

BAB II

TINJAUAN PUSTAKA

2.1 Studi Literatur

Adapun beberapa penelitian sebelumnya yang masih berhubungan dengan topik penelitian yang dilakukan saat ini sehingga peneliti menggunakan beberapa penelitian tersebut sebagai referensi dalam pembuatan laporan pada tugas akhir ini.

Penelitian Totok Sugiarto yang berjudul **Pengaruh Lingkungan, Media Massa dan Masyarakat Sebagai Penyebab Anak-Anak Melakukan Tindakan Kriminal** dikatakan bahwa Terdapat pengaruh antara perilaku kriminal remaja dengan surat kabar cetak dan elektronik. Media cetak dan elektronik juga turut andil dalam proses pembentukan kepribadian anak. Jika media cetak dan elektronik bisa menyajikannya dengan baik, maka dampaknya juga bisa positif. Sebaliknya jika media cetak dan elektronik tidak dikemas dengan baik maka akan berdampak buruk juga karena mudah diserap dan ditiru oleh anak-anak[9].

Adapun penelitian lainnya yang berjudul **Dampak Tayangan Film di Televisi Terhadap Perilaku Anak** yang dilakukan oleh Sri Desti mengatakan bahwa banyak yang beranggapan bahwa media televisi adalah penyebab dari berbagai tindak kekerasan seperti perampokan, pembunuhan, pemerkosaan, tawuran, dan sebagainya di lingkungan masyarakat. Namun anggapan ini kurang tepat karena setiap kejadian atau tindakan yang ditiru oleh penonton terutama anak-anak, tidak bisa sepenuhnya menyalahkan televisi. Tindakan tersebut mungkin juga dipengaruhi oleh lingkungan tempat tinggal anak, seperti di rumah, sekolah atau lingkungan tempat anak tersebut bermain. Meskipun demikian, televisi tetap memiliki peran dalam perkembangan anak yang menonton, oleh karena itu, peran aktif orang tua dalam membimbing dan mengarahkan anak-anak saat menonton tayangan televisi sangat penting agar dampak negative televisi tidak merusak setiap individu penontonnya[10].

2.2 Data Mining

Dalam jurnal yang berjudul **Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau** karya Insanul Kamila, Ulya Khairunnisa, Mustakim menjelaskan bahwa *Data Mining*

merupakan Proses mencari informasi atau pola yang menarik pada data didalam sebuah database dengan menggunakan metode tertentu[11].

2.3 Klasterisasi

Studi terdahulu yang diambil oleh penulis adalah jurnal yang berjudul **Klasterisasi Judul Buku dengan Menggunakan Metode K-Means** karya Deka Dwinavinta Candra Nugraha, Makhfuzi Fahmi, Zumrotun Naimah, dan Novi Setiani. Sekumpulan metode yang dipakai untuk mendapatkan sebuah informasi dari sekumpulan data adalah dengan mempelajari pola dari data itu sendiri disebut dengan Data Mining[12].

Penelitian terdahulu selanjutnya berjudul **Penerapan Clustering Pada Penduduk yang Mempunyai Keluhan Kesehatan Dengan Data Mining K-Means** yang ditulis oleh Nurul Rofiqo, Agus Perdana Windarto, dan Dedy Hartama mengatakan bahwa suatu metode pengolahan data untuk menemukan sebuah pola tersembunyi dari sekelompok data disebut Data Mining yang hasilnya bisa digunakan untuk membantu melakukan pengambilan keputusan dimasa yang akan datang[13].

Pada jurnal karya Yulia Darmi dan Agus Setiawan yang berjudul **Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk** mengatakan dalam knowledge discovey in database (KDD) data mining merupakan suatu langkah untuk menganalisa sebuah data untuk mendapatkan informasi dalam jumlah besar yang menghasilkan output berupa pola atau karakteristik data tersebut[14].

2.4 Decision Tree Algorithm

Penelitian tentang decision tree yang dilakukan oleh Harsh H. Patel, dan Purvi Prajapati yang berjudul **Study and Analysis of Decision Tree Based Classification Algorithms** mengatakan bahwa beberapa algoritma decision tree telah dibangun berdasarkan akurasi dan biaya efektivitasnya untuk berperan penting dalam setiap situasi pengambilan keputusan, berbeda situasi berbeda juga decision tree yang digunakan. Dalam penelitian ini Harsh H. Patel, dan Purvi Prajapati menemukan bahwa

decision tree memiliki keunggulan dalam hal akurasi, waktu, dan presisi dibandingkan dengan algoritma lain[15].

Pada penelitian yang berjudul **Liver Disease Prediction by Using Different Decision Tree Techniques** Nazmun Nahar dan Ferdous Ara mengatakan algoritma decision tree memberikan hasil yang lebih akurat dibandingkan dengan algoritma lainnya dan juga dapat dengan mudah mengklasifikasikan dataset yang berukuran besar untuk mempermudah dipahami oleh manusia. Dari beberapa algoritma decision tree yang digunakan untuk memprediksi penyakit hati pada tahap awal dalam penelitian ini memberikan hasil bahwa Decision Stump merupakan algoritma decision tree yang memiliki performa yang lebih baik dibandingkan dengan algoritma lain dengan akurasi sebesar 70,67%[16].

Dalam penelitian lain yang berjudul **Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes** yang dilakukan oleh Achmad Bayhaqy, Kaman Nainggolan, Sfenrianto, dan Emil R. Kaburuan mengatakan bahwa decision tree merupakan metode yang cukup efisien dalam hal pengklasifikasikan data. Hasil dari penelitian ini menunjukkan bahwa tingkat akurasi dari algoritma decision tree merupakan yang tertinggi diantara dua algoritma lainnya dengan tingkat akurasi sebesar 80% namun memiliki presisi yang lebih rendah sebesar 79,96% serta recall yang lebih tinggi sebesar 84% hal ini terjadi karena dipengaruhi oleh karakteristik dataset serta proses yang berbeda[17].

Penelitian yang dilakukan oleh Nasiba M. Abdulkareem, Adnan Mohsin Abdulazeez, Diyar Qader Zeebaree, dan Dathar A. Hasan yang berjudul **COVID-19 World Vaccination Progress Using Machine Learning Classification Algorithms** menemukan bahwa algoritma terbaik untuk melakukan klasifikasi adalah decision tree dengan tingkat akurasi sebesar 99,98% dan untuk membangun model total waktu yang diperlukan adalah 0,08 detik. Hasil ini mereka temukan setelah melakukan uji coba menggunakan dataset yang sama pada beberapa metode lain seperti K-Nearest Neighbors (KNN), Random Tree, dan Naive Bayes yang masing-masing metode tersebut menghasilkan akurasi sebesar 81,26% untuk KNN, 96,17% untuk Random Tree, dan 69,74% untuk Naïve bayes[18].

Algoritma C5.0 merupakan algoritma yang cocok digunakan ketika bekerja dengan jumlah data yang sangat besar, algoritma C5.0 juga dikenal sebagai penyempurnaan atau tingkat lanjutan dari algoritma C4.5. Pada jurnal **The comparison of the model performances of Naïve Bayes, C4.5 and C5.0: Implementation on fish consumption habits**, Senor Celik sebagai penulisnya mengatakan bahwa algoritma C5.0 lebih cepat dan lebih efisien dalam penggunaan memori disbanding algoritma C4.5. Hal ini didasarkan dengan beberapa percobaan yang dilakukan ketika menggunakan algoritma pada dataset yang ada. Dalam percobaan ini hasil klasifikasi algoritma naïve bayes menghasilkan tingkat akurasi sebesar 46,35%; algoritma C4.5 menghasilkan tingkat akurasi sebesar 53,86%; dan algoritma C5.0 mendapatkan hasil tingkat akurasi tertinggi dibandingkan dengan dua algoritma lainnya sebesar 56,2% [19].

Pada penelitian berjudul **Perbandingan Performansi Algoritma Decision Tree C5.0, CART, dan CHAID: Kasus Prediksi Status Resiko Kredit DI BANK X**, Yogi Yusuf W sebagai penulis melakukan sebuah percobaan dalam pembuatan model pohon keputusan dengan 3 algoritma berbeda untuk memprediksi status resiko kredit, dimana hasil akurasi dari model pohon keputusan tersebut memberikan tingkat akurasi rata-rata sebesar 87,72% untuk algoritma C5.0 dengan standar deviasi 1,56; 87,28% untuk CART dengan standar deviasi sebesar 1,51; dan 87,15% untuk CHAID dengan standar deviasi sebesar 2,19 [3].

Pada penelitian yang dilakukan oleh V. Rahmayanti, Y. Azhar, dan A. E. Pramudita yang berjudul **“Penerapan Algoritma C5.0 Pada Analisis Faktor-Faktor Pengaruh Kelulusan Tepat Waktu Mahasiswa Teknik Informatika Universitas Muhammadiyah Malang”**, dilakukan implementasi dan pengujian algoritma klasifikasi C5.0 yang mampu menghasilkan tingkat akurasi yang lebih tinggi ketika menggunakan seleksi fitur algoritma C5.0 dibandingkan dengan menggunakan keseluruhan fitur, dimana tingkat akurasi ketika menggunakan seleksi fitur algoritma C5.0 adalah 69,8% sedangkan hasil tingkat akurasi pada pengujian dengan menggunakan keseluruhan fitur sebesar 67,6% [20].

2.5 Feature selection Algorithm

Penelitian berjudul **Analisis Sentimen Movie Review Menggunakan Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square** yang dilakukan oleh Ahmad Zuli Amrullah, Andi Sofyan Anas, dan Muh. Adrian Juniarta Hidayat menyatakan bahwa untuk melakukan penyeleksian fitur *Chi Square*, peranan penting terletak pada frekuensi fitur kategori yang diharapkan dan kategori yang tidak diharapkan. Oleh karena itu dalam melakukan seleksi fitur, *Chi Square* baik untuk digunakan[21].

Penelitian lain yang berjudul **Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square** yang dilakukan oleh Juen Ling, I Putu Eka N. Kencana, dan Tjokorda Bagus Oka, mengatakan dalam seleksi fitur *Chi Square* apabila ada fitur yang muncul beberapa kali pada kategori yang tidak diharapkan, maka frekuensi pada fitur tersebut menjadi kurang penting. Sehingga jika dibandingkan dengan fitur *frequency-based*, akan lebih baik menggunakan metode seleksi fitur *Chi Square*[22].

Pada penelitian yang dilakukan oleh Nova Tri Romadloni, dan Hilman F pardede dalam Jurnal Yang berjudul **Seleksi Fitur Berbasis Pearson Correlation Untuk Optimasi Opinion Mining Review Pelanggan** terdapat peningkatan hasil akurasi ketika adanya penerapan seleksi fitur berbasis *pearson correlation* pada masing – masing metode yang sedang digunakan untuk penelitian[23].