

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Laporan keuangan berperan penting dalam memberikan gambaran menyeluruh mengenai kondisi ekonomi dan kinerja perusahaan publik. Informasi dalam laporan tersebut tidak hanya berupa angka, tetapi juga mencakup narasi yang menjelaskan risiko, strategi, serta prospek usaha yang dihadapi. Menurut (Zhu et al., 2024) pengungkapan risiko dalam laporan keuangan terbukti memiliki nilai informatif yang signifikan karena dapat mencerminkan perubahan nyata pada profil risiko perusahaan, khususnya selama periode krisis ekonomi global. Dalam penelitian terhadap lebih dari 14.000 laporan keuangan perusahaan di Amerika Serikat, ditemukan bahwa perusahaan yang terdampak krisis cenderung menggunakan bahasa yang lebih negatif dan spesifik, yang menunjukkan adanya korelasi antara gaya bahasa dengan tingkat risiko finansial aktual. Hal ini membuktikan bahwa analisis berbasis teks memiliki potensi besar untuk mengungkap dimensi risiko yang tidak terlihat dalam data numerik tradisional.

Di Indonesia, transparansi risiko perusahaan publik menjadi sorotan utama seiring meningkatnya kompleksitas bisnis dan regulasi dari Otoritas Jasa Keuangan (OJK). Berdasarkan penelitian (Largyta Grahawijaya et al., 2025) terhadap 90 bank di Indonesia selama 2020–2022, faktor ukuran perusahaan dan proporsi komisaris independen memiliki pengaruh positif terhadap tingkat pengungkapan manajemen risiko. Sebaliknya, kepemilikan manajerial menunjukkan efek negatif yang signifikan, yang mengindikasikan adanya konflik kepentingan antara pemegang saham internal dan publik dalam keterbukaan informasi risiko. Penelitian tersebut menegaskan bahwa tata kelola perusahaan dan karakteristik organisasi berperan penting dalam mendorong kualitas pengungkapan risiko, yang pada akhirnya meningkatkan kepercayaan investor dan stabilitas sektor keuangan.

Analisis risiko dalam laporan keuangan semakin kompleks dengan adanya dominasi teks naratif yang panjang dan ambigu. Penggunaan metode Natural NLP dan *Machine Learning* menjadi solusi untuk mengambil informasi tersembunyi dalam teks keuangan. Dalam hal ini, algoritma seperti SVM dan IndoBERT sebagai model transformer yang telah dilatih khusus untuk Bahasa Indonesia banyak digunakan untuk klasifikasi teks berbasis risiko. Menurut Rahma dan Suadaa (2023), salah satu

tantangan dalam klasifikasi teks di Indonesia adalah masalah ketidakseimbangan data yang dapat menurunkan kinerja model. Dalam penelitiannya, penggunaan teknik text augmentation seperti back translation dan synonym replacement mampu meningkatkan F1-score hingga 8% pada kelas minoritas. Hal ini menunjukkan bahwa kualitas dan distribusi data sangat berpengaruh terhadap performa model NLP.

Pemilihan SVM dan IndoBERT pada penelitian ini didasarkan pada perbedaan karakteristik keduanya. SVM umumnya bekerja cukup baik pada dataset berukuran sedang dengan fitur berbasis frekuensi seperti TF-IDF serta memiliki kemampuan generalisasi yang stabil. Sementara itu, IndoBERT lebih unggul dalam memahami konteks dua arah dan hubungan makna antar kata dalam bahasa Indonesia. (Rahma & Suadaa, 2023) juga menemukan bahwa model deep learning seperti BERT dapat meningkatkan akurasi sekitar 3–5% dibandingkan metode tradisional. Hal ini menunjukkan bahwa model kontekstual memiliki kelebihan dalam menangani bahasa yang kompleks, termasuk teks keuangan.

Selain aspek teknis, penggunaan NLP dalam analisis risiko juga memiliki nilai strategis dalam mendukung tata kelola perusahaan. Penelitian Marfiah et al. (2025) menjelaskan bahwa pengungkapan risiko bukan hanya kewajiban regulasi, tetapi juga menjadi sinyal bagi investor. Perusahaan dengan tingkat leverage dan kompleksitas bisnis tinggi cenderung mengungkapkan risiko secara lebih luas untuk menjaga kepercayaan pasar. Oleh karena itu, analisis teks berbasis Machine Learning dapat membantu menilai sejauh mana perusahaan menyampaikan informasi risiko secara terbuka.

Dalam skala global, penerapan text mining dan *deep language models* seperti BERT telah terbukti efektif untuk mendeteksi pola bahasa yang berhubungan dengan risiko keuangan dan potensi kecurangan. (Zhu et al., 2024) menunjukkan bahwa nada dan intensitas bahasa dalam laporan risiko berkaitan dengan tingkat ketidakpastian pasar dan volatilitas saham. Temuan ini menjadi dasar bagi penelitian ini untuk membandingkan SVM dan IndoBERT dalam konteks laporan keuangan Indonesia. Melalui analisis teks laporan tahunan, NLP diharapkan dapat membantu meningkatkan transparansi informasi dan memberikan gambaran risiko yang lebih jelas bagi investor.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, penelitian ini berfokus pada pemanfaatan teknologi NLP dan *Machine Learning* untuk menganalisis risiko dalam laporan keuangan perusahaan publik di Indonesia. Laporan keuangan umumnya memuat narasi yang cukup kompleks mengenai kondisi perusahaan, kebijakan manajemen, serta berbagai faktor risiko yang dapat memengaruhi kinerja usaha. Banyaknya teks dan penggunaan bahasa yang formal serta panjang membuat proses analisis secara manual menjadi kurang efisien dan berpotensi menimbulkan subjektivitas. Oleh karena itu, diperlukan pendekatan berbasis algoritma pembelajaran mesin untuk membantu mengidentifikasi dan mengklasifikasikan informasi risiko secara lebih sistematis dan akurat.

Dalam penelitian ini digunakan dua algoritma NLP, yaitu SVM dan IndoBERT. SVM mewakili pendekatan machine learning klasik yang dikenal stabil dan efisien, sedangkan IndoBERT merupakan model deep learning berbasis transformer yang mampu memahami konteks bahasa secara lebih mendalam. Melalui penggunaan kedua metode tersebut, penelitian ini bertujuan untuk membandingkan kinerja keduanya dalam mengenali pola risiko yang terdapat pada teks laporan keuangan perusahaan publik di Indonesia.

Berdasarkan uraian tersebut, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan metode NLP dalam mengidentifikasi dan mengklasifikasikan informasi risiko pada laporan keuangan perusahaan publik Indonesia?
2. Bagaimana perbandingan kinerja algoritma SVM dan IndoBERT dalam menganalisis teks laporan keuangan perusahaan publik Indonesia?
3. Sejauh mana hasil perbandingan kedua algoritma tersebut dapat menunjukkan efektivitas model NLP dalam mendeteksi potensi risiko yang termuat dalam teks laporan keuangan perusahaan publik?

1.3. Batasan Masalah

Agar penelitian ini dapat berjalan secara terarah, fokus, dan sesuai dengan tujuan yang telah ditetapkan, maka diperlukan batasan ruang lingkup penelitian.

Penetapan batasan ini bertujuan untuk memperjelas arah analisis serta menghindari pembahasan yang terlalu luas atau keluar dari konteks penelitian.

Penelitian ini difokuskan pada analisis teks laporan keuangan perusahaan publik Indonesia yang tergabung dalam indeks LQ45. Pemilihan LQ45 didasarkan pada reputasi dan stabilitas perusahaan-perusahaan yang termasuk di dalamnya, serta ketersediaan laporan keuangan yang lengkap dan mudah diakses melalui situs resmi Bursa Efek Indonesia (BEI). Jenis laporan keuangan yang digunakan mencakup laporan tahunan (annual report) yang dipublikasikan dalam kurun waktu 2024.

Data yang dianalisis hanya terbatas pada bagian naratif laporan keuangan, khususnya pada segmen *Management's Discussion and Analysis* (MD&A) dan Faktor Risiko. Bagian tersebut dianggap paling relevan dalam menggambarkan pengungkapan risiko yang bersifat kualitatif. Elemen numerik seperti rasio keuangan, laba rugi, atau arus kas tidak termasuk dalam cakupan analisis.

Dari sisi metode, penelitian ini menggunakan pendekatan Natural Language Processing (NLP) dengan dua algoritma utama, yaitu SVM sebagai metode klasik dan IndoBERT sebagai model deep learning berbasis transformer yang dilatih untuk Bahasa Indonesia. Proses analisis mencakup tahap praproses teks, pelatihan model, dan evaluasi kinerja.

Hasil penelitian ini dibatasi pada perbandingan performa kedua algoritma dalam klasifikasi informasi risiko pada teks laporan keuangan, tanpa membahas faktor eksternal seperti kondisi makroekonomi, kebijakan pemerintah, atau variabel non-teknis lainnya. Dengan demikian, penelitian ini berfokus pada analisis berbasis teks untuk mendeteksi dan mengklasifikasikan risiko yang terungkap dalam laporan keuangan perusahaan publik Indonesia.

1.4. Tujuan Penelitian

Tujuan penelitian ini disusun untuk menjawab rumusan masalah yang telah dijelaskan sebelumnya dan memberikan arah yang jelas terhadap pelaksanaan penelitian. Penelitian ini bertujuan untuk menerapkan metode Natural Language Processing (NLP) dalam menganalisis narasi laporan keuangan perusahaan publik Indonesia, khususnya dalam mendeteksi dan mengklasifikasikan informasi yang berkaitan dengan risiko.

Secara lebih rinci, tujuan penelitian ini adalah sebagai berikut:

1. Menganalisis penerapan metode NLP dalam proses identifikasi dan klasifikasi informasi risiko yang terdapat pada teks laporan keuangan perusahaan publik Indonesia. Tujuan ini mencakup tahapan praproses teks, ekstraksi fitur, serta pembentukan model klasifikasi berbasis pembelajaran mesin untuk memahami struktur dan makna dari narasi laporan keuangan.
2. Membandingkan kinerja dua algoritma Machine Learning, yaitu SVM sebagai metode klasik dan IndoBERT sebagai model deep learning berbasis transformer yang dilatih untuk Bahasa Indonesia sebagai metode berbasis *deep learning*, dalam menganalisis teks laporan keuangan perusahaan publik Indonesia. Perbandingan dilakukan berdasarkan metrik evaluasi seperti akurasi, presisi, recall, dan *F1-score* guna mengetahui algoritma yang paling efektif untuk klasifikasi informasi risiko.
3. Menilai efektivitas model NLP dalam mendeteksi potensi risiko yang terdapat pada teks laporan keuangan perusahaan publik, sehingga dapat memberikan kontribusi dalam mendukung pemahaman risiko investasi, khususnya bagi investor pemula. Melalui hasil penelitian ini, diharapkan diperoleh model analisis berbasis teks yang dapat digunakan sebagai alat bantu dalam pengambilan keputusan investasi secara lebih objektif dan berbasis data.

1.5. Kontribusi Penelitian

Penelitian ini diharapkan memberikan kontribusi yang signifikan, baik dari sisi akademis maupun praktis, dalam pengembangan ilmu pengetahuan di bidang analisis teks laporan keuangan menggunakan NLP dan *Machine Learning*. Secara akademis, penelitian ini memperkaya literatur mengenai penerapan algoritma SVM dan IndoBERT dalam konteks analisis risiko keuangan di Indonesia. Selain itu, penelitian ini memberikan pemahaman baru tentang efektivitas model klasik dan model berbasis *deep learning* dalam mengklasifikasikan narasi risiko dari laporan keuangan perusahaan publik.

Dari sisi metodologis, penelitian ini berkontribusi dalam pengembangan model analisis berbasis NLP untuk bahasa Indonesia, yang selama ini masih didominasi oleh penelitian berbahasa Inggris. Model yang dibangun dapat menjadi acuan bagi penelitian selanjutnya yang berfokus pada pemrosesan teks keuangan, baik untuk deteksi risiko, analisis sentimen, maupun pengungkapan informasi non-keuangan. Penelitian ini juga dapat menjadi dasar dalam pengembangan sistem pendukung

keputusan (*Decision Support System*) di bidang keuangan berbasis analisis teks otomatis.

Dari sisi praktis, penelitian ini memberikan manfaat bagi investor pemula dalam memahami risiko investasi melalui analisis otomatis terhadap laporan keuangan perusahaan publik. Hasil penelitian diharapkan mampu membantu pengguna non-teknis untuk memperoleh gambaran risiko dengan cara yang lebih mudah, cepat, dan objektif. Selain itu, penelitian ini juga dapat menjadi rujukan bagi perusahaan dan regulator dalam meningkatkan transparansi pengungkapan risiko di laporan keuangan, sehingga kualitas informasi keuangan di pasar modal Indonesia semakin baik dan kredibel.

