

**PERBANDINGAN ALGORITMA C4.5 DAN K-MEANS
DALAM MEMPREDIKSI KATEGORI INDEKS PRESTASI
MAHASISWA**

Skripsi



Diajukan oleh:

GABRIELLA AMELIA PRASETYO

71150064

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
YOGYAKARTA

2019

**PERBANDINGAN ALGORITMA C4.5 DAN K-MEANS
DALAM MEMPREDIKSI KATEGORI INDEKS PRESTASI
MAHASISWA**

Skripsi



Diajukan kepada Fakultas Teknologi Informasi Program Studi Informatika
Universitas Kristen Duta Wacana
Sebagai salah satu syarat dalam memperoleh gelar Sarjana Komputer

Diajukan oleh:
GABRIELLA AMELIA PRASETYO
71150064

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA
2019

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

PERBANDINGAN ALGORITMA C4.5 DAN K-MEANS DALAM MEMPREDIKSI KATEGORI INDEKS PRESTASI MAHASISWA

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 17 Oktober 2019



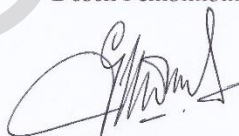
GABRIELLA AMELIA PRASETYO
71150064

HALAMAN PERSETUJUAN

Judul Skripsi : PERBANDINGAN ALGORITMA C4.5 DAN K-
MEANS DALAM MEMPREDIKSI KATEGORI
INDEKS PRESTASI MAHASISWA
Nama Mahasiswa : GABRIELLA AMELIA PRASETYO
N I M : 71150064
Matakuliah : Skripsi (Tugas Akhir)
Kode : TIW276
Semester : Gasal
Tahun Akademik : 2019/2020

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 17 Oktober 2019

Dosen Pembimbing I



R. Gunawan Santosa, Drs. M.Si.

Dosen Pembimbing II



Antonius Rachmat C., S.Kom., M.Cs.

HALAMAN PENGESAHAN

PERBANDINGAN ALGORITMA C4.5 DAN K-MEANS DALAM MEMPREDIKSI KATEGORI INDEKS PRESTASI MAHASISWA

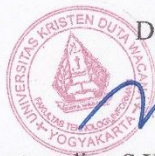
Oleh: GABRIELLA AMELIA PRASETYO / 71150064

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 15 Oktober 2019

Yogyakarta, 17 Oktober 2019
Mengesahkan,

Dewan Penguji:

1. R. Gunawan Santosa, Drs. M.Si.
2. Antonius Rachmat C., S.Kom., M.Cs.
3. Maria Nila Anggia Rini, S.T, M.T.I
4. Danny Sebastian, S.Kom., M.M., M.T.



Dekan

(Restyandito, S.Kom., MSIS., Ph.D.)

Ketua Program Studi


(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul Perbandingan Algoritma C4.5 dan k-Means Dalam Memprediksi Kategori Indeks Prestasi Mahasiswa. Laporan Tugas Akhir ini dibuat untuk memenuhi salah satu syarat kelulusan untuk memperoleh gelar Sarjana Komputer.

Dalam menyelesaikan Tugas Akhir ini, penulis mendapatkan banyak bimbingan, dan saran dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan banyak terima kasih kepada:

1. Papa, mama, dan keluarga yang selalu memberikan dukungan dan doa
2. Bapak Drs. R. Gunawan Santosa, M.Si. selaku dosen pembimbing pertama
3. Bapak Antonius Rachmat Chrismanto, S.Kom., M.Cs. selaku dosen pembimbing kedua
4. Teman-teman yang selalu menemani, mendukung, dan memberi semangat (Chika, Rico, Albert, Kevin, Yadi, Dimas, Audri, Zenas, dan yang tidak bisa disebutkan satu per satu)

Laporan Tugas Akhir ini tidak lepas dari kesalahan dan kekurangan, oleh karena itu saran dan kritik yang bersifat membangun sangat penulis harapkan. Semoga laporan Tugas Akhir ini dapat bermanfaat bagi penulis dan pembaca.

DAFTAR ISI

PERNYATAAN KEASLIAN SKRIPSI.....	Error! Bookmark not defined.
HALAMAN PERSETUJUAN.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN.....	Error! Bookmark not defined.
UCAPAN TERIMA KASIH.....	vi
INTISARI.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xiii
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian.....	2
1.5. Metodologi Penelitian	3
1.6. Sistematika Penulisan.....	4
BAB 2 TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	5
2.1. Tinjauan Pustaka	5
2.2. Landasan Teori	6
2.2.1. Data Mining	6
2.2.2. Decision Tree	8
2.2.3. Clustering	8
2.2.4. Algoritma C4.5.....	9
2.2.5. Algoritma k-Means Clustering.....	11
2.2.6. Cross-tabulation	12
BAB 3 METODOLOGI PENELITIAN.....	14
3.1. Spesifikasi Sistem.....	14
3.2. Sumber Data	14
3.3. Diagram Alir Sistem.....	15

3.4.	Perancangan Skema Basis Data	21
3.5.	Perancangan Antarmuka.....	22
3.6.	Rancangan Pengujian Sistem	31
3.7.	Contoh Kasus	34
3.7.1.	Contoh Kasus Algoritma C4.5	34
3.7.2.	Contoh Kasus Algoritma k-Means.....	42
3.7.3.	Contoh Kasus Cross-tabulation.....	47
BAB 4 IMPLEMENTASI DAN PENGUJIAN		49
4.1.	Implementasi Sistem	49
4.1.1.	Pengolahan Data.....	49
4.1.2.	Instalasi	49
4.1.3.	Implementasi Algoritma C4.5.....	50
4.1.4.	Implementasi Algoritma k-Means.....	51
4.1.5.	Antarmuka Sistem.....	52
4.2.	Pengujian dan Analisis Sistem	61
BAB 5 KESIMPULAN DAN SARAN		85
5.1.	Kesimpulan.....	85
5.2.	Saran	85
DAFTAR PUSTAKA		87
LAMPIRAN A		89
LAMPIRAN B		93

DAFTAR GAMBAR

Gambar 2.1. Data mining sebagai bagian dari proses KDD	7
Gambar 2.2. Decision tree untuk pembelian komputer.....	8
Gambar 2.3. Clustering pada data ke dalam 4 cluster	9
Gambar 3.1. Diagram alir secara keseluruhan	15
Gambar 3.2. Diagram alir prediksi menggunakan skenario	16
Gambar 3.3. Diagram alir prediksi data satuan	17
Gambar 3.4. Diagram alir algoritma C4.5	18
Gambar 3.5. Diagram alir algoritma k-Means clustering	19
Gambar 3.6. Diagram alir penghitungan akurasi dengan cross-tabulation	20
Gambar 3.7. Rancangan basis data	21
Gambar 3.8. Rancangan antarmuka halaman utama	22
Gambar 3.9. Rancangan antarmuka halaman data latih jalur prestasi	22
Gambar 3.10. Rancangan antarmuka halaman data latih jalur non-prestasi	23
Gambar 3.11. Rancangan antarmuka halaman data uji jalur prestasi	24
Gambar 3.12. Rancangan antarmuka halaman data uji jalur non-prestasi	24
Gambar 3.13. Rancangan antarmuka halaman prediksi algoritma C4.5 jalur prestasi	25
Gambar 3.14. Rancangan antarmuka halaman prediksi algoritma C4.5 jalur non-prestasi	25
Gambar 3.15. Rancangan antarmuka halaman prediksi algoritma k-Means jalur prestasi	26
Gambar 3.16. Rancangan antarmuka halaman prediksi algoritma k-Means jalur non-prestasi	26
Gambar 3.17. Rancangan antarmuka halaman perbandingan akurasi prediksi ..	27
Gambar 3.18. Rancangan antarmuka halaman data latih satuan	28

Gambar 3.19. Rancangan antarmuka halaman data uji satuan jalur prestasi	28
Gambar 3.20. Rancangan antarmuka halaman data uji satuan jalur non-prestasi	29
Gambar 3.21. Rancangan antarmuka halaman cek data uji satuan jalur prestasi	29
Gambar 3.22. Rancangan antarmuka halaman cek data uji satuan jalur non-prestasi.....	30
Gambar 3.23. Rancangan antarmuka halaman hasil prediksi satuan jalur prestasi	30
Gambar 3.24. Rancangan antarmuka halaman hasil prediksi satuan jalur non-prestasi	31
Gambar 3.25. Decision tree dengan lokasi sekolah sebagai root node	37
Gambar 3.26. Lokasi sekolah di Jawa memprediksi kategori IPS “0”	37
Gambar 3.27. Kategori sekolah sebagai node tambahan	39
Gambar 3.28. Level bahasa Inggris sebagai node tambahan	41
Gambar 3.29. Decision tree yang dibangun dengan algoritma C4.5	41
Gambar 4.1. File CSV data mahasiswa FTI UKDW angkatan 2008 sampai 2017	49
Gambar 4.2. Cara import library Chefboost	50
Gambar 4.3. Cara menggunakan library Chefboost untuk membuat decision tree	50
Gambar 4.4. Antarmuka halaman utama	52
Gambar 4.5. Antarmuka halaman data latih	53
Gambar 4.6. Antarmuka halaman data uji	54
Gambar 4.7. Antarmuka halaman hasil prediksi C4.5	55
Gambar 4.8. Antarmuka halaman hasil prediksi k-Means	56
Gambar 4.9. Antarmuka halaman cross-tabulation	57
Gambar 4.10. Antarmuka halaman data latih satuan	57

Gambar 4.11. Antarmuka halaman form data uji satuan jalur prestasi	58
Gambar 4.12. Antarmuka halaman form data uji satuan jalur non-prestasi	58
Gambar 4.13. Antarmuka halaman cek form data uji satuan jalur prestasi	59
Gambar 4.14. Antarmuka halaman cek form data uji satuan jalur non-prestasi.	60
Gambar 4.15. Antarmuka halaman hasil prediksi data satuan jalur prestasi	60
Gambar 4.16. Antarmuka halaman hasil prediksi data satuan jalur non-prestasi	61
Gambar 4.17. Grafik hasil pengujian berdasarkan skenario	80
Gambar 4.18. Grafik hasil pengujian berdasarkan algoritma	80
Gambar 4.19. Grafik hasil pengujian berdasarkan jalur	81
Gambar 4.20. Grafik hasil pengujian berdasarkan keseimbangan data	82
Gambar 4.21. Grafik hasil pengujian berdasarkan binning	82
Gambar 4.22. Grafik hasil pengujian berdasarkan kategori.....	83

DAFTAR TABEL

Tabel 2.1. Cross-tabulation antara data aktual dengan data prediksi	12
Tabel 3.1. Jumlah mahasiswa FTI UKDW angkatan 2008-2017	14
Tabel 3.2. Skenario pengujian sistem	32
Tabel 3.3. Binning untuk atribut numerik	33
Tabel 3.4. Pembagian 2 kategori IPS	33
Tabel 3.5. Pembagian 3 kategori IPS	33
Tabel 3.6. Data latih jalur prestasi	34
Tabel 3.7. Data uji jalur prestasi	35
Tabel 3.8. Data dengan lokasi sekolah di Jawa	37
Tabel 3.9. Data dengan lokasi sekolah di luar Jawa	38
Tabel 3.10. Data dengan lokasi sekolah di luar Jawa dan kategori sekolah SMA	40
Tabel 3.11. Hasil prediksi kategori IPS mahasiswa dengan algoritma C4.5	42
Tabel 3.12. Centroid yang dipilih secara acak dari data latih	43
Tabel 3.13. Hasil perhitungan euclidian distance dan pengelompokkan dalam cluster baru	43
Tabel 3.14. Hasil perhitungan centroid baru setelah iterasi pertama	44
Tabel 3.15. Hasil perhitungan iterasi kedua	45
Tabel 3.16. Hasil penghitungan jarak sesungguhnya dan prediksi	46
Tabel 3.17. Hasil pengelompokkan dalam cluster berdasarkan nilai sesungguhnya dan prediksi	46
Tabel 3.18. Centroid dan anggota C1	46
Tabel 3.19. Centroid dan anggota C2	47
Tabel 4.1. Total data latih dan data uji pada setiap skenario	62

Tabel 4.2. Cross-tabulation skenario 1 (C4.5, jalur prestasi, data seimbang, 2 kategori IPS)	62
Tabel 4.3. Cross-tabulation skenario 2 (C4.5, jalur prestasi, data seimbang, 3 kategori IPS)	63
Tabel 4.4. Cross-tabulation skenario 3 (C4.5, jalur prestasi, data tidak seimbang, 2 kategori IPS)	63
Tabel 4.5. Cross-tabulation skenario 4 (C4.5, jalur prestasi, data tidak seimbang, 3 kategori IPS)	63
Tabel 4.6. Cross-tabulation skenario 5 (k-Means, jalur prestasi, data seimbang, 2 kategori IPS)	64
Tabel 4.7. Cross-tabulation skenario 6 (k-Means, jalur prestasi, data seimbang, 3 kategori IPS)	64
Tabel 4.8. Cross-tabulation skenario 7 (k-Means, jalur prestasi, data tidak seimbang, 2 kategori IPS)	64
Tabel 4.9. Cross-tabulation skenario 8 (k-Means, jalur prestasi, data tidak seimbang, 3 kategori IPS)	65
Tabel 4.10. Cross-tabulation skenario 9 (C4.5, jalur non-prestasi, data seimbang, binning, 2 kategori IPS)	65
Tabel 4.11. Cross-tabulation skenario 10 (C4.5, jalur non-prestasi, data seimbang, binning, 3 kategori IPS)	66
Tabel 4.12. Cross-tabulation skenario 11 (C4.5, jalur non-prestasi, data seimbang, tidak binning, 2 kategori IPS)	66
Tabel 4.13. Cross-tabulation skenario 12 (C4.5, jalur non-prestasi, data seimbang, tidak binning, 3 kategori IPS)	66
Tabel 4.14. Cross-tabulation skenario 13 (C4.5, jalur non-prestasi, data tidak seimbang, binning, 2 kategori IPS)	67
Tabel 4.15. Cross-tabulation skenario 14 (C4.5, jalur non-prestasi, data tidak seimbang, binning, 3 kategori IPS)	67
Tabel 4.16. Cross-tabulation skenario 15 (C4.5, jalur non-prestasi, data tidak seimbang, tidak binning, 2 kategori IPS)	68

Tabel 4.17. Cross-tabulation skenario 16 (C4.5, jalur non-prestasi, data tidak seimbang, tidak binning, 3 kategori IPS)	68
Tabel 4.18. Cross-tabulation skenario 17 (k-Means, jalur non-prestasi, data seimbang, binning, 2 kategori IPS)	68
Tabel 4.19. Cross-tabulation skenario 18 (k-Means, jalur non-prestasi, data seimbang, binning, 3 kategori IPS)	69
Tabel 4.20. Cross-tabulation skenario 19 (k-Means, jalur non-prestasi, data seimbang, tidak binning, 2 kategori IPS)	69
Tabel 4.21. Cross-tabulation skenario 20 (k-Means, jalur non-prestasi, data seimbang, tidak binning, 3 kategori IPS)	70
Tabel 4.22. Cross-tabulation skenario 21 (k-Means, jalur non-prestasi, data tidak seimbang, binning, 2 kategori IPS)	70
Tabel 4.23. Cross-tabulation skenario 22 (k-Means, jalur non-prestasi, data tidak seimbang, binning, 3 kategori IPS)	70
Tabel 4.24. Cross-tabulation skenario 23 (k-Means, jalur non-prestasi, data tidak seimbang, tidak binning, 2 kategori IPS)	71
Tabel 4.25. Cross-tabulation skenario 24 (k-Means, jalur non-prestasi, data tidak seimbang, tidak binning, 3 kategori IPS)	71
Tabel 4.26. Akurasi setiap skenario dan rata-ratanya	72
Tabel 4.27. Akurasi skenario dengan algoritma C4.5 dan rata-ratanya	73
Tabel 4.28. Akurasi skenario dengan algoritma k-Means dan rata-ratanya	73
Tabel 4.29. Akurasi skenario jalur prestasi dan rata-ratanya	74
Tabel 4.30. Akurasi skenario jalur non-prestasi dan rata-ratanya	75
Tabel 4.31. Akurasi skenario data latih seimbang dan rata-ratanya	76
Tabel 4.32. Akurasi skenario data latih tidak seimbang dan rata-ratanya	76
Tabel 4.33. Akurasi skenario dengan binning dan rata-ratanya	77
Tabel 4.34. Akurasi skenario tanpa binning dan rata-ratanya	78
Tabel 4.35. Akurasi skenario dengan 2 kategori dan rata-ratanya	79

Tabel 4.36. Akurasi skenario dengan 3 kategori dan rata-ratanya	79
Tabel 4.37. Perbandingan persentase nilai akurasi terbaik algoritma C4.5 (Gain Ratio), k-Means, C4.5 (Gain), CART, Regresi Logistik, K-Nearest Neighbor, dan Naïve Bayes Classifier	84

©UKPDN

BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

Setiap tahunnya, Universitas Kristen Duta Wacana (UKDW) menerima sejumlah mahasiswa baru. Jalur penerimaan mahasiswa baru dibagi menjadi dua, yaitu: jalur prestasi dan jalur non-prestasi. Jalur prestasi adalah jalur penerimaan mahasiswa baru berdasarkan nilai rapor calon mahasiswa, sedangkan jalur non-prestasi adalah jalur penerimaan mahasiswa baru berdasarkan hasil tes seleksi (Universitas Kristen Duta Wacana, n.d.). Data yang didapatkan dari jalur prestasi adalah kategori sekolah (SMA atau SMK), status sekolah (Negeri atau Swasta) dan lokasi sekolah (Jawa atau luar Jawa), sedangkan data yang didapatkan dari jalur non-prestasi adalah data yang sama dengan jalur prestasi, ditambah dengan kemampuan Spatial, Verbal, Numerik dan Analogi. Selain itu, kemampuan bahasa Inggris mahasiswa juga menjadi pertimbangan penerimaan mahasiswa, dimana calon mahasiswa diklasifikasikan ke dalam empat level kemampuan bahasa Inggris, yaitu: level 1, 2, 3 atau ESP (*English for Special Purpose*) (Santosa & Chrismanto, 2018).

Data penerimaan mahasiswa yang didapatkan oleh UKDW mengandung banyak informasi. Dari data tersebut dapat dilakukan pencarian pola dengan metode *data mining* atau penambangan data. Hasil dari penambangan data adalah informasi mengenai pola dari suatu data dan dapat digunakan untuk memprediksi data yang akan datang. Dengan kata lain, data penerimaan mahasiswa yang sudah ada dapat digunakan untuk memprediksi data mahasiswa yang akan datang. Salah satu implementasi dari penambangan data ini adalah untuk memprediksi indeks prestasi mahasiswa baru pada semester 1. Dari data mahasiswa yang sudah ada akan dicari pola yang menarik untuk memprediksi nilai indeks prestasi mahasiswa dengan data pribadi sesuai dengan jalur penerimaannya (jalur prestasi atau non-prestasi). Hasil prediksi ini dapat dimanfaatkan sebagai acuan baru dalam penerimaan mahasiswa,

juga membantu pengajar memberi perhatian lebih kepada mahasiswa baru dengan nilai indeks prestasi tertentu.

Penelitian ini merupakan lanjutan dari penelitian Santosa dan Chrismanto (2017), Santosa dan Chrismanto (2018), Alverina (2017), dan Sari (2017), dimana perbedaannya terletak pada algoritma dan data mahasiswa yang digunakan. Penelitian terdahulu menggunakan data mahasiswa angkatan 2008-2014 sebagai data latih dan angkatan 2015 sebagai data uji, sedangkan pada penelitian ini menggunakan data mahasiswa angkatan 2008-2016 sebagai data latih dan angkatan 2017 sebagai data uji.

1.2. Rumusan Masalah

Permasalahan yang akan dibahas pada penelitian ini adalah:

1. Berapa tingkat akurasi algoritma C4.5 dan k-Means *clustering* dalam memprediksi indeks prestasi mahasiswa FTI UKDW?
2. Algoritma manakah yang lebih akurat dalam memprediksi indeks prestasi mahasiswa FTI UKDW?

1.3. Batasan Masalah

Pada penelitian ini, penulis memberikan batasan masalah, antara lain:

1. Penelitian dilakukan terhadap indeks prestasi mahasiswa Fakultas Teknologi Informasi (FTI) di Universitas Kristen Duta Wacana Yogyakarta melalui jalur prestasi dan non-prestasi
2. Data yang digunakan adalah data indeks prestasi mahasiswa FTI UKDW pada tahun 2008 hingga 2017
3. Data yang digunakan adalah indeks prestasi mahasiswa pada semester 1
4. Data mahasiswa angkatan 2008-2016 akan digunakan sebagai data latih, dan data mahasiswa angkatan 2017 akan digunakan sebagai data uji

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah menerapkan dan membandingkan akurasi algoritma C4.5 dan k-Means *clustering* dalam memprediksi kategori IPK mahasiswa FTI UKDW.

1.5. Metodologi Penelitian

Dalam penyelesaian penelitian ini, penulis menggunakan beberapa metode, yaitu:

1. Studi Pustaka

Studi Pustaka dilakukan dengan mempelajari literatur berupa buku, jurnal dan tulisan ilmiah yang memuat informasi berkaitan dengan *data mining*, algoritma C4.5, algoritma k-Means *clustering* dan cara implementasinya ke dalam sistem.

2. Pengumpulan Data

Pada tahap ini, penulis akan mengumpulkan data dari penelitian yang sudah dilakukan sebelumnya. Data yang digunakan adalah data mahasiswa FTI UKDW angkatan 2008-2017 yang didapatkan dari penelitian Santosa dan Chrismanto (2018). Data mahasiswa dibedakan menjadi dua, yaitu jalur prestasi dan jalur non-prestasi. Jalur prestasi memiliki empat atribut, yaitu kategori sekolah (SMA atau SMK), status sekolah (Negeri atau Swasta), lokasi sekolah (Jawa atau luar Jawa), dan level kemampuan bahasa Inggris (level 1, 2, 3 atau ESP). Jalur non-prestasi memiliki delapan atribut, dimana empat atribut adalah atribut sama dengan jalur prestasi ditambah dengan empat atribut lainnya, yaitu kemampuan Spatial, Verbal, Numerik dan Analogi.

3. Perancangan Sistem

Dalam tahapan ini, dilakukan perancangan antarmuka dan basis data yang akan digunakan dalam sistem yang dibuat. Data yang digunakan sebagai data latih adalah data mahasiswa FTI UKDW angkatan 2008-2016, sedangkan data yang digunakan sebagai data uji adalah data mahasiswa FTI UKDW angkatan 2017.

4. Pembangunan Sistem

Pembangunan sistem dilakukan dengan menggunakan perangkat lunak Django dengan bahasa pemrograman Python dan SQLite untuk basis data.

5. Pengujian Sistem

Tahapan terakhir dari penelitian ini adalah pengujian sistem. Pengujian keakuratan hasil prediksi algoritma C4.5 dan k-Means *clustering* akan dilakukan dengan menggunakan *cross-tabulation*. Prediksi akan dilakukan terhadap nilai IPK semester 1 mahasiswa FTI UKDW angkatan 2017.

1.6. Sistematika Penulisan

Sistematika laporan tugas akhir ini dapat dibagi ke dalam 5 bab, yaitu:

Bab 1 Pendahuluan yang menjelaskan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian, dan sistematika penulisan laporan tugas akhir.

Bab 2 Tinjauan Pustaka dan Landasan Teori yang menjelaskan tinjauan pustaka dan landasan teori yang digunakan oleh penulis sebagai dasar dalam melakukan penelitian. Tinjauan Pustaka berisi rangkuman dari penelitian-penelitian terdahulu yang berkaitan dengan penelitian dalam penggunaan algoritma C4.5 dan k-Means *clustering*. Landasan Teori berisi penjelasan tentang teori dasar dalam melakukan penelitian yaitu, *data mining*, *decision tree*, *clustering*, algoritma C4.5, algoritma k-Means *clustering*, dan *cross-tabulation*.

Bab 3 Metodologi Penelitian yang menjelaskan tentang metode, data, dan alat yang digunakan dalam penelitian. Bab ini juga mencakup perancangan penelitian dan perancangan pengujian penelitian.

Bab 4 Implementasi dan Pengujian yang memuat tahap implementasi, hasil, pembahasan, dan pengujian sistem yang dilakukan pada penelitian.

Bab 5 Kesimpulan dan Saran yang berisi kesimpulan dari hasil penelitian yang dilakukan dan saran-saran untuk kegiatan penelitian dengan yang berkaitan dengan penelitian ini.

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Kesimpulan yang didapat dari penelitian ini adalah:

1. Pada skenario jalur prestasi dengan data latih sebanyak 905 data dan data uji sebanyak 102 data, algoritma C4.5 mendapat hasil terbaik sebesar 77,45%, sedangkan algoritma k-Means mendapat hasil terbaik sebesar 60,78%.
2. Pada skenario jalur non-prestasi dengan data latih sebanyak 1322 data dan data uji sebanyak 99 data, algoritma C4.5 mendapat hasil terbaik sebesar 46,46%, sedangkan algoritma k-Means mendapat hasil terbaik sebesar 43,43%.
3. Berdasarkan skenario dengan jalur penerimaannya, algoritma C4.5 mendapatkan hasil yang lebih baik dibandingkan dengan algoritma k-Means.
4. Skenario dengan jalur prestasi memiliki hasil yang lebih baik dibandingkan dengan jalur non-prestasi, dimana rata-rata akurasi skenario jalur prestasi adalah 55,27% sedangkan rata-rata akurasi skenario jalur non-prestasi adalah 38,95%.
5. Skenario dengan 2 kategori IPS memiliki hasil yang lebih baik dibandingkan dengan 3 kategori IPS.
6. Angkatan 2017 FTI UKDW (Data Uji) menggunakan kurikulum yang berbeda dari angkatan-angkatan sebelumnya, sehingga mempengaruhi hasil prediksi.

5.2. Saran

Saran yang dapat penulis berikan untuk pengembangan penelitian selanjutnya adalah:

1. Dapat ditambahkan atribut apakah sekolah tersebut favorit atau tidak di lingkungan sekolah tersebut. Mahasiswa lulusan sekolah favorit bisa saja

memiliki nilai IPS yang lebih bagus dibanding dengan lulusan sekolah bukan favorit.

2. Dapat ditambahkan nilai rapor sekolah mahasiswa untuk menjadi atribut tambahan, dimana asumsinya mahasiswa dengan nilai rapor sekolah yang baik akan mendapatkan nilai IPS yang baik pula.
3. Penelitian selanjutnya dapat mempertimbangkan pergantian kurikulum antar tahun angkatan.

©UKDW

DAFTAR PUSTAKA

- Alverina, D. (2017). *Perbandingan Algoritma C4.5 Dan Cart Dalam Memprediksi Kategori indeks prestasi mahasiswa*. (Undergraduate thesis, Duta Wacana Christian University, 2017). Retrieved from <http://sinta.ukdw.ac.id>
- Berry, M. J., & Linoff, G. S. (2000). *Mastering Data Mining: The Art and Science of Customer Relationship Management*. New York: Wiley Computer Publishing.
- Bramer, M. (2013). *Principles of Data Mining* (2nd ed.). London: Springer-Verlag London.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and Techniques* (3rd ed.). Amsterdam: Elsevier.
- Hssina, B., Merbouha, A., Ezzikouri, H., & Erritali, M. (2014). A comparative study of decision tree ID3 and C4.5. *International Journal of Advanced Computer Science and Applications*, 4(2). doi: 10.14569/specialissue.2014.040203
- Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, and Algorithms* (2nd ed.). John Wiley & Sons.
- Oyelade, O. J., Oladipupo, O. O., & Obagbuwa, I. C. (2010). Application of k-Means Clustering algorithm for prediction of Students' Academic Performance. *International Journal of Computer Science and Information Security*, 7, 292-295.
- Universitas Kristen Duta Wacana. (n.d.). Retrieved from <https://www.ukdw.ac.id/pendaftaran-program-sarjana/>
- Ramadhani, R. D. (2014). *Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro*. (Undergraduate thesis, Universitas Dian Nuswantoro, 2014). Retrieved from <http://eprints.dinus.ac.id>
- Santosa, R. G., & Chrismanto, A. R. (2017). Logistic Regression Model for Predicting First Semester Students GPA Category Based on High School Academic Achievement. *Journal of Arts, Science & Commerce*, VIII(2), 1st ser., 58-66.

Santosa, R. G., & Chrismanto, A. R. (2018). Perbandingan Akurasi Model Regresi Logistik untuk Prediksi Kategori IP Mahasiswa Jalur Prestasi dengan Non Jalur Prestasi. *Jurnal Teknik & Ilmu Komputer*, 7, 25th ser., 107-121.

Sari, V. H. A. (2017). *Perbandingan Algoritma K-nearest Neighbor Dan Naïve Bayes Classifier Dalam Memprediksi Kategori Indeks Prestasi Mahasiswa*. (Undergraduate thesis, Duta Wacana Christian University, 2017). Retrieved from <http://sinta.ukdw.ac.id>

©UKDW