**BAB I**

**PENDAHULUAN**

* 1. **Latar Belakang**

Berita saat ini sudah dapat diakses melalui berbagai macam platform, salah satunya yang umum yaitu melalui situs berita di internet. Kelebihan dari berita dalam situs di internet ini adalah beritanya yang selalu diperbaharui secara berkelanjutan (continuous updates), memiliki interaktivitas, hypertext, dan multimedia. Situs penyedia berita pun sudah banyak, beberapa diantaranya kompas.com, tribunnews.com, detik.com. Umunya berita-berita yang terdapat pada media cetak maupun media online seperti situs berita biasanya terdiri dari berbagai macam kategori, seperti berita olahraga, politik, sosial, ekonomi, kesehatan dan sebagainya. Semakin banyak nya volume berita yang dihasilkan dan semakin banyak nya konten berita yang ada maka semakin sulit juga kita dalam memilah berita tersebut.

*Corona Virus Disease 2019* atau biasa kita sebut COVID-19 adalah penyakit jenis baru yang belum pernah diidentifikasi sebelumnya pada manusia. Tanda dan gelaja umum infeksi COVID-19 antara lain gejala gangguan pernapasan akut seperti demam, batuk, dan sesak napas. Masa inkubasi rata-rata 5- 6 hari dengan masa inkubasi terpanjang 14 hari. Pada tanggal 30 Januari 2020 WHO telah menetapkan sebagai kedaruratan kesehatan masyarakat yang meresahkan dunia. Pada tanggal 2 Maret 2020, Indonesia melaporkan kasus konfirmasi COVID-19 sebanyak 2 kasus. Sampai dengan tanggal 16 Maret 2020 ada 10 orang yang dinyatakan positif corona. (Yurianto, Ahmad, Bambang Wibowo, 2020).

Dengan adanya virus COVID-19 di Indonesia saat ini berdampak bagi seluruh masyarakat. Menurut kompas, 28/03/2020 dampak virus COVID-19 terjadi diberbagai bidang seperti sosial, ekonomi, politik, pariwisata dan pendidikan. Banyaknya jumlah platform berita online yang memuat tentang berita COVID-19, seringkali membuat masyarakat sulit untuk menemukan informasi yang relevan dengan waktu yang singkat. Masyarakat tidak mungkin mengunjungi satu per satu situs untuk mendapatkan informasi yang diinginkan. Dari masalah ini akan dibuat suatu sistem yang dapat mengumpulkan semua data dari situs-situs berita resmi dengan satu web service dan pengambilan data dari website menggunakan teknik scrapping untuk bahan klasifikasinya.

Dari permasalahan tersebut, solusi yang akan dilakukan yaitu pembuatan aplikasi *web service* dengan melakukan akuisisi atau penggabungan tiga situs berita online menggunakan teknik *crawling* untuk pengambilan datanya, yang kemudian dilakukan kategorisasi berita COVID-19 pada masing-masing bidang seperti sosial, ekonomi, politik, pariwisata dan pendidikan.

Beberapa penelitian sebelumnya dalam melakukan klasifikasi, umumnya menggunakan metode Naïve Bayes, K-Nearest Neighbor, serta Support Vector. Naïve Bayes digunakan untuk proses klasifikasi karena memiliki beberapa kelebihan antara lain, caranya yang sederhana, cepat tetapi memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan metode yang lain. Contoh penelitan yang pernah dilakukan menggunakan metode Naïve Bayes oleh S. Hassan, M. Rafi and M. S. Shaikh, (2011), yang berjudul “*Comparing SVM And Naïve Bayes Classifiers For Text Categorization With Wikitology As Knowledge Enrichment*”. Penelitian tesebut membandingkan *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC) dibawah *text enrichment* pada *Wikitology.* Memberikan hasil peningkatan +28,78% pada metode *Naïve Bayes Classifier* (NBC) dengan menggunakan *10-fold cross validation*, sedangkan metode *Support Vector Machine* (SVM) hanya memberikan peningkatan sebesar 6,36% jika dibandingkan dengan hasil aslinya. Menunjukan metode klasifikasi *Naïve Bayes* adalah pilihan yang lebih baik ketika *external enriching* digunakan pada *external knowledge base.*

Metode Naive Bayes yang akan digunakan ini nantinya akan mencari nilai probabilitas dari teks dan kalimat dalam konten berita yang terdapat pada setiap kategori berita dan probabilitas dari masing – masing kategori berita. Berdasarkan hal tersebut, maka penulis akan melakukan penelitian dengan judul “Klasifikasi Berita COVID-19 Secara Otomatis Pada Tiga Portal Berita Online Menggunakan Metode *Naïve Bayes Classification*”.

* 1. **Rumusan Masalah**

Berdasarkan latar belakang tersebut, maka perumusan masalah yang didapat adalah bagaimana metode *Naïve Bayes Classifier* (NBC) dapat melakukan klasifikasikan berita COVID-19 berdasarkan kontennya ke dalam sub topik kategori berita yang ada.

* 1. **Batasan Masalah**

Batasan masalah pada penelitian ini adalah:

1. Informasi yang di ambil hanya COVID-19 dari tiga situs berita Indonesia.

2. Berita yang digunakan dalam penelitian ini hanya berita berbahasa Indonesia.

3. Situs berita yang akan di *crawling* datanya yaitu Kompas, Tribunnews, dan Detik.

4. Proses *crawling* membutuhkan koneksi Internet.

5. Melakukan *crawling* data dari situs berita tanpa mengedit data.

6. Ketersedian data yang di cari dan ditampilkan tergantung dari situs berita.

7. Pencarian berita COVID-19 berdasarkan bidang sosial, ekonomi, politik, pariwisata dan pendidikan.

8. Kategorisasi dilakukan dengan metode *Naïve Bayes*.

* 1. **Tujuan Tugas Akhir**

Berdasarkan permasalahan yang dijelaskan pada latar belakang, maka tujuan penelitian ini adalah mengkelompokkan berita COVID-19 berdasarkan bidang sosial, ekonomi, politik, pariwisata dan pendidikan menggunakan metode *Naïve bayes* dalam melakukan kategorisasinya disitus berita Indonesia.

* 1. **Manfaat Tugas Akhir**

Manfaat yang diharapkan dari penelitian ini adalah:

1. Memberikan kemudahan masyarakat untuk mencari berita terkait COVID-19.

2. Mendapatkan informasi terkait COVID-19 dari tiga situs berita sekaligus.

2. Dapat dijadikan sebagai referensi dalam pengembangan aplikasi yang serupa.

* 1. **Metodologi Penelitian dan Pengembangan Sistem**

Metode yang digunakan pada penelitian ini dibagi dua yaitu metode pengumpulan

data dan metode pengembangan sistem. Adapun penjelasan dari kedua metode tersebut

yaitu :

1. Metodologi Pengumpulan Data

Pengumpulan data yang digunakan dalam penelitian ini yaitu observasi, kuisioner, dan studi pustaka. Observasi yang dilakukan yaitu menganalisis tingkah laku seseorang atau kelompok dalam membeli sepatu. Selanjutnya kuisioner, dengan membagikan kuisioner. Penyebaran kuisioner dilakukan secara online melalui media social whatshapp dan line dengan bantuan google form dalam pembuatan formulir kuisioner dan perhitungan hasil kuisioner. Kuisioner ini dilakukan untuk mendukung dalam pengambilan data yang akan diolah menjadi informasi. Selanjutnya adalah studi pustaka yaitu mencari bahan pendukung dan pendefinisian masalah melalui buku-buku maupun jurnal yang berkaitan dengan objek. Penelitian ini menggunakan data crawling langsung dari tiga situs e-commerce tanpa diubah ataupun diedit.

1. Metodologi Pengembangan Sistem

Metode pengembangan sistem yang digunakan pada penelitian ini adalah metode *Prototyping* yang berarti dalam pengembangannya dilakukan dengan cara interaksi antara pengembang dan pengguna, dengan tahapan sebagai berikut (Pressman, 2010):

1. *Communication*
2. *Quick plan and modeling quick design*
3. *Construction of prototype*
4. *Deployment delivery and feedback*
   1. **Sistematika Penulisan**

Adapun sistematika penulisan penelitian ini adalah sebagai berikut:

**BAB I PENDAHULUAN**

Dalam bab ini dijelaskan tentang Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, Metodologi Penelitian, serta Sistematika Penulisan dari tugas akhir ini.

**BAB II TINJAUAN PUSTAKA**

Bab ini menjelaskan teori-teori yang menadi landasan dalam proses penelitian. Tinjauan pustaka yang akan disajikan meliputi *web crawling, web service,* PHP, *naïve bayes, text mining.*

**BAB III METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM**

Bab ini membahas tentang tahap perancangan kebutuhan, tahap analisis, serta menampilkan gambaran garis besar struktur program serta pengambilan data seperti metodologi riset, arsitektur, rekayasa,pemodelan sistem, serta sistem pengujian.

**BAB IV HASIL, PENGUJIAN DAN PEMBAHASAN**

Bab ini menjelaskan tentang aplikasi yang dibuat, serta implementasi dari teori- teori yang digunakan dalam pembangunan aplikasi.

**BAB V PENUTUP**

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilaksanakan serta saran- saran terkait aplikasi yang telah dibuat untuk dikembangkan menjadi sistem yang lebih baik.