

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Kemajuan teknologi kini merevolusi bagaimana manusia sekarang saling berinteraksi satu sama lain, aplikasi komunikasi jarak jauh (*online*) seperti whatsApp, instagram, twitter, facebook dan lain-lain kini sudah menjadi bagian dari keseharian masyarakat Indonesia. Meskipun perkembangan teknologi dan internet menjadi semakin pesat dan penting untuk kehidupan sekarang ini, namun teknologi juga membawa ancaman buruk yang lebih tinggi lagi perihal keamanan dan privasi setiap penggunanya. Menjaga keamanan data merupakan tindakan yang perlu dilakukan oleh suatu individu atau perusahaan untuk tetap menjaga atau melindungi ekosistem teknologi informasi dari kerusakan data, modifikasi data maupun penyebaran data yang dilakukan sengaja ataupun tidak sengaja. Kebocoran data adalah masalah yang semakin meningkat dalam era digital saat ini, kebocoran data terjadi ketika data yang seharusnya bersifat pribadi dan rahasia diungkapkan atau disebarkan ke pihak yang tidak berwenang. Hal ini dapat berdampak negatif pada individu atau perusahaan, termasuk hilangnya kepercayaan dari pelanggan dan merugikan finansial.

Pada 12 Maret 2023, terjadi insiden pembobolan data yang menimpa salah satu Badan Hukum Milik Negara (BHMN) yaitu BPJS Ketenagakerjaan. Seorang pengguna dengan nama @bjorkanesian berhasil membobol data BPJS Ketenagakerjaan sebanyak

19 juta data. Data yang dibobol mencakup NIK, nama lengkap, tanggal lahir, alamat, nomor ponsel, alamat email, jenis pekerjaan dan nama perusahaan. Tidak hanya itu, Bjorka juga memperjualbelikan data tersebut pada forum *hacker* “*Breached Forums*” [1]. Semua data yang dibobol diperjualbelikan seharga USD 10.000 atau kisaran Rp 154 juta dalam bitcoin. Pakar keamanan *cyber* *Chairman Communication & Information System Security Research Center* (CISSReC) Pratama Persadha, menyatakan bahwa data yang tersebar di forum gelap tersebut valid. Sebelum kasus ini terjadi, Bjorka juga sempat membobol data pendaftaran kartu SIM (*SIM Card*) sebanyak 1.3 miliar data pada bulan September 2022 lalu, data yang dibobol mencakup Nomor Induk Penduduk (KTP), nomor telepon, tanggal registrasi, dan informasi provider dari pelanggan. Dengan munculnya kembali Bjorka, menyebabkan banyak lapisan masyarakat mulai kehilangan kepercayaan kepada pemerintah. Bahkan, banyak dari masyarakat yang beropini bahwa munculnya kembali Bjorka adalah sebuah pengalihan isu semata saja. Dalam insiden ini pemerintah juga harus mempertimbangkan opini masyarakat dalam melakukan perbaikan untuk mendapatkan kepercayaan masyarakat kembali.

Bjorka, pertama kali muncul di Indonesia pada April 2020 ketika ia membocorkan 11 GB data pengguna *market place* Tokopedia. Kemudian pada tanggal 20 Agustus 2022, ia membocorkan 270 juta data pengguna Wattpad dan 26 juta data pelanggan Indihome. Pada tanggal 31 Agustus 2022, ia membocorkan 1.3 miliar data registrasi *SIM Card*, dan pada tanggal 6 September 2022, ia membocorkan 105 juta

data Komisi Perlindungan Umum. Pada tanggal 9 September 2022, ia membocorkan dokumen kerahasiaan kepresidenan periode 2019-2024 dan surat tertutup Badan Intelijen Negara milik Presiden Republik Indonesia, Joko Widodo. Tidak hanya itu, pada tanggal 10 September 2022, Bjorka kembali membocorkan data Menteri Koinfo, Jhonny G. Plate, dan data Ketua DPR RI, Puan Maharani. Pada tanggal 11 September 2020, ia membocorkan data Menkomarves RI, Luhut Binsar Pandjaitan, dan data Gubernur DKI Jakarta, Anies Baswedan. Terakhir, Bjorka membocorkan data Menteri Polhukam, Mahfud MD, dan data Ketua PKB, Muhaimin Iskanda [1]. Kemunculan *hacker* Bjorka ini menuai berbagai banyak tanggapan dari masyarakat Indonesia, jumlah opini yang sangat banyak dan terus-menerus meningkat menjadi permasalahan untuk mengetahui sentimen masyarakat sehingga diperlukan analisis terhadap opini masyarakat. Salah satu solusi yang dapat digunakan dalam mengatasi permasalahan tersebut adalah *text mining* [2].

Sebelumnya penelitian terhadap Bjorka sudah pernah dilakukan oleh Adhitya Karel Maulaya & Junadhi (2022) dengan judul penelitian Analisis sentimen menggunakan *Support Vector Machine* masyarakat indonesia di Twitter terkait Bjorka menghasilkan *accuracy* sebesar 62.33%. Pada penelitian tersebut dikatakan bahwa SVM sulit untuk menyelesaikan himpunan data berskala besar dan sulit mengklasifikasikan terhadap dua kelas saja [3].

Penelitian lain yang dilakukan oleh Ridho Sholehurrohman dan Igit Sabda Ilman (2022) melakukan analisis sentimen terhadap kasus kebocoran data pengguna

*Facebook* oleh Cambridge Analytica. Penelitian ini menggunakan metode *Naïve Bayes* yang menghasilkan akurasi sebesar 83.06% dengan nilai rata-rata *error precision* 0.873, *error recall* 0.851 dan *error accuracy* 1.693 [4].

Beberapa penelitian sebelumnya menjadi riset yang akan dilakukan peneliti terkait topik sentimen analisis yang peneliti ambil. Berdasarkan penelitian terdahulu yang peneliti paparkan di atas, maka penelitian yang peneliti ambil akan merujuk pada penelitian [3]. Peneliti terdahulu menggunakan algoritma *Support Vector Machine*, mereka menyatakan bahwa algoritma *Support Vector Machine* memiliki kelemahan sulit untuk untuk menyelesaikan himpunan data berskala besar dan sulit mengklasifikasikan data terhadap dua kelas saja.

Berangkat dari penelitian inilah, maka peneliti akan melakukan penelitian serupa, namun yang membedakan, disini peneliti akan melakukan perbandingan algoritma menggunakan metode *Naïve Bayes* dan C4.5. Maka peneliti bermaksud untuk membuat penelitian baru yang berjudul **“Analisis Sentimen Masyarakat terhadap Hacker Bjorka dalam Kasus Kebocoran Data BPJS Ketenagakerjaan Menggunakan Algoritma *Naïve Bayes* dan C4.5”**.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan pada latar belakang, dapat disimpulkan identifikasi masalahnya yaitu:

1. Kebocoran data yang dialami BPJS Ketenagakerjaan membuat banyak lapisan masyarakat Indonesia resah dan masyarakat Indonesia mulai kehilangan kepercayaan kepada pemerintah.
2. Jumlah opini masyarakat yang sangat banyak dan terus-menerus meningkat sehingga diperlukan sistem analisis terhadap sentimen masyarakat.
3. Belum adanya teknik pengklasifikasian data untuk sentimen Bjorka pada kasus kebocoran data BPJS Ketenagakerjaan pada data *Facebook*.

### 1.3 Rumusan Masalah

Berdasarkan masalah yang sudah dijelaskan pada latar belakang, maka dapat disimpulkan permasalahan yang akan menjadi kajian penelitian ini adalah bagaimana cara mengimplementasikan algoritma *Naïve Bayes* dan C4.5 untuk analisis sentimen opini masyarakat terhadap *hacker* Bjorka dalam kasus kebocoran data BPJS Ketenagakerjaan pada data *Facebook*?

### 1.4 Batasan Masalah

Berdasarkan perumusan masalah di atas dan agar terhindar dari permasalahan yang menyimpang, maka dapat disimpulkan batasan masalah pada penelitian ini adalah sebagai berikut.

1. Klasifikasi sentimen masyarakat terhadap kasus Bjorka menggunakan data yang diambil dari postingan-postingan *Facebook*.

2. *Range* data sentimen masyarakat yang digunakan pada bulan Maret – April 2023, dikarenakan pada bulan Maret - April berita Bjorka melakukan aksi pembobolan data BPJS Ketenagakerjaan.

### 1.5 Tujuan Penelitian

Berdasarkan latar belakang yang sudah penulis jelaskan maka tujuan dari penelitian ini adalah sebagai berikut.

1. Untuk melihat seberapa baik performa algoritma *Naïve Bayes* dan C4.5 dalam pengklasifikasian sentimen analisis terhadap Bjorka dalam kasus kebocoran data BPJS Ketenagakerjaan.
2. Untuk mengetahui tanggapan masyarakat apakah lebih kearah positif atau negatif terhadap kejadian kebocoran data yang dialami BPJS Ketenagakerjaan yang dilakukan oleh *hacker* Bjorka dengan melakukan sentimen analisis menggunakan algoritma *Naïve Bayes* dan C4.5.

### 1.6 Manfaat Penelitian

Berdasarkan analisis sentimen yang dilakukan, manfaat dari penelitian yang peneliti lakukan adalah sebagai berikut.

1. Mahasiswa dapat menerapkan ilmu-ilmu yang didapat selama penulis melakukan perkuliahan.
2. Bertambahnya wawasan dan pengalaman penulis dalam melakukan sentimen analisis.



## 1.7 Sistematika Penulisan

Sistematika penulisan merupakan uraian tentang susunan dari penulisan yang dibuat secara teratur dan terperinci sehingga dapat memberikan gambaran secara menyeluruh. Adapun sistematika penulisan terdiri dari lima bab, yaitu sebagai berikut.

### **BAB I PENDAHULUAN**

Pada bab ini peneliti menjelaskan tentang latar belakang dari judul penelitian yang penulis buat. Bab ini terdiri dari latar belakang, identifikasi masalah, tujuan dan manfaat, metode penelitian dan sistematika penulisan.

### **BAB II LANDASAN TEORI**

Pada bab ini peneliti menjelaskan tentang teori-teori yang berhubungan dengan pokok pembahasan dalam penelitian dan secara garis besar berisi tentang analisis sentimen.

### **BAB III METODOLOGI PENELITIAN**

Pada bab ini peneliti membahas penjelasan secara bertahap dan terperinci tentang langkah-langkah yang digunakan untuk membuat kerangka berfikir dan kerangka kerja dalam menyelesaikan penelitian ini.

### **BAB IV HASIL DAN PEMBAHASAN**

Pada bab ini peneliti membahas hasil *data selection*, *preprocessing*, *transformation*, *data mining*, *evaluation* dan *knowledge presentation*.

## **BAB V PENUTUP**

Pada bab ini berisi kesimpulan dan saran dari penelitian yang telah dilakukan.

