

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Kartu Indonesia Pintar (KIP) Kuliah merupakan transformasi dari Bidikmisi yang telah berjalan sejak tahun 2010. Program ini merupakan salah satu dari Program Indonesia Pintar (PIP) di bawah naungan Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi (Kemendikbudristek). Pada 2021, KIP Kuliah telah menjamin biaya pendidikan dan bantuan biaya hidup lebih dari 150.000 mahasiswa penerima yang masuk ke perguruan tinggi melalui beragam jalur masuk perguruan tinggi dan politeknik di seluruh PTN dan PTS. Adanya KIP Kuliah membantu untuk meningkatkan potensi ekonomi dan mobilitas sosial bagi mahasiswa dari keluarga kurang mampu untuk berkuliah pada program studi unggulan di perguruan tinggi terbaik di seluruh Indonesia (Setiawan, 2023).

Sejak diluncurkan, program ini telah membantu dan memberikan dampak positif bagi banyak mahasiswa. Hal ini terbukti dengan adanya penerima dari latar belakang yang sederhana, mampu melanjutkan pendidikan tinggi berkat bantuan tersebut (Ekaptiningrum, 2024). Namun, di sisi lain, terdapat masalah serius terkait mengenai ketepatan penyaluran bantuan. Masalah ini tersebut memicu kontroversi dan memunculkan perdebatan di kalangan masyarakat, khususnya di media sosial terutama pada platform X. Diskusi-diskusi tersebut mencerminkan kekhawatiran masyarakat mengenai penerapan program ini terutama terkait dengan beberapa penerima yang dianggap kurang sesuai dengan tujuan awal program, (Puspapertiwi & Nugroho, 2024).

Kekhawatiran tersebut tidak hanya muncul dari opini masyarakat, tetapi juga didukung oleh laporan resmi yang menunjukkan adanya penyaluran bantuan yang tidak tepat sasaran. Berdasarkan informasi yang diperoleh dari Badan Pemeriksa Keuangan (BPK) terdapat catatan bahwa dana bantuan yang disalurkan melalui Program Indonesia Pintar kepada 5.364.986 siswa dengan total dana sebesar Rp 2,86 triliun, tidak tepat sasaran (Yuliantri P, 2021). Masalah ini terjadi karena dana bantuan tersebut diberikan kepada siswa yang

tidak layak menerima. Salah satu contohnya adalah kasus mahasiswa yang menjadi penerima Kartu Indonesia Pintar Kuliah di sebuah universitas di Semarang. Bantuan yang seharusnya diterima oleh mahasiswa dengan keterbatasan finansial justru diberikan kepada mahasiswi yang sering membagikan gaya hidup mewah di akun sosial medianya (Negara, 2024). Situasi ini menimbulkan dampak negatif terhadap kepercayaan masyarakat terhadap program tersebut dan menimbulkan pertanyaan tentang seberapa efektif program KIP dalam menargetkan individu yang benar-benar membutuhkan. Untuk mengatasi masalah ini, diperlukan upaya untuk memahami persepsi masyarakat sebagai langkah awal untuk mengevaluasi dan meningkatkan pelaksanaan program.

Salah satu cara untuk memahami persepsi masyarakat adalah dengan melakukan analisis sentimen pada bidang ilmu *text mining*. Metode yang kerap digunakan dalam penelitian analisis sentimen adalah *Naive Bayes*, *Decision Tree*, *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN) dan lainnya. Metode-metode tersebut telah digunakan dalam beberapa penelitian sebelumnya untuk menganalisis sentimen masyarakat terhadap program KIP Kuliah. Salah satu penelitian tersebut dilakukan oleh (Pramudita dkk., 2024), yang menggunakan algoritma *Naive Bayes*. Penelitian tersebut menghasilkan nilai akurasi 84.99%. Namun, penelitian tersebut menyarankan agar menggunakan penelitian selanjutnya menambah jumlah dataset agar lebih bervariasi serta mempertimbangkan penggunaan algoritma lain untuk membandingkan performa dan memperoleh gambaran yang lebih komprehensif mengenai persepsi masyarakat terhadap program KIP-K.

Selain itu, terdapat juga penelitian yang membahas mengenai tanggapan pengguna media sosial X terhadap program beasiswa KIP-Kuliah dengan menggunakan algoritma *Support Vector Machine* (Amelia & Sarimole, 2024). Dataset yang digunakan berbahasa Indonesia dengan kata kunci “KIP-Kuliah”. Berdasarkan hasil analisis, algoritma SVM mencapai hasil akurasi 86,27%. Namun, pada penelitian tersebut algoritma SVM tidak berhasil memprediksi kelas positif dengan baik dan terdapat bias secara signifikan

terhadap sentimen negatif, hal tersebut disebabkan karena ketidakseimbangan jumlah data pada kelas positif.

Ketidakseimbangan data dapat memengaruhi kelas minoritas sehingga sulit diklasifikasikan, karena jumlahnya yang terbatas dalam dataset. Dataset dengan distribusi kelas yang tidak merata cenderung menghasilkan model yang bias terhadap kelas mayoritas. Untuk mengatasi permasalahan ketidakseimbangan pada data dapat dilakukan dengan menggunakan teknik *resampling*. Salah satu teknik *resampling* yang dapat digunakan adalah Synthetic Minority Over-sampling Technique (SMOTE). Teknik SMOTE adalah salah satu pendekatan *resampling* yang digunakan untuk menyetarakan jumlah data dengan pengambilan data dari kelas minoritas lalu menambahkan sampel tambahan.

Terdapat juga beberapa penelitian yang sebelumnya telah melakukan komparasi antara berbagai algoritma. Penelitian oleh (Iskandar & Nataliani, 2021) melakukan perbandingan metode *Naive Bayes*, *Support Vector Machine* (SVM), dan K-NN dalam analisis sentimen komentar YouTube terkait Samsung Galaxy Z Flip 3, metode SVM mencapai akurasi tertinggi yaitu 96,43%, diikuti oleh *Naive Bayes* sebesar 83,54% dan K-NN sebesar 59,68%. Selain itu, penelitian yang dilakukan oleh (Hakim dkk., 2021) meneliti sentimen ulasan myIndihome dengan menggunakan *Naive Bayes* dan SVM, dimana SVM kembali unggul dengan akurasi 86,54% dibandingkan *Naive Bayes* yang memperoleh akurasi 84,69%.

Metode *Support Vector Machine* dipilih karena memiliki performa yang lebih baik saat melakukan klasifikasi teks, *Support Vector Machine* memiliki kelebihan dalam kemampuan klasifikasi yang tinggi dalam menangani data berdimensi tinggi dibantu fungsi kernel dan kemampuan resisten terhadap *overfitting*. SVM bekerja dengan cara mencari anggota kelas terluar yang disebut *support vector* untuk dapat mencari batas atau *hyperplane* yang menjadi pemisah antara kelas-kelas dengan jarak terbesar. Sementara itu, metode *Naive Bayes* dipilih karena memiliki keunggulan dalam proses klasifikasi dengan jumlah data yang besar, selain itu algoritma *Naive Bayes Classifier* yang

sederhana dan kecepatannya yang tinggi dalam proses pelatihan dan klasifikasi (Ilmawan & Mude, 2020).

Berdasarkan pemaparan latar belakang di atas, penelitian ini dirancang dengan judul “Komparasi Algoritma *Naive Bayes* dan *Support Vector Machine* (SVM) pada Analisis Sentimen Implementasi Program Kartu Indonesia Pintar Kuliah” untuk dapat mengetahui perbandingan performa antara metode *Naive Bayes* dan *Support Vector Machine*.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan adapun permasalahan yang diidentifikasi adalah:

1. Ketidakakuratan penyaluran KIP Kuliah, dengan adanya penerima yang tidak memenuhi kriteria, menimbulkan kontroversi perdebatan dan menurunkan kepercayaan masyarakat. Hal ini bisa berdampak negatif pada keberlanjutan program sosial pemerintah di masa depan, meskipun program tersebut telah memberikan manfaat bagi banyak mahasiswa.
2. Penempatan KIP Kuliah di bawah beberapa kementerian dapat menciptakan tantangan dalam koordinasi antar lembaga. Kurangnya komunikasi dan sinergi antara Kementerian Pendidikan, Kementerian Sosial, dan Kementerian Keuangan dapat menyebabkan kebingungan mengenai tanggung jawab dan prosedur penyaluran bantuan.
3. Penelitian sebelumnya menunjukkan bahwa performa *Naive Bayes* dan *Support Vector Machine* memiliki akurasi yang cukup baik, yaitu metode *Naive Bayes* mencapai 84.44% dan metode *Support Vector Machine* 86.27%, namun pada penelitian yang menggunakan metode *support vector machine* model tidak berhasil memprediksi kelas positif dan terdapat bias pada sentimen negatif.

Berdasarkan latar belakang yang dipaparkan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana performa algoritma *Naive Bayes* dan *Support Vector Machine* untuk analisis sentimen Program Kartu Indonesia Pintar Kuliah dengan dan tanpa menggunakan metode SMOTE?

2. Bagaimana hasil analisis sentimen dari algoritma *Naive Bayes* dan *Support Vector Machine* untuk analisis sentimen Program Kartu Indonesia Pintar Kuliah?

### 1.3 Tujuan Penelitian

Berdasarkan uraian rumusan masalah, tujuan penelitian ini adalah:

1. Untuk mengetahui performa algoritma *Naive Bayes* dan *Support Vector Machine* untuk analisis sentimen Program Kartu Indonesia Pintar Kuliah dengan dan tanpa menggunakan metode SMOTE.
2. Untuk mengetahui hasil analisis sentimen algoritma *Naive Bayes* dan *Support Vector Machine* terhadap Program Kartu Indonesia Pintar Kuliah.

### 1.4 Ruang Lingkup Penelitian

Adapun ruang lingkup dalam penelitian ini adalah sebagai berikut:

1. Pada penelitian ini, digunakan sekumpulan *tweet* berbahasa Indonesia sebagai dataset yang relevan dengan implementasi Program Kartu Indonesia Pintar Kuliah. Dataset tersebut diperoleh dengan menggunakan kata kunci yang merujuk pada dua artikel penelitian (Amelia & Sarimole, 2024; Pramudita dkk., 2024) yaitu “KIP-K”, serta ditambah dengan kata kunci lainnya seperti “KIP Kuliah”, dan tagar #KIPK serta #KIPKuliah
2. Penelitian ini hanya menggunakan dua kategori kelas label yaitu positif dan negatif. Fokus utamanya adalah pada opini yang mendukung program Kartu Indonesia Pintar Kuliah dan opini yang menentang program tersebut. Opini dengan sentimen netral tidak dimasukkan karena tidak memberikan informasi signifikan dalam analisis sentimen.
3. Penelitian ini berfokus melakukan komparasi algoritma *Naive Bayes* dengan jenis *Multinomial Naive Bayes* dan algoritma *Support Vector Machine* dengan kernel RBF.
4. Rentang waktu *tweet* yang digunakan pada penelitian ini adalah *tweet* yang diunggah dalam rentang waktu tahun 20 Februari 2020 sampai dengan tahun 30 September 2024.

5. Tidak dilakukan validasi langsung terhadap opini penulis pada data *tweet* yang dikumpulkan dari platform X.

### 1.5 Manfaat Penelitian

Penelitian Komparasi Algoritma *Naive Bayes* dan *Support Vector Machine* (SVM) pada Analisis Sentimen Implementasi Program Kartu Indonesia Pintar Kuliah diharapkan memberikan manfaat sebagai berikut:

#### 1. Manfaat Teoritis

Hasil penelitian ini diharapkan memberikan kontribusi dalam meningkatkan pemahaman peneliti terhadap perbedaan kinerja antara metode *Naive Bayes* dan *Support Vector Machine* untuk analisis sentimen program Kartu Indonesia Pintar Kuliah.

#### 2. Manfaat Praktis

##### a. Manfaat bagi peneliti

Peneliti mampu mengimplementasikan ilmu yang dipelajari selama perkuliahan dan dapat meningkatkan wawasan dalam bidang data mining, khususnya dalam analisis sentimen Program Kartu Indonesia Pintar Kuliah dengan menggunakan metode *Naive Bayes* dan *Support Vector Machine*.

##### b. Manfaat bagi Masyarakat Umum

Masyarakat memperoleh informasi yang lebih jelas mengenai persepsi publik terhadap Program Kartu Indonesia Pintar Kuliah, berdasarkan hasil analisis sentimen dengan menggunakan dua metode yang diujikan. Hal ini membantu masyarakat untuk memahami bagaimana program Kartu Indonesia Pintar Kuliah diterima oleh berbagai kalangan.

##### c. Manfaat bagi Penyelenggara Program

Penyelenggara Program dapat menggunakan hasil analisis sentimen ini untuk memahami persepsi masyarakat terhadap Program Kartu Indonesia Pintar Kuliah, sehingga penyelenggara bisa mengidentifikasi dan mengevaluasi aspek-aspek positif dan negatif dari implementasi program berdasarkan opini publik, hal ini juga mendukung pengambilan

keputusan strategis untuk memperbaiki program di masa depan sehingga program tersebut lebih efektif.

