

BAB II

TINJAUAN PUSTAKA

2.1. Studi Sebelumnya

Penelitian yang dilakukan oleh Wahidna & Nerisafitra [20] menganalisis sentimen pengguna terhadap layanan *paylater* dengan memanfaatkan data dari Twitter. Penelitian ini bertujuan untuk mengidentifikasi kecenderungan opini pengguna terhadap layanan *paylater* di *platform e-commerce*. Studi ini menggunakan pendekatan *lexicon-based* untuk *preprocessing* dan metode klasifikasi sentimen dengan algoritma *Support Vector Machine* (SVM). Penelitian ini berhasil membangun model klasifikasi SVM yang mampu memprediksi sentimen dengan akurasi 89,74% untuk Shopee PayLater dan 90,27% untuk GoPay Later. Hasil penelitian ini menunjukkan bahwa dari 255 data terkait Shopee PayLater, 55 di antaranya merupakan sentimen positif, 255 data dengan sentimen netral, dan 196 sentimen negatif. Sedangkan untuk GoPay Later diperoleh 34 sentimen positif, 87 sentimen netral, dan 61 sentimen negatif. Hasil ini menunjukkan bahwa sentimen pengguna terhadap Shopee PayLater dan GoPay Later cenderung netral.

Safira & Hasan [21] melakukan penelitian analisis sentimen terkait pandangan masyarakat terhadap *paylater*. Data yang digunakan pada penelitian ini diambil dari Twitter. Metode yang digunakan adalah metode *Naïve Bayes Classifier* dan *TextBlob*. Jumlah data yang berhasil dikumpulkan sebanyak 405 data. Berdasarkan metode *Naïve Bayes* diperoleh hasil sentimen negatif sebesar 70,62%, disusul sentimen positif sebesar 22,72%, dan 6,67% sentimen netral. Sedangkan dengan metode *TextBlob*, sebanyak 55,8% sentimen negatif, 33,09% sentimen positif, dan 11,11% sentimen netral. Berdasarkan analisis sentimen dengan kedua metode tersebut dapat disimpulkan bahwa pandangan masyarakat terhadap *paylater* didominasi oleh sentimen negatif.

Penelitian lainnya dilakukan oleh Mahayani dkk. [22] pada tahun 2020 bertujuan untuk menganalisis sentimen pengguna terhadap layanan Shopee PayLater melalui data yang diambil dari Twitter. Klasifikasi sentimen dilakukan dengan menggunakan metode *lexicon-based* dan *Naïve Bayes Classifier*. Dari hasil penelitian ini ditemukan bahwa dari 344 data yang digunakan, mayoritas pengguna menilai pembayaran Shopee PayLater dengan sentimen positif dan tingkat akurasi sebesar 82,52%.

Penelitian-penelitian tersebut memanfaatkan data dari X (sebelumnya Twitter) untuk memahami gambaran umum perspektif publik terhadap layanan BNPL. Namun, pendekatan tersebut umumnya hanya berhenti pada klasifikasi sentimen, tanpa mengeksplorasi lebih lanjut mengenai topik-topik utama yang dibicarakan dalam ulasan pengguna. Padahal, identifikasi isu-isu spesifik yang sering dibahas pengguna sangat penting untuk memperkaya pemahaman terhadap persepsi publik secara menyeluruh.

Salah satu contoh penelitian yang menggunakan pendekatan eksplorasi topik adalah penelitian yang dilakukan oleh Widianoro & Harnadi [23]. Penelitian ini lebih menekankan pada pemodelan topik menggunakan metode *Latent Dirichlet Allocation (LDA)* untuk menganalisis ulasan pengguna aplikasi OVO, sebuah *platform* dompet digital yang juga menyediakan fitur *paylater*. Tujuan dari penelitian ini adalah untuk mengeksplorasi topik-topik dominan yang dibahas pengguna terkait layanan aplikasi tersebut. Metode LDA berhasil mengelompokkan topik-topik utama yang diurutkan berdasarkan probabilitas tertinggi, seperti *“transaction difficulties on premium services”*, *“the balance hasn’t arrived at the time of top up”*, *“fun app”*, *“slow in responding to reports”*, *“good services but busy app”*, dan *“easy transaction”*. Namun, penelitian ini tidak mengintegrasikan hasil topik dengan analisis sentimen, sehingga tidak memberikan informasi mengenai persepsi pengguna terhadap masing-masing topik.

Sementara itu, dalam penelitian yang dilakukan oleh Setiawan dkk. [24] telah mengintegrasikan pendekatan analisis sentimen dan pemodelan topik terhadap layanan

bank digital di Indonesia melalui data Twitter dan Instagram. Tujuan dari penelitian ini adalah untuk mengklasifikasikan dan mengidentifikasi topik utama yang sering dibicarakan terkait layanan bank digital, khususnya Jenius, Jago, dan Line Bank. Penelitian ini menganalisis ulasan pengguna dengan menggabungkan analisis sentimen dengan algoritma *Naïve Bayes Classifier (NBC)* dan pemodelan topik menggunakan metode *Latent Dirichlet Allocation (LDA)*. Hasil penelitian ini menunjukkan bahwa ulasan didominasi dengan sentimen positif sebesar 92,48% dan sentimen negatif sebesar 7,52%. Berdasarkan hasil pemodelan topik, topik yang paling sering dibahas pada sentimen positif adalah fitur integrasi dengan *platform* lain dan biaya transfer yang relatif kecil, bahkan gratis. Sedangkan untuk sentimen negatif kebanyakan membahas mengenai fitur *login* dan layanan *customer services*. Namun, penelitian ini lebih berfokus pada aplikasi perbankan digital, bukan pada layanan BNPL.

Berdasarkan studi-studi sebelumnya, dapat disimpulkan bahwa penelitian yang mengkaji persepsi publik terhadap layanan BNPL dengan pendekatan yang mengintegrasikan analisis sentimen dan eksplorasi topik masih cukup terbatas. Sebagian besar penelitian terkait BNPL hanya menekankan pada satu pendekatan saja, baik itu analisis sentimen ataupun pemodelan topik. Oleh karena itu, penelitian ini bertujuan untuk mengisi kekosongan ini dengan menggabungkan kedua pendekatan tersebut untuk menganalisis persepsi publik terhadap layanan BNPL di media sosial dengan lebih komprehensif. Dengan pendekatan ini, penelitian tidak hanya mengukur sentimen pengguna, tetapi juga mengeksplorasi isu-isu utama yang sering dibicarakan.

Tabel 2. 1. Tabel Perbandingan Penelitian

No	Penulis	Tahun	Tujuan	Metode	Hasil
1	Wahidna & Nerisafitra [20]	2023	Mengidentifikasi opini pengguna terhadap Shopee PayLater dan GoPay Later	<i>Support Vector Machine (SVM)</i>	Penelitian ini menunjukkan sentimen cenderung netral, baik untuk Shopee PayLater maupun GoPay Later. Nilai akurasi untuk Shopee PayLater sebesar 89,74% dan GoPay Later sebesar 90,27%.
2	Safira & Hasan [21]	2023	Menilai persepsi masyarakat terhadap layanan <i>paylater</i>	<i>Naïve Bayes Classifier, TextBlob</i>	Hasil klasifikasi sentimen dengan metode <i>Naïve Bayes</i> dan <i>TextBlob</i> berbeda, meskipun begitu hasil sentimen keduanya didominasi oleh sentimen negatif.
3	Mahayani dkk. [22]	2020	Menganalisis sentimen pengguna terhadap <i>Shopee PayLater</i>	<i>Lexicon-Based, Naïve Bayes Classifier</i>	Berdasarkan hasil pengujian menggunakan <i>Naïve Bayes</i> , didapatkan nilai akurasi sebesar 92% dengan mayoritas sentimen yang didapatkan bersifat positif.
4	Widianoro & Harnadi [23]	2022	Mengeksplorasi topik dari ulasan pengguna aplikasi OVO	<i>Latent Dirichlet Allocation (LDA)</i>	Didapatkan enam topik utama yang paling sering dibahas, yaitu “ <i>transaction</i>

No	Penulis	Tahun	Tujuan	Metode	Hasil
					<i>difficulties on premium services”, “the balance hasn’t arrived at the time of top up”, “fun app”, “slow in responding to reports”, “good services but busy app”, dan “easy transaction”.</i>
5	Setiawan [24]	2022	Mengklasifikasikan sentimen dan mengidentifikasi topik-topik utama pada layanan bank digital	<i>Latent Dirichlet Allocation (LDA), Naïve Bayes Classifier</i>	Hasil klasifikasi sentimen didominasi oleh ulasan positif sebesar 92,48%. Pemodelan topik berhasil menunjukkan topik yang paling sering dibahas pada ulasan positif dan negatif.
6	Viola (Penelitian ini)	2025	Mengklasifikasikan sentimen pengguna dan mengeksplorasi lebih lanjut mengenai topik yang paling sering dibahas mengenai layanan BNPL	<i>Latent Dirichlet Allocation (LDA), Naïve Bayes Classifier</i>	

2.2. Dasar Teori

2.2.1 Buy Now, Pay Later (BNPL)

BNPL merupakan salah satu inovasi *fintech* dengan skema pembayaran yang memungkinkan konsumen untuk membeli barang atau jasa dan membayarnya di kemudian hari dalam bentuk cicilan tanpa bunga atau dengan bunga yang relatif rendah [25]. Layanan ini berkembang pesat karena kemudahan akses, proses yang cepat, dan minimnya persyaratan administratif jika dibandingkan dengan kredit konvensional [4]. Beberapa penelitian menunjukkan bahwa BNPL banyak diminati di kalangan generasi muda karena memberikan fleksibilitas finansial jangka pendek, meskipun di sisi lain dapat memiliki potensi risiko terkait kontrol keuangan dan kemungkinan keterlambatan pembayaran [13], [26].

2.2.2 Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah salah satu cabang dari *Artificial Intelligence (AI)* yang berfokus pada interaksi antara komputer dan bahasa manusia [27]. NLP memungkinkan komputer untuk memahami, menafsirkan, dan memproses bahasa manusia [28]. NLP digunakan untuk beragam tujuan, seperti penerjemahan otomatis, analisis sentimen, pemodelan topik, *chatbot*, dan pengenalan suara [29]. Dalam penelitian ini, NLP digunakan untuk mendukung proses analisis sentimen dan pemodelan topik dari data teks di media sosial.

2.2.3 Text Mining

Text mining adalah proses mengekstrak informasi berharga dari kumpulan data teks yang tidak terstruktur [30]. Proses ini mencakup beberapa langkah, seperti pembersihan data, *preprocessing data*, analisis frekuensi data, dan penerapan metode *machine learning* seperti klasifikasi dan klusterisasi [27]. *Text mining* memungkinkan pengelolaan data teks dalam skala besar untuk menemukan pola, mengidentifikasi kata kunci dominan, serta mengklasifikasikan teks berdasarkan topik [31].

2.2.4 Analisis Sentimen

Analisis sentimen adalah proses mengidentifikasi dan mengelompokkan opini, sikap, atau emosi pengguna dalam sebuah teks terhadap suatu topik, produk, atau layanan [32]. Analisis sentimen digunakan untuk mengetahui persepsi positif, netral, atau negatif publik terhadap suatu fenomena [33]. Analisis ini akan memberikan wawasan dalam mengukur kepuasan, ekspektasi, dan risiko yang dirasakan pengguna. Terdapat beberapa pendekatan dalam analisis sentimen, yaitu:

1. Pendekatan berbasis kamus (*lexicon-based*)

Pendekatan ini menggunakan kamus sentimen (*sentiment lexicon*) yang berisi kata-kata beserta nilai polaritasnya (positif atau negatif) [34]. Setiap kata diberi skor positif atau negatif, dan sentimen sebuah teks ditentukan berdasarkan skor kumulatif dari kata-kata dalam teks tersebut [35]. Pendekatan ini cocok untuk teks pendek, namun kurang efektif menangani konteks atau sarkasme [35], [36].

2. Pendekatan berbasis pembelajaran mesin (*machine learning-based*)

Pendekatan ini menggunakan algoritma klasifikasi, seperti *Naïve Bayes*, *SVM*, atau *deep learning* untuk mengenali pola sentimen dalam data yang telah diberi label sentimen [37]. Pendekatan ini memiliki fleksibilitas dan akurasi yang tinggi, khususnya pada dataset besar [38].

3. Pendekatan hibrida (*hybrid-based*)

Pendekatan ini menggabungkan pendekatan *lexicon* dan *machine learning* untuk menghasilkan klasifikasi yang lebih akurat [39]. *Lexicon-based* digunakan sebagai fitur tambahan dalam model klasifikasi *machine learning* [40].

2.2.5 Preprocessing Text

Preprocessing text merupakan tahap penting dalam analisis data teks yang bertujuan untuk membersihkan dan menyiapkan data sebelum dianalisis lebih lanjut [28]. Proses ini penting untuk meningkatkan kualitas data yang akan dianalisis sehingga memperoleh hasil model analisis yang akurat [41]. Tahapan ini meliputi beberapa bagian, yaitu:

1. *Cleaning* adalah tahapan untuk menghapus elemen-elemen yang tidak relevan di dalam teks, seperti tanda baca, angka, *URL*, simbol, dan lain-lain [28]. Selain itu data yang duplikat dan yang mengandung *missing value* juga dihapus [42]
2. *Case folding* adalah tahapan mengubah semua karakter dalam teks menjadi huruf kecil (*lowercase*) [43].
3. *Normalization* adalah proses mengubah kata-kata tidak baku (*slangword*) menjadi bahasa yang baku [44].
4. *Tokenizing* adalah proses memecah teks menjadi potongan yang lebih kecil, berupa kata atau token [41].
5. *Stopword removal* merupakan proses penghapusan kata-kata umum yang tidak memiliki makna penting dalam proses analisis, seperti dan, yang, di, dan lain-lain. [28].
6. *Stemming* adalah proses mengubah sebuah kata menjadi kata dasarnya dengan menghapus imbuhan [44].

2.2.6 Klasifikasi Sentimen

Klasifikasi sentimen merupakan proses dalam analisis teks yang bertujuan untuk mengidentifikasi dan mengelompokkan opini atau emosi yang terkandung dalam data teks ke dalam kategori tertentu [45]. Teknik ini banyak diterapkan dalam memahami persepsi publik terhadap isu, layanan, atau produk tertentu, terutama melalui data yang bersumber dari media sosial [46]. Dalam konteks analisis media sosial, klasifikasi sentimen berperan penting untuk menangkap persepsi publik secara *real-time* dan dalam skala besar [47].

Secara umum, klasifikasi sentimen dibagi menjadi beberapa jenis, yaitu:

1. *Binary Classification*, teks diklasifikasikan ke dalam dua kategori sentimen, yaitu positif dan negatif [48].

2. *Multi-Class Classification*, teks diklasifikasikan ke dalam tiga kategori, yaitu sentimen positif, netral, dan negatif [48].
3. *Fine-Grained Sentimen Analysis*, teks diklasifikasikan dengan lebih rinci, seperti sangat positif, positif, netral, negatif, dan sangat negatif [49].
4. *Aspect-Based Sentimen Analysis (ABSA)*, klasifikasi ini menilai sentimen berdasarkan aspek tertentu dari sebuah entitas. ABSA memecah opini menjadi aspek-aspek tertentu dan mengklasifikasikannya [50].
5. *Emotional Classification*, klasifikasi ini mengelompokkan sentimen berdasarkan emosi dalam teks, seperti senang, marah, dan lain-lain [51].

Selain jenis klasifikasi, terdapat juga berbagai metode yang dikembangkan untuk melakukan klasifikasi sentimen. Beberapa metode yang umum digunakan antara lain:

1. *Naïve Bayes Classifier (NBC)*, metode berbasis probabilistik yang sederhana dan cepat, sering digunakan untuk data teks [52].
2. *Support Vector Machine (SVM)*, algoritma berbasis *supervised learning* yang mencari *hyperline* terbaik untuk memisahkan data antar kategori [53].
3. *Random Forest*, algoritma *ensemble learning* yang membangun banyak *decision tree* dan menggabungkan hasilnya menjadi interpretasi hasil untuk meningkatkan akurasi prediksi [54].
4. *Logistic Regression*, metode ini digunakan untuk klasifikasi biner yang menghitung probabilitas suatu teks termasuk dalam kategori tertentu dengan fungsi logistic (sigmoid) [55].
5. *Long-Short Term Memory (LSTM)*, algoritma *deep learning* yang mampu mengingat informasi jangka panjang pada data sekuensial, seperti teks panjang [56].

2.2.7 Topic Modeling

Topic modeling merupakan pendekatan yang digunakan untuk mengidentifikasi struktur tersembunyi dari kumpulan dokumen yang besar dan tidak terstruktur dengan menemukan kelompok kata yang sering muncul bersama, atau yang merepresentasikan suatu tema tertentu [57]. Pemodelan topik bekerja dengan mengelompokkan kata-kata yang sering muncul bersama menjadi satu topik, sehingga mempermudah dalam memahami dan mengorganisir dokumen berdasarkan tema-tema yang terkandung di dalamnya, tanpa perlu membaca setiap dokumen [58]. Pemodelan topik umumnya digunakan untuk mengidentifikasi pola pembahasan, minat pengguna, dan dinamika opini publik terhadap suatu fenomena [59].

Beberapa metode pemodelan topik yang umum digunakan antara lain:

1. *Latent Dirichlet Allocation (LDA)*, pendekatan probabilistik yang mengasumsikan setiap dokumen tersusun dari beberapa topik, dan setiap topik terdiri dari sekumpulan kata [57].
2. *Non-negative Matrix Factorization (NMF)*, metode faktorisasi matriks yang menguraikan representatif dokumen teks menjadi dua matriks non-negatif untuk mengidentifikasi topik-topik utama [60].
3. *Latent Semantic Analysis (LSA)*, teknik dekomposisi matriks yang mengeksplorasi hubungan semantik antar kata dalam dokumen [61].
4. *BERTopic*, teknik yang menggabungkan *embedding* dan *transformer* (seperti *BERT*) dengan *clustering* untuk menghasilkan topik-topik yang lebih semantic dan akurat [58].
5. *Top2Vec*, pemodelan topik yang menggunakan *embedding* dokumen dan kata untuk secara langsung menemukan topik berdasarkan kedekatan spasial dalam ruang vector [62].