

# BAB I

## PENDAHULUAN

### 1.1. Latar belakang

Pasar saham merupakan sebuah tempat bertemunya antara pembeli dengan penjual suatu perusahaan yang telah mendaftarkan sahamnya di Bursa Efek Indonesia yang nantinya akan disebut emiten. Kinerja atau keadaan yang sedang terjadi di pasar saham diukur dengan menggunakan Indeks Harga Saham Gabungan (IHSG). Sedangkan saham itu sendiri adalah sebuah bukti kepemilikan suatu perusahaan. Investasi saham telah menjadi pilihan investasi yang lumrah dilakukan oleh masyarakat khususnya di Indonesia karena dengan melakukan investasi di instrumen ini, investor dapat memperoleh keuntungan yang cukup besar dalam waktu yang cenderung singkat [1].

Pandemi covid-19 telah menyebabkan berbagai macam dampak buruk bagi Indonesia, salah satunya adalah pasar saham. Hal ini ditandai dengan merosotnya nilai IHSG. Menurut laporan Badan Pusat Statistik (2021) di tahun 2020, nilai IHSG berada di angka 4.539 pada Q1 (Maret), 4.095 pada Q2 (Juni), 4.878 pada Q3 (September), dan 5.979 pada Q4 (Desember)[2]. Hal ini berarti nilai IHSG terkoreksi atau turun sebesar 5,1 persen bila dibandingkan dengan tahun sebelumnya. Namun, turunnya IHSG ini justru menjadi peluang bagi para investor saham untuk menginvestasikan pendapatannya di produk-produk pasar saham, salah satunya saham. Berdasarkan data dari Kustodian Sentral Efek Indonesia (2022), jumlah *Single Investor Identification* (SID) mencapai 7,5 juta yang berarti meningkat 92,99 persen sejak tahun 2019 [3]. Hal ini tidak luput dari kemajuan teknologi yang mempermudah akses transaksi dalam pembelian dan penjualan emiten suatu perusahaan [4].

Berinvestasi di saham memiliki konsep “*high risk, high return*”, yang berarti investor memiliki kesempatan untuk mendapatkan nilai keuntungan yang cukup besar, akan tetapi diikuti dengan risiko yang cukup besar pula. Salah satu risiko yang ada adalah fluktuasi harga dari saham suatu perusahaan. Fluktuasi harga saham dapat disebabkan oleh kondisi pasar yang dipengaruhi oleh kondisi ekonomi, politik, ataupun sentimen dari para investor. Oleh karena itu, untuk memantau hal

tersebut, setiap investor memerlukan sumber informasi sebagai analisis tahap pertama mereka. Hal ini karena informasi merupakan kebutuhan dasar dari seorang calon investor untuk menentukan pilihan investasi. Semakin banyak informasi yang dikumpulkan maka sebuah peluang investasi akan lebih meyakinkan bagi calon investor untuk menanamkan modalnya di sebuah perusahaan [5]. Contoh dari sumber informasi tersebut adalah berita ekonomi dan bisnis dan pengumuman perusahaan.

Menganalisis informasi saham melalui berita membutuhkan waktu yang signifikan, terutama karena volume dan kompleksitas informasi yang tersebar di berbagai sumber berita. Dalam konteks ini, para analis harus menilai dan mengekstrak informasi yang relevan dari berbagai artikel berita untuk mendapatkan wawasan yang berguna dalam pengambilan keputusan investasi. Proses ini menjadi semakin sulit ketika jumlah berita yang perlu dianalisis terus berkembang, yang mengarah pada kebutuhan untuk mempercepat dan meningkatkan efisiensi proses analisis. Oleh karena itu, diperlukan metode yang efektif untuk menangani dan menganalisis informasi tersebut secara otomatis, sehingga analisis dapat dilakukan dengan lebih cepat dan akurat.

Salah satu pendekatan yang efisien untuk mengatasi tantangan ini adalah dengan mengidentifikasi ide pokok atau tema utama dari setiap berita yang berkaitan dengan saham tertentu. Teknik pemodelan topik (*topic modeling*) diterapkan untuk tujuan ini, di mana metode ini dapat secara otomatis mengekstraksi tema atau topik yang terkandung dalam sekumpulan teks berita. Dengan menggunakan pemodelan topik, kita dapat mengidentifikasi pola kata-kata yang sering muncul bersama dalam konteks yang sama, yang membantu untuk memahami topik utama yang dibahas dalam setiap artikel. Pendekatan ini memungkinkan analisis yang lebih cepat dan efisien karena tidak memerlukan pembacaan manual seluruh teks berita, melainkan mengandalkan algoritma untuk mengenali hubungan antar kata dan konsep yang relevan. Selain itu, pemodelan topik dapat memperkaya wawasan analisis saham dengan mengelompokkan berita-berita yang memiliki tema serupa, sehingga memudahkan dalam pemantauan perkembangan pasar dan pengambilan keputusan berbasis informasi yang lebih terstruktur. Dengan demikian, teknik pemodelan topik tidak hanya mempercepat

proses analisis, tetapi juga meningkatkan akurasi dan efektivitas dalam memahami dinamika pasar saham.

Pemodelan topik (*topic modeling*) adalah salah satu *natural language processing* (NLP) yang digunakan untuk mengidentifikasi tema utama dalam kumpulan dokumen. Pemodelan topik biasa digunakan untuk mendapatkan topik dari sekumpulan dokumen, dimana dokumen ini berisikan teks [6]. Cara kerja pemodelan topik secara singkat melibatkan pemrosesan teks untuk menemukan topik-topik yang sering muncul bersama dalam dokumen. Terdapat berbagai macam metode dapat digunakan untuk melakukan pemodelan topik, yaitu *Latent Dirichlet Allocation* (LDA), *Latent Semantic Analysis* (LSA), *Probabilistic Latent Semantic Analysis* (PLSA), *Non-Negative Matrix Factorization* (NMF), dan *Hierarchical Dirichlet Process* (HDP).

Berdasarkan penelitian yang dilakukan dalam jurnal yang berjudul “*Topic Modeling for Evolving Textual Data Using LDA, HDP, NMF, BERTopic, and DTM With a Focus on Research Papers*” menunjukkan bahwa *latent dirichlet allocation* (LDA) lebih unggul daripada metode *hiearchial dirichlet process* (HDP) dan *negative matrix factorization* (NMF). Metode LDA unggul dengan *score precision* sebesar 0.46254 dibandingkan dengan HDP dan NMF yang meraih *score precision* sebesar 0.45254 dan 0.45651 dengan data berjumlah 15 topik. Penulis juga menambahkan LDA merupakan model probabilitas yang sering digunakan, berhasil dalam mengidentifikasi topik dengan tingkat koherensi yang moderat, dengan catatan memperhatikan secara seksama jumlah topik yang berperan penting untuk mencegah terjadinya *overfitting* [7].

Penelitian lain yang dilakukan dalam jurnal yang berjudul “Perbandingan metode *Non-Negative Matrix Factorization* dan *Latent Dirichlet Allocation* pada permasalahan *question retrieval* untuk studi kasus forum kesehatan” menunjukkan hasil *mean average precision* (MAP) LDA untuk topik non-medis sebesar 0.588 sedangkan MAP NMF untuk topik non-medis sebesar 0.584. Disamping MAP, terdapat pula *mean reciprocal rank* (MRR) LDA untuk topik non-medis sebesar 0.746 sedangkan MRR NMF untuk topik non-medis sebesar 0.747. Dari hal ini, dapat disimpulkan bahwa LDA merupakan metode yang lebih unggul dibandingkan

NMF untuk melakukan *topic modeling* meskipun selisih MAP dan MRR diantara keduanya tidak signifikan [8].

Berdasarkan hal yang telah disampaikan, penelitian ini menawarkan solusi untuk mencari melakukan analisis pemodelan topik (*topic modeling*) dengan menggunakan metode *latent dirichlet allocation* (LDA) untuk berita saham berindeks LQ45. Sumber data pada penelitian ini didapatkan dari website kontan.id dengan kata kunci "LQ45". Harapan dari penelitian ini adalah dapat mempermudah pengguna dalam melakukan riset saham-saham yang terindeks di LQ45 dengan hasil analisis topik-topik dari berita indeks saham terkait.

### **1.2. Rumusan Masalah**

Berdasarkan latar belakang yang telah disampaikan pada bagian sebelumnya, maka rumusan masalah yang menjadi fokus pada penelitian ini adalah:

- 1) Bagaimana proses implementasi pemodelan topik (*topic modeling*) dengan metode *latent dirichlet allocation* (LDA) dalam mencari topik untuk berita saham?
- 2) Bagaimana hasil dari pemodelan topik dari sekumpulan berita mengenai indeks saham LQ45?
- 3) Bagaimana performa *latent dirichlet allocation* (LDA) dalam melakukan pemodelan topik bila dibandingkan dengan metode lain untuk dataset ukuran sedang-besar seperti berita saham?

### **1.3. Batasan masalah**

Dari permasalahan yang telah disebutkan, berikut adalah hal-hal yang menjadi batasan masalah dalam penelitian ini:

- 1) Topik yang menjadi sumber data pada penelitian ini berfokus pada saham berindeks LQ45.
- 2) Sumber data didapatkan dengan cara *web-scraping* dari portal berita kontan.id.
- 3) Pemodelan topik digunakan dengan metode *latent dirichlet allocation* (LDA).
- 4) Menggunakan metode *non-negative matrix factorization* (NMF) dan *hierarchical dirichlet process* (HDP) sebagai metode pembanding.

- 5) Evaluasi masing-masing metode dilakukan dengan menggunakan *coherence score*.

#### **1.4. Tujuan Penelitian**

Berdasarkan latar belakang diatas, berikut merupakan hal-hal yang menjadi tujuan dari penelitian ini:

- 1) Mengaplikasikan pemodelan topik dengan menggunakan LDA pada sekumpulan berita.
- 2) Mengetahui topik yang dihasilkan dari pemodelan topik dengan menggunakan LDA.
- 3) Mengukur kinerja dan performa LDA dalam melakukan pemodelan topik.
- 4) Memberi perbandingan kepada metode LDA terhadap metode pemodelan topik lain.

#### **1.5. Manfaat Penelitian**

Adapun manfaat yang ingin dicapai dari penelitian ini, yaitu:

- 1) Bagi penulis
  - a. Mengimplementasikan ilmu-ilmu yang dipelajari selama masa studi, khususnya di bidang sistem cerdas rekayasa.
  - b. Sebagai salah satu persyaratan untuk memenuhi gelas sarjana komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, UPN “Veteran” Jawa Timur.
- 2) Bagi pembaca
  - a. Memberikan informasi mengenai penerapan pemodelan topik dengan menggunakan *latent dirichlet allocation* (LDA) terhadap sekumpulan berita.
  - b. Mengetahui kinerja dan performa *latent dirichlet allocation* (LDA) dalam melakukan pemodelan topik,

*Halaman ini sengaja dikosongi*