

BAB II. TINJAUAN PUSTAKA

Penelitian tentang penerapan algoritma *decision tree* C4.5 pernah dilakukan oleh Mashlahah (2013), dengan judul "Prediksi Kelulusan Mahasiswa Menggunakan Metode *Decision Tree* Dengan Penerapan Algoritma C4.5". Penelitian tersebut menggunakan teknik *data mining* klasifikasi agar dapat mengolah data yang sangat besar untuk menemukan pola pada data mahasiswa. Algoritma yang digunakan pada penelitian tersebut adalah algoritma *decision tree* C4.5, yang digunakan untuk memprediksi kelulusan mahasiswa yang belum diketahui dari data mahasiswa. Atribut yang digunakan untuk melakukan klasifikasi adalah asal daerah, asal sekolah, jalur masuk, keterangan pesantren, Indeks Prestasi Kumulatif (IPK), Indeks Prestasi Semester (IPS) mulai dari semester 1 sampai 5, dan keterangan lulus sebagai label data. Objek yang diteliti pada penelitian tersebut adalah data mahasiswa yang sudah lulus pada tahun 2005 sampai 2008. Hasil penelitian pada penerapan algoritma *decision tree* C4.5 pada penelitian tersebut memiliki akurasi mencapai 82.79% [7].

Adapun penelitian lain dilakukan oleh Heryana (2019), dengan judul "*Data Mining* Untuk Memprediksi Kelulusan Mahasiswa Pendidikan Matematika UIN Raden Intan Lampung Menggunakan Naive Bayes". Pada penelitian tersebut digunakan teknik *data mining* klasifikasi untuk memprediksi kelulusan mahasiswa, dimana prediksi tersebut adalah tepat atau tidak tepat waktunya mahasiswa dalam menyelesaikan masa studinya. Algoritma yang digunakan pada penelitian tersebut adalah algoritma Naive Bayes. Algoritma ini digunakan untuk memprediksi masa depan menggunakan probabilitas yang menerapkan aturan Bayes. Aturan Bayes memprediksikan masa yang akan datang dengan menggunakan pengalaman sebelumnya. Atribut yang digunakan untuk melakukan klasifikasi adalah jenis kelamin, kota kelahiran, tipe sekolah, lokasi sekolah, ekonomi, Indeks Prestasi Kumulatif (IPK), dan Keputusan sebagai label data. Objek yang diteliti pada penelitian tersebut adalah data mahasiswa angkatan 2013 sebagai data latih dan data mahasiswa angkatan 2015 sebagai data uji. Hasil yang diperoleh pada penelitian tersebut menunjukkan nilai akurasi pada data latih sebesar 74,67% dari data yang

berjumlah 51. Hasil lain yang diperoleh pada penelitian tersebut menunjukkan bahwa data uji dari 184 mahasiswa angkatan 2015, diprediksi akan lulus tepat waktu sebanyak 42 mahasiswa atau sekitar 22,8%, dan mahasiswa yang lulus tidak tepat waktu sebanyak 142 mahasiswa atau sekitar 77,2% [8].

Penelitian lain dilakukan oleh Rohman dan Rufiyanto (2019), dengan judul "Implementasi *Data Mining* Dengan Algoritma *Decision Tree* C4.5 Untuk Prediksi Kelulusan Mahasiswa Di Universitas Pandanaran". Pada penelitian tersebut digunakan teknik *data mining* klasifikasi untuk menemukan pola dan memprediksi kelulusan mahasiswa. Algoritma yang digunakan pada penelitian tersebut adalah algoritma *decision tree* C4.5, yang digunakan untuk menyusun sistem yang mempunyai kemampuan melihat pola kelulusan mahasiswa. Atribut yang digunakan untuk melakukan klasifikasi adalah jurusan, umur, jenis kelamin, pekerjaan, Indeks Prestasi Kumulatif (IPK) 1 sampai 4, dan atribut label sebagai kelas data. Objek yang diteliti pada penelitian tersebut adalah data kelulusan mahasiswa di Universitas Pandanaran Semaarang jenjang pendidikan D3 di Fakultas Teknik. Hasil yang diperoleh pada penelitian tersebut menghasilkan nilai akurasi sebesar 65,98% dan nilai Area Under Curve (AUC) sebesar 0,874, dimana hasil ini dapat dikategorikan sebagai klasifikasi data yang baik [9].

Rohmawan juga melakukan penelitian tentang prediksi kelulusan mahasiswa pada tahun 2018, dengan judul "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Decision Tree* Dan *Artificial Neural Network*". Pada penelitian tersebut teknik *data mining* yang digunakan adalah klasifikasi untuk memprediksi kelulusan mahasiswa. Algoritma yang digunakan pada penelitian tersebut adalah algoritma *decision tree* dan *artificial neural network*, yang akan diimplementasikan untuk memprediksi kelulusan dan menemukan pola kelulusan mahasiswa. Atribut yang digunakan untuk melakukan klasifikasi adalah jenis kelamin, IP semester 1 sampai semester 6, IPK semester 1 sampai semester 6, total sks semester 1 sampai semester 6, semester 2 sampai semester 6, dan atribut keterangan sebagai label data. Objek yang diteliti pada penelitian ini adalah mahasiswa angkatan 2006 sampai 2010 sebagai data latih pada proses *data mining* dan mahasiswa angkatan 2011 sampai 2012 yang diasumsikan belum lulus sebagai

data uji pada proses *data mining*. Hasil yang diperoleh pada penelitian ini yaitu dapat digunakannya algoritma *decision tree* dengan akurasi 74.51% dan *artificial neural network* dengan akurasi 79.74%, untuk melakukan prediksi kelulusan mahasiswa tepat waktu.[10].

Penelitian lain dilakukan oleh Hakim, Rizal, dan Rantasari (2019), dengan judul "Aplikasi Prediksi Kelulusan Mahasiswa Berbasis *K-Nearens Neighbor* (K-NN)". Pada penelitian ini dibuat sebuah aplikasi yang digunakan untuk melakukan prediksi kelulusan mahasiswa. Teknik *data mining* yang digunakan adalah klasifikasi yang diimplementasikan ke dalam aplikasi. Algoritma yang digunakan pada pembuatan aplikasi prediksi kelulusan mahasiswa adalah algoritma *K-Nearest Neighbor*. Penggunaan algoritma ini bertujuan untuk mengidentifikasi kelulusan mahasiswa pada kasus baru, dengan cara mengadaptasi solusi dari kasus sebelumnya yang memiliki kedekatan dengan kasus baru. Atribut yang digunakan untuk melakukan klasifikasi adalah NIM, Nama Mahasiswa, Jenis Kelamin, Indeks Prestasi Semester 1 sampai 5, dan Keterangan Lulus sebagai label data. Objek yang diteliti pada penelitian ini adalah data – data kelulusan mahasiswa dan data mahasiswa yang masih aktif. Hasil yang diperoleh pada penelitian ini menghasilkan nilai akurasi 98% pada K=1 untuk klasifikasi "Tepat Waktu" [11].

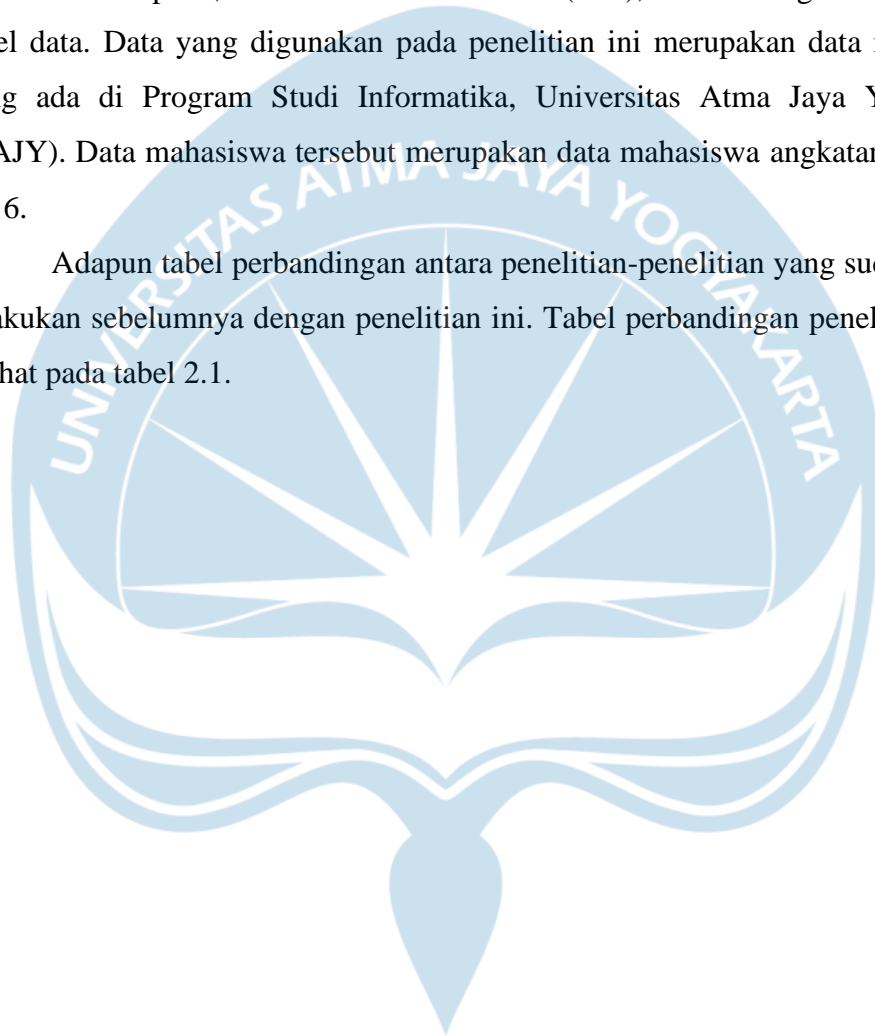
Penelitian lain dilakukan oleh Armansyah dan Kurniawan (2021), dengan judul "Model Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Naive Bayes*". Pada penelitian tersebut digunakan teknik *data mining* klasifikasi untuk melakukan prediksi kelulusan mahasiswa. Algoritma yang digunakan adalah algoritma *Naive Bayes*. Penelitian tersebut menggunakan sembilan atribut dan satu label yang meliputi JK (jenis kelamin), Indeks Prestasi Semester (IPS) dari semester 1 sampai 4, SKS 4 (jumlah sks yang diambil pada semester 4), lama TA (tugas akhir), masa studi, jalur masuk, dan KL (kelulusan) sebagai label data. Objek yang diteliti adalah data mahasiswa mulai dari angkatan 2015 sampai 2019 yang telah lulus pada oktober 2020. Data yang digunakan adalah sebanyak 44 data mahasiswa yang kemudian dibagi menjadi 2, dengan 38 data mahasiswa sebagai data latih dan 6 data mahasiswa sebagai data uji. Hasil yang diperoleh pada penelitian ini menghasilkan nilai akurasi prediksi sebesar 100% [12].

Penelitian tentang prediksi menggunakan algoritma *decision tree* C4.5 juga dilakukan oleh Noviandi (2018), dengan judul "Implementasi Algoritma *Decision Tree* C4.5 Untuk Prediksi Penyakit Diabetes". Pada penelitian tersebut Teknik *data mining* yang digunakan adalah klasifikasi, yang digunakan untuk mengklasifikasikan dan memprediksi penyakit diabetes. Algoritma yang digunakan adalah algoritma *decision tree* C4.5 yang diimplementasikan untuk memprediksi penyakit diabetes. Atribut yang digunakan untuk melakukan klasifikasi adalah jumlah kehamilan pasien, kadar gula darah, tekanan darah, tingkat insulin, *Body Mass Index* (BMI), usia, dan *outcome* sebagai label data. Objek yang diteliti pada penelitian tersebut adalah data kesehatan *Pima Indians Diabetes Dataset* (PIDD). Hasil yang diperoleh pada penelitian ini menghasilkan nilai akurasi sebesar 70,32% yang menghasilkan 9 *rule* [13].

Penelitian lain dilakukan oleh Pratama, Wihandika, dan Ratnawati (2018), dengan judul "Implementasi Algoritme *Support Vector Machine* (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa". Teknik *data mining* yang digunakan pada penelitian tersebut adalah klasifikasi, yang digunakan untuk mengklasifikasikan dan memprediksi ketepatan waktu kelulusan mahasiswa. Algoritma yang digunakan pada penelitian tersebut adalah algoritma *Support Vector Machine* (SVM). Algoritma ini berusaha menemukan *hyperplane* (pemisah) terbaik untuk memisahkan klasifikasi dua kelas dan memaksimalkan margin antara dua kelas tersebut. Objek yang diteliti pada penelitian tersebut adalah *dataset* dari bagian akademik Filkom yang berjumlah 188 data. Proses pengujian dilakukan beberapa kali dengan setiap pengujian mengubah jumlah data latih dan data uji. Hasil yang diperoleh pada penelitian ini menghasilkan nilai akurasi rata-rata sebesar 80,55% [14].

Penelitian yang diteliti pada penelitian ini adalah menerapkan algoritma *decision tree* C4.5 dan metode *AdaBoost* pada data mahasiswa untuk mendapatkan informasi yang belum diketahui pada data tersebut. Teknik *data mining* yang digunakan pada penelitian ini adalah klasifikasi. Atribut yang digunakan untuk melakukan klasifikasi pada penelitian ini adalah tahun masuk, sks kumulatif, IP Semester 1 sampai 8, Indeks Prestasi Kumulatif (IPK), dan keterangan lulus sebagai label data. Data yang digunakan pada penelitian ini merupakan data mahasiswa yang ada di Program Studi Informatika, Universitas Atma Jaya Yogyakarta (UAJY). Data mahasiswa tersebut merupakan data mahasiswa angkatan 2012 dan 2016.

Adapun tabel perbandingan antara penelitian-penelitian yang sudah pernah dilakukan sebelumnya dengan penelitian ini. Tabel perbandingan penelitian dapat dilihat pada tabel 2.1.



Tabel 2.1. Tabel Perbandingan Penelitian

Peneliti	Algoritma	Objek	Atribut	Metode <i>Ensemble</i>	Tempat Objek yang diteliti	Hasil
[1]	<i>Decision Tree C4.5</i>	Mahasiswa	Atribut Klasifikasi: asal daerah, asal sekolah, jalur masuk, keterangan pesantren, IPK, dan IPS 1 sampai 5; Label Data: Keterangan lulus	Tidak Ada	UIN Maulana Malik Ibrahim Malang	Akurasi = 82,79% (90 data sampel)
[2]	<i>Naive Bayes</i>	Mahasiswa	Atribut Klasifikasi: jenis kelamin, kota kelahiran, tipe sekolah, lokasi sekolah, ekonomi, dan IPK; Label Data: Keputusan	Tidak Ada	UIN Raden Intan Lampung	Akurasi = 74,67%
[3]	<i>Decision Tree C4.5</i>	Mahasiswa	Atribut Klasifikasi: jurusan, umur, jenis kelamin, pekerjaan, dan IPK semester 1 sampai 4; Label Data: Label	Tidak Ada	Universitas Pandanaran	Akurasi = 65,98% <i>Area Under Curve (AUC) = 0,874</i>

[4]	<i>Decision Tree</i> dan <i>Artificial Neural Network</i>	Mahasiswa	Atribut Klasifikasi: jenis kelamin, IPS 1 sampai 6, IPK semester 1 sampai 6, total sks semester 1 sampai 6, dan semester 2 sampai 6; Label Data: Keterangan	Tidak Ada	Tidak Diketahui	Akurasi (<i>Decision Tree</i>) = 74,51% Akurasi (<i>Artificial Neural Network</i>) = 79,74%
[5]	<i>K-Nearest Neighbor</i> (K-NN)	Mahasiswa	Atribut Klasifikasi: nim, nama mahasiswa, jenis kelamin, dan IPS 1 sampai 5; Label Data: Keterangan Lulus	Tidak Ada	Universitas Bumigora	Akurasi tertinggi (nilai K = 1) = 98%
[6]	<i>Naive Bayes</i>	Mahasiswa	Atribut Klasifikasi: jenis kelamin, IPS 1 sampai 4, SKS 4, lama TA, lama studi, dan jalur masuk; Label Data: KL (Kelulusan)	Tidak Ada	UIN Sumatera Utara Medan	Akurasi = 100%
[7]	<i>Decision Tree</i> C4.5	Penyakit Diabetes	Atribut Klasifikasi: jumlah kehamilan pasien, kadar gula	Tidak Ada	<i>Pima Indians Diabetes Dataset</i> (PPID)	Akurasi = 70,32%

			darah, tekanan darah, tingkat insulin, <i>Body Mass Index</i> (BMI), dan usia; Label Data: <i>Outcome</i>			
[8]	<i>Support Vector Machine</i> (SVM)	Mahasiswa	-	Tidak Ada	Universitas Brawijaya	Akurasi = 80,55%

Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada digunakannya algoritma *AdaBoost*. Penelitian ini menggunakan metode *ensemble* yaitu algoritma *AdaBoost* untuk memperkuat hasil performa dari menerapkan algoritma *decision tree* C4.5. Untuk aspek perbandingan lainnya tidak memiliki perbedaan yang signifikan.