

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Dalam penelitian ini, peneliti juga menyertakan beberapa uraian singkat mengenai penelitian terdahulu, tentang prediksi lama masa studi mahasiswa, klasifikasi, metode *Naïve Bayes* dan *Adaboost*.

2.1.1 Penelitian Yang Terkait Dengan Metode *Naïve Bayes*

Penelitian yang dilakukan oleh (Amornsinlaphachai, 2016) yang berjudul “*Efficiency of data mining models to predict academic performance and a cooperative learning model*” dengan tujuan Memilih model Data Mining untuk memprediksi prestasi akademik terhadap peserta didik program computer untuk membandingkan model efisiensi *Data Mining* dengan teknik klasifikasi dan membangun model pembelajaran berbasis web terhadap partisipasi peserta didik (Amornsinlaphachai, 2016). Sedangkan pada penelitiannya (Devasia, P and Hegde, 2016) yang berjudul “*Prediction of Students Performance using Educational Data Mining*” Mengusulkan sebuah sistem berbasis web dengan memanfaatkan teknik data mining untuk mengekstraksi informasi yang berguna. Percobaan yang dilakukan menggunakan metode *Naïve Bayes* terhadap 700 data siswa dengan 19 atribut (Devasia, P and Hegde, 2016).

Sedangkan Peneliatian lain dilakukan oleh (Guleria and Sood, 2015) dengan judul “*Predicting student placements using Bayesian classification*” permintaan

siswa agar institusi menempatkan pendidikan tinggi yang berkualitas dan bermanfaat bagi para siswa. Didalam aplikasi teorema bayesian pada dataset pendidikantelah diusulkan untuk memprediksi hasil penempatan. Hasilnya yang didapatkan akan dievaluasi menggunakan *Naive Bayes Classifier* dengan alat bantu WeKa dan RapidMiner. Hasil yang dievaluasi akan membantu para pendidik agar meningkatkan aspek-aspek dalam hal penempatan para siswa (Guleria and Sood, 2015). Penelitian berikut yang dilakukan oleh (Liu *et al.*, 2017) dengan topic “*A Preprocessing Method of AdaBoost for Mislabeled Data Classification*” Mengusulkan metode *Adaboost konvensional* untuk memperbaiki data yang nois, *Decision Stump* merupakan algoritma pembelajaran lemah untuk klasifikasi data (Liu *et al.*, 2017). Sedangkan penelitian yang dilakukan oleh peneliti (Jacob *et al.*, 2017) dengan judul “*Combination of Naive Bayes and Adaboost methods in predicting the study period of the students*” tujuan dari penelitian ini dapat memprediksi lama studi dari setiap mahasiswa atas dasar data kelulusan sebelumnya, hasilnya dapat digunakan sebagai bahan pertimbangan pimpinan dalam proses pengambilan kebijakan dan keputusan sebagai peringatan dini (*early warning*) bagi mahasiswa yang berdasarkan hasil prediksi dinyatakan berpotensi lulus dengan melampaui ketentuan lama studinya.

Selain itu, penelitian yang dilakukan oleh (Ridwan, 2013), dalam penelitiannya tentang evaluasi kinerja akademik mahasiswa, bahwa faktor yang paling berpengaruh dalam penentuan klasifikasi adalah IPK, IP semester 1-4 dan jenis kelamin. Data yang diuji adalah angkatan 2005 – 2009, sehingga dengan menggunakan algoritma *Naive Bayes* menghasilkan nilai *Precision*, *Recall* dan

Accuracy masing-masing 83%, 50% dan 70%. Begitu pula yang dilakukan oleh (Jananto, 2013) dalam penelitiannya mencari perkiraan waktu studi mahasiswa dengan menggunakan algoritma *Naïve Bayes* dapat dilakukan prediksi terhadap ketepatan waktu studi mahasiswa. Data latih dan data uji diambil secara random kemudian dilakukan pengujian terhadap fungsi yang dibuat untuk menentukan kelas pada data testing dengan tingkat kesalahan klasifikasi sebesar 20 *record* dari 66 data uji. Jadi tingkat kesalahan prediksi sebesar 34%. Jadi hasil prediksi dari ketepatan waktu lama studi mahasiswa angkatan 2008 terdapat 254 mahasiswa diprediksi “Tepat Waktu” dan 4 orang diprediksi “Tidak Tepat Waktu”.

Penelitian lain yang dilakukan oleh (Yetli Oslan, 2013) Implementasi *Naïve Bayes* untuk melakukan eksperimen data penjurusan SMA dan menghasilkan tingkat Keakuratan tertinggi dari hasil proses penjurusan dengan menggunakan range yang ditentukan secara manual berada pada range dengan interval 20 dan 25. Pada interval 20, angkatan 2009 mendapatkan sebesar 59%, 2010 sebesar 66%, 2011 sebesar 61%. Sedangkan pada Interval 25, angkatan 2009 mendapatkan sebesar 62%, 2010 sebesar 58%, 2011 sebesar 54%. Hasil proses penjurusan dengan menggunakan range yang didapatkan dari proses Box Plot rata-rata memiliki tingkat keakuratan lebih tinggi dibanding dengan range yang ditentukan secara manual. Pada range Box Plot, tingkat keakuratan angkatan 2009 adalah 62%, tingkat keakuratan angkatan 2010 adalah 63% dan angkatan 2011 adalah 66%. Berdasarkan proses yang telah dilakukan, hasil penjurusan dengan cara tersebut rata-rata memiliki keakuratan lebih tinggi karena seluruh siswa mendapatkan saran jurusan tanpa terkecuali.

Menurut (Manjusha K. K, 2014) Pada penelitian menyatakan bahwa metode *Naïve Bayes* juga digunakan untuk memprediksi penyakit Dermatology yang diabaikan tetapi dapat menyebabkan kematian dimana metode *Naïve Bayes* digunakan untuk mengenal pola data dan mengungkap kemungkinan penyakit dermatology. Selain itu (S.L. Ting, 2011) metode *Naïve Bayes* juga dinilai berpotensi baik dalam mengklasifikasi dokumen dibandingkan metode pengklasifikasian yang lain dalam hal ini akurasi dan efisiensi saat komputasi.

2.1.2 Penelitian Yang Terkait Dengan Metode *Adaboost*

Penelitian yang dilakukan oleh (Adi Suwondo, 2013) dengan menerapkan algoritma C4.5 berbasis *Adaboost* untuk prediksi penyakit jantung coroner, berdasarkan hasil klasifikasinya algoritma C4.5 menunjukkan nilai akurasi 91,38%, kemudian ditambahkan metode *Adaboost* nilai akurasi menjadi 94,84%. Sehingga dapat disimpulkan bahwa penerapan *Adaboost* pada algoritma C4.5 dapat meningkatkan akurasi prediksi penyakit jantung. Begitu pula yang dilakukan oleh (Lila D. Utami, 2015) yang penelitiannya menggunakan tiga metode yakni metode Informasi Gain, *Adaboost* dan *Naïve Bayes*. Ketiga metode tersebut digunakan untuk analisis Sentiment review restoran. Hasil dari penelitian ini menunjukkan bahwa peningkatan akurasi *Naïve Bayes* dari 73% menjadi 81,50% dan nilai AUC dari 0,500 jadi 0,887 sehingga dapat disimpulkan bahwa integrasi metode *Informasi Gain* dan *Adaboost* pada analisis sentiment review restoran ini mampu meningkatkan akurasi algoritma *Naïve Bayes*. *Adaboost* diterapkan pada *Naïve Bayes* dapat meningkatkan kinerja sebesar 33,33% dan memberikan hasil yang

akurat dengan mengurangi nilai kesalahan klasifikasi (Korada, 2012). Algoritma ini juga sangat dikenal dengan sebutan *idiot's Bayes*, *simple Bayes* dan *independent Bayes*.

Selain uraian diatas, dalam penelitian ini juga menyajikan beberapa penelitian-penelitian sebelumnya yang memiliki keterkaitan dengan penelitian yang dilakukan saat ini. Dengan melakukan perbandingan berdasarkan penggunaan metode, objek penelitian, serta hasil penelitian. Penelitian sebelumnya dapat dilihat pada tabel 2.1.

Tabel 2.1 Tabel perbandingan

Peneliti	(Amornsinlaphachai, 2014)
Judul	Efficiency of data mining models to predict academic performance and a cooperative learning model
Metode yang digunakan	Artificial Neural Network, K-Nearest Neighbor, Naive Bayes, Bayesian Belief Network, JRIP, ID3 and C4.5
Objek Penelitian	<ol style="list-style-type: none"> 1. Menggunakan model data mining untuk memprediksi prestasi akademik peserta didik dalam pemrograman computer. Membandingkan efisiensi model yang dibuat dari data mining dengan teknik klasifikasi. 2. Untuk mengembangkan Model pembelajaran kooperatif via web dengan menggunakan model data mining yang dipilih untuk mengelompokkan peserta didik
Hasil Penelitian	Model <i>Data Mining</i> yang dibangun menggunakan teknik klasifikasi dari ketujuh algoritma yaitu: <i>Artificial Neural</i>

	<p><i>Network, Tetangga K-Nearest, Naive Bayes, Bayesian Belief Network, JRIP, ID3 dan C4.5.</i> dari ketujuh metode tersebut algoritma C4.5 memiliki nilai akurasi yang lebih tinggi yaitu sebesar 74, 89 %.</p>
Peneliti	(Devasia, P and Hegde, 2016)
Judul Penelitian	Prediksi Kinerja Mahasiswa dengan menggunakan teknik <i>Data mining</i>
Metode yang digunakan	Algoritma <i>Naive Bayes</i>
Objek Penelitian	Membuat aplikasi berbasis web dengan memanfaatkan teknik Data Mining menggunakan algoritma <i>Naive Bayes</i> untuk mengekstraksi informasi yang berguna.
Hasil Penelitian	Percobaan dilakukan terhadap dataset mahasiswa sebanyak 700 rekor dengan 19 atribut dengan membandingkan beberapa algoritma antara lain yaitu: <i>Naive Bayes, Regression, Decision Tree, Neural network.</i> Dari hasil prediksi algoritma <i>Naive Bayes</i> lebih baik dibandingkan dengan metode lain dalam hal prediksi.
Peneliti	(Liu <i>et al.</i> , 2017)
Judul	A Preprocessing Method of AdaBoost for Mislabeled Data Classification
Metode yang digunakan	Adaboost, Decision Stump

Objek Penelitian	Melakukan preprocessing untuk memperbaiki terhadap data yang noise dan missing value dengan menggunakan <i>Adaboost konvensional</i> .
Hasil Penelitian	Hasil perbandingan antara simulasi <i>Adaboost</i> konvensional dan metode yang diusulkan menunjukkan bahwa algoritma yang diusulkan telah meningkatkan akurasi pengujian terhadap kumpulan data yang bising.
Peneliti	Arief, Jananto 2013
Judul Penelitian	Algoritma <i>Naive Bayes</i> Untuk Mencari Perkiraan Waktu Studi Mahasiswa
Metode yang digunakan	Algoritma <i>Naive Bayes</i>
Object Penelitian	Prediksi, Lama Studi, <i>Data Mining</i> , <i>Naive bayes</i> , <i>MySql</i>
Hasil Penelitian	Dari hasil pengujian diperoleh tingkat kesalahan sebesar 20% dari data latih dan data uji yang diambil secara acak. Tetapi tinggi rendahnya tingkat kesalahan rata-rata berkisar 20% sampai 34% karena dipengaruhi oleh jumlah data latih dan data uji yang digunakan.