

BAB I

PENDAHULUAN

1.1. Latar Belakang Penelitian

Revolusi industri 4.0 mengidentifikasi teknologi internet sebagai alat penggerak utama ekonomi untuk mendapat momentum inovasi teknologi dalam perekonomian [1]. Istilah *Financial Technology (fintech)* atau teknologi keuangan mengacu pada cabang teknologi yang telah ada selama beberapa waktu dan telah dicirikan sebagai model bisnis yang moderat, yang mana sebelumnya hanya dapat dilakukan melalui transaksi dengan metode tradisional tatap-muka dan membawa sejumlah uang tunai [2]. Pinjaman *online, peer-to-peer (P2P)* atau sering disebut pinjol adalah salah satu jenis jasa keuangan di bidang *fintech* yang dioperasikan secara online dan merupakan model bisnis yang berkembang pesat di era kini [3]. Pinjaman *online* menjadi alternatif keuangan bagi masyarakat ketika berada pada kondisi yang mendesak dan membutuhkan dana tunai segera, dimana hanya dengan mengunduh aplikasi dari ponsel atau mengakses *web* penyedia layanan pinjaman, dengan mudah masyarakat bisa mendapatkan pinjaman uang [4].

Pertumbuhan perusahaan pinjaman *online* yang semakin banyak dan marak terjadinya kasus pelanggaran yang dilakukan perusahaan pinjaman *online* ilegal atau tidak berizin [4]. Per 3 Januari 2022, data statistik Otoritas Jasa Keuangan (OJK) menyebutkan sebanyak 103 perusahaan menjadi penyelenggara *fintech peer-to-peer lending* berizin di Otoritas Jasa Keuangan (OJK) [5]. Dalam rentang waktu Maret 2019 - Oktober 2021, Saji Sonjaya sebagai Ketua Satgas Anti Rentenir kota Bandung mengatakan adanya peningkatan laporan pengaduan *fintech* ilegal sebesar 30% dengan 6.000 laporan tentang pengaduan pinjol ilegal [6]. Berdasarkan prinsip syariah, pinjaman *online* adalah pinjaman yang dalam penerapannya terbebas dari *gharar* (ketidakjelasan akad), *riba*, *dharar* (bahaya), *maysir* (spekulasi), *zhulm* (kerugian), *tadlis* (tidak transparan), dan *haram* (tidak transparan), secara *sharih* ini dijelaskan pada Firman Allah dalam Surat Al-Baqarah ayat 275 [10].

Berbagai permasalahan berdasarkan uraian diatas memunculkan banyak perbincangan, komentar, tanggapan, atau pendapat masyarakat yang diungkapkan di berbagai media sosial, salah satunya *Twitter* yang sering kali digunakan sebagai sumber data untuk melakukan kegiatan analisis sentimen [7]. *Twitter* memberikan akses kepada penggunanya untuk dapat membagikan pesan baik dalam bentuk teks ataupun foto yang disebut dengan *tweet*, sehingga ini memungkinkan pengguna berinteraksi lebih dekat dengan pengguna lainnya dengan mengirimkan apa yang sedang dipikirkan atau apa yang terjadi di sekitar [7]. Kemudahan yang diberikan *twitter* bagi peneliti adalah adanya akses dari fitur API (*Application Programming Interface*) yang dapat mengoleksi dan mengumpulkan data seperti *tweet*, sehingga dengan data itu peneliti bisa melakukan *mining* data dari data berjumlah besar [7].

Untuk menyaring berbagai komentar dan pendapat masyarakat yang ada di media sosial, kegiatan Analisis Sentimen sangat diperlukan, karena kegiatan ini akan menganalisa kalimat pada tiap komentar atau pendapat seseorang apakah memiliki makna yang positif, negatif, atau netral [8]. Kegiatan analisis sentimen seperti testimoni atau *opinion mining* adalah proses untuk mendapatkan informasi yang termuat pada suatu kalimat opini dengan cara memahami, mengekstrak, dan mengolah data tekstual secara otomatis [9]. Penelitian-penelitian tentang analisis sentimen telah banyak dilakukan, dengan menggunakan dua pendekatan yang biasa digunakan, yaitu pendekatan berbasis pembelajaran mesin (*machine learning*) dan pendekatan berbasis leksikal (*Lexicon-Based*). Penelitian analisis sentimen ini menggunakan kedua pendekatan tersebut, metode *Lexicon-Based* untuk pendekatan berbasis leksikal dan algoritma *SVM* untuk pendekatan berbasis *machine learning*.

Kedua metode tersebut layak dibandingkan, sebagaimana keduanya adalah masing-masing perwakilan dari pendekatan untuk kegiatan analisis, *Lexicon-Based* untuk pendekatan berbasis kamus atau korpus, dan *SVM* untuk pendekatan berbasis pembelajaran mesin. Selain itu, kelebihan *Lexicon-Based* ialah bekerja berdasarkan kamus kata, sehingga adanya perbandingan secara langsung antara kata dari kalimat dengan kamus kata opini yang termuat dalam *lexicon*, maka bila suatu kalimat memiliki kata berupa opini, maka kalimat itu akan dianggap sebagai kalimat opini, tidak heran jika metode ini disebut sebagai ekstraksi fitur yang dapat meningkatkan performansi sistem. Sedangkan untuk metode *SVM* sebagai salah satu contoh dari

pendekatan *machine learning* yang sering digunakan dalam kegiatan klasifikasi teks, deteksi spam, penentuan kategori, dan sentimen analisis. Meski dilatih dengan relatif sedikit data, *SVM* dapat menghasilkan model klasifikasi dan akurasi yang baik, ini dibuktikan dengan rata-rata tingkat akurasi yang dihasilkan adalah diatas 60% berdasarkan referensi yang peneliti gunakan. *CRISP-DM (Cross Industry Standard Process for Data Mining)* digunakan sebagai standar proses data mining yang menyediakan strategi dan langkah pemecahan masalah penelitian.

Pada tahun 2018, Hidayatulah Himawan dkk meneliti analisis rekomendasi favoritisme oleh-oleh pada media sosial dengan metode *Lexicon-Based* dan *Support Vector Machine* dengan akurasi terbesar didapatkan pada metode *Lexicon-Based* [17]. Lalu untuk penelitian perbandingan metode atau algoritma pada analisis sentimen telah dilakukan sebelumnya pada tahun 2019 oleh Ahmad Choirun Najib dkk yaitu perbandingan *Lexicon-Based* dan *Support Vector Machine (SVM)* untuk analisis sentimen berbasis ontologi pada kampanye pilpres Indonesia 2019 di *twitter*, penelitian tersebut menggabungkan ontologi dan sentimen analisis dengan hasil penelitian menunjukkan metode *SVM* memiliki performa yang lebih baik apabila dibandingkan dengan algoritma *Lexicon-Based*, dengan persentase akurasi yang diperoleh berdasarkan metode *SVM* mencapai 83% sedangkan metode *Lexicon-Based* sebesar 39% [11]. Pada tahun 2019 juga dilakukan penelitian oleh Azhar dkk terkait pengaruh pendekatan *Lexicon* pada kinerja algoritma *Naïve Bayes* dan *K-NN* untuk analisis sentimen di media *twitter* [16].

Pada tahun 2020, Gupta Sanjaya dkk meneliti penggunaan *Lexicon-Based* untuk analisis sentimen komentar konten *Youtube* terkait terpilihnya menteri Kabinet Indonesia Maju Jokowi-Ma'ruf [18]. Pada tahun 2021, Dian Siti Utami, dkk menganalisis sentimen pinjaman *online* di media sosial *Twitter* menggunakan algoritma *Support Vector Machine* [7]. Kemudian pada tahun 2021 juga terdapat penelitian yang menggabungkan metode *SVM* dan *Lexicon-Based* untuk Analisis Sentimen *Twitter* dengan objek yang diteliti adalah tanggapan pengguna *twitter* untuk PP Tapera (Tabungan Perumahan Rakyat) Nomor 25 tahun 2020 [33]. Pada tahun 2022, dilakukan penelitian perbandingan metode *Lexicon-Based* dan *Naïve Bayes Classifier* pada analisis sentimen pendapat masyarakat terhadap aplikasi investasi pada media *twitter* [32].

Polaritas dari sebuah kalimat yang ada pada sebuah *tweet* menjadi hal yang diperhatikan berdasarkan penelitian sebelumnya. Pada penelitian ini, peneliti akan melakukan analisis sentimen tanpa penggabungan ontologi, dan bertujuan untuk mencari tingkat akurasi dari penggunaan kedua metode, serta fokus utamanya adalah membandingkan hasil yang didapat berdasarkan kedua metode dengan parameter yang digunakan yaitu hasil akurasi, performa metrik pada langkah uji, dan asumsi faktor yang mempengaruhi hasil akurasi dan performa sehingga bisa dibandingkan performa mana yang lebih unggul dari kedua metode tersebut. Penelitian ini mengambil studi kasus topik tentang pinjaman *online* atau pinjol.

Dari uraian tersebut, untuk dapat mengetahui apakah metode *Lexicon-Based* dan *Support Vector Machine* dapat melakukan klasifikasi teks dengan baik dan bagaimana perbandingan akurasi dan performa yang dihasilkan dari kedua metode tersebut, maka perlu dilakukan penelitian. Peneliti merumuskan penelitian dengan judul *Perbandingan Metode Lexicon-Based dan Support Vector Machine untuk Analisis Sentimen Pinjaman Online pada Media Sosial Twitter*. Hasil akhir dari penelitian ini berupa tanggapan positif, netral, dan negatif.

1.2. Perumusan Masalah Penelitian

Berdasar pada latar belakang penelitian yang telah diuraikan sebelumnya, berikut perumusan masalah dari penelitian yang akan dilakukan:

1. Bagaimana metode *Lexicon-Based* dan *Support Vector Machine (SVM)* dapat melakukan analisis sentimen?
2. Bagaimana tingkat akurasi yang diperoleh dari metode *Lexicon-Based* dan metode *Support Vector Machine (SVM)* pada analisis sentimen dengan topik Pinjaman *Online*?
3. Bagaimana perbandingan dari metode *Lexicon-Based* dengan *Support Vector Machine (SVM)* berdasarkan hasil tingkat akurasi yang diperoleh dari analisis sentimen yang dilakukan?

1.3. Tujuan Penelitian

Beberapa hal yang menjadi tujuan dari penelitian ini diantaranya sebagai berikut.

1. Mengetahui penggunaan metode *Lexicon-Based* dan *Support Vector Machine (SVM)* dalam analisis sentimen.
2. Mengetahui tingkat akurasi yang diperoleh dari metode *Lexicon-Based* dan *Support Vector Machine (SVM)* pada analisis sentimen Pinjaman *Online*.
3. Mengetahui perbandingan dari metode *Lexicon-Based* dan *Support Vector Machine (SVM)* berdasarkan hasil tingkat akurasi yang diperoleh dari analisis sentimen yang dilakukan

Selanjutnya, manfaat yang didapat dari penelitian ini yaitu sebagai berikut:

1. Membantu penelitian sebelumnya yang berhubungan dengan perbandingan metode atau algoritma *Lexicon-Based* dan *Support Vector Machine (SVM)* maupun yang berhubungan dengan analisis sentimen.
2. Menjadi harapan dan awal bagi penelitian lain untuk meneliti tentang perbandingan metode atau algoritma dalam analisis sentimen.
3. Memberikan pengetahuan kepada masyarakat terkait pinjaman *online* yang menjadi tren di era digital saat ini, sehingga diharapkan dapat membantu masyarakat untuk mempertimbangkan segala risiko tentang pinjaman *online*, memilih mana pinjaman *online* yang berizin (legal) dan meminimalisir kasus korban pinjaman *online* tidak berizin (ilegal) yang sering terjadi.

1.4. Batasan Masalah Penelitian

Batasan masalah penelitian bertujuan supaya penelitian menjadi memiliki arah yang jelas dan sesuai dengan apa tujuan yang diinginkan. Beberapa batasan masalah dari penelitian tentang Perbandingan Metode *Lexicon-Based* dan *Support Vector Machine* untuk Analisis Sentimen Pinjaman *Online* pada Media Sosial *Twitter* yaitu sebagai berikut:

1. Sumber data yang digunakan pada penelitian analisis sentimen ini adalah dari media sosial *twitter*.
2. *Tweet* berbahasa Indonesia adalah *tweet* yang digunakan dalam penelitian analisis sentimen ini, dan mengandung kata kunci “Pinjaman *Online*”, “Pinjol”, “aplikasi pinjaman *online*”, dan “aplikasi pinjol”.
3. Pengambilan data dilakukan sebanyak tiga kali, dengan rentang waktu yaitu percobaan pengambilan data pertama dilakukan pada tanggal 1 September

2022, percobaan pengambilan data kedua dilakukan pada tanggal 8 November 2022, dan percobaan pengambilan data ketiga dilakukan pada 30 November 2022, dimana data *tweet* yang diambil adalah data *tweet* yang relevan.

4. Metode atau Algoritma yang digunakan pada penelitian analisis sentimen ini adalah menggunakan *Lexicon-Based* dan *Support Vector Machine (SVM)*
5. Dalam pembagian data, sebanyak 4.136 data diolah dalam tahap pemodelan untuk metode *Lexicon-Based* maupun *Support Vector Machine (SVM)*.
6. Pada metode *Support Vector Machine (SVM)*, sebanyak 9.710 data *tweet* (keseluruhan dataset) akan digunakan sebagai data pada tahap pengujian atau analisis sentimen.
7. Penelitian ini membandingkan bagaimana penggunaan kedua metode atau algoritma yaitu *Lexicon-Based* dengan *Support Vector Machine (SVM)* berdasarkan dari hasil penelitian yang diperoleh.
8. Komentar atau pendapat dari *tweet* pada *twitter* yang dianalisis hanya berkaitan dengan topik Pinjaman *Online*.

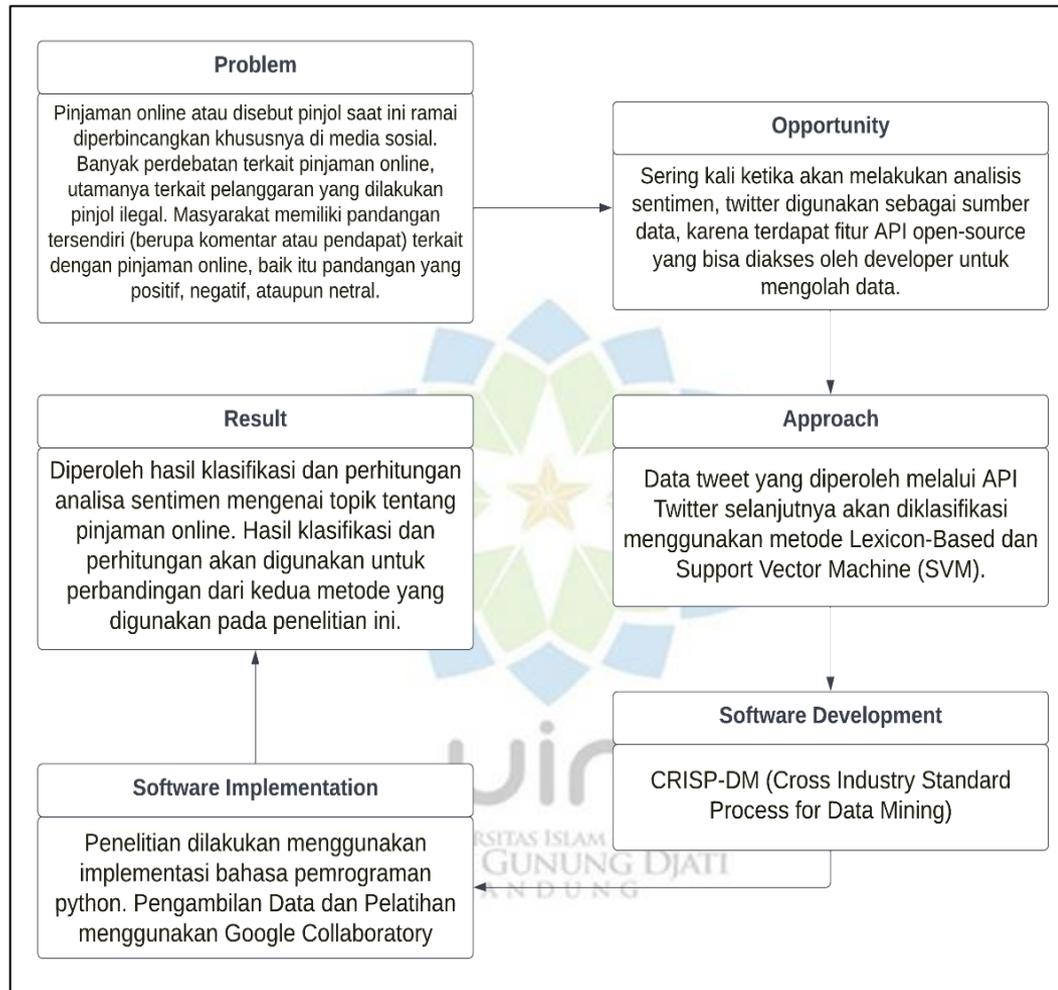
1.5. Kerangka Pemikiran Penelitian

Peer to peer (P2P) lending atau layanan pinjam uang berbasis aplikasi atau teknologi informasi (*online*) atau pinjaman *online* merupakan jenis dari penyelenggaraan *fintech* yang menjadi alternatif solusi bagi masyarakat apabila berada pada kondisi yang mendesak sehingga memaksa mereka membutuhkan pinjaman dana tunai segera. Pinjaman *online* atau pinjol saat ini cukup ramai diperbincangkan khususnya di media sosial. Fakta yang faktual dan terkini diperlukan untuk dapat mengetahui bagaimana pandangan masyarakat (berupa komentar atau pendapat) terkait dengan pinjaman *online*. Oleh karena itu, peneliti menggunakan data yang diperoleh melalui *tweet* dari media sosial *twitter* dengan memanfaatkan API *twitter* mengenai topik pinjaman *online* yang dapat dijadikan sebagai sampel untuk seluruh sentimen mayoritas masyarakat.

Algoritma atau metode *Lexicon-Based* dan *Support Vector Machine (SVM)* akan digunakan. Keduanya akan dibandingkan untuk dapat mengetahui performa mana yang lebih unggul dari kedua algoritma tersebut berdasarkan parameter yang digunakan antara lain hasil akurasi, performa metrik pengujian model kedua

metode, dan asumsi peneliti terkait faktor yang mempengaruhi performa dan hasil akurasi algoritma atau metode tersebut.

Berikut ini merupakan kerangka pemikiran dari penelitian tugas akhir dapat dilihat pada gambar berikut.



Gambar 1. 1. Kerangka Pemikiran

1.6. Metodologi Penelitian

1.6.1. Tahap Perencanaan

Tahapan perencanaan merupakan tahap analisa untuk segala hal yang dibutuhkan dalam mempelajari literatur berupa jurnal ataupun buku yang berhubungan dengan penelitian sehingga dapat dilakukan kegiatan analisis sentimen dengan penggunaan metode *Lexicon-Based* dan *Support Vector Machine (SVM)*.

1.6.2. Tahap Penelitian

Tahapan penelitian dilakukan melalui serangkaian tahapan-tahapan sesuai dengan metode *Cross Industry Standard Process for Data Mining (CRISP-DM)*. Metodologi ini menyediakan proses standar data mining sebagai strategi untuk pemecahan masalah penelitian. Serangkaian tahapan tersebut meliputi tahap *Business Understanding, Data Understanding, Data Preparation, Modeling, dan Evaluation*.

1.6.2.1. Business Understanding

Tahapan ini dilakukan proses memahami bagaimana objek yang akan diteliti untuk dapat menghasilkan keluaran atau output yang diinginkan atau sesuai dengan harapan. Tujuan dilakukan penelitian pada objek *Twitter* dan studi kasus Pinjaman *Online* dan batasan penelitian dibahas pada tahap pemahaman bisnis ini.

1.6.2.2. Data Understanding

Apa yang menjadi sumber data, berapa banyak data, dan bagaimana kualitas data yang diklasifikasi akan dibahas pada proses pemahaman data ini, tahapan ini penting supaya tujuan dari penelitian analisis sentimen bisa tercapai. Seringkali terdapat data yang bersifat duplikasi dan tidak bernilai atau kosong.

1. Sumber Data

Media sosial *Twitter* digunakan sebagai sumber data, dilengkapi akses *API (Application Programming Interface)*. Pinjaman *Online* pernah menjadi trending topik di *twitter*, sehingga memungkinkan banyak data yang dapat diambil untuk digunakan pada penelitian.

2. Jumlah Data

Pengambilan data menghasilkan sebanyak 9.710 data *tweet*. Besarnya jumlah data yang didapatkan pada proses selanjutnya akan melalui pembersihan sehingga akan banyak data yang dihilangkan karena data tersebut bersifat duplikat, data bernilai kosong, dan terdapat karakter pada data yang tidak diperlukan pada tahap selanjutnya.

1.6.2.3. Data Preparation

Langkah ini perlu diselesaikan sebelum data diproses selama tahap pemodelan (*modelling*). Dikenal sebagai tahap persiapan data, mengubah data yang sebelumnya tidak diproses menjadi data yang bisa digunakan. Beberapa tahapan untuk proses ini meliputi tahap *cleansing*, *case folding*, *tokenizing*, *filtering* atau *stopwords removal*, dan *stemming*.

1.6.2.4. Modeling

Metode klasifikasi yang digunakan pada tahap pemodelan adalah dengan pendekatan berbasis leksikal (*Lexicon-Based*) dan pendekatan berbasis pembelajaran mesin yaitu dengan algoritma *Support Vector Machine (SVM)*. Implementasi menggunakan bahasa pemrograman python, kegiatan pelatihan serta pengujian model dilakukan melalui *Google Collaboratory*. Data yang telah dibersihkan pada tahap persiapan data kemudian diklasifikasi oleh *Lexicon-Based* dan *SVM* sehingga didapat label akhir yang terdiri dari “positif”, “netral”, dan “negatif”. Didapat juga hasil akurasi dan performa metrik pada pengujian model.

1.6.2.5. Evaluation

Hasil klasifikasi proses pemodelan sebelumnya diinterpretasikan dengan mengkontraskan hasil pemodelan berupa nilai akurasi, *recall*, dan presisi. Model yang digunakan selama tahap pemodelan dievaluasi untuk memastikan model yang dibangun sesuai dengan tujuan yang ingin dicapai pada pemahaman bisnis. Dilakukan perbandingan metode berdasarkan hasil pengujian model yang telah diperoleh.

1.7. Sistematika Penulisan

Untuk laporan tugas akhir ini, sistematika penulisan meliputi gambaran bagaimana isi pada tiap bab, urutan penulisan, dan keterkaitan satu bab dengan bab lain pada laporan. Berikut adalah contoh sistem penulisan dari laporan sebelumnya.

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang penelitian, rumusan masalah penelitian, tujuan penelitian, batasan masalah penelitian, kerangka penelitian, dan sistematika penulisan.

BAB II KAJIAN LITERATUR

Pada bab ini dijabarkan apa saja literatur atau penelitian terdahulu, konsep-konsep, teori-teori, dan model-model yang menjadi landasan dalam perancangan sistem untuk analisa sentimen.

BAB III METODOLOGI PENELITIAN

Bab ini berisi penjelasan bagaimana rangkaian metodologi yang digunakan atau implementasinya pada penelitian.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini memaparkan dua hal utama, pertama adalah pemaparan tentang temuan atau hasil penelitian berdasarkan tahapan penelitian yang dilakukan. Kedua adalah pembahasan hasil temuan penelitian, yang mana ini menjawab pertanyaan dari perumusan penelitian.

BAB V SIMPULAN DAN SARAN

Pada bab ini berisi bagaimana kesimpulan dari hasil penelitian yang telah ditemukan dan pembahasan yang telah dijabarkan, selain itu diberikan saran-saran untuk peningkatan pada penelitian selanjutnya.