

BAB 2

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Syed Akib Anwar Hridoy (2015) telah melakukan penelitian mengenai penambangan opini menggunakan analisis sentimen di *Twitter* (Hridoy, et al., 2015). Penelitian ini bertujuan untuk menganalisa informasi publik dari jejaring sosial untuk menentukan opini publik. Dalam penelitian ini digunakan metode penambangan opini yaitu *sentimen analysis* untuk menganalisa opini publik.

Data yang digunakan dalam penelitian ini diambil dari jejaring sosial *Twitter*. Dari berbagai informasi yang didapat dalam satu *tweet* yang digunakan untuk penelitian ini hanya informasi teks dan *tweet id*. Ruang lingkup penelitian ini difokuskan secara umum di negara Amerika Serikat dan secara khusus di kota New York, Los Angeles, Boston, Chicago, Dallas, San Francisco dan Philadelphia. Untuk kata kunci pencarian *tweet* yang dicari untuk penelitian ini adalah iPhone 6 dengan kata kunci pencarian pendukung yaitu "camera", "iOS", "iTunes", "screen", "sound" dan "touch". Untuk melakukan beberapa tahapan penelitian digunakan beberapa aplikasi pihak ketiga yaitu *twitteroauth by Williams* untuk ekstraksi data, *Stanford Natural Language Processing tool by The Stanford NLP Group* untuk *preprocessing* data, *NamSor* untuk membagi data berdasarkan jenis kelamin dan *SentiWordNet* untuk melakukan penilaian untuk menentukan kata opini positif dan negatif. Untuk mendapatkan hasil akhir penelitian, peneliti menghitung skor sentimen dari satu *tweet* utuh

dengan cara menghitung rata-rata dari nilai *SentiWord* keseluruhan dari setiap kata penyusun *tweet*.

Hasil akhir penelitian ini adalah pemetaan nilai sentimen opini publik berdasarkan negara, kota dan jenis kelamin. Kelebihan dari hasil penelitian ini adalah dihasilkan pemetaan hasil yang jelas dan kecocokan hasil penelitian dengan hasil deduksi di kehidupan nyata. Kelemahan penelitian ini adalah jumlah data yang digunakan dalam penelitian masih sangat sedikit sehingga hasil tidak optimal, penggunaan aplikasi pihak ketiga sebagai pendukung menyebabkan kesulitan dalam menghasilkan hasil yang optimal karena kesalahan pemrosesan dari aplikasi pihak ketiga tidak dapat dihindari.

Harlili (2011) telah melakukan penelitian mengenai cara analisis opini berbahasa Indonesia di media *Microblogging* (Harlili & Wibisono, 2011). Penelitian ini bertujuan untuk membangun sistem yang dapat membantu melakukan analisis opini *microblogging* berbahasa Indonesia. Dalam penelitian ini digunakan data dari jejaring sosial *Twitter* dengan jumlah 35.000 *tweet* berbahasa Indonesia dengan kata kunci pencarian "Telkomsel" dan "Indosat" yang diambil pada bulan Agustus 2010. Tahapan awal yang dilakukan dalam penelitian ini adalah dengan menentukan *tweet* yang merupakan opini. Dari data *tweet* opini tersebut data dipisahkan berdasarkan dua jenis yaitu opini negatif dan opini positif. Tahap terakhir adalah menentukan kata kunci yang menunjukkan aspek apa yang mendapat opini. Metode yang digunakan dalam penelitian ini adalah metode *Naïve Bayes Classifier* untuk mengklasifikasikan *tweet* berdasarkan opini dan nonopini

serta mengklasifikasikan berdasarkan opini negatif dan opini positif, dan metode TF-IDF (*Term Frequency - Inverse Document Tfrequency*) untuk ekstraksi data kata kunci.

Hasil akhir penelitian adalah pemetaan jumlah opini positif dan negatif berdasarkan waktu. Kelebihan dari hasil penelitian ini adalah dapat dicapai akurasi sebesar 91,83% untuk pengklasifikasian *tweet* opini dan nonopini serat 93,40% untuk pengklasifikasian opini positif dan negatif. Kelemahan dari hasil penelitian ini adalah rendahnya akurasi penentuan kata kunci yang membandingkan hasil sistem dengan deduksi manusia yaitu 34,48%.

Sari (2013) telah melakukan penelitian mengenai pembangunan aplikasi klasifikasi mahasiswa baru Universitas Atma Jaya Yogyakarta dengan metode *Naïve Bayes Classifier*. Penelitian ini bertujuan untuk memprediksi IPK, rata-rata SKS, dan persentase kehadiran mahasiswa baru di Universitas Atma Jaya Yogyakarta dengan variabel masukan yaitu program studi, jalur pendaftaran, jenis kelamin, nilai masuk dan kabupaten atau kota. Dalam penelitian ini digunakan data mahasiswa dari basis data Universitas Atma Jaya Yogyakarta. Metode yang digunakan dalam penelitian ini adalah *Naïve Bayes Classifier*.

Hasil akhir dari penelitian didapatkan prediksi IPK, rata-rata SKS, dan persentase kehadiran mahasiswa baru dengan rata-rata keakuratan sebesar 81.67%, dengan catatan bahwa persentase keakuratan dipengaruhi oleh data yang diujikan. Kelebihan dari hasil penelitian ini adalah dapat dilakukan penyesuaian klasifikasi data berdasarkan masukan variabel yang akan digunakan dan

hasil keluaran yang diinginkan. Kelemahan dari hasil penelitian ini adalah rata-rata keakuratan klasifikasi dipengaruhi oleh data yang diujikan, sehingga tingkat keakuratannya tidak selalu tinggi.

Perbandingan penelitian yang akan penulis lakukan dengan penelitian sejenis yang pernah dibuat sebelumnya :

Tabel 2.1 Tabel Perbandingan Penelitian

Pembanding	Penulis (2015)	Hridoy (2015)	Harlili (2011)	Sari (2013)
Algoritma	<i>Naïve Bayes Classifier dan Perkiraan Relasi Kata</i>	<i>Sentimen Analysis</i>	<i>Naïve Bayes dan TF-IDF</i>	<i>Naïve Bayes Classifier</i>
Klasifikasi Pada	<i>Tweets Pengguna Twitter</i>	<i>Twitter Data</i>	<i>Opini dalam Microblogging</i>	Mahasiswa Baru
Klasifikasi yang Dicari	Topik dan Sentimen	Sentimen	Sentimen	IPK, Rata-rata SKS, Persentase Kehadiran