

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi dalam jasa keuangan atau *finansial technology* (*fintech*) saat ini telah mengalami perkembangan yang meningkat dengan pesat. Salah satu layanan jasa keuangan yang perkembangannya cepat di masyarakat adalah pinjaman online atau yang lebih dikenal dengan istilah pinjol. Pinjaman online adalah layanan pinjam meminjam berbasis teknologi informasi yang dilakukan secara online menggunakan gawai tanpa perlu adanya tatap muka. Cara ini tentu lebih memberikan kemudahan dan kecepatan dalam proses pengajuan kredit, sebab aplikasi pinjol dapat diunduh dengan mudah melalui Google Play atau AppStore (Abrianti, Anggraini, & Probondaru, 2024).

Fitur utama yang dimiliki oleh pinjaman online yang tidak dimiliki oleh pinjaman konvensional yaitu proses pengajuan, pencairan serta pengembalian uang pinjaman dapat dilakukan lewat aplikasi hanya dengan gawai secara cepat dan dengan syarat yang mudah seperti KTP, data pribadi, nomor kontak dan swafoto (Hakim & Setyabudi, 2020). Karena mudahnya alur dan akses dalam pencairan dana tersebut, pinjol dijadikan cara alternatif untuk memenuhi kebutuhan pokok. Seiring dengan pertumbuhannya yang cepat dan tingginya angka penggunaan pinjaman online, sayangnya masih banyak masyarakat Indonesia yang kurang pengetahuan dalam menggunakan produk tersebut. Masih banyak masyarakat Indonesia yang tidak paham terkait untung, rugi, syarat, serta akibat jika melanggar ketentuan dari pinjaman online.

Akibat dari melanggar ketentuan tentu akan menimbulkan berbagai masalah serius, salah satunya adalah penagihan saat pengguna pinjaman online terlambat membayar cicilan. Namun, dalam praktik penagihan pinjol saat ini masih banyak cara penagihan yang dianggap melanggar etika dan privasi pengguna. Berdasarkan data yang dihimpun dari Otoritas Jasa Keuangan (OJK), terhitung mulai 1 Januari 2025 hingga 17 November 2025, setidaknya sudah tercatat 470.678 permintaan layanan melalui Aplikasi Portal Perlindungan Konsumen (APPK).

Sedangkan data lain dari OJK menyatakan mulai Januari hingga Juni 2025, terdapat sekitar 3.858 aduan terkait penagihan pinjaman online yang dilakukan oleh pihak ketiga yang tidak sesuai ketentuan. Para penagih pinjaman atau *debt collector* (DC) diduga banyak melakukan tindak pidana, mulai dari *spam* panggilan telepon, ancaman, intimidasi, teror, hingga berbagai tindakan yang dapat memperlakukan seperti penyebaran data pribadi.

Keluhan terkait penagihan pinjol ini banyak disuarakan penggunanya melalui berbagai media sosial, salah satunya X yang sebelumnya dikenal dengan nama Twitter. Hal ini menjadikan Twitter sebagai sumber referensi bagi ilmuwan sosial dan akademisi yang tertarik untuk mempelajari nuansa emosi manusia dari berbagai sudut (Öztürk & Ayvaz, 2018). X adalah wadah publik untuk menyampaikan berbagi informasi secara *real time*, termasuk keluhan terkait penagihan pinjaman online. Namun, tantangan utama dalam mengolah data dari platform ini adalah penggunaan bahasa yang tidak baku, singkatan, serta slang. Hal ini dapat menyebabkan munculnya kata-kata yang tidak terdaftar dalam kamus (*out-of-vocabulary*), sehingga diperlukan teknik representasi kata seperti FastText yang mampu menangani keterbatasan tersebut dengan memecah kata menjadi unit sub-kata (*n-grams*).

Selain kendala pada ragam bahasa, tantangan dalam mengolah data keluhan dari media sosial X adalah adanya ketidakseimbangan kelas data atau *imbalanced data*. Secara alamiah, jumlah sentimen negatif terkait penagihan lebih mendominasi dibandingkan dengan sentimen netral atau positif. Kondisi ini dapat menyebabkan algoritma klasifikasi bias terhadap kelas mayoritas, sehingga informasi penting pada kelas minoritas menjadi terabaikan. penelitian ini. Menurut Sipahutar dkk. (2025), penerapan teknik *Synthetic Minority Over-sampling Technique* atau SMOTE diperlukan untuk meningkatkan akurasi klasifikasi pada dataset yang tidak seimbang, seperti pada kasus kelayakan donor darah. Oleh karena itu, penelitian ini menerapkan teknik SMOTE untuk menyeimbangkan distribusi data secara sintesis pada ruang vektor FastText. Dalam penelitiannya, penerapan SMOTE berhasil meningkatkan rata-rata akurasi sebesar 2,55%, yakni dari 93,50% menjadi 96,05% melalui kombinasi algoritma tambahan. Bahkan, pada pengujian menggunakan *K-Fold Cross Validation*, model yang dihasilkan mampu mencapai akurasi tertinggi

hingga 96,50%. Penelitian tersebut juga membuktikan bahwa dengan menyeimbangkan jumlah sampel kelas minoritas dan mayoritas SMOTE berhasil meningkatkan nilai *recall* dan *f1-score* pada kelas minoritas secara stabil. Hasil evaluasi menunjukkan model menjadi lebih sensitif dalam mendeteksi kelas target dengan jumlah *true positive* mencapai 31 dan menekan *false negative* hingga titik terendah sebesar 13. Temuan ini menunjukkan bahwa penerapan SMOTE pada ruang vektor FastText dalam penelitian ini merupakan langkah tepat untuk memastikan algoritma Bayesian Network dan C4.5 dapat mengenali setiap aspek keluhan penagihan pinjaman online secara lebih presisi, stabil, dan objektif.

Sehubungan dengan semakin banyak keluhan terkait penagihan pinjaman online yang ditemukan di media sosial X, sehingga diperlukan suatu analisis sentimen yang komprehensif untuk mengidentifikasi masalah apa saja yang terjadi secara spesifik. Oleh karena itu, dibutuhkan pendekatan *aspect based sentiment analysis* atau ABSA untuk menganalisis sentimen berdasarkan aspeknya. Tujuan dari analisis sentimen berbasis aspek adalah untuk mengidentifikasi sentimen yang terkait dengan entitas atau aspeknya (Putri, Sumarjaya, & Wijayakusuma, 2025).

Untuk meningkatkan efektivitas penerapan ABSA, penelitian ini memanfaatkan teknik *data mining* dengan membandingkan dua algoritma klasifikasi, yaitu C4.5 dan *Bayesian Network*. *Bayesian Network* merupakan algoritma yang melakukan pemodelan hubungan probabilistik antara variabel-variabel yang menarik serta dapat menunjukkan probabilitas hubungan antara kejadian-kejadian yang saling berhubungan maupun tidak berhubungan (Hasniati, Arianti, & Philip, 2018). *Bayesian Network* dapat memperkirakan probabilitas sentimen terhadap suatu aspek tertentu dengan menggunakan *Teorema Bayes*. Hal ini sangat berguna terutama saat mengolah data berupa teks dengan bahasa alami atau bahasa manusia. Pemilihan Bayesian Network dalam penelitian ini karena unggul dibandingkan algoritma klasifikasi sejenis. Dalam pemodelan probabilistik, algoritma seperti *Naive Bayes* sering kali dijadikan standar. Namun, *Naive Bayes* memiliki kelemahan mendasar berupa asumsi yang menganggap setiap kata berdiri sendiri tanpa dipengaruhi kata lain. Padahal, struktur kalimat keluhan penagihan pinjaman online sangat kompleks dan saling berkaitan. Oleh karena itu, Bayesian Network dipilih karena kemampuannya memodelkan hubungan ketergantungan

antar variabel secara lebih realistis dan terstruktur melalui *directed acyclic graph* (DAG).

Keunggulan *Bayesian Network* telah dibuktikan oleh beberapa penelitian terdahulu. Dalam penelitian Regina dkk. (2023) yang menganalisis sentimen di Twitter menggunakan fitur TF-IDF, *Bayesian Network* mendapatkan hasil akurasi 76,5556%, presisi 77,4%, dan *recall* 76,6%. Selain itu, keunggulan *Bayesian Network* juga terlihat saat dilakukan perbandingan dengan algoritma *Naïve Bayes* dalam penelitian Windarti (2018), meskipun *Bayesian Network* dan *Naïve Bayes* menghasilkan nilai akurasi yang setara pada pengujian *percentage split* 90 yaitu sebesar 80%, tetapi pada pengujian *percentage split* 80, *Bayesian Network* memperoleh akurasi 75%, lebih unggul dibandingkan *Naïve Bayes* yang hanya memperoleh 70%.

Sementara itu, C4.5 adalah algoritma pembelajaran mesin yang digunakan untuk membangun pohon keputusan (*decision tree*) dari data pelatihan (Nugraha & Said, 2024). Struktur *decision tree* yang dihasilkan oleh C4.5 dapat digunakan untuk mencari fitur-fitur teks yang berkaitan dengan klasifikasi aspek keluhan dan polaritas sentimennya. Algoritma ini bekerja dengan menetapkan aturan-aturan yang diperoleh dari data untuk mengklasifikasikan ulasan pengguna berdasarkan pola yang ditemukan dalam dataset (Satria, Adinugroho, & Suprpto, 2020). Selain akurasi yang tinggi, C4.5 unggul dalam mengidentifikasi akar permasalahan melalui visualisasi pohon keputusan.

Berdasarkan penelitian *systematic literature review*, Nazifah dan Prianto (2023) membuktikan bahwa akurasi C4.5 lebih unggul bahkan mampu mencapai nilai akurasi hingga 98% pada karakteristik data tertentu dibandingkan algoritma lain seperti KNN. Selain itu, keunggulan C4.5 didukung oleh penelitian Nurkholis dkk. (2024), yang menunjukkan bahwa dalam analisis sentimen di media sosial, C4.5 mendapatkan akurasi 98,42%, hasil tersebut lebih unggul dibandingkan algoritma KNN 97,33%, dan Naive Bayes 57,32%. Oleh karena itu, C4.5 dipilih untuk mengatasi keterbatasan metode seperti KNN yang memakan waktu komputasi tinggi saat memproses vektor kata dari FastText dan bersifat *black-box*. Berbeda dengan KNN, algoritma C4.5 mampu menyeleksi fitur kata yang paling relevan secara otomatis dan menghasilkan pohon keputusan yang transparan.

Pemilihan algoritma Bayesian Network dan C4.5 dalam penelitian ini karena perbedaan karakteristik dan keunggulan masing-masing pendekatan dalam mengklasifikasikan teks. Bayesian Network menggunakan pendekatan berbasis peluang dalam menganalisis keterkaitan antar kata pada kalimat yang tidak terstruktur. Sementara itu, algoritma C4.5 menggunakan pendekatan berbasis aturan yang mampu menghasilkan pohon keputusan, sehingga alasan mengapa suatu teks masuk ke dalam kategori keluhan tertentu dapat ditelusuri dengan jelas. Karena data teks keluhan penagihan pinjaman online di media sosial X memiliki karakteristik yang kompleks, penuh ekspresi, dan banyak menggunakan bahasa tidak baku, sehingga belum dapat dipastikan algoritma mana yang memiliki kinerja lebih baik.

Oleh karena itu, perbandingan antara Bayesian Network dan C4.5 yang digabungkan dengan teknik FastText dan SMOTE menjadi sangat penting untuk diteliti lebih lanjut. Penelitian ini tidak sekadar bertujuan untuk mencari nilai akurasi tertinggi, tetapi juga untuk mengevaluasi algoritma mana yang memiliki keandalan klasifikasi paling optimal dalam melakukan *aspect based sentiment analysis*. Hasil penelitian ini diharapkan tidak hanya berkontribusi secara keilmuan, tetapi juga dapat menjadi rujukan bagi pihak regulator dan industri *fintech* dalam mengevaluasi standar operasional prosedur penagihan, guna menciptakan ekosistem pinjaman online yang lebih sehat, aman, dan beretika di Indonesia.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah disebutkan di atas, maka didapatkan permasalahan yang dapat diidentifikasi yaitu:

- 1) Dalam pengolahan data teks dari media sosial X, terdapat kendala berupa penggunaan bahasa tidak baku, singkatan, dan *slang*, sehingga diperlukan teknik *FastText* agar dapat merepresentasikan kata yang belum pernah dilihat (OOV).
- 2) Masih minimnya penelitian yang secara spesifik membandingkan kinerja antara algoritma *Bayesian Network* dan C4.5 dalam penerapan *aspect based sentiment analysis*.

- 3) Adanya karakteristik ketidakseimbangan kelas data pada dataset keluhan penagihan pinjaman online, di mana sentimen negatif sangat mendominasi, yang berpotensi menyebabkan bias pada hasil klasifikasi model.

1.3 Pembatasan Masalah

Berdasarkan latar belakang dan identifikasi masalah yang telah disampaikan, maka didapatkan pembatasan masalah, yaitu:

- 1) Penelitian ini menggunakan algoritma *Bayesian Network* dan C4.5 untuk tugas *aspect based sentiment analysis* terhadap keluhan penagihan pinjaman online.
- 2) Sumber data yang dianalisis berasal dari media sosial X dengan kata kunci terkait keluhan penagihan pinjaman online.
- 3) Representasi teks menggunakan metode *word embedding* dengan model *FastText* untuk menangani keterbatasan kosakata.
- 4) Penanganan ketidakseimbangan kelas data dilakukan dengan menerapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE).
- 5) Pengukuran kinerja model dilakukan menggunakan *confusion matrix* yang berfokus pada nilai *accuracy*, *precision*, *recall*, dan *f1-score*.

1.4 Rumusan Masalah

Berdasarkan latar belakang, identifikasi masalah, dan pembatasan masalah tersebut, maka didapatkan sebuah rumusan masalah yaitu "Bagaimana perbandingan kinerja algoritma *Bayesian Network* dan C4.5 dalam melakukan *aspect based sentiment analysis* untuk mengidentifikasi keluhan penagihan pinjaman online di media sosial X?"

1.5 Tujuan Penelitian

Berdasarkan uraian di atas, maka tujuan dari penelitian ini adalah untuk menganalisis perbandingan kinerja algoritma *Bayesian Network* dan C4.5 dalam melakukan klasifikasi sentimen berbasis aspek pada keluhan penagihan pinjaman online.

1.6 Manfaat Penelitian

Dari penelitian ini maka didapatkan manfaat dari penelitian yaitu:

1.6.1 Manfaat Teoritis

- 1) Memberikan kontribusi dalam literatur *natural language processing* melalui penerapan algoritma *Bayesian Network* dan C4.5 pada klasifikasi keluhan penagihan pinjaman online di media sosial X.
- 2) Menambah pemahaman mengenai efektivitas *aspect based sentiment analysis* dibandingkan metode analisis sentimen tradisional dalam mengelola data.
- 3) Menjadi dasar pengembangan kajian akademik mengenai perbandingan kinerja antara algoritma *Bayesian Network* dan C4.5.

1.6.2 Manfaat Praktis

- 1) Menyediakan model klasifikasi berbasis data yang dapat digunakan sebagai masukan dalam membuat kebijakan terkait pengawasan dan perbaikan etika penagihan bagi regulator maupun perusahaan *fintech*.
- 2) Menawarkan saran analisis yang lebih objektif dan efektif dalam memetakan peluang perbaikan layanan pinjaman online berdasarkan pola keluhan yang ada di media sosial.
- 3) Menjadi referensi praktis bagi penelitian dan pengembangan *aspect based sentiment analysis* di masa mendatang.

Intelligentia - Dignitas