

BAB I

PENDAHULUAN

1.1. Latar Belakang

KPK (Komisi Pemberantasan Korupsi) adalah garda terdepan dalam penanganan kasus korupsi serta lembaga negara yang mempunyai kuasa penuh dalam memberantas korupsi. Kinerja KPK dalam memberantas korupsi banyak mendapat sorotan dari masyarakat, tak sedikit yang memberi opini positif terhadap kinerjanya tetapi tidak sedikit pula yang memberi opini negatif terhadap kinerja KPK. Sosial media pun tak luput dari sasaran polemik opini masyarakat terhadap kinerja KPK, salah satunya *Twitter*. Opini masyarakat terhadap KPK pada media sosial yang masif ini belum terakomodasi atau digambarkan secara sederhana sehingga KPK tidak mengetahui persentase sentimen masyarakat terhadap kinerjanya, oleh karena itu belum bisa menjadi bahan pertimbangan dalam perbaikan kinerja yang diperoleh melalui opini atau kritik dan saran masyarakat pada media sosial khususnya *twitter* . Seperti yang ditulis oleh akun @wiriantiwidjaya pada *twitter* “Silahkan, saya dukung @KPK_RI memberantas korupsi tanpa pandang bulu!”. Lalu tulisan oleh @fahrihamzah yaitu “Pertanyaannya kok #15TahunKPK OTT makin banyak? Bukankah ini pengakuan korupsi tambah banyak? Lalu sukses KPK di mana?”

Dengan ketenaran *Twitter* tersebut maka tidak heran banyak orang yang memberikan pendapatnya tentang berbagai hal termasuk pendapat kinerja KPK dalam memberantas korupsi di Indonesia dan juga *Twitter* berfokus pada pendekatan analisis sentimen yang dirancang khusus untuk *tweet*. Analisis sentimen adalah proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini baik yang bersifat positif dan negatif.

(Falahah, Pengembangan Aplikasi Sentiment Analysis Menggunakan Metode *Naïve Bayes*: 2015) dalam hasil penelitiannya menyimpulkan bahwa Metoda *Naïve Bayes Classifier* dapat diterapkan sebagai metode untuk

melakukan klasifikasi sentiment analysis dan proses klasifikasi akan semakin akurat apabila jumlah keyword training yang digunakan semakin banyak . (Nabila Safina, Analisis Sentimen Pada *Twitter* Terhadap Jasa Transportasi Online Di Indonesia Dengan Metode Support Vector Mechine: 2016) mengambil data dari kata kunci mention terhadap @gojekindonesia dan @GrabID, lalu mengelompokkannya menjadi sentimen negatif dan positif dengan metode SVM menggunakan LibSVM bahwa akurasi klasifikasi *tweet* tentang transportasi online yaitu 87%. Berdasar uraian diatas, penulis tertarik untuk mengangkat permasalahan tersebut kedalam penelitian skripsi dengan judul “ANALISIS SENTIMEN PADA *TWITTER* TERHADAP KINERJA KOMISI PEMBERANTASAN KORUPSI (KPK) DI INDONESIA DENGAN METODE NAIVE BAYES”.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, rumusan masalah penelitian ini adalah sebagai berikut :

1. Bagaimana hasil klasifikasikan sentimen positif dan sentimen negatif masyarakat Indonesia terhadap kinerja KPK pada *Twitter* menggunakan metode *Naive Bayes Classifier*?
2. Bagaimana hasil persentase *confusion matrix* sentimen masyarakat Indonesia terhadap KPK pada *Twitter* menggunakan metode *Naive Bayes Classifier*?

1.3. Tujuan

Berdasarkan permasalahan yang telah dikemukakan, tujuan penelitian skripsi ini adalah sebagai berikut :

1. Mengetahui sentimen masyarakat Indonesia terhadap kinerja KPK pada *Twitter* menggunakan metode *Naive Bayes Classifier*.
2. Mengetahui persentase *confusion matrix* sentimen masyarakat Indonesia terhadap KPK pada *Twitter* menggunakan metode *Naive Bayes Classifier*.

1.4. Manfaat

Manfaat dari penelitian ini antara lain :

1. Dengan mengetahui sentimen masyarakat dapat dijadikan bahan pertimbangan dalam upaya perbaikan KPK.
2. Dengan mengetahui persentase *confussion matrix* dapat dijadikan bahan acuan sebagai kritik dan saran yang valid serta dapat digunakan untuk mengukur sentimen analisis yang lain seperti pemilihan presiden, legislatif dan kepala daerah.

1.5. Batasan Masalah

Adapun batasan masalah yang akan dibahas adalah :

1. Klasifikasi sentimen masyarakat berdasar sentimen positif dan negatif.
2. Sumber data yang diambil melalui *Twitter* yang ditulis dari tanggal 31 Januari 2018 sampai 31 Mei 2018 dengan jumlah data 803 *tweet* dengan data latih 451 data uji 58 *tweet* serta data tidak terklasifikasi berjumlah 294 *tweet*.
3. Data *Twitter* yang dianalisis tidak mencakup data gambar/foto, video dan link.
4. Tidak mengklasifikasi *emoticon*.