

## BAB II

### TINJAUAN PUSTAKA

Penelitian ini berjalan dengan memanfaatkan berbagai referensi sebagai acuan, dimana referensi tersebut berasal dari penelitian-penelitian yang telah dibuat sebelumnya dengan topik penelitian yang berkaitan dengan penelitian yang sedang dilakukan. Penelitian pertama dilakukan oleh Fitriana pada tahun 2021 menghasilkan sebuah kesimpulan perbandingan kinerja pada dua metode yang berbeda. Dengan memanfaatkan klasifikasi yang merupakan bagian dari teks mining, sentimen yang menggunakan metode *Support Vector Machine* dan *Naïve Bayes* dari hasil klasifikasi tersebut digunakan sebagai penilaian opini masyarakat terhadap Vaksin Covid-19. Pada proses penelitian menggunakan pemrograman PHP 7.4 dan JavaScript 1.8.5 serta untuk menampung data kedalam database menggunakan MySQL [5].

Dalam tahap *action planning* digunakan algoritma *Support Vector Machine* dan *Naïve Bayes* untuk membuat pemodelan data. Pengujian dilakukan sebanyak 5 kali pengulangan guna mendapatkan validasi pengujian. Penelitian tersebut jika menggunakan algoritma *Naïve Bayes* menghasilkan akurasi sebesar 88,64%, presisi 87,32% *recall* 88,13% dengan total waktu 8,1 detik. Dengan menggunakan algoritma SVM menghasilkan akurasi sebesar 90,47%, presisi 90,23% *recall* 90,78% dengan total waktu 11 detik. Dimana menghasilkan kesimpulan algoritma SVM kelihatan lebih unggul dalam hal akurasi, presisi dan *recall* serta algoritma *Naïve Bayes* lebih unggul dalam perihal eksekusi waktu *training*. Hal tersebut yang menimbulkan *Naïve Bayes* lebih unggul dari segi waktu *training* dibandingkan SVM.

Berikutnya penelitian yang dilakukan oleh Buntoro pada tahun 2017. Penelitian tersebut bertujuan untuk mengetahui opini masyarakat mengenai Pemilihan Gubernur DKI Jakarta pada tahun 2017 yang

disampaikan melalui media sosial twitter. Melihat banyaknya atensi dari masyarakat Indonesia khususnya warga DKI Jakarta dilakukanlah penelitian guna mengetahui opini masyarakat kemudian diklasifikasi menjadi 3 kelas yaitu positif, negatif, dan netral dengan menggunakan metode *Lexicon Based*. Proses klasifikasi dilakukan untuk menguji akurasi metode *Lexicon Based* dalam menentukan sentimen pada sebuah opini. Proses tersebut dilakukan menggunakan WEKA 3.8.1, metode *Naïve Bayes Classifier* dan metode *Support Vector Machine (SVM)* juga digunakan dalam proses klasifikasi.

Dataset dibagi menjadi 2 bagian yaitu *training* dan *testing*. Data diuji menggunakan *10-fold cross validation*. Hasilnya dari 100 data untuk kata kunci AHY menghasilkan 62 bernilai positif, 25 bernilai netral, dan 13 bernilai negatif. Dengan menggunakan metode *Naïve Bayes Classifier* menghasilkan *accuracy* sebesar 95%, 95% *precision*, 95% *recall*, 96,8% *TP rate*, dan 84,6% *TN rate*. Sedangkan menggunakan metode *Support Vector Machine* menghasilkan tingkat *accuracy* sebesar 90%, 89,9% *precision*, 90% *recall*, 98,4% *TP rate*, dan 38,5% *TN rate*. Kata kunci selanjutnya adalah Ahok dengan jumlah data set yang sama yaitu 100 data menghasilkan 33 bernilai positif, 10 bernilai netral, dan 57 bernilai negatif. Menggunakan metode *Naïve Bayes Classifier* menghasilkan tingkat *accuracy* sebesar 82%, 81,6% *precision*, 82% *recall*, 87,9% *TP rate*, dan 86% *TN rate*. Sedangkan menggunakan metode *Support Vector Machine* menghasilkan tingkat *accuracy* sebesar 84%, 86,1% *precision*, 84% *recall*, 72,7% *TP rate*, dan 98,2% *TN rate*. Kata kunci terakhir adalah Anies dengan jumlah data yang sama menghasilkan 62 bernilai positif, 26 bernilai netral, dan 12 bernilai negatif. Hasil klasifikasi menggunakan *Naïve Bayes Classifier* menghasilkan tingkat *accuracy* sebesar 82%, 82,7% *precision*, 82% *recall*, 87,1% *TP rate*, dan 75% *TN rate*. Sedangkan menggunakan metode *Support Vector Machine* menghasilkan tingkat *accuracy* sebesar 87%, 87,5% *precision*, 87% *recall*, 95,2% *TP rate*, dan 75% *TN rate* [6].

Analisis sentimen yang dilakukan oleh Tuhuteru pada tahun 2020 mengenai Pembatasan Sosial Berskala Besar atau yang biasanya kita sebut dengan PSBB dimana data tersebut diambil melalui postingan masyarakat melalui media sosial facebook dan twitter. Pada penelitian ini ingin diketahui bagaimana opini masyarakat mengenai terhadap penerapan PSBB di Kota Ambon serta menghasilkan sebuah sentimen yang bernilai positif, negatif, maupun netral. Adapun software yang digunakan dalam penelitian ini adalah RapidMiner Studio dengan ekstensi *Text Processing* untuk mengelola data *training*. Proses awal dilakukan *preprocessing* guna membersihkan data, kemudian dengan menggunakan ekstensi *text processing* pada RapidMiner dilakukan pembobotan untuk setiap kata yang ada, klasifikasi penelitian dilakukan menggunakan metode SVM, dan evaluasi performa dengan menggunakan *k-fold cross validation*. Penelitian tersebut menghasilkan sentimen netral lebih mendominasi daripada sentimen positif ataupun negatif dengan hasil sebesar 45% sedangkan sentimen positif sebesar 28% dan sentimen negatif 27% dari total data sebesar 725 data dengan tingkat validasi yang cukup tinggi pada *9-fold cross validation* dengan nilai 82,5% untuk *accuracy*, 82,07% *recall*, 83,20% *precision* dengan waktu eksekusi 5 menit 41 detik [7].

Penelitian mengenai jasa ojek *online* dibuat oleh Nugroho pada tahun 2016. Menggunakan metode *Naïve Bayes* peneliti berusaha menghasilkan tingkat akurasi dari metode tersebut. Pada tahap awal yaitu *pre-processing* dilakukan beberapa proses diantaranya *case folding*, *tokenizing*, *stemming*, dan *filtering* guna menghasilkan fitur yang akan digunakan sebagai data *machine learning* oleh *Naïve Bayes Classifier*. Pada tahap klasifikasi menggunakan metode *Naïve Bayes Classifier* dapat diketahui kategori dari teks tersebut berdasarkan data latih. Pembentukan data latih dilakukan pelatihan bagi data latih berdasarkan kata kunci pada dokumen (*tweet*) untuk diklasifikasi menjadi 3 kelas yang telah ditentukan yaitu positif, negatif, dan netral. Kemudian akan dilakukan perhitungan Vmap agar dapat diketahui probabilitas tertinggi dari data uji dari masing-

masing kelas. Data sebanyak 800 tweet yang dibagi menjadi 500 data uji dan 300 data latih. Dimana dengan menggunakan metode *Naïve Bayes Classifier* menghasilkan tingkat akurasi sebesar 80%. Dengan 100 data latih positif, 100 data latih negatif, dan 100 data latih netral [8].

Selanjutnya penelitian yang dilakukan oleh Sipayung mengambil judul tentang komentar pelanggan mengenai penilaian berupa *feedback* dan komentar dari hotel XYZ yang dipesan secara online melalui agoda.com, booking.com, dan tripadvisor.com. Penelitian tersebut dilakukan pada tahun 2016 dengan menggunakan metode *Naïve Bayes Classifier* yang akan mengklasifikasi penilaian tersebut menjadi 2 kelas yaitu positif dan negatif. Implementasi dari perancangan dari penelitian dibuat dengan menggunakan bahasa pemrograman Java dan database MySql yang kemudian ditampilkan dalam bentuk diagram pie. Berdasarkan pengujian tersebut, diketahui jika tiap fungsi pada sistem sanggup memproses data dan membagikan respon sesuai dengan permasalahan yang terdapat pada pengujian seperti data input yang benar, salah ataupun tidak lengkap. Kemampuan sistem dalam melangsungkan analisis dengan mengelola pendapat dengan menghasilkan *accuracy* kategori, serta untuk sentimen *precision*, *recall*, serta *accuracy*. Klasifikasi sentimen menghasilkan 155 komentar yang bernilai positif serta 20 komentar bernilai negatif. Perhitungan keandalan sistem analisis sentimen dengan metode *Naïve Bayes Classifier* ini dari 175 data latih dapat dipisah menjadi 2 bagian, perhitungan klasifikasi kategori serta perhitungan klasifikasi sentimen, untuk hasil dari akurasi sistem memiliki nilai *accuration* kategori 77,14%, *precision* sentimen 99,12%, *recall* sentimen 72,9%, serta *accuration* sentimen 75,42%. Sehingga sistem mampu mengembalikan dokumen dan kecocokan data yang tinggi [9].

Pada penelitian berikutnya dibuat oleh Pitria mengenai sentimen customer samsung mengenai *feedback* sehingga dapat dimanfaatkan untuk evaluasi produk samsung yang disampaikan customer melalui akun resmi twitter Samsung Indonesia. Pada tahun 2014 penelitian dilakukan

menggunakan metode *Naïve Bayes Classifier* sehingga dapat menghasilkan klasifikasi sentimen. Tahap pertama dilakukan *preprocessing* dengan mengambil data menggunakan *crawling* dengan stream API, dimana data akan dibersihkan dari noise. Penelitian dibuat menggunakan bahasa pemrograman PHP dan code editor berupa Sublime Text3Portable. Dalam proses pengujian digunakan 1350 data *training* dan 150 data *testing*. Metode *Naïve Bayes Classifier* dapat diterapkan dengan baik pada analisis sentimen dalam mengklasifikasikan sentimen positif dan negatif, hal ini bisa dilihat dari akurasi yang didapatkan cukup tinggi dengan menggunakan pengujian *10-fold cross validation* menghasilkan rata-rata akurasi sebesar 98,87% [10].

Pada tahun 2013 Nurhuda melakukan penelitian mengenai calon presiden Indonesia tahun 2014. Metode *Naïve Bayes Classifier* digunakan dalam penelitian ini guna melakukan klasifikasi sentimen menjadi 3 kelas yaitu positif, negatif, dan netral. Banyak masyarakat mengutarakan opininya melalui media sosial twitter, sehingga API twitter dapat dimanfaatkan sebagai tempat pengambilan dataset. Setelah dilakukan *streaming* data kemudian data tersebut disimpan kedalam MySQL. *Preprocessing* data dilakukan guna membersihkan data serta menjadikan sederhana dimensi dari dataset. Selanjutnya dataset tersebut kembali dipilih melalui ekstraksi fitur dengan cara memanfaatkan *part-of-speech tagging* untuk mendapatkan kelas kata. Pengujian akurasi menggunakan *Naïve Bayes Classifier* dilakukan dengan cara eksperimen tingkat akurasi untuk klasifikasi polaritas sentimen terhadap calon presiden Indonesia 2014. Hasil dari pengamatan sentimen publik melalui twitter menunjukkan jumlah obrolan terhadap capres serta cawapres di twitter menjelang pemilu 2014 terus menjadi bertambah. Hasil dari pengamatan polaritas sentimen warga terhadap pasangan calon presiden serta wakil presiden menampilkan pasangan Prabowo Subianto- Hatta Rajasa memperoleh 47,7% untuk sentimen positif, 26,4% sentimen negatif serta 25,9% sentimen netral. Sedangkan pasangan Joko Widodo- Jusuf Kalla memperoleh total 37,6%

sentimen positif, 34,4% sentimen negatif, serta 27,9% sentimen netral. Dari hal ini bisa disimpulkan bahwa pasangan Prabowo Subianto– Hatta Rajasa lebih unggul dari pasangan Joko Widodo–Jusuf Kalla perihal jumlah obrolan serta sentimen positif pada jejaring sosial twitter [11].

*Cyberbullying* merupakan tindakan *bullying* yang sering terjadi di dunia maya. Oleh karena itu Luqyana melakukan penelitian berupa Analisis sentimen terkait dengan *cyberbullying* yang terdapat pada kolom komentar instagram. Karena analisis sentimen sangat diperlukan guna menyaring komentar-komentar di media sosial agar dapat mengetahui apakah komentar tersebut merupakan komentar negatif ataupun komentar positif. Penelitian ini dilakukan pada tahun 2018 dengan menggunakan metode klasifikasi *Support Vector Machine*. *Preprocessing* dilakukan dalam penelitian ini guna mengubah data sesuai dengan format yang dibutuhkan dan untuk menggali informasi. Ada beberapa proses yang dilakukan pada tahap *preprocessing* diantaranya adalah *case folding*, *data cleaning*, normalisasi bahasa, *stopword removal*, *stemming*, dan tokenisasi. Algoritma *Support Vector Machine* pada penelitian ini digunakan untuk melakukan klasifikasi dengan menggunakan bobot indeks *term* sebagai fitur. Berdasarkan hasil pengujian yang dilakukan didapatkan tingkat akurasi terbaik sebesar 90%, *precision* sebesar 94,44%, 85% *recall* sebesar dan *f-measure* sebesar 89,47% dengan komposisi data latih 50% dan data uji 50% [12].

Penelitian yang dilakukan oleh Buntoro pada tahun 2016 dengan membandingkan metode *Naïve bayes Classifier* dan *Support Vector Machine* sebagai basis dari penelitian analisis sentimen. Penelitian ini melakukan klasifikasi terhadap *hatespeech* atau ujaran kebencian yang terjadi pada media sosial twitter. Setelah dilakukannya *preprocessing* data, data tweet diubah menjadi bentuk vektor dengan cara *StringToWorld* pada tool WEKA 3.7.11. Pemberian bobot juga dilakukan dengan menggunakan metode TF-IDF (*Term Frequency – Invers Document Frequency*). Pada proses klasifikasi data diuji dengan *10-fold cross validation* dimana data

dibagi menjadi 2 9/10 digunakan untuk *training* dan 1/10 digunakan sebagai *testing* dengan jumlah dataset 522 tweet. Hasilnya Nilai akurasi tertinggi didapat dengan metode klasifikasi *Support Vector Machine (SVM)* dengan tokenisasi *unigram*, *stopword list* Bahasa Indonesia dan *emoticons*, dengan nilai rata-rata akurasi mencapai 66,6%, nilai presisi 67,1%, nilai *recall* 66,7% nilai *TP rate* 66,7% dan nilai *TN rate* 75,8%. Dalam penelitian ini juga dapat diketahui metode klasifikasi *Support Vector Machine (SVM)* lebih tinggi akurasinya untuk klasifikasi sentiment tweet *hatespeech* dibandingkan metode klasifikasi *Naïve Bayes Classifier (NBC)* [13].

Ratnawati mencoba melakukan penelitian terhadap penerapan dari algoritma *Naïve Bayes* pada tahun 2018. Dengan metode tersebut peneliti ingin melakukan klasifikasi sentimen terhadap opini publik mengenai film yang pernah mereka tonton dan disampaikan melalui media sosial twitter. Pada pengujian ini digunakan 200 data negatif dan 200 data positif. Metode ini diimplementasikan menggunakan python dan library NLTK. Dengan tujuan untuk dapat secara otomatis mengklasifikasikan *tweet* sebagai sentimen *tweet* yang positif dan negatif. Akurasi tertinggi didapat pada fold kedua yaitu 90%, *precision* 92%, *Recall* 90% dan *f measure* 90%. Dimana hasil tersebut didapatkan dengan menggunakan metode *Naïve Bayes Classifier* dan *k-fold cross validation* dengan pengulangan sebanyak 5 kali [14].

Pada tahun 2019 Sari melakukan penelitian terhadap toko online JD.ID. Dengan menggunakan metode *Naïve Bayes Classifier* peneliti melakukan klasifikasi sentimen berbasis konversi dan ikon emosi. Data yang digunakan adalah 900 tweet dengan pembagian 300 tweet positif, 300 tweet negatif, dan 300 tweet netral. Pada evaluasi dan pengujian menggunakan *confused matrix* untuk mengukur tingkat akurasi dan *k-fold cross validation*. Metode *Naïve Bayes Classifier* diimplementasikan menggunakan bahasa pemrograman PHP dan HTML yang digunakan untuk mengambil tweet adalah Rstudio. Sedangkan *tools* yang digunakan untuk

melakukan eksperimen adalah *RapidMiner* dengan *text preprocessing* yang dilakukan yaitu *case folding*, *convert emoticon*, *cleansing*, *text transformation*, *stopword removal*, *tokenizing* dan *stemming*. *Naïve Bayes Classifier* tanpa pembobotan *tf-idf* dan konversi ikon emosi (*convert emoticon*) memiliki nilai akurasi 96,44% sedangkan *Naïve Bayes Classifier* dengan pembobotan *tf-idf* dan konversi ikon emosi (*convert emoticon*) memiliki nilai akurasi 98%, dengan demikian terjadi peningkatan akurasi sebesar 1,56% [15].

Penelitian yang dilakukan oleh Zuhri terhadap brand smartfren dengan melakukan klasifikasi sentimen menggunakan metode *Naïve Bayes Classifier* pada tahun 2017. Dataset yang digunakan dalam penelitian ini merupakan opini berbahasa Indonesia mengenai *brand* Smartfren di forum Kaskus. Dataset berupa komentar yang merupakan ungkapan masyarakat mengenai suatu hal melalui media sosial. Proses *crawling* dilakukan secara otomatis melalui website import.io. Dataset yang didapat oleh peneliti yaitu sebanyak 7386 kalimat opini. Hasil dari klasifikasi nya adalah 4049 sentimen positif, 2339 sentimen negatif, dan 1001 sentimen netral. Namun, proses pengklasifikasian data uji menjadi 2 kelas, yaitu, sentimen positif dan negatif. Hasil klasifikasi data uji sangat dipengaruhi oleh model yang telah dibuat dari data latih. Dari proses evaluasi dan validasi data latih, didapatkan precision sebesar 98.42 %, recall sebesar 98.40 %, dan tingkat akurasi sebesar 98.40 %. Hal ini menunjukkan bahwa sistem dapat memisahkan data bersentimen positif dan negatif dengan sangat baik. Dan juga memperoleh nilai koefisien kappa sebesar 0.948, maka menunjukkan kesepakatan yang sangat kuat antar penguji karena mendekati nilai 1, maka instrumen data latih dapat dikatakan reliabel [16].

Winarko pada tahun 2015 melakukan penelitian mengenai nilai rating dan ulasan pengguna aplikasi pada Google Play Store. Tujuannya adalah agar dapat dijadikan pertimbangan bagi calon pengguna sebuah aplikasi maupun untuk para pengembang dari aplikasi tersebut. Metode

*Naïve Bayes Classifier* dilakukan klasifikasi sentimen dengan membagi menjadi 2 kelas yaitu positif dan negatif. Jumlah data yang digunakan dalam penelitian ini adalah 9000 komentar dengan tingkat akurasi 83,87%. Evaluasi *classifier* menggunakan metode *3-folds cross-validation*. Selama proses pengujian digunakan 3 partisi data yang berukuran sama, salah satu dari partisi dipilih untuk *testing*, sedangkan dua partisi lainnya digunakan untuk *training*. Prosedur ini diulangi sebanyak 3 kali. Berdasarkan penggunaan fitur, nilai akurasi paling tinggi didapatkan dengan menggunakan fitur *negation tag* tanpa *semantic handling*. Pada penelitian ini, pembuatan Aplikasi *Mobile* untuk Analisis Sentimen pada Google Play diimplementasikan dengan bahasa pemrograman java, android SDK dan python. Bahasa pemrograman java dan python digunakan untuk server sedangkan untuk pembuatan aplikasi mobile digunakan android SDK [17].

Analisis sentimen yang dilakukan oleh Ferryawan pada tahun 2019. Penelitian mengangkat tentang wisata Jawa Tengah dengan menggunakan metode *Naïve Bayes Classifier*, peneliti mengklasifikasi informasi dan opini seputar fasilitas obyek wisata yang disampaikan oleh pengunjung melalui media sosial twitter. Penelitian ini dilakukan dengan mengambil opini publik tentang obyek wisata Karimunjawa di Jepara, candi Borobudur di Magelang, Gua Maria Kerep di Ambarawa, dataran tinggi Dieng di Wonosobo, kawasan wisata Baturaden di Purwokerto, Lawang Sewu di Semarang, Grojogan Sewu di Karanganyar, Umbul Ponggok di Klaten, Masjid Agung Jawa Tengah di Semarang dan pantai Karang Bolong di Kebumen. Perhitungan yang dihasilkan oleh aplikasi yaitu 98 sentimen positif, 90 sentimen negatif dan 27 sentimen netral. Hasil tersebut menunjukkan masyarakat lebih menunjukkan sikap positif pada wisata di Jawa Tengah, akan tetapi hasil perhitungan antara sentimen positif dan sentimen negatif hanya berbeda 8 data. Perbedaan yang tidak terlalu signifikan ini patut menjadi pertimbangan bagi pemerintah untuk memperbaiki obyek wisata yang ada sebagai investasi pemerintah [18].

Penelitian yang terakhir yaitu analisis sentimen mengenai *review* dari aplikasi Ruangguru. Data tersebut diambil dari website Google Play Store dengan jumlah 2000 *review*. Penelitian dilakukan oleh Irfani dengan menggunakan algoritma *Support Vector Machine* pada tahun 2020. Data tersebut dibagi menjadi dua yaitu data latih dan data uji. Dengan menggunakan metode tersebut memiliki tingkat akurasi 90%. Pengujian menggunakan seluruh data latih dan data uji. Dari pengujian tersebut diperoleh sebuah sentimen, dari 1000 data yang diuji berhasil diklasifikasikan menjadi 892 *review* positif dan 108 *review* negatif. Hasil dari pengujian menyatakan bahwa sentimen masyarakat terhadap aplikasi Ruangguru cenderung positif [19].

Berdasarkan referensi penelitian yang telah ada sebelumnya, bahwa analisis sentimen digunakan untuk mengetahui persentase mengenai nilai dari sebuah topik permasalahan. Kolom ulasan yang terdapat di *Google Play Store* dapat digunakan oleh pengguna untuk menyampaikan bentuk apresiasi, kritik, dan saran terhadap suatu aplikasi. Dimana dengan adanya komentar yang diberikan oleh pengguna dapat digunakan untuk memperbaiki serta melakukan inovasi untuk suatu aplikasi yang telah ada. Dengan menggunakan metode *Naïve Bayes Classifier* kita dapat melakukan klasifikasi terhadap komentar pengguna. Seperti yang telah tertera pada tinjauan pustaka, belum ada penelitian yang membahas mengenai aplikasi MyBluebird. Oleh karena itu, penelitian ini akan membahas mengenai komentar yang telah disampaikan oleh pengguna melalui kolom ulasan aplikasi MyBluebird di *Google Play Store* dengan menggunakan algoritma pemrograman Python dengan mengakses *librarynya* sebagai metode untuk melakukan *collecting* data. Kemudian opini tersebut dapat dikelompokkan menjadi sentimen positif, netral, dan negatif setelahnya dapat divalidasi. Sehingga dapat menghasilkan tingkat akurasi dari performa uji menggunakan metode *k-fold cross validation*.

Pembanding	Fitriana [5]	Buntoro [6]	Tuhuteru [7]	Winarko [17]	Nadila (2022)*)
Judul	Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes	Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter	Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berskala Besar Menggunakan Algoritma Support Vector Machine	Aplikasi Mobile untuk Analisis Sentimen pada Google Play	Analisis Sentimen Komentar Pengguna Terhadap Aplikasi MyBluebird Pada Play Store Menggunakan Metode Naive Bayes Classifier
Objek	Tweet pada media sosial Twitter	Tweet pada media sosial Twitter	Tweet dan komentar di platform media sosial Facebook dan Twitter	Ulasan tekstual aplikasi di Google Play	Ulasan tekstual aplikasi di Play Store
Metode	Support vector machine, naïve bayes, preprocessing, k-fold cross validation, PHP	Preprocessing, support vector machine, naïve bayes classifier, lexicon based	Preprocessing, crawling, Support vector machine, naïve bayes classifier, snipping	Support vector machine, naïve bayes classifier, Preprocessing, k-fold cross validation	Web Scraping, Preprocessing, k-fold cross validation, naïve bayes classifier
Tools/Code	PHP, Java	WEKA	Rapid Miner	Java, Python	Python

**Table 2. 1** Perbandingan Penelitian Terdahulu

*\*) Penelitian yang dilakukan*