

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Di era revolusi industri 4.0, perkembangan teknologi menjadi semakin pesat yang dapat menunjang segala aspek, salah satunya yaitu dalam peran media. Dalam era digital seperti saat ini terdapat banyak media yang mampu menyebarkan informasi dengan mudah dan cepat untuk diakses, seperti media sosial. *Facebook*, *Instagram*, dan *Twitter* merupakan beberapa media sosial kategori *social network* yang populer di masyarakat saat ini [1].

Twitter menjadi salah satu media paling banyak digunakan oleh masyarakat. Di Indonesia sendiri *twitter* menempati peringkat empat yaitu sebesar 52% pengguna internet Indonesia pada tahun 2019 [1]. Melalui *twitter* pertukaran informasi akan semakin cepat dan baru, dan juga sangat mudah untuk didapatkan. Konten-konten atau komentar pada *twitter*, dapat digunakan untuk mengukur presentase partisipasi netizen baik itu mendukung atau tidak mendukung. Salah satu topik atau opini yang sering dibicarakan saat ini yaitu tentang kebijakan PPDB 2020 selama pandemi Covid-19 yang diterapkan di Indonesia .

Banyak dampak yang ditimbulkan akibat pandemi Covid-19. Salah satunya adalah pendidikan (sekolah). Pemerintah mengeluarkan kebijakan baru untuk PPDB 2020, yaitu PPDB secara daring yang proses seleksinya dilakukan melalui jalur zonasi, afirmasi, dan jalur perpindahan tugas orang tua/wali. Jika masih terdapat sisa kuota dari pelaksanaan tiga jalur tersebut, maka dilakukan seleksi berdasarkan jalur prestasi [2]. Kebijakan tersebut menuai banyak kontroversi di masyarakat. Pemerintah perlu memperhatikan persepsi masyarakat terhadap kebijakan tersebut untuk mendapatkan kebijakan publik yang ideal, dengan media sosial *twitter* pemerintah dapat menjadikannya sebagai sumber data sentimen dan pendapat masyarakat.

Dalam suatu sentimen dan pendapat masyarakat pasti ada yang pro dan kontra terhadap kebijakan PPDB 2020. Untuk bisa menyelesaikan kasus tersebut, membutuhkan klasifikasi sentimen yang bersifat positif dan negatif, dengan menggunakan *text mining*. *Text mining* merupakan suatu proses untuk mengolah teks menjadi suatu data [3]. Dalam penelitian sebelumnya terdapat beberapa metode klasifikasi yang sudah digunakan, yaitu penelitian yang berjudul Analisis Sentimen Penerapan Psbb di Dki Jakarta dan Dampaknya Terhadap Pergerakan Ihsg. Pada penelitiannya membandingkan hasil klasifikasi beberapa metode yaitu, *Naïve bayes*, *K-nearest neighbor*, *Random forest algorithm*, dan *Logistic regression*. Dalam penelitian tersebut *logistic regression* menunjukkan akurasi paling tinggi dibanding dengan metode lainnya, 75% untuk *Logistic regression*, 70% untuk *K-nearest neighbor*, 60% untuk *Random forest*, dan 70% untuk *Naïve bayes* [4].

Penelitian lain yaitu berjudul Perbandingan Algoritma Klasifikasi untuk Analisis Sentimen terhadap Komentar Film. Pada penelitian ini melakukan perbandingan algoritma klasifikasi *K-nearest neighbor*, *Naïve bayes*, dan *Logistic regression*. Data yang digunakan yaitu data sentimen terhadap komentar film. Terdapat beberapa proses justifikasi awal yang dilakukan yaitu : load data set (menetapkan nilai positif dan negatif), *text preprocessing*, transformasi data teks menjadi *tf-idf*. Dari hasil eksperimen dengan menggunakan 25.000 dataset menunjukkan bahwa *Logistic regression* memiliki tingkat akurasi yang paling tinggi dengan akurasi sebesar 89%, dibandingkan akurasi *Naïve bayes* yakni 86%, sedangkan *K-nearest neighbor* 65,22% [5].

Penelitian yang berjudul *Analysis of the Grain Loss in Harvest Based on Logistic Regression* juga telah melakukan penelitian perbandingan model *Logistic regression* sebelum menggunakan algoritma SGD (*Stochastic Gradient Descent*) dan sesudah menggunakan SGD. Penelitian ini menghasilkan tanpa menggunakan algoritma SGD mendapatkan hasil prediksi 86,25% dan mengalami peningkatan hasil prediksi menjadi 92,53% pada saat menggunakan algoritma SGD. Berdasarkan ulasan penelitian

sebelumnya, dalam penelitian ini peneliti mengusulkan metode *Logistic regression* untuk mengklasifikasi persepsi netizen terhadap kebijakan PPDB 2020. Dengan menggunakan data *tweet* yang didapatkan dari media sosial twitter. Sebelum data digunakan harus melalui tahap *preprocessing* atau mengolah teks menjadi sebuah data agar bisa diproses [6]. Setelah melalui *preprocessing*, data diolah menggunakan *TF-IDF* yang berguna untuk pembobotan kata (*term*) pada suatu teks [7]. Proses berikutnya yaitu mengestimasi parameter (koefisien) menggunakan algoritma SGD (*Stochastic Gradient Descent*) untuk membentuk suatu model yang akan digunakan pada model regresi logistik [8]. Setelah mendapatkan model langkah selanjutnya yaitu memprediksi menggunakan model Regresi logistik [9].

Berdasarkan uraian diatas serta masih terbatasnya penelitian tentang persepsi netizen terhadap PPDB 2020, oleh karena itu peneliti tertarik untuk mengangkat topik yang berjudul “Klasifikasi sentimen publik pada *twitter* tentang kebijakan ppdb menggunakan metode *logistic regression*”, dengan tujuan memanfaatkan metode *logistic regression* dan mengukur tingkat akurasi hasil klasifikasi *tweet* melalui *twitter* terhadap kebijakan pemerintah di Indonesia tentang PPDB 2020.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Mengetahui tingkat akurasi penerapan metode *logistic regression* dalam mengklasifikasi *tweet* tentang kebijakan PPDB 2020

1.2.2 Metode Usulan

Berdasarkan permasalahan yang telah dijabarkan, maka penelitian ini menggunakan metode *Logistic Regression* untuk mengukur nilai akurasi hasil klasifikasi data *tweet* tentang kebijakan PPDB 2020.

1.2.3 Pertanyaan Penelitian

Berapa tingkat akurasi yang diperoleh dari penerapan metode *Logistic regression* dalam mengklasifikasi *tweet* melalui *twitter* terhadap kebijakan pemerintah di Indonesia tentang PPDB 2020 ?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Mengukur tingkat akurasi yang diperoleh dari penerapan metode *Logistic regression* dalam mengklasifikasi *tweet* melalui *twitter* terhadap kebijakan pemerintah di Indonesia tentang PPDB 2020.

1.3.2 Manfaat

Manfaat pada penelitian ini adalah:

1. Secara praktis penelitian ini memberikan kontribusi untuk mengetahui persepsi netizen terhadap kebijakan pemerintah di Indonesia tentang PPDB 2020.
2. Secara teoritik, penelitian ini memberikan kontribusi dalam pengembangan keilmuan *computer science*, *data science*, *text mining*, politik, pemerintah dan kebijakan publik.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Data yang digunakan merupakan *tweet* dari seluruh indonesia dari twitter yang menggunakan bahasa Indonesia yang diambil selama masa PPDB tahun 2020 dari bulan Januari sampai bulan Desember sebanyak 600 *tweet*. Terdiri dari 300 komentar negatif dan 300 komentar positif.
2. Data yang akan digunakan menggunakan kata kunci 'PPDB'
3. Klasifikasi hanya menggunakan dua kelas, positif dan negatif

1.5 Sistematika Laporan

Sistematika penulisan proposal ini terbagi ke dalam beberapa bab, yaitu :

BAB I PENDAHULUAN

Pada bab ini menjelaskan dasar-dasar penulisan laporan tugas akhir, yaitu tentang latar belakang, rumusan masalah, tujuan dan manfaat, batasan-batasan masalah dan sistematika laporan.

BAB II KAJIAN PUSTAKA

Pada bab ini membahas landasan teori serta penelitian-penelitian sebelumnya yang pernah dilakukan.

BAB III METODE USULAN

Pada bab ini membahas tentang tahapan penelitian metode yang disulkan, penjelasan desain proses, data yang digunakan serta skenario pengujian.

BAB IV HASIL DAN PEMBAHASAN

Bagian ini menjelaskan tentang penerapan/implementasi sistem yang dibuat berdasarkan hasil rancangan sebelumnya.

BAB V PENUTUP

Pada bab ini berisi kesimpulan yang didapat dari hasil dan pembahasan serta saran untuk pengembangan sistem selanjutnya.