

# Analisis Sentimen dan Pemodelan Topik Terhadap Aplikasi Pembelajaran Online pada Platform Google Play (Studi Kasus: QUIPPER)

Yose Timothy Gosal, Oktariani Nurul Pratiwi, Iqbal Yulizar Mukti, dan Deden Witarsyah

Program Studi S1 Sistem Informasi, Fakultas Rekayasa Industri,  
Universitas Telkom, Bandung  
E-mail: yosehimothygosal@gmail.com, onurulp@telkomuniversity.ac.id\*),  
iqbal.yulizar@gmail.com, dedenw@telkomuniversity.ac.id

## Abstrak

Kemajuan teknologi informasi dan komunikasi pada era Industri 4.0 telah mengubah metode pengajaran di berbagai lembaga, termasuk melalui aplikasi e-learning seperti Quipper. Quipper memanfaatkan teknologi untuk mendukung pembelajaran. Analisis sentimen ulasan pengguna di Google Play diperlukan untuk memahami persepsi pengguna terhadap aplikasi ini. Penelitian ini menggunakan algoritma K-Nearest Neighbor (KNN) untuk klasifikasi sentimen dan Latent Dirichlet Allocation (LDA) untuk identifikasi topik utama dalam ulasan. Algoritma KNN, dipilih karena kesederhanaannya dan akurasi tinggi, menghasilkan akurasi 84.68%, precision 86.40%, recall 85.87%, dan F1-score 86.09% dengan nilai K terbaik 14. Sebagian besar ulasan bersifat positif (55.1%), menunjukkan kepuasan pengguna terhadap Quipper. Analisis LDA mengidentifikasi topik utama dari ulasan positif dan negatif, yang mengungkap aspek-aspek spesifik yang perlu diperbaiki oleh Quipper untuk meningkatkan kepuasan pengguna.

**Kata kunci** : Analisis Sentimen, K-NN, LDA, Pemodelan Topik, dan Quipper

## Pendahuluan

Kemajuan teknologi informasi dan komunikasi pada era Industri 4.0 memiliki dampak signifikan terhadap metode pengajaran dan pembelajaran di berbagai lembaga, termasuk lembaga pelatihan dan pendidikan [1]. Metode pembelajaran perlu disesuaikan dengan kemajuan dan perkembangan teknologi, khususnya teknologi informasi. Dalam praktiknya, metode ini sebaiknya dapat memanfaatkan berbagai media yang telah berkembang guna mengefektifkan kualitas pembelajaran. Salah satunya adalah dengan menggunakan aplikasi berbasis E-learning. E-Learning ialah sistem pembelajaran melalui aplikasi elektronik yang berperan dalam mendukung perkembangan kegiatan pembelajaran. Saat ini, E-Learning telah mengalami pertumbuhan yang pesat [2].

Akses pembelajaran di Indonesia saat ini juga telah memanfaatkan e-learning, dimulai dari adanya aplikasi yang gratis maupun yang menetapkan biaya berlangganan. Salah satu media belajar yang beragam, fleksibel, dan variatif adalah Quipper [3]. Quipper merupakan media pembelajaran yang berbasis *e-learning* yang memadukan

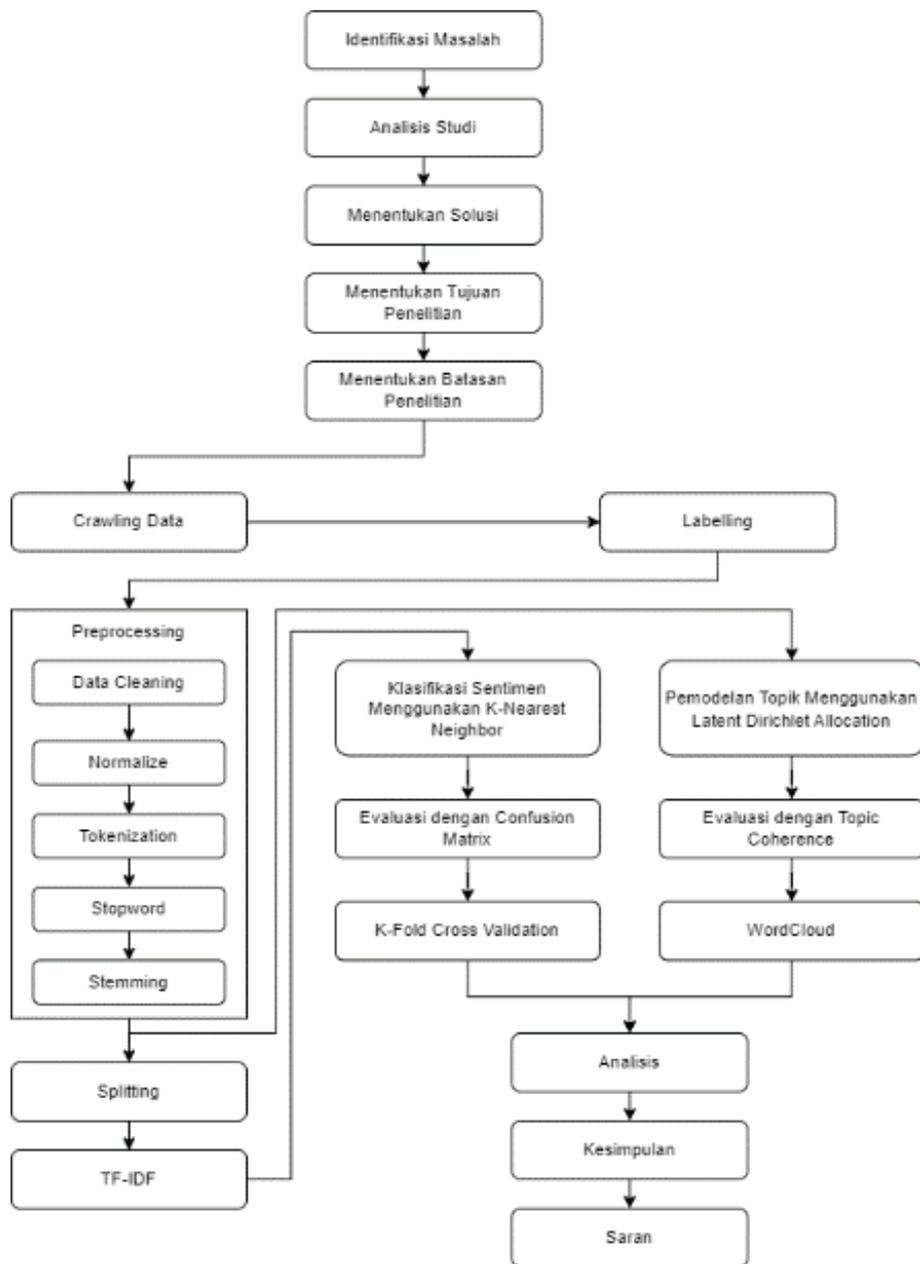
antara media pembelajaran *continue* dengan bantuan teknologi canggih [4]. Quipper adalah salah satu LMS (*Learning Management System*) yang pertama kali dirintis tahun 2010 di Inggris, kemudian berkembang lebih luas di negara-negara Asia dan Amerika seperti Mexico, Jepang, Filipina dan Indonesia. Produk dari Quipper terdiri dari dua jenis, yaitu Quipper Video dan Quipper School [5].

Respons kepuasan pelanggan terhadap produk di Google Play merupakan sumber informasi yang dapat membantu Quipper dalam mengevaluasi pencapaian tujuan perusahaan dan juga memantau kinerja produk mereka. Namun karena opini pengguna pada kolom ulasan tersebut sangat beragam dan ada pada jumlah yang banyak, membuat pengembang sulit untuk membacanya, akan menghabiskan terlalu banyak waktu dan tenaga jika harus membaca dan menganalisisnya secara manual dan cara seperti ini tidak disarankan karena tidak efektif [6]. Untuk itu perlu dilakukannya analisis sentimen yang memungkinkan sistem untuk secara otomatis menganalisis dan mengklasifikasikan sentimen tersebut.

Menurut [7], metode K-NN digunakan dalam proses penjabaran, hal ini disebabkan oleh al-

goritmanya memiliki prosedur dan logika yang sederhana, mudah dimengerti dan diaplikasikan, serta mencapai tingkat akurasi yang cukup tinggi untuk mengatasi permasalahan dalam klasifikasi atau proses pengidentifikasian. Berdasarkan tinjauan literatur dari penelitian sebelumnya, penulis memilih metode *K-Nearest Neighbor* untuk digunakan dalam penelitian ini. Salah satu metode topic modeling yang dapat digunakan dalam penelitian ini adalah *Latent Dirichlet Allocation (LDA)*. Menurut [8], LDA baru-baru ini muncul sebagai metode pilihan untuk bekerja dengan koleksi dokumen teks yang besar. LDA menghasilkan daftar topik yang diberi bobot untuk masing-masing dokumen, yang memungkinkannya digunakan un-

tuk meringkas, klusterisasi, menghubungkan, dan memproses jumlah data yang sangat besar. *Latent Dirichlet Allocation (LDA)* memiliki kinerja yang lebih baik dibandingkan metode pemodelan topik lainnya dan dapat diterapkan untuk mengidentifikasi topik dalam jurnal ilmiah, serta untuk klasifikasi dan pengelompokan [9]. Dengan menerapkan metode analisis sentimen menggunakan *K-Nearest Neighbor* dan pemodelan topik dengan *Latent Dirichlet Allocation*, penelitian ini diharapkan dapat memberikan wawasan yang berharga bagi Quipper dan pemangku kepentingan dalam memahami persepsi kepuasan pelanggan terhadap aplikasi.



Gambar 1: Sistematika Penyelesaian Masalah

Penelitian ini memiliki beberapa sistematika yang harus dilakukan untuk mengefektifkan pelaksanaan penelitian. Gambar 1 menjelaskan terdapat lingkungan penelitian yang melibatkan pengguna aplikasi Quipper yang memberikan ulasan berisi opini terkait aplikasi Quipper. Ulasan tersebut akan dikumpulkan dengan menggunakan data crawling menggunakan tools Python yang memanfaatkan *library google play scraper*. Penelitian ini menerapkan pendekatan *Text Mining* dengan menggunakan algoritma klasifikasi *K-Nearest Neighbor* (KNN) dan *Latent Dirichlet Allocation* sebagai algoritma untuk dilakukannya pemodelan topik.

## Metode Penelitian

Penelitian ini memiliki beberapa tahapan, yaitu :

### 1. Pengumpulan Data

Pada penelitian ini, data dikumpulkan menggunakan library '*google\_play\_scraper*' dalam bahasa pemrograman Python, berupa ulasan pengguna aplikasi Quipper dalam Bahasa Indonesia, diurutkan dari yang paling relevan. Sebanyak 12.442 ulasan dikumpulkan dari Januari 2017 hingga Maret 2024. Setelah pengumpulan data selesai, setiap ulasan akan diberi label sentimen—positif untuk sentimen positif atau kepuasan pengguna, dan negatif untuk sentimen negatif atau ketidakpuasan. Label ini bertujuan untuk melatih model analisis sentimen agar lebih akurat dalam mengidentifikasi pola-pola sentimen.

### 2. Pelabelan Data

Dalam proses labelling, data yang telah diproses terlebih dahulu ditandai secara manual menjadi tiga kategori: positif, negatif, dan netral. Data netral kemudian dihapus karena tidak relevan untuk klasifikasi. Hanya data positif dan negatif yang digunakan dalam proses klasifikasi. Komentar netral berbeda dari komentar positif dan negatif karena biasanya tidak mengandung kata-kata khas, sehingga sulit untuk diidentifikasi. Akibatnya, komentar netral sering kali susah dikenali sebagai kelas netral, karena kalimat dengan sentimen netral biasanya memiliki variasi kata yang lebih luas dibandingkan dengan kelas komentar lainnya.

### 3. Preprocessing

Proses preprocessing menjadi aspek krusial dalam penelitian analisis sentimen karena melibatkan sejumlah langkah yang berpotensi mempengaruhi tingkat akurasi algoritma machine learning[10]. Umumnya text preprocessing memiliki beberapa langkah yaitu case folding, normalisasi slang word, stopword removal, dan tokenizing [11]. Tahapan-tahapan preprocessing yang dilakukan adalah:

#### (a) *Data Cleaning*

Proses pembersihan data mencakup langkah-langkah seperti menghapus data dengan nilai null, menghapus data yang tidak memiliki nilai, dan menghapus data yang duplikat, case folding, menghapus angka, karakter, spasi kosong, tanda baca, baris baru, dan emoticon. Dalam tahap case folding akan dilakukan perubahan terhadap kumpulan kalimat menjadi huruf kecil[12]. Ini dilakukan untuk memastikan bahwa kata yang sama dengan huruf kapital atau huruf kecil tidak berbeda.

#### (b) *Tokenization*

Menurut [11], tokenisasi melibatkan pembagian urutan karakter menjadi fragmen terpisah yang disebut token. Langkah ini melibatkan proses memecah kata-kata dalam setiap kalimat, mengubahnya menjadi kumpulan token terpisah. Memisahkan kata-kata menjadi entitas terpisah yang akan memudahkan proses pelatihan model klasifikasi di masa mendatang.

#### (c) *Stopwords Removal*

Kata-kata seperti "yang", "saja", dan "para" tidak memiliki kontribusi makna signifikan dalam kalimat. Oleh sebab itu, kata-kata tersebut kemudian dihapus, selanjutnya akan membuat teks yang tersisa lebih relevan untuk keperluan analisis sentimen [14].

#### (d) *Stemming*

Stemming merupakan proses menghilangkan imbuhan pada masing-masing kata sehingga menjadi kata dasar dan dalam proses ini bertujuan untuk membersihkan suatu kata dari pengejaan yang kurang tepat.

## Pembagian Data

Data splitting adalah metode validasi model dengan membagi data menjadi dua, yaitu data training dan data testing. Data training digunakan untuk melatih model, sementara data testing menguji performa model. Rasio pembagian data perlu ditentukan sebelum proses ini, dengan rasio umum 80:20, di mana 80% data digunakan untuk pelatihan dan 20% untuk pengujian.

## Implementasi *K-Nearest Neighbor*

KNN bekerja dengan mengklasifikasikan data uji berdasarkan jarak terdekatnya dengan data latih. Setiap data pada data latih memiliki atribut kelas, dan langkah berikutnya adalah menguji model menggunakan data uji yang tidak memiliki atribut kelas. Adapun tahapan yang dilakukan untuk

menghitung K-Nearest Neighbor (KNN) antara lain [14]:

1. Tahapan awal pada KNN ini adalah menentukan nilai K, misal  $k=23$  artinya 23 dokumen terdekat dengan dokumen uji yang akan diambil.
2. Menghitung jarak antara data baru di setiap label data (jarak euclidean) dengan jarak semua data training. Untuk menghitung tingkat kesamaan dalam dokumen ini menggunakan euclidean distance. Cara perhitungan euclidean distance adalah dengan rumus berikut

$$d(P,Q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

Ket:

d = Jarak antara dua titik P dan Q

P = Data uji

Q = Sampel data n = Dimensi data

3. Kemudian urutkan hasil euclidean distance berdasarkan jarak K yang ditentukan jika  $K=3$  artinya akan dipilih 3 jarak terkecil dari hasil euclidean distance.
4. Kemudian, gunakan mayoritas atribut kelas dari tetangga terdekat yang telah dipilih untuk menentukan prediksi kelas pada data baru tersebut. Misalnya, jika dari 3 tetangga terdekat terdapat 2 atribut kelas positif dan 1 atribut kelas negatif, maka kelas yang diprediksi untuk data baru tersebut adalah positif.

### Evaluasi Dengan *Confusion Matrix*

Confusion Matrix digunakan untuk mengevaluasi performa dari model klasifikasi, termasuk dalam analisis sentimen. Dimana metode yang digunakan dalam menghitung akurasi dengan membandingkan hasil klasifikasi yang dilakukan secara aktual dengan hasil klasifikasi dari metode [12]. Evaluasi confusion matrix menghasilkan metrik yaitu accuracy, precision, recall dan f-measure serta berdasarkan analisis nilai True Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN) yang didapatkan melalui identifikasi terkait model.

### *K-fold Cross Validation*

Dalam proses *K-Fold cross validation*, seluruh dataset diacak dan kemudian dibagi menjadi bagian-bagian yang sama besar. Sebagian dari

bagian tersebut digunakan sebagai data uji, sementara sisanya menjadi data latih [15]. Dalam pendekatan ini, dataset dipisahkan menjadi K subset. Pada situasi ini, nilai K yang digunakan adalah 5.

### *Metode Latent Dirichlet Allocation (LDA)*

LDA mengenali informasi dalam bentuk topik tersembunyi yang mewakili suatu dokumen sebagai distribusi probabilitas terhadap beberapa topik. Setiap topik, pada gilirannya, diwakili sebagai distribusi probabilitas terhadap sejumlah kata [16]. Menurut [17] menyebutkan bahwa Latent Dirichlet Allocation (LDA) saat ini menjadi salah satu metode paling umum yang digunakan untuk melakukan pemodelan topik dan analisis topik. Hal ini disebabkan oleh kemampuan LDA dalam menghasilkan daftar topik yang diberi bobot, memungkinkan penggunaan untuk meringkas, mengklasifikasikan, menghubungkan, serta memproses data yang berskala besar.

### *Evaluasi Dengan Topic Coherence*

Dengan menggunakan topic coherence, evaluasi model dapat digunakan untuk menilai kualitas topik yang konsisten dengan perspektif manusia. Nilai topic coherence yang lebih tinggi menunjukkan bahwa model dapat menghasilkan kelompok kata yang lebih baik dalam topik tertentu. Terdapat beberapa jenis metrik dalam pengukuran pada Coherence yaitu cv, C\_UMass, C\_UCI, C\_NPMI [18].

### *Visualisasi Wordcloud*

*Wordcloud* adalah metode visualisasi data yang digunakan untuk mempermudah pengenalan atau identifikasi konten dari sejumlah besar dokumen teks dengan lebih cepat. *Word clouds* dapat menampilkan istilah atau kata-kata kunci yang paling penting atau sering muncul dengan variasi warna dan font untuk menarik perhatian [19].

## Hasil dan Pembahasan

### Pengumpulan Data

Data yang diperlukan dalam penelitian ini meliputi teks ulasan mengenai aplikasi Quipper pada Google Play Store. Data ini diperoleh menggunakan teknik scraping dengan jumlah total ulasan sebanyak 12.442 data pada tanggal 29 Maret 2024. Untuk melakukan scraping, digunakan tools python dengan menjalankan konfigurasi pada google colab. Setelah proses scraping selesai, data tersebut dapat disimpan dalam format CSV atau Excel, lihat Gambar 2.

reviewId	userName	UserImage	content	Score	thumbUpCount	reviewCreatedAtVersion	at	replyContent	replyDate	appVersion
ccc268ae63-4240-acec-7748d66ff653	Andhika Makhfud Ridwan	https://play-lh.googleusercontent.com/a-/ALV-UjVetvqlvfeEEP9uhfKHeRPLZGENgsDKEC5eRzGKhr6sncFCFU	Rating 1 karena ingin menautkan akun sering terjadi kendala padahal alamat email sudah benar sandi sudah benar tapi masih tidak bisa di tautkan akun nya gimana ini 😞	1	8	6.14.6	2024-03-30 14:53:23	Hai Andhika! Maaf untuk kendalanya ya 😊 Jika email dan kata sandinya benar, seharusnya akan bisa ditautkan Andhika. Agar berhasil menautkan akun, pastikan kembali untuk emailnya aktif dan kata sandi emailnya benar ya. Jika masih belum bisa juga, boleh infokan detail kendalanya ke WA 0811 1902 7360 untuk kami bantu pengecekan lebih lanjut 😊	2024-04-01 11:37:30	6.14.6
75e773f9-c0a6-4a48-a223-1d55bd088415	Pengguna Google	https://play-lh.googleusercontent.com/EGemol2NTXmTsBVUJk8jxF9rb8Ap	Terimakasih quipper, sudah membantu saya untuk belajar, tapi mohon agar durasi videonya di singkat ya agar bisa belajar banyak materi dan tidak membosankan, dan juga tambahkan animasi video pelajarannya, filter dan tampilannya juga diperbaiki. Terimakasih quipper.	3	0	6.1.7	2019-12-04 03:14:01	Halo, terima kasih untuk masukannya, masukan kamu akan menjadi pucuan kami untuk mengembangkan Quipper Video kedepannya. :)	2019-11-25 12:45:22	6.1.7

Gambar 2: Sampel Data Ulasan Aplikasi Quipper

Tabel 1: Sampel Data Hasil Pelabelan

content	at	sentimen
Rating 1 karena ingin menautkan akun sering terjadi kendala padahal alamat email sudah benar sandi sudah benar tapi masih tidak bisa ditautkan akun nya gimana ini 😞	2024-03-30 14:53:23	Negatif
Terimakasih quipper, sudah membantu saya untuk belajar, tapi mohon agar durasi videonya di singkat ya agar bisa belajar banyak materi dan tidak membosankan, dan juga tambahkan animasi video pelajarannya, filter dan tampilannya juga diperbaiki. Terimakasih quipper.	2019-12-04 03:14:01	Positif

## Pelabelan Data

Data yang telah diproses terlebih dahulu ditandai secara manual menjadi tiga kategori: positif, negatif, dan netral. Data netral kemudian dihapus karena tidak relevan untuk klasifikasi. Hanya data positif dan negatif yang digunakan dalam proses klasifikasi, lihat Tabel 1.

## Preprocessing

**Data Cleaning** Proses ini meliputi beberapa langkah yaitu mengubah kalimat menjadi huruf kecil, menghapus angka dari kalimat, dan menghapus tanda baca dari kalimat. Selain itu, spasi pada awal dan akhir kalimat juga dihapus. Proses ini juga mencakup penghilangan tab, baris baru, serta potongan teks yang tidak diinginkan, lihat Tabel 2.

## Tokenization

Dalam tahapan tokenization, memiliki tujuan yaitu untuk memecah teks atau kalimat menjadi unit-unit

yang lebih kecil atau dapat disebut juga dengan istilah “token”. Hal ini dapat membantu dalam pengolahan teks menjadi lebih mudah karena membagi kumpulan kata menjadi kata per kata sehingga lebih mudah dan bermakna, lihat Tabel 3.

Tabel 2: Sampel Data Hasil Cleaning

Sebelum	Sesudah
Terimakasih Quipper, aplikasi yang sangat bagus, karna sudah membantu dan memudahkan kami belajar daring disaat masa pandemi seperti ini👍👍👍	Terimakasih quipper aplikasi yang sangat bagus karna sudah membantu dan memudahkan kami belajar daring disaat masa pandemi seperti ini

Tabel 3: Sampel Data Hasil Tokenization

Sebelum	Sesudah
terima kasih quipper aplikasi yang sangat bagus karena sudah membantu dan memudahkan kami belajar daring disaat masa pandemi seperti ini	['terima', 'kasih', 'quipper', 'aplikasi', 'yang', 'sangat', 'bagus', 'karena', 'sudah', 'membantu', 'dan', 'memudahkan', 'kami', 'belajar', 'daring', 'disaat', 'masa', 'pandemi', 'seperti', 'ini']

### Stopwords Removal

Tahapan stopwords removal merupakan adanya proses filtering, dari hasil tokenization kemudian selanjutnya akan memilih kata-kata yang akan digunakan untuk mewakili dokumen, lihat Tabel 4.

Tabel 4: Sampel Data Hasil Stopwords Removal

Sebelum	Sesudah
['terima', 'kasih', 'quipper', 'aplikasi', 'yang', 'sangat', 'bagus', 'karena', 'sudah', 'membantu', 'dan', 'memudahkan', 'kami', 'belajar', 'daring', 'disaat', 'masa', 'pandemi', 'seperti', 'ini']	terima kasih quipper aplikasi bagus membantu memudahkan belajar daring disaat pandemi

### Steaming

Stemming adalah sebuah proses mengembalikan menjadi kata dasar dengan cara membuang awalan, akhiran, dan sisipan atau dapat disebut kata yang berimbuhan. Pada tahapan stemming ini menggunakan library sastrawi stemmer factory, lihat Tabel 5.

Tabel 5: Sampel Data Hasil Steaming

Sebelum	Sesudah
terima kasih quipper aplikasi yang sangat bagus membantu memudahkan belajar daring disaat pandemi	['terima', 'kasih', 'quipper', 'aplikasi', 'bagus', 'membantu', 'memudahkan', 'belajar', 'daring', 'disaat', 'pandemi']

### Pembagian Data

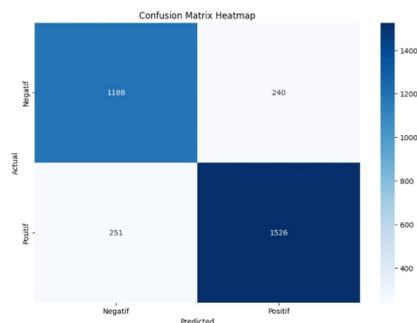
Split atau pembagian data selanjutnya akan dilakukan pada tahap ini, data akan dibagi ke dalam dua kategori yaitu data training dan data testing. Penelitian ini membandingkan tiga rasio 70:30, 80:20, 90:10.

### Implementasi *K-Nearest Neighbor*

Pada tahap klasifikasi, data dibagi menjadi dua proses: pelatihan dan pengujian. Model dibangun melalui pelatihan, lalu diuji dengan Euclidean Distance sebagai metrik untuk mengukur jarak antara data baru dan data pelatihan. Data dibagi dalam tiga rasio: 70:30, 80:20, dan 90:10, lalu diuji untuk menentukan rasio dengan akurasi tertinggi yang akan digunakan dalam penelitian ini. Dalam perbandingan ketiga rasio yaitu 70:30, 80:20, 90:10 didapatkan perbandingan nilai K dengan percobaan nilai K yaitu 1 sampai 20.

### Evaluasi Dengan Confusion Matrix

Evaluasi dalam pemodelan K-NN ini dilakukan dengan menggunakan metode *Confusion Matrix* dengan rasio yang sudah ditentukan, yaitu 70:30 dengan nilai K tertinggi yaitu K = 14. Gambar 3 memperlihatkan visualisasi dalam bentuk heatmap untuk metode yang diuji.



Gambar 3: Visualisasi Heatmap

Gambar 3 menunjukkan *heatmap* dari klasifikasi menggunakan algoritma *K-Nearest Neighbor* dengan rasio 70:30. *Heatmap* ini memperlihatkan persentase dan jumlah metrik: 85.88% *true positive* (1526 data), 83.19% *true negative* (1188 data), 16.81% *false positive* (240 data), dan 14.12% *false negative* (251 data).

Berikut hasil perhitungan confusion matrix:

1.

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{1526 + 1188}{1526 + 1188 + 240 + 251} \times 100\% \\
 &= \frac{2714}{3205} \times 100\% = 0.8468
 \end{aligned}$$

2.

$$\begin{aligned}
 \text{Presisi} &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{1526}{1526 + 240} \times 100\% \\
 &= \frac{1526}{1766} \times 100\% = 0.8640
 \end{aligned}$$

3.

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{1526}{1526 + 251} \times 100\% \\
 &= \frac{1526}{1777} \times 100\% = 0.8587
 \end{aligned}$$

4.

$$\begin{aligned}
 F1 - \text{Score} &= 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \\
 &= 2 * \frac{0.8640 * 0.8587}{0.8640 + 0.8587} \\
 &= 2 * \frac{0.7419}{1.7227} = 0.8609
 \end{aligned}$$

### ***K-fold Cross Validation***

*K-Fold Cross Validation* adalah suatu pengujian yang berfungsi untuk menilai kinerja suatu model algoritma dengan membagi sampel data menggunakan mengelompokkan data sejumlah nilai K. Pada proses *K-Fold Cross Validation* peneliti membagi data menjadi 10 bagian dan menjalankan 10 skenario.

Tabel 6: Hasil *K-Fold Cross Validation*onn

Pengujian	Akurasi
<b>K1</b>	83.72%
<b>K2</b>	82.60%
<b>K3</b>	85.21%
<b>K4</b>	84.55%
<b>K5</b>	84.92%
<b>K6</b>	84.55%
<b>K7</b>	85.11%
<b>K8</b>	84.17%
<b>K9</b>	86.51%
<b>K10</b>	85.11%
<b>Rata – Rata</b>	84.64%

Dari tabel, peneliti melakukan 10 kali pengujian menggunakan *K-Fold Cross Validation*. Hasil akurasi berkisar antara 82.60% hingga 86.51%, dengan rata-rata 84.64%. Hasil ini menunjukkan bahwa model K-NN pada data ulasan Quipper memiliki performa yang baik dan konsisten, tanpa variasi besar antar fold. Rata-rata akurasi 84.64% menegaskan bahwa model KNN dapat mengklasifikasikan data dengan tingkat keakuratan yang baik.

### ***Implementasi Metode Latent Dirichlet Allocation (LDA)***

Pemodelan topik metode *Latent Dirichlet Allocation* dilakukan dengan menggunakan library *ldamodel* dari *gensim*. Pembentukan model dilakukan dengan menentukan jumlah topik yang menjadi kelompok klasterisasi kata dan penentuan jumlah passes dalam pemodelan topik yang merupakan jumlah iterasi proses pembelajaran dari training model. Nilai koherensi untuk setiap jumlah topik pada kelas sentimen positif dan negatif ditampilkan dalam Tabel 7. Berdasarkan hasil perhitungan, penelitian ini menggunakan 4 sebagai model topik terbaik pada sentimen positif dan 3 sebagai topik terbaik pada sentimen negatif.

Tabel 7: Nilai Koherensi Topik

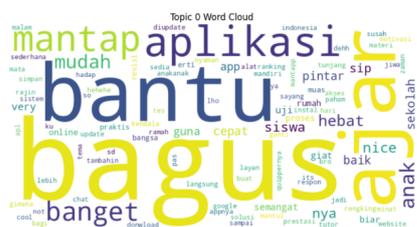
Num of Topic	Coherence score	
	Label Positif	Label Negatif
2	0.4502	0.3709
3	0.4361	0.4277
4	0.5081	0.4163
5	0.4565	0.4012
6	0.4543	0.3972
7	0.4389	0.4273
8	0.4552	0.3923
9	0.4471	0.4153
10	0.4442	0.3817

### Evaluasi Dengan Topic Coherence

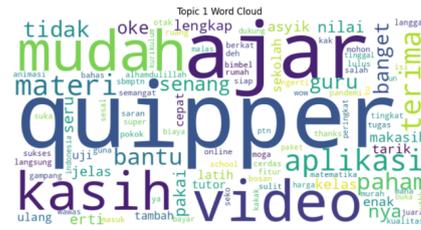
Hasil koherensi topik menunjukkan bahwa nilai koherensi topik tertinggi pada sentimen positif terdapat pada jumlah topik ke-4 dengan koherensi topik sebesar 0.5081 dan nilai koherensi topik tertinggi pada sentimen negatif terdapat pada jumlah topik ke-3 dengan skor koherensi sebesar 0.4277. Berdasarkan skor koherensi ini, jumlah topik terbaik yang diperoleh akan digunakan sebagai acuan dalam pembuatan model. Oleh karena itu, penelitian ini menggunakan 4 model topik terbaik pada sentimen positif dan 3 topik terbaik pada sentimen negatif.

### Visualisasi Wordcloud

Berdasarkan hasil topik koherensi tertinggi dengan nilai sebesar 0.5081 dengan jumlah topik 4 maka hasil tersebut dapat divisualisasikan dan ditampilkan dalam visualisasi data yaitu *wordcloud*. Visualisasi *wordcloud* pada kelas positif dapat dilihat pada Gambar 4 sampai 7, dan untuk kelas negatif pada Gambar 8 sampai 10.



Gambar 4: Visualisasi Topik “0” Kelas Positif



Gambar 5: Visualisasi Topik “1” Kelas Positif



Gambar 6: Visualisasi Topik “2” Kelas Positif

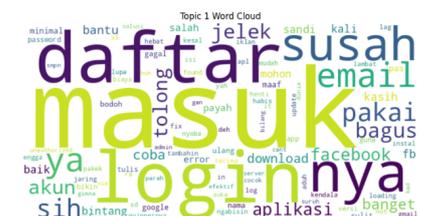


Gambar 7: Visualisasi Topik “3” Kelas Positif

Berdasarkan visualisasi dalam setiap topik terdapat empat topik dari visualisasi wordclouds mengenai ulasan aplikasi Quipper. Topik "0" membahas kemudahan penggunaan dan efektivitas aplikasi yang diapresiasi pengguna. Topik "1" berfokus pada kualitas materi dan video pembelajaran yang dianggap mudah dipahami dan membantu, bahkan oleh guru. Topik "2" mengulas kepuasan pengguna terhadap fitur-fitur.



Gambar 8: Visualisasi Topik “0” Kelas Negatif



Gambar 9: Visualisasi Topik “1” Kelas Negatif



- [10] W. Sriyanong, N. Moungmingsuk, and N. Khamphakdee, "A text preprocessing framework for text mining on big data infrastructure", In 2018 2nd International Conference on Imaging, Signal Processing and Communication (ICISPC) IEEE, (pp. 169-173), DOI: 10.1109/ICISPC44900.2018.9006718. 2018.
- [11] V. Yolanda, R. Mas'ud, S. M. Hasanah, "Pengaruh Teknologi Keuangan dan Kualitas Layanan Digital Terhadap Kepuasan Nasabah Pada Bank Syariah Indonesia KCP Masbagik", journal.uinmataram, vol. 13, no. 1, hal. 63, <https://doi.org/10.20414/iqtishaduna.v13i1.5373>. 2022.
- [12] F. Amaliah dan Dwi Nuryana, "Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter", Journal of Informatics and Computer Science. vol. 3, no. 3, <https://doi.org/10.26740/jinacs.v3n03.p384-393>. 2022.
- [13] S. Pradha, M. N. Halgamuge, and N. Tran Quoc Vinh, "Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data", 2019 11th International Conference on Knowledge and Systems Engineering (KSE), Da Nang, Vietnam, 2019, pp. 1-8, DOI: 10.1109/KSE.2019.8919368. 2019.
- [14] J. A. Septian, T. M. Fachrudin, dan A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor", IN-SYST: Journal of Intelligent System and Computation, vol. 1, no. 1, hal. 43-49, <https://doi.org/10.52985/insyst.v1i1.36>. 2019.
- [15] M. R. A. Nasution dan M. Hayaty, "Perbandingan akurasi dan waktu proses algoritma K-NN dan SVM dalam analisis sentimen twitter", Journal Inform, vol. 6, no. 2, hal. 226-235, <https://doi.org/10.31294/ji.v6i2.5129>, 2019.
- [16] Liangjie Hong and Brian D. Davison, "Empirical study of topic modeling in Twitter", In Proceedings of the First Workshop on Social Media Analytics (SOMA '10). Association for Computing Machinery, New York, NY, USA, pp. 80-88. <https://doi.org/10.1145/1964858.1964870>, 2010.
- [17] A. I. Alfanzar, "Topic Modelling Skripsi Menggunakan Metode Latent Dirichlet Allocation", Jurnal Sistem Informasi (JSiI). vol. 7, no. 1, hal. 7-13, <https://doi.org/10.30656/jsii.v7i1.2036>. 2019.
- [18] M. Röder, A. Both, and A. Hinneburg, "Exploring the Space of Topic Coherence Measures", Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Shanghai: Association for Computing Machinery, hal. 399-408, DOI:10.1145/2684822.2685324. 2015.
- [19] Y. Kalmukov, "Using word clouds for fast identification of papers' subject domain and reviewers' competences", Proceedings of University of Ruse, vol. 60, book 3.2, hal. 114-119, <https://doi.org/10.48550/arXiv.2112.14861>. 2021.