

BAB I PENDAHULUAN

1.1 Latar Belakang

Corona Virus Disease 2019 atau biasa disingkat COVID-19 merupakan penyakit menular yang disebabkan oleh SARS-CoV-2. Infeksi dari COVID-19 menyebar melalui percikan dari batuk atau bersin dan ditandai dengan demam, batuk kering, dan kesulitan bernafas. Munculnya COVID-19 juga menyebabkan penggunaan jejaring sosial meningkat pesat. Salah satu jejaring sosial yang meningkat yaitu X. X adalah layanan jejaring sosial yang dapat digunakan untuk berinteraksi, promosi, maupun mencari berita. Interaksi yang terjadi pada media X dapat melalui *Direct Message* (DM) maupun melalui *Tweet* atau kicauan. *Tweet* atau kicauan pengguna dapat dilihat maupun diikuti oleh pengguna lain.

Penelitian ini dilakukan dengan menggunakan *Apache Spark*. *Apache Spark* dikenal dengan kecepatan dan efisiensinya dalam menangani *Big Data* serta penerapan algoritma *Machine Learning*. Spark menyediakan berbagai library pendukung, termasuk Spark MLlib, yang menyertakan implementasi Latent Dirichlet Allocation (LDA). LDA digunakan untuk melakukan ekstraksi topik secara otomatis, dengan dukungan dari Spark NLP pada tahap pra-pemrosesan data.

Studi ini mengambil kasus pada percakapan dan penyebaran informasi seputar COVID-19 di Indonesia pada periode 2020-2022 melalui Media X. Dengan menggunakan pemodelan topik, penelitian ini bertujuan untuk mengidentifikasi dan menganalisis isu-isu utama yang berkembang di media sosial selama pandemi, sehingga dapat memberikan pemahaman yang lebih mendalam mengenai dinamika komunikasi dan penyebaran informasi dalam situasi krisis.

Berdasarkan latar belakang diatas, peneliti mengambil judul "**Pemodelan Topik Melalui Media X Menggunakan Apache Spark (Studi Kasus Covid-19 di Indonesia Tahun 2020-2022)**".

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka permasalahan yang dapat disimpulkan yaitu :

1. Seberapa banyak topik yang akan muncul yang berkaitan dengan COVID-19 di Indonesia Tahun 2020-2022 ?
2. Seberapa besar akurasi yang diperoleh dari hasil pemodelan topik yang dilakukan?

1.3 Batasan Masalah

Berdasarkan latar belakang dan identifikasi masalah. Maka penelitian ini hanya dibatasi pada variabel-variabel berikut :

1. Penelitian ini membatasi waktu studi kasus pada tahun 2020-2022, sehubungan dengan pandemic COVID-19.
2. Studi kasus dilakukan hanya pada wilayah Indonesia.
3. Data yang dianalisis adalah tweet mengenai COVID-19 di Indonesia pada tahun 2020-2022.
4. Penelitian ini hanya membatasi pemodelan topik menggunakan Apache Spark.
5. Penelitian ini membatasi hasil yang dicapai hanya mengenai seberapa banyak topik yang akan muncul yang berkaitan dengan COVID-19 di Indonesia Tahun 2020-2022.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitiannya adalah menghasilkan identifikasi seberapa banyak topik yang akan muncul yang berkaitan dengan COVID-19 di Indonesia Tahun 2020-2022.

Tujuan untuk masyarakat umum dapat digunakan sebagai sumbangan pemikiran guna menambah wawasan pengetahuan dan menjadi bahan referensi dan masukan bagi peneliti selanjutnya sehingga akan lebih baik untuk kedepannya.

1.5 Manfaat Penelitian

1. Menambah wawasan dan memperkaya literatur tentang pemodelan topik melalui media sosial.
2. Memberikan informasi mengenai kemampuan Apache Spark dalam membantu proses pemodelan topik data tweet.
3. Memberikan masukan bagi pengembangan lebih lanjut dalam proses pemodelan topik melalui media sosial.

1.6 Sistematika Penulisan

Sub bab ini menguraikan tentang sistematika penulisan tugas akhir/skripsi yang akan penulis lakukan diantaranya :

BAB I : PENDAHULUAN

Pada bab ini berisikan penjelasan tentang latar belakang masalah, perumusan masalah, pembatasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan laporan skripsi.

BAB II : TINJAUAN PUSTAKA

Pada bab ini berisikan teori berupa pengertian dan definisi yang diambil dari kutipan buku yang berkaitan dengan penyusunan skripsi serta beberapa *literature review* yang berhubungan dengan penelitian.

BAB III : METODE PENELITIAN

Bab ini menguraikan tentang objek penelitian, alur penelitian, serta alat dan bahan yang digunakan untuk memecahkan masalah dalam penelitian ini.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini menguraikan tentang hasil dan analisis dari penelitian yang dilakukan terhadap variabel Seberapa banyak topik yang akan muncul yang berkaitan dengan COVID-19 di Indonesia Tahun 2020-2022 dan Seberapa besar akurasi yang diperoleh dari hasil pemodelan topik yang dilakukan.

BAB V : PENUTUP

Bab ini akan menjelaskan kesimpulan dari hasil penelitian yang telah dilakukan oleh peneliti pada bagian sebelumnya juga saran yang merupakan masukan untuk pihak-pihak terkait.

