

BAB II

TINJAUAN PUSTAKA

Beberapa aplikasi sejenis yang menjadi perbandingan dan referensi bagi peneliti dalam melakukan penelitian ini dijelaskan dalam bab ini.

Adityawan (2014) membangun aplikasi analisis sentimen untuk situs microblogging Twitter dengan menggunakan algoritma *Naïve Bayes* dengan data seimbang. Sebelum proses klasifikasi, dilakukan praproses terlebih dahulu yaitu normalisasi teks dan pembuangan *stopword*. Untuk memilih data latih yang terbaik menggunakan *K-fold Cross Validation*. Setelah itu, untuk pengklasifikasian dilakukan dengan menggunakan *Naïve Bayes* model Multinomial dan model Bernoulli. Penelitian ini menghasilkan tingkat akurasi sebesar 66.42% untuk model Multinomial dan 71.09% untuk model Bernoulli.

Ling, et al (2014) membangun sebuah aplikasi analisis sentimen dengan algoritma *Naïve Bayes* dan seleksi fitur *Chi Square*. Praproses dalam pembangunan aplikasi ini juga dilakukan dengan melakukan *tokenization*, *stemming*, dan pembuangan *stopwords*. *Tokenization* adalah tugas pemotongan urutan karakter dan sebuah set dokumen yang diberikan menjadi potongan-potongan kata atau karakter yang sesuai dengan kebutuhan sistem. Potongan-potongan tersebut dikenal dengan istilah token. *Stemming* adalah proses mengubah token yang berlebihan menjadi kata dasar. Setelah itu, kata yang dianggap tidak memiliki tendensi sentimen atau *stopword* dibuang. Seleksi

fitur *chi square* digunakan untuk mereduksi fitur-fitur yang tidak relevan dalam proses pengklasifikasian NBC. Seleksi fitur *chi square* menggunakan teori statistika untuk menguji independensi sebuah *term* dengan kategorinya. Dari penelitian ini diperoleh hasil analisis dengan tingkat akurasi sebesar 83% dan rata-rata harmonik sebesar 90.713%.

Anugroho, et al (2011) membangun aplikasi pendeteksi email spam dengan menggunakan algoritma *Naïve Bayes* pada bahasa pemrograman Java. Praproses tidak dijelaskan di dalam karya tulis namun dari hasil yang didapat dapat ditarik kesimpulan bahwa klasifikasi email spam dengan *Naïve Bayes classifier* yang dilakukan secara offline dan online tidak memiliki perbedaan dalam mengklasifikasikan email spam secara tepat dengan tingkat error yang kecil.

Ginting & Trinanda (2010) membangun aplikasi perpustakaan dengan fitur pengklasifikasian buku dengan menggunakan *Naïve Bayes classifier*. Praproses tidak dijelaskan dalam karya tulis. Pengklasifikasian dengan *Naïve Bayes classifier* digunakan untuk mengklasifikasi buku berdasarkan deskripsi. Aplikasi ini dibangun dengan menggunakan Visual Basic.

Hamzah (2012) membangun aplikasi untuk mengklasifikasi teks dan abstrak akademis dengan menggunakan *Naïve Bayes classifier*. Aplikasi ini dibangun dengan bahasa pemrograman Java. Praproses yang dilakukan adalah pembuangan *stopwords*. Kemudian terdapat proses pelatihan yaitu untuk membentuk

vocabulary. Tingkat akurasi yang didapat adalah 91% untuk dokumen berita dan 82% untuk dokumen akademik.



Tabel 2.1. Perbandingan Perangkat Lunak

No.	Pembanding	Adityawan (2014)	Ling, et al. (2014)	Anugroho (2011)	Ginting & Trinanda (2010)	Hamzah (2012)	Penulis (2017)
1.	Judul	Analisis Sentimen dengan Klasifikasi Naïve Bayes pada Pesan Twitter dengan Data Seimbang	Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier dengan Seleksi Fitur Chi Square	Klasifikasi Email Spam dengan Metode Naïve Bayes Classifier Menggunakan Java Programing	Penggunaan Metode Naïve Bayes Classifier Pada Aplikasi Perpustakaan	Klasifikasi Teks dengan Naïve Bayes Classifier (NBC) untuk Pengelompokan Teks Berita dan Abstract Akademis	Implementasi Metode Naive Bayes dalam Pembangunan Analisis Sentimen Berita

No.	Pembanding	Adityawan (2014)	Ling, et al. (2014)	Anugroho (2011)	Ginting & Trinanda (2010)	Hamzah (2012)	Penulis (2017)
2.	Tujuan	Klasifikasi sentiment tweet dengan data seimbang	Menggabungkan NBC Chi Square sebagai seleksi fitur	Klasifikasi Email dengan metode NBC menggunakan Java Programming	Klasifikasi buku perpustakaan dengan NBC	Klasifikasi teks berita dan abstrak akademis dengan NBC	Klasifikasi sentimen berita dengan NBC berbasis web
3.	Basis Data	ya	-	Tidak	Ya	Ya	Ya
4.	Platform	Web	-	Mail Client & Server	Desktop	Desktop	Web
5.	Bahasa yang diproses	Indonesia	Inggris	Inggris	Indonesia	Indonesia	Indonesia
6.	Bahasa Pemrograman	PHP	-	Java	C#	Java	PHP