

BAB II

TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Sri Widaningsih, dkk (2018) melakukan penelitian berjudul “Klasifikasi Jurnal Ilmu Komputer Berdasarkan Pembagian Web Of Science dengan menggunakan Text Mining”. Tujuan dari penelitian ini yaitu untuk mengategorikan artikel berdasarkan topik di bidang ilmu komputer yaitu *hardware and architecture, interdisciplinary applications, artificial intelligence, cybernetics, information system, science and photography, software engineering, theory and method*. Teknik pengklasifikasian menggunakan dua metode yaitu support vector machine dan naïve bayes classifier. Proses pengklasifikasian menggunakan tahapan yang ada pada *text mining*. Hasil dari penelitian ini yaitu ditemukan bahwa pengklasifikasian dengan metode naïve bayes classifier memberikan nilai lebih tinggi dibandingkan dengan metode *support vector machine* sehingga pengklasifikasian otomatis diputuskan akan menggunakan metode naïve bayes (Widaningsih and Suheri 2018).

Angga Pradikdo, dkk (2018) melakukan penelitian berjudul “Model Klasifikasi Abstrak Skripsi Menggunakan Text Mining untuk Pengkategorian Skripsi Sesuai Bidang Kajian”. Penelitian ini dilatarbelakangi oleh dokumen skripsi yang bertambah setiap tahun dan pada website Universitas Nusanara PGRI Kediri terdapat banyak skripsi yang pengkategorianya tidak sesuai dengan isi naskah. Tujuan dari penelitian ini untuk memudahkan mahasiswa

dalam mencari referensi skripsi sesuai dengan bidangnya dan pihak pustakawan dapat dengan cepat melakukan *upload file* pada *repository*. Metode yang digunakan dalam penelitian ini yaitu *text mining* diantaranya dengan *preprocessing* dan *transformation*. Hasil penelitian menunjukkan bahwa siklus terbaik memiliki akurasi 82,76% dapat digunakan dalam pengklasifikasian skripsi pada program studi sistem informasi universitas nusantara PGRI Kediri (Pradikdo and Ristyawan 2018).

Effrida Manalu, dkk melakukan penelitian dalam jurnal ilmiah yang berjudul “Penenerapan Algoritma Naïve Bayes untuk Memprediksi Jumlah Produksi Baarang berdasarkan Data Persediaan dan Jumlah Pesanan pada CV Papadan Mama Pastries”. Penelitian ini dilatarbelakangi oleh sebuah perusahaan dalam bidang industry toko roti yang memasarkan hasil produksi mereka ke berbagai tempat. Permasalahannya yaitu pada perusahaan tersebut lalu lintas data antara pesanan dan stok produksi terkadang tidak sejalan karena data yang kurang akurat antara pesanan dan hasil produksi, padahal hal tersebut dapat membuat proses bisnis menjadi kacau dan merugikan produsen dan konsumen. Tujuan dari penelitian ini yaitu penulis ingin memberikan solusi terkait keakuratan data dan proses penrekan daa pesanan dan stok barang yang tersedia serta produksi yang harus dilakukan untuk memnuhi permintaan setiap harinya dengan membuat sistem informasi menggunakan visual basic 2010 dengan menerapkan naïve bayes dalam memprediksi persediaan roti, stok dan keharusan produksi berdasarkan penjualan tahun lalu. Hasil dari penelitian ini yaitu sistem informasi yang dibangun dapat memberikan solusi alternative bagi keakuratan data sebagai acuan produksi (Manalu, Sianturi, and Manalu 2017).

Kesimpulan dari penelitin terdahulu yang diijadikan referensi yaitu algoritma naïve bayes dapat digunakan untuk mengidentifikasi kata dari data yang tersedia. Maka penulis akan melakukan perancangan klasifikasi teks pada artikel ilmiah Universitas Muhammadiyah Ponorogo dengan menerapkan algoritma naïve bayes yang dibangun menggunakan bahasa pemrograman web python dan database MySQL.

2.2. Klasifikasi Text

Klasifikasi teks merupakan proses pengkategorian teks berdasarkan pola atau pattern tertentu yang muncul dan disaring atau ditemukan oleh sistem pengklasifikasian bergantung pada metode apa yang digunakan (Ridok and Latifah 2015). Klasifikasi dokumen bertujuan untuk mengelompokan dokumen-dokumen yang tidak terstruktur menjadi bersatu dengan kelompok yang memiliki topik sama serta menggambarkan isi dari dokumen tersebut (Widaningsih and Suheri 2018).

2.3. Naïve Bayes

Metode Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Rifqo and Wijaya 2017).

2.4. Data Mining

Data mining atau penambangan data merupakan proses yang dilakukan untuk dengan pengambilan pola data yang tersedia guna menemukan atau mencari suatu informasi yang berguna dan penting (Setyawan and Suradi 2017). Data mining memiliki beberapa jenis dan metode serta algoritma yang digunakan untuk membantu menemukan informasi yang sesuai dan akurat. Jenis data mining yaitu asosiasi, clustering, klasifikasi, prediksi dan estimasi (Muhammad, Sudrajat, and Rosadi 2016)

2.5. Python

Sisi utama yang membedakan Python dengan bahasa lain adalah dalam hal aturan penulisan kode program. Bagi para programmer di luar python siap-siap dibingungkan dengan aturan indentasi, tipe data, tuple, dan dictionary. Python memiliki kelebihan tersendiri dibandingkan dengan bahasa lain terutama dalam hal penanganan modul, ini yang membuat beberapa programmer menyukai python. Selain itu python merupakan salah satu produk yang opensource, free, dan multiplatform (Dini 2012). Python adalah salah satu bahasa pemrograman tingkat tinggi yang bersifat interpreter, interaktif, object-oriented dan dapat beroperasi di hampir semua platform, seperti keluarga UNIX, Mac, Windows, dan lainnya. Sebagai bahasa tingkat tinggi, Python termasuk salah satu bahasa pemrograman yang mudah untuk dipelajari karena sintaks yang jelas dan elegan, dikombinasikan dengan penggunaan module-module siap pakai dan struktur data tingkat tinggi yang efisien (Rosmala and Dwipa 2012).