learning Framework*). Dimana teknik ini digunakan untuk melakukan estimasi seberapa baik model IndoBERT dapat belajar dari dataset yang diberikan, dalam sistem pembobotan CLF nilai -0 ini memiliki arti sentimen *Negative*, 0 memiliki arti sentimen *Neutral*, dan 1 memiliki arti sentimen *Positive*. Setelah pembobotan CLF pada dataset selesai, hasil pembobotan dan hasil pengubahan tipe data menjadi numerik dipisah menjadi 2 variabel yang berbeda. Dimana setelah dipisah menjadi 2 variabel yang berbeda, lalu dilakukan *splitting data* dari kedua variabel tersebut sebesar 80:20. 80% dimasukan kedalam dataset yang akan dilakukan *training* untuk model IndoBERT dan 20% sisanya dimasukan kedalam dataset yang akan dilakukan *testing* untuk prediksi dari sentimen yang sudah dipelajari dari data *training* menggunakan IndoBERT sebelumnya.

- **Evaluasi Model**

Dalam tahap evaluasi model, digunakanlah beberapa parameter seperti *Accuracy*, *Recall*, *Precision*, *F1-Score*, *ROC AUC*, *Classification Report*, dan terakhir *Confusion Matrix*. Parameter ini memiliki fungsinya tersendiri seperti *Accuracy* yang dipakai untuk mengukur prediksi yang benar dari total jumlah prediksi, *Precision* yang dipakai untuk mengukur seberapa banyak dari prediksi *Positive* yang benar-benar *Positive*, *Recall* yang dipakai untuk mengukur seberapa banyak dari semua sampel *Positive* yang berhasil dilakukan identifikasi dari model, *F1-Score* yang dipakai untuk memberikan skor dimana menyeimbangkan nilai *Precision* dan *Recall*, *ROC AUC* yang dipakai untuk mengukur kemampuan model dalam membedakan antara kelas *Positive* dengan *Negative* diberbagai ambang batasan klasifikasi, *Classification Report* yang dipakai untuk ringkasan dalam menyediakan nilai evaluasi dari setiap kelas label yang ada, dan terakhir *Confusion Matrix* adalah visualisasi yang dipakai untuk melihat hasil kinerja model dalam melakukan klasifikasi sentimen. Dalam penelitian ini, dihasilkanlah hasil evaluasi seperti dalam Tabel 1 dan Gambar 4 untuk hasil *Confusion Matrix* dari hasil implementasi dengan model IndoBERT dalam melakukan analisis sentimen.

Accuracy	Recall	Precision	F1-Score	ROC AUC
65.0%	65.0%	63.81%	64.35%	0.6245

Tabel 1. Tabel hasil evaluasi model IndoBERT



Gambar 4. Hasil *confusion matrix* dari model IndoBERT

Dapat dilihat dari hasil Tabel 1, nilai *Accuracy* sampai *ROC AUC* sudah lumayan tinggi diatas 50%. Namun dapat dilihat pada Gambar 4 hasil dari *Confusion Matrix*, model lebih tepat melakukan klasifikasi pada kelas sentimen label *Positive* dari pada kelas label yang lain. Dimana model dapat melakukan klasifikasi kelas *Positive* dengan tepat sebanyak 122 sentimen. Jika dibanding dengan kelas label seperti *Neutral* dan *Negative*, kelas label *Neutral* hanya dapat melakukan 4 klasifikasi sentimen dengan benar dan kelas label *Negative* hanya dapat melakukan 17 klasifikasi sentimen dengan benar. Tentu, hasil ini sangat timpang balik jika dibandingkan dengan kelas label dari *Positive*. Ini dikarenakan, dataset yang didapat lebih banyak mempunyai kelas label *Positive* yang dapat dilihat pada Gambar 3 sebanyak 739 sentimen. Hasil dari visualisasi dengan *Confusion Matrix* dalam penelitian ini, dapat didukung dengan hasil yang ditunjukkan dari *Classification Report* pada Gambar 5 berikut.

Classification Report:				
	precision	recall	f1-score	support
Positive	0.78	0.80	0.79	153
Neutral	0.39	0.40	0.40	42
Negative	0.21	0.16	0.18	25
accuracy			0.65	220
macro avg	0.46	0.45	0.45	220
weighted avg	0.64	0.65	0.64	220

Gambar 5. Hasil *classification report* dari model IndoBERT

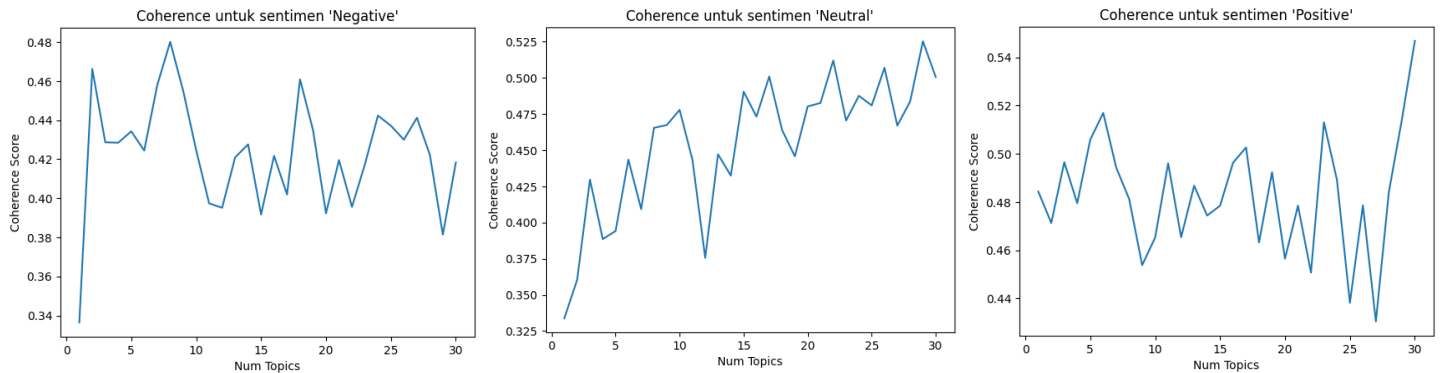
- **Model LDA**

Model LDA (*Latent Dirichlet Allocation*) digunakan untuk mengelompokkan kata-kata menjadi berbentuk topik. Tujuannya adalah untuk mencari frekuensi kata-kata dari topik tertentu yang sering muncul. Pada penelitian ini, digunakan model LDA untuk mencari topik pada kata-kata yang sering muncul pada semua kelas label sentimen dan pada setiap sentimen dari fasilitas yang disediakan oleh UKSW.

- **Model LDA Pada Kelas Label Sentimen**

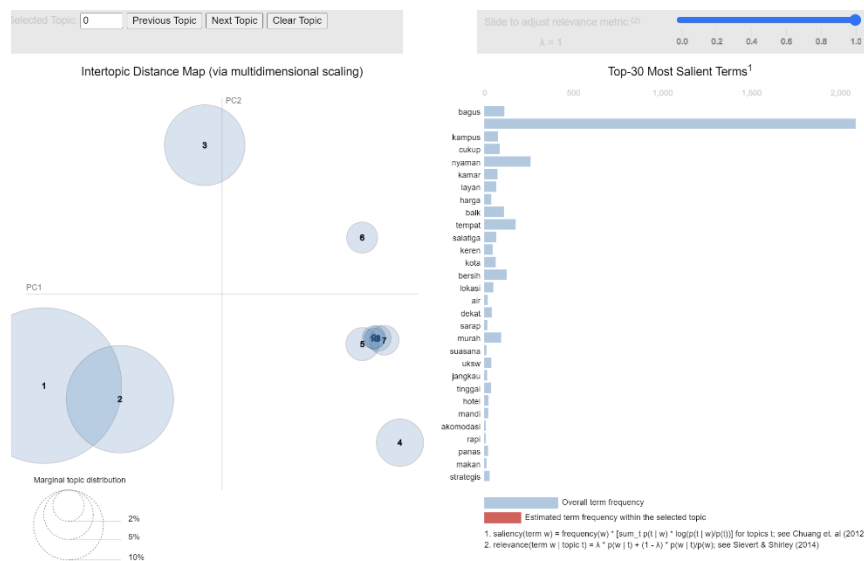
Sebelum membentuk hasil visualisasi dari model LDA, dilakukan terlebih dahulu menentukan jumlah topik untuk menghasilkan nilai *Coherence* tertinggi dan stabil dari ketiga kelas label sentimen tersebut. Maka dibuatlah visualisasi *line chart* untuk memberikan gambaran umum dengan total jumlah 30 topik untuk dapat melihat topik

mana yang menghasilkan nilai *Coherence* tertinggi, yang dapat dilihat pada Gambar 6 berikut.

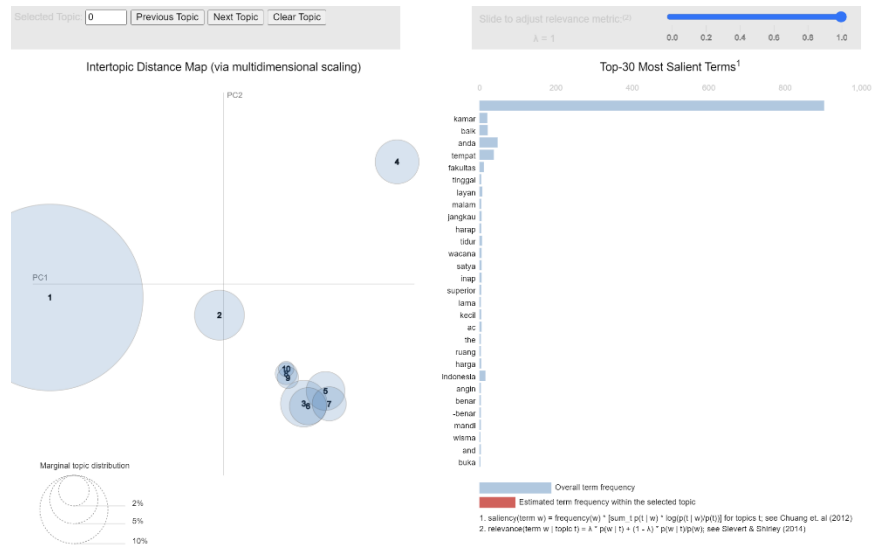


Gambar 6. Visualisasi *line chart* untuk pembagian topik pada model LDA pada sentimen

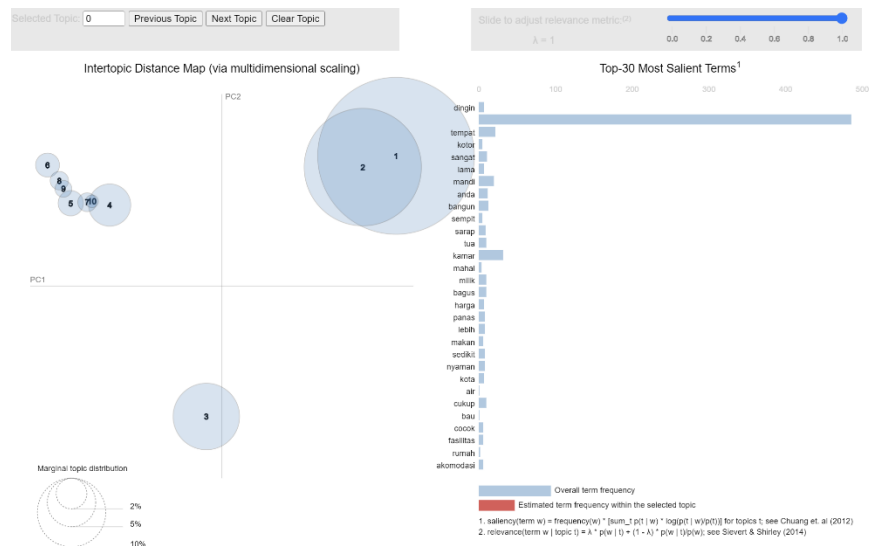
Dari hasil visualisasi pada Gambar 6, diambil nilai tertinggi dan seimbang untuk ketiga kelas label sentimen yaitu 10. Dikarenakan dari ketiga kelas label sentiment setelah model LDA dibagi menjadi 10 topik, ditopik selanjutnya nilai *Coherence* turun ini menunjukkan bahwa nilai 10 pada pembagian topik sudah dapat dianggap seimbang dan tertinggi untuk dijadikan patokan pembagian topik pada model LDA. Setelah ditentukan jumlah topik untuk membuat visualisasi dari model LDA, visualisasi dengan model LDA dibuat yang dapat dilihat pada Gambar 7, Gambar 8, dan Gambar 9 berikut.



Gambar 7. Visualisasi model LDA untuk sentimen *Positive*



Gambar 8. Visualisasi model LDA untuk sentimen *Neutral*



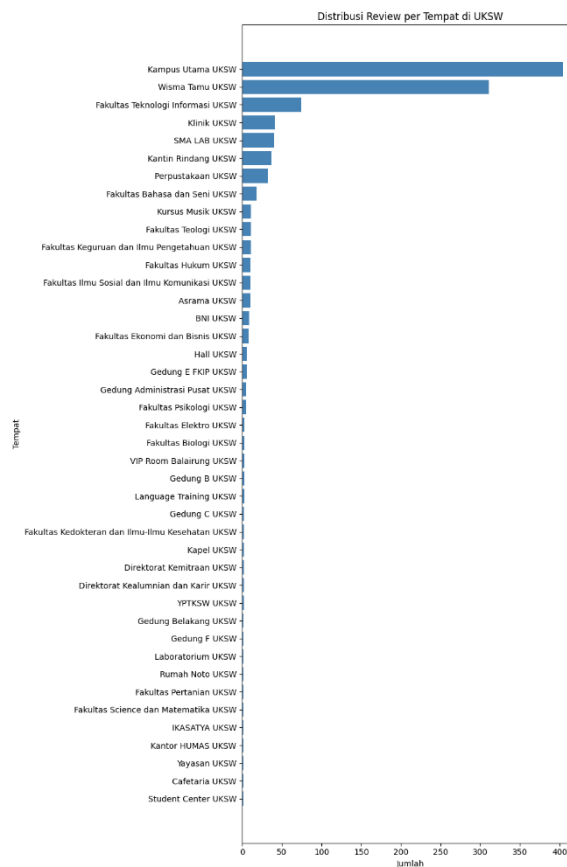
Gambar 9. Visualisasi model LDA untuk sentimen *Negative*

Dapat dilihat, dari hasil model LDA sentimen *Positive* kebanyakan muncul kata seperti bagus, kampus, cukup, nyaman, kamar, dan lainnya ini dapat disimpulkan visualisasi dari model LDA kelas sentimen *Positive* sudah dapat dengan akurat menggambarkan kata-kata yang sering muncul dalam sentimen *Positive*. Pada kelas sentimen *Neutral*, sering muncul kata-kata seperti kamar, baik, anda, tempat, fakultas, dan lainnya yang menunjukkan bahwa memang benar kata-kata ini bersifat *Neutral*. Dan

pada kelas sentimen *Negative* sering muncul kata-kata seperti dingin, tempat, kotor, sangat, lama, dan lainnya yang memang menunjukkan konotatif kata-kata bersifat *Negative*. Hal ini juga ditunjukkan dari nilai *Coherence* dari model LDA, untuk model LDA sentimen *Positive* mendapatkan nilai sebesar 0.4653, sentimen *Negative* sebesar 0.4246, dan terakhir sentimen *Neutral* sebesar 0.4778.

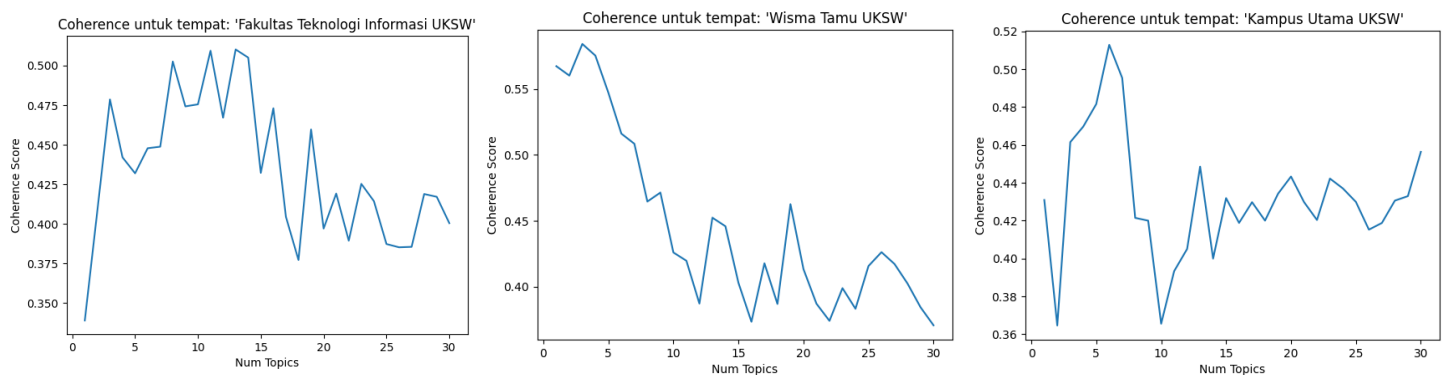
○ Model LDA Pada Sentimen Fasilitas UKSW

Untuk tahapan melakukan implementasi model LDA terhadap sentimen dari tiap fasilitas UKSW, tahapan implementasi sama seperti pada tahap implementasi model LDA pada kelas label sentimen. Namun, terdapat perbedaan pada tahapan implementasinya. Dikarenakan tidak semua komentar *review* dari *Google Maps* terbagi rata bagi semua fasilitas yang disediakan oleh UKSW. Oleh karena itu, dibuatlah visualisasi dari *bar chart* untuk dapat melihat persebaran sentimen pada tiap fasilitas di UKSW yang dapat dilihat pada Gambar 10 berikut.



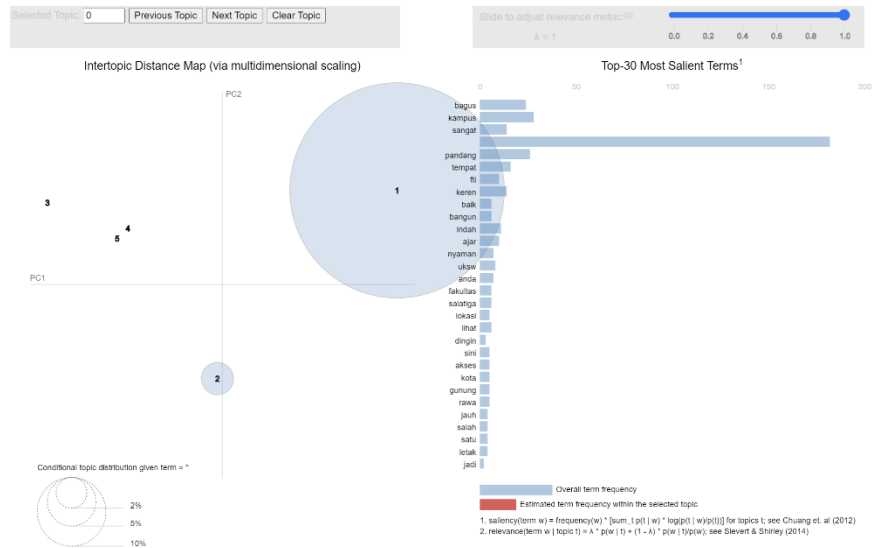
Gambar 10. Visualisasi *bar chart* distribusi *review* fasilitas UKSW

Pada visualisasi *bar chart* diatas, diambil peringkat 3 besar dari distribusi *reviews* terbanyak yang didapat dari tiap fasilitas UKSW yaitu Kampus Utama UKSW, Wisma Tamu UKSW, dan Fakultas Teknologi Informasi UKSW. Alasan hanya diambil peringkat 3 teratas, ini dikarenakan banyak hasil *reviews* yang tidak merata pada semua fasilitas yang disediakan oleh UKSW yang menyebabkan model LDA tidak menjadi akurat atau valid. Setelah ditentukan sentimen dari fasilitas UKSW mana saja yang ingin diteliti, ditentukan jumlah topik kembali yang menghasilkan nilai *Coherence* tinggi dalam model LDA yang dapat dilihat pada visualisasi *line chart* pada Gambar 11 berikut.

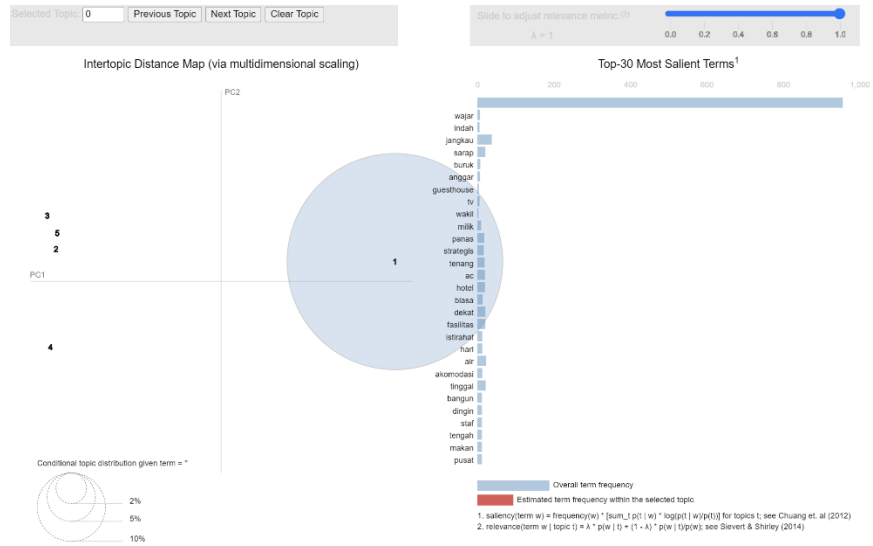


Gambar 11. Visualisasi *line chart* untuk pembagian topik pada model LDA pada fasilitas UKSW

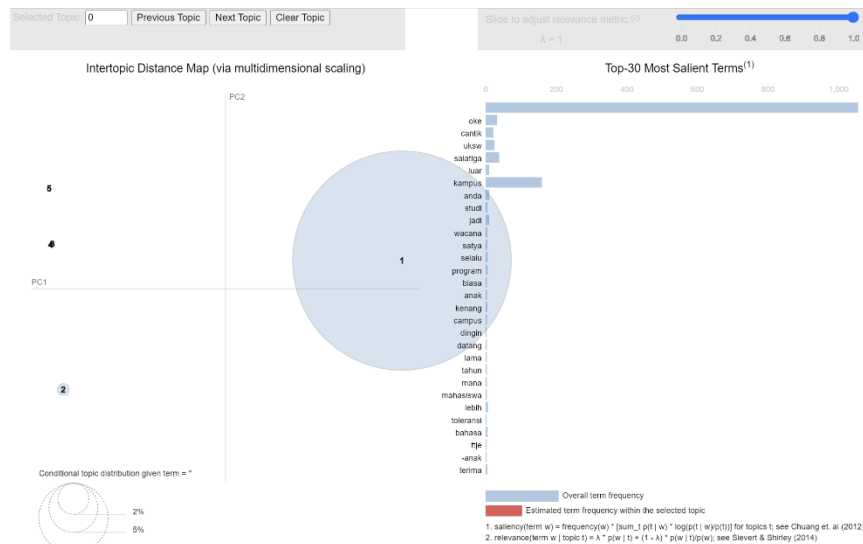
Dapat dilihat dari hasil visualisasi pada Gambar 11, ditentukan kembali jumlah topik yang menghasilkan nilai *Coherence* tertinggi yaitu 5. Dikarenakan dari ketiga visualisasi pada Gambar 11, setelah jumlah topik 5 rata-rata nilai *Coherence* turun. Oleh karena itu, nilai jumlah topik 5 adalah nilai yang seimbang dan tinggi bagi ketiga sentimen dari fasilitas UKSW tersebut. Setelah ditentukan jumlah topik yang tinggi dan seimbang dari model LDA, lalu selanjutnya adalah membuat visualisasi dari model LDA berdasarkan sentimen yang didapat dari fasilitas UKSW. Untuk visualisasi model LDA dapat dilihat pada Gambar 12, Gambar 13, dan Gambar 14 berikut.



Gambar 12. Visualisasi Model LDA pada sentimen fasilitas Fakultas Teknologi Informasi UKSW



Gambar 13. Visualisasi Model LDA pada sentimen fasilitas Wisma Tamu UKSW



Gambar 14. Visualisasi Model LDA pada sentimen fasilitas Kampus Utama UKSW

Dari hasil visualisasi model LDA terhadap sentimen dari ketiga fasilitas UKSW, dapat dilihat bahwa sering muncul kata-kata seperti bagus, kampus, sangat, pandang, tempat, dan kata-kata yang berkonotasi *Positive* dari fasilitas Fakultas Teknologi Informasi UKSW. Ini menandakan bahwa rata-rata fasilitas ataupun pelayanan yang diberikan oleh UKSW dalam Fakultas Teknologi Informasi, mendapatkan sentimen yang *Positive* dari pengguna yang pernah menggunakan fasilitas atau merasakan pelayanan dari Fakultas Teknologi Informasi UKSW. Pada fasilitas Wisma Tamu UKSW, dapat dilihat bahwa sering muncul kata-kata seperti wajar, indah, jangkau, sarap, buruk, dan kata-kata lainnya. Ini menunjukkan terdapat pengguna yang mengeluarkan sentimen *Positive* bagi fasilitas maupun pelayanan UKSW, namun juga ada beberapa pengguna yang mengeluarkan sentimen *Negative* terhadap fasilitas maupun pelayanan dari Wisma Tamu UKSW. Dengan mempelajari *feedback* dari pengguna fasilitas Wisma Tamu UKSW, hal ini bisa dijadikan pembelajaran lebih dalam memperbaiki fasilitas maupun pelayanan dalam Wisma Tamu bersama. Pada fasilitas Kampus Utama UKSW, dapat dilihat sering muncul kata-kata seperti oke, cantik, uksw, salatiga, luar, kampus, dan kata-kata berkonotasi lebih *Neutral*. Ini menunjukkan bahwa rata-rata fasilitas ataupun pelayanan yang diberikan oleh UKSW dalam Kampus Utama UKSW, mendapatkan sentimen *Neutral* dari pengguna.

Kesimpulan

Dengan melakukan penelitian mengenai sentimen yang didapat dari fasilitas yang disediakan oleh UKSW, dapat diketahui bahwa rata-rata sentimen yang didapat adalah *Positive*. Namun, terdapat juga beberapa sentimen *Negative* yang dapat dijadikan dasar sebagai perubahan dan perbaikan fasilitas maupun pelayanan yang diberikan oleh UKSW. Untuk beberapa fasilitas yang mendapatkan sentimen *Neutral*, dapat dijadikan pedoman dalam pengembangan fasilitas maupun pelayanan yang disediakan oleh UKSW agar dapat terus berkembang dan berkompetisi dengan universitas-universitas lainnya. Tidak hanya hasil sentimen saja, namun juga dalam penelitian ini dapat disimpulkan bahwa model IndoBERT sangatlah cocok untuk digunakan dalam mengolah data teks Bahasa Indonesia menjadi suatu kesimpulan sentimen analisis. Juga dapat disimpulkan, bahwa menggunakan model LDA untuk melakukan penelitian mengenai sentimen analisis juga cocok untuk digunakan. Dikarenakan dengan menggunakan model LDA, hasil kesimpulan analisis sentimen dapat dibuat menjadi lebih spesifik fokus kedalam domain topik tertentu dan tidak secara menyeluruh. Dimana, hal ini sangat berguna jika ingin melakukan penelitian mengenai suatu topik tertentu saja. Diharapkan dari dilakukannya penelitian ini, dapat dijadikan pedoman sebagai perubahan dan perbaikan fasilitas maupun pelayanan dalam UKSW menjadi lebih baik lagi. Diluar itu, diharapkan juga dibuatnya penelitian ini dapat menjadi inspirasi dalam pihak UKSW untuk menerapkan sistem analisis sentimen agar dapat mengetahui *feedback* dari pengguna fasilitas UKSW.

Daftar Pustaka

- [1] E. Yulianti and N. K. Nissa, “ABSA of Indonesian customer reviews using IndoBERT: single-sentence and sentence-pair classification approaches,” *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 5, pp. 3579–3589, Oct. 2024, doi: 10.11591/eei.v13i5.8032.
- [2] S. Riyadi, L. K. Salsabila, C. Damarjati, and R. A. Karim, “INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi 233 Sentiment Analysis of YouTube Users on Blackpink Kpop Group Using IndoBERT 1 *,” *INTENSIF*, vol. 8, no. 2, pp. 2549–6824, 2024, doi: 10.29407/intensif.v8n2.22678.
- [3] S. Aras, M. Yusuf, R. Y. Ruimassa, E. A. B. Wambrauw, and E. B. Pala’langan, “Sentiment Analysis on Shopee Product Reviews Using IndoBERT,” *Journal of*

- Information Systems and Informatics*, vol. 6, no. 3, pp. 1616–1627, Sep. 2024, doi: 10.51519/journalisi.v6i3.814.
- [4] Fransiscus and A. S. Girsang, “Sentiment Analysis of COVID-19 Public Activity Restriction (PPKM) Impact using BERT Method,” *International Journal of Engineering Trends and Technology*, vol. 70, no. 12, pp. 281–288, Dec. 2022, doi: 10.14445/22315381/IJETT-V70I12P226.
 - [5] Y. A. Singgalen, “Performance Analysis of IndoBERT for Sentiment Classification in Indonesian Hotel Review Data,” *Article in Journal of Information System Research*, vol. 6, no. 2, 2025, doi: 10.47065/josh.v6i2.6505.
 - [6] Y. A. Singgalen, “IndoBERT-Based Sentiment Analysis for Understanding Hotel Guests’ Preferences,” *Article in Journal of Computer System and Informatics*, vol. 6, no. 2, pp. 508–520, 2025, doi: 10.47065/josyc.v6i2.6864.
 - [7] H. Imaduddin, F. Yusfida A’la, and Y. S. Nugroho, “Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach.” [Online]. Available: www.ijacsa.thesai.org
 - [8] P. Sayarizki and H. Nurrahmi, “Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates,” *Journal on Computing*, vol. 9, no. 2, pp. 61–72, 2024, doi: 10.34818/indojc.2024.9.2.934.
 - [9] M. Z. Arifin, S. Y. Maulana, A. Noertjahyana, and A. Mohamed Asghaiyer, “Analyzing the Indonesian sentiment to rohingya refugees using IndoBERT model,” *Bulletin of Social Informatics Theory and Application*, vol. 8, no. 2, pp. 180–191, Dec. 2024, doi: 10.31763/businta.v8i2.749.
 - [10] A. S. Rizky and E. Y. Hidayat, “Emotion Classification in Indonesian Text Using IndoBERT”.
 - [11] kamus-alay/colloquial-indonesian-lexicon.csv at master · nasalsabila/kamus-alay. GitHub. Published 2025 <https://github.com/nasalsabila/kamus-alay/blob/master/colloquial-indonesian-lexicon.csv>
 - [12] NLP_bahasa_resources/combined_slang_words.txt at master · louisowen6/NLP_bahasa_resources. GitHub. Published 2025. https://github.com/louisowen6/NLP_bahasa_resources/blob/master/combined_slang_words.txt