

**DEPARTEMEN ILMU KOMPUTER**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**INSTITUT PERTANIAN BOGOR**

**BOGOR**

**2017**

**ILHAM TRI MULYAWAN**

**MODEL PREDIKSI CAPAIAN AKADEMIK MAHASISWA PROGRAM SARJANA ILMU KOMPUTER IPB**

**DENGAN ALGORITME C5.0**

**PERNYATAAN MENGENAI SKRIPSI DAN  
SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA**

Dengan ini saya menyatakan bahwa skripsi berjudul Model Prediksi Capaian Akademik Mahasiswa Program Sarjana Ilmu Komputer IPB dengan Algoritme C5.0 adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Mei 2017

*Ilham Tri Mulyawan*

NIM G64144026

**ABSTRAK**

ILHAM TRI MULYAWAN. Model Prediksi Capaian Akademik Mahasiswa Program Sarjana Ilmu Komputer IPB dengan Algoritme C5.0. Dibimbing oleh AHMAD RIDHA.

Penelitian ini mengembangkan prediksi capaian akademik mahasiswa Program Sarjana Ilmu Komputer IPB di tingkat 2 berdasarkan capaian di tingkat 1 dengan pohon keputusan menggunakan algoritme C5.0. Data yang digunakan ialah nilai mahasiswa yang masuk pada tahun 2012 hingga 2014 pada mata kuliah mayor dan interdepartemen. Penelitian ini menggunakan 2 kelas (*low risk* dan *high risk*) dan 3 kelas (*low risk*, *medium risk*, dan *high risk*). Jumlah data tidak imbang sehingga penelitian ini menggunakan metode *undersampling* dan *oversampling.* Penelitian ini juga melihat keterhubungan antara mata kuliah dan prasyaratnya dengan menggunakan korelasi Spearman dan pohon keputusanalgoritme C5.0. Hasil uji model prediksi menghasilkan akurasi terbaik sebesar 94.92% untuk percobaan 3 kelas (2.76 ≤ IPK ≤ 4.00 untuk kelas *low risk,* 2.00 ≤ IPK< 2.76 untuk kelas *medium risk*, dan 0 ≤ IPK< 2.00 untuk kelas *high risk*). Percobaan tersebut menggunakan metode *oversampling*. Korelasi Spearman menunjukkan adanya keterhubungan antara mata kuliah dan prasyaratnya dengan nilai korelasi 0.308 - 0.839.

Kata kunci: Algoritme C5.0, keberhasilan studi, klasifikasi, korelasi Spearman, pohon keputusan

**ABSTRACT**

ILHAM TRI MULYAWAN. Academic Achievement Prediction Model for Computer Science Undergraduate Students in IPB using C5.0 Algorithm. Supervised by AHMAD RIDHA.

The aim of this research is to develop academic achievement prediction model for computer science undergraduate students in IPB in the second year based on their achievement in the first year using decision tree with C5.0 algorithm. The dataset used in this research are grades of students admitted in 2012 until 2014 on major and interdepartment courses. This research uses 2 classes (low risk and high risk) and 3 classes (low risk, medium risk, and high risk). The distribution of data is imbalanced so this research uses undersampling and oversampling method. This research also shows the correlation between courses and their prerequisites by using Spearman’s correlation and C5.0 algorithm. The result of prediction model obtained the best accuracy of 94.92% for experiment of 3 classes (2.76 ≤ IPK ≤ 4.00 for low risk*,* 2.00 ≤ IPK< 2.76 for medium risk, and 0 ≤ IPK< 2.00 for high risk). The experiment uses oversampling method. Spearman’s correlation shows correlation between courses and their prerequisites with correlation values 0.308 - 0.839.

Keywords: Algorithm C5.0, classification, decision tree, Spearman’s correlation, study success

Skripsi  
sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Komputer   
pada  
Departemen Ilmu Komputer

**ILHAM TRI MULYAWAN**

**DEPARTEMEN ILMU KOMPUTER**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**INSTITUT PERTANIAN BOGOR**

**BOGOR**

**2017**

**MODEL PREDIKSI CAPAIAN AKADEMIK MAHASISWA PROGRAM SARJANA ILMU KOMPUTER IPB**

**DENGAN ALGORITME C5.0**

Penguji:

1. Dr Imas Sukaesih Sitanggang, SSi MKom
2. Vektor Dewanto, ST MEng

Judul Skripsi : Model Prediksi Capaian Akademik Mahasiswa Program Sarjana Ilmu Komputer IPB dengan Algoritme C5.0

Nama : Ilham Tri Mulyawan

NIM : G64144026

Disetujui oleh

|  |
| --- |
| Ahmad Ridha, SKom MS  Pembimbing |

Diketahui oleh

Dr Ir Agus Buono, MSi MKom

Ketua Departemen

Tanggal Lulus:

**PRAKATA**

Puji dan syukur penulis panjatkan kepada Allah *subhanahu wa ta’ala* atas segala karunia-Nya sehingga penulis dapat menyelesaikan penelitian ini dengan baik.

Penulis menyadari bahwa dalam proses penulisan skripsi ini banyak mengalami kendala dan masalah, namun berkat bantuan, bimbingan, kerja sama dari berbagai pihak dan berkah dari Allah *subhanahu wa ta'ala,* kendala-kendala yang dihadapi tersebut dapat diatasi. Untuk itu penulis menyampaikan ungkapan terima kasih kepada Bapak Herry Poerwanto selaku ayah, Ibu Etty Mulhetty selaku ibu, serta seluruh keluarga atas segala doa dan kasih sayangnya. Serta ucapan terima kasih dan penghargaan kepada Bapak Ahmad Ridha, S.Kom MS selaku pembimbing yang telah dengan sabar, tekun, tulus dan ikhlas meluangkan waktu, tenaga, dan pikiran memberikan bimbingan, motivasi, arahan, dan saran-saran yang sangat berharga kepada penulis selama menyusun skripsi.

Penulis juga menyampaikan terima kasih kepada:

1. Ibu Dr Imas Sukaesih Sitanggang, SSi MKom dan Bapak Vektor Dewanto, ST MEng selaku penguji.
2. Bapak Dr Ir Agus Buono, MSi MKom selaku Ketua Departemen Ilmu Komputer IPB.
3. Seluruh dosen dan tenaga kependidikan Departemen Ilmu Komputer IPB.
4. Raden Asri Ramadhina Fitriani yang selalu memberi dukungan dan bantuan selama menyusun skripsi ini.
5. Teman-teman yang selalu mendukung yaitu Amanda Rizka, Faiz, Arnold, Bintang, Denny, Ican, Pandu, Fauzan, dan Mutiara.
6. Seluruh teman-teman Program S1 Alih Jenis Ilmu Komputer IPB Angkatan 9.

Semoga segala bantuan, bimbingan, motivasi, dan dukungan yang telah diberikan kepada penulis senantiasa dibalas oleh Allah *subhanahu wa ta’ala*. Semoga karya ilmiah ini bermanfaat bagi semua pihak yang membutuhkan.

Bogor, Mei 2017

*Ilham Tri Mulyawan*

**DAFTAR ISI**

DAFTAR TABEL ii

DAFTAR GAMBAR ii

[PENDAHULUAN 1](#_Toc482000814)

[Latar Belakang 1](#_Toc482000815)

[Perumusan Masalah 2](#_Toc482000816)

[Tujuan Penelitian 2](#_Toc482000817)

[Manfaat Penelitian 2](#_Toc482000818)

[Ruang Lingkup Penelitian 2](#_Toc482000819)

[METODE 3](#_Toc482000820)

[Data Penelitian 3](#_Toc482000821)

[Tahapan Penelitian 3](#_Toc482000822)

[Praproses Data 3](#_Toc482000823)

[Data Latih dan Data Uji 6](#_Toc482000824)

[Pemodelan Klasifikasi Pohon Keputusan dengan Algoritme C5.0 6](#_Toc482000825)

[Pengujian Model Prediksi 7](#_Toc482000826)

[Keterhubungan Mata Kuliah dan Prasyaratnya 8](#_Toc482000827)

[Lingkungan Pengembangan 9](#_Toc482000828)

[HASIL DAN PEMBAHASAN 10](#_Toc482000829)

[Praproses Data 10](#_Toc482000830)

[Data Latih dan Data Uji 10](#_Toc482000831)

[Pemodelan Klasifikasi Pohon Keputusan dengan Algoritme C5.0 11](#_Toc482000832)

[Keterhubungan Mata Kuliah dan Prasyaratnya 19](#_Toc482000833)

[SIMPULAN DAN SARAN 22](#_Toc482000834)

[Simpulan 22](#_Toc482000835)

[Saran 22](#_Toc482000836)

[DAFTAR PUSTAKA 22](#_Toc482000837)

LAMPIRAN 24

RIWAYAT HIDUP 42

**DAFTAR TABEL**

1. [Jumlah data nilai mahasiswa yang tersedia 3](file:///E:\FIX%20TERKAIT%20PREDIKSI%20PPKU\SIDANG%20SEMINAR\Burn%20CD\SKRIPSI-ILHAM%20TRI%20MULYAWAN-G64144026.docx#_Toc482828827)
2. [Rentang IPK mata kuliah mayor dan interdepartemen dengan 2 kelas pada tingkat 2 4](file:///E:\FIX%20TERKAIT%20PREDIKSI%20PPKU\SIDANG%20SEMINAR\Burn%20CD\SKRIPSI-ILHAM%20TRI%20MULYAWAN-G64144026.docx#_Toc482828828)
3. [Rentang IPK mata kuliah mayor dan interdepartemen dengan 3 kelas versi 1 pada tingkat 2 5](file:///E:\FIX%20TERKAIT%20PREDIKSI%20PPKU\SIDANG%20SEMINAR\Burn%20CD\SKRIPSI-ILHAM%20TRI%20MULYAWAN-G64144026.docx#_Toc482828829)
4. [Rentang IPK mata kuliah mayor dan interdepartemen dengan 3 kelas versi 2 pada tingkat 2 5](file:///E:\FIX%20TERKAIT%20PREDIKSI%20PPKU\SIDANG%20SEMINAR\Burn%20CD\SKRIPSI-ILHAM%20TRI%20MULYAWAN-G64144026.docx#_Toc482828830)
5. [*Confusion matrix* 8](#_Toc482828831)
6. [Pembagian kelas untuk keterhubungan yang memiliki lebih dari 2 mata kuliah prasyarat 9](file:///E:\FIX%20TERKAIT%20PREDIKSI%20PPKU\SIDANG%20SEMINAR\Burn%20CD\SKRIPSI-ILHAM%20TRI%20MULYAWAN-G64144026.docx#_Toc482828832)
7. [Penyetaraan mata kuliah dari tahun masuk 2012 hingga 2014 10](#_Toc482828833)
8. [Contoh bentuk data olahan dengan label kelas 10](#_Toc482828834)
9. [Jumlah mahasiswa pada setiap kelas (LR = *Low Risk,* MR = *Medium Risk,* dan HR = *High Risk*) 11](file:///E:\FIX%20TERKAIT%20PREDIKSI%20PPKU\SIDANG%20SEMINAR\Burn%20CD\SKRIPSI-ILHAM%20TRI%20MULYAWAN-G64144026.docx#_Toc482828835)
10. [Akurasi untuk percobaan ke-1 12](#_Toc482828836)
11. [Penggunaan atribut pada model terbaik di percobaan ke-1 dengan metode *oversampling* 12](file:///E:\FIX%20TERKAIT%20PREDIKSI%20PPKU\SIDANG%20SEMINAR\Burn%20CD\SKRIPSI-ILHAM%20TRI%20MULYAWAN-G64144026.docx#_Toc482828837)
12. [Nilai *precision* dan *recall* untuk percobaan ke-1 13](#_Toc482828838)
13. [Akurasi untuk percobaan ke-2 13](#_Toc482828839)
14. [Penggunaan atribut pada model terbaik di percobaan ke-2 dengan metode *oversampling* 14](file:///E:\FIX%20TERKAIT%20PREDIKSI%20PPKU\SIDANG%20SEMINAR\Burn%20CD\SKRIPSI-ILHAM%20TRI%20MULYAWAN-G64144026.docx#_Toc482828840)
15. [Nilai *precision* dan *recall* untuk percobaan ke-2 15](#_Toc482828841)
16. [Akurasi untuk percobaan ke-3 15](#_Toc482828842)
17. [Penggunaan atribut pada model terbaik di percobaan ke-3 dengan metode *undersampling* 16](#_Toc482828843)
18. [Nilai *precision* dan *recall* untuk percobaan ke-3 16](#_Toc482828844)
19. [Hasil akurasi terbaik untuk semua percobaan 17](#_Toc482828845)
20. [Prasyarat mata kuliah tahun masuk 2012 20](#_Toc482828846)
21. [Akurasi terbaik masing-masing RMK 20](#_Toc482828847)

**DAFTAR GAMBAR**

1. [Tahapan penelitian 4](#_Toc482686537)
2. [Rata-rata akurasi dari setiap kategori kelas 17](#_Toc482686538)
3. [Potongan pohon keputusan percobaan ke-2 dengan metode *oversampling* 17](#_Toc482686539)
4. [Kurva *precision recall* pada model terbaik percobaan ke-2 dengan metode *oversampling* 18](file:///C:\Users\ASUS%20K43TA\Desktop\Revisi%20ke-10%20After%20Sidang-IlhamTM.docx#_Toc482686540)
5. [Rata-rata area *precision recall* setiap kategori kelas 18](#_Toc482686541)
6. [Kurva ROC pada model terbaik percobaan ke-2 dengan metode *oversampling* 19](file:///C:\Users\ASUS%20K43TA\Desktop\Revisi%20ke-10%20After%20Sidang-IlhamTM.docx#_Toc482686542)
7. [Rata-rata AUC setiap kategori kelas 19](#_Toc482686543)

**DAFTAR LAMPIRAN**

1. [Atribut *dataset* mahasiswa tingkat 1 24](#_Toc482686585)
2. [Mata Kuliah Semester 3 dan selanjutnya di Program Sarjana Ilmu Komputer 24](#_Toc482686586)
3. [Struktur prasyarat mata kuliah tahun masuk 2012 27](#_Toc482686587)
4. [Jumlah pembagian data latih dan data uji untuk percobaan ke-1 dengan data tidak imbang 27](#_Toc482686588)
5. [Jumlah pembagian data latih dan data uji untuk percobaan ke-2 dengan data tidak imbang 28](#_Toc482686589)
6. [Jumlah pembagian data latih dan data uji untuk percobaan ke-3 dengan data tidak imbang 28](#_Toc482686590)
7. [Jumlah pembagian data latih dan data uji untuk percobaan ke-1 dengan metode *undersampling* 29](#_Toc482686591)
8. [Jumlah pembagian data latih dan data uji untuk percobaan ke-1 dengan metode *oversampling* 29](#_Toc482686592)
9. [Jumlah pembagian data latih dan data uji untuk percobaan ke-2 dengan metode *oversampling* 30](#_Toc482686593)
10. [Jumlah pembagian data latih dan data uji untuk percobaan ke-3 dengan metode *undersampling* 30](#_Toc482686594)
11. [Jumlah pembagian data latih dan data uji untuk percobaan ke-3 dengan metode *oversampling* 30](#_Toc482686595)
12. [Model prediksi pohon keputusan terbaik dengan *oversampling* untuk percobaan ke-1 dengan algoritme C5.0 31](#_Toc482686596)
13. [Model prediksi pohon keputusan terbaik dengan *oversampling* untuk percobaan ke-2 dengan algoritme C5.0 32](#_Toc482686597)
14. [Model prediksi pohon keputusan terbaik dengan *undersampling* untuk percobaan 3 kelas versi 2 dengan algoritme C5.0 34](#_Toc482686598)
15. [Hasil akurasi pada percobaan ke-1 36](#_Toc482686599)
16. [Hasil akurasi pada percobaan ke-2 36](#_Toc482686600)
17. [Hasil akurasi pada percobaan ke-3 37](#_Toc482686601)
18. [Kurva *precision recall* pada percobaan ke-1 dengan metode *oversampling* 37](#_Toc482686602)
19. [Kurva *precision recall* pada model terbaik percobaan ke-3 dengan metode *undersampling* 38](#_Toc482686603)
20. [Kurva ROCpada model terbaik percobaan ke-1 dengan metode *oversampling* 38](#_Toc482686604)
21. [Kurva ROCpada model terbaik percobaan ke-3 dengan metode *undersampling* 39](#_Toc482686605)
22. [Model prediksi pohon keputusan pada korelasi data 1 (MAT100, MAT221, dan KOM208) 39](#_Toc482686606)
23. [Model prediksi pohon keputusan pada korelasi data 2 (KOM203, KOM206, KOM311, dan KOM312) 39](#_Toc482686607)
24. [Model prediksi pohon keputusan pada korelasi data 3 (KOM202, KOM207, KOM321, dan KOM323) 40](#_Toc482686608)
25. [Model prediksi pohon keputusan pada korelasi data 4 (MAT100, MAT221, KOM321, dan KOM323) 40](#_Toc482686609)
26. [Model prediksi pohon keputusan pada korelasi data 5 (MAT103, MAT217, dan MAT321) 40](#_Toc482686610)
27. [Model prediksi pohon keputusan pada korelasi data 6 (KOM202, KOM331, dan KOM334) 40](#_Toc482686611)
28. [Model prediksi pohon keputusan pada korelasi data 7 (MAT100, MAT215, dan KOM301) 40](#_Toc482686612)
29. [Model prediksi pohon keputusan pada korelasi data 8 (KOM206, KOM311, dan KOM312) 41](#_Toc482686613)

# PENDAHULUAN

## Latar Belakang

Mata kuliah dalam kurikulum program sarjana Institut Pertanian Bogor (IPB) terdiri atas mata kuliah Program Pendidikan Kompetensi Umum (PPKU), mata kuliah mayor, mata kuliah interdepartemen, mata kuliah minor dan mata kuliah penunjang (*supporting courses*) (IPB 2014). Mata kuliah PPKU diselenggarakan ketika mahasiswa IPB memasuki tingkat 1 (semester 1 dan 2). Pada semester 3 dan selanjutnya, mahasiswa mengambil mata kuliah mayor, interdepartemen, minor, dan penunjang (*supporting courses*). Mata kuliah mayor merupakan mata kuliah berdasarkan keilmuan utama pada suatu departemen. Mata kuliah interdepartemen merupakan mata kuliah wajib dari departemen lain di luar departemen utama. Mata kuliah minor atau penunjang (*supporting courses*) merupakan mata kuliah pilihan yang diselenggarakan oleh departemen lain yang dapat diambil oleh mahasiswa untuk menunjang kompetensi utamanya.

Penelitian ini berfokus pada mahasiswa Program Sarjana Ilmu Komputer (PSIK) Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) IPB. Banyak mahasiswa di PSIK yang mengalami penurunan Indeks Prestasi Kumulatif (IPK) di tingkat 2. Oleh sebab itu, perlu diketahui mata kuliah pada tingkat 1 yang dapat menjadi indikator bagi keberhasilan mahasiswa di tingkat selanjutnya. Prediksi nilai mata kuliah dapat dilakukan dengan menggunakan pendekatan teknik *data mining*. Dengan *data mining,* pengetahuan dan pola yang menarik dapat ditemukan dari data yang berjumlah besar (Han *et al*. 2012).

Penelitian sebelumnya telah dilakukan oleh Swastina (2005) yang memprediksi kesesuaian jurusan mahasiswa STMIK Indonesia Banjarmasin tahun masuk 2008 dan 2009 dengan pohon keputusan menggunakan algoritme C4.5. Data sampel mahasiswa yang digunakan terdiri atas atribut nama, jenis kelamin, umur, asal sekolah, jurusan asal sekolah, nilai Ujian Akhir Nasional (UAN), Indeks Prestasi (IP) semester 1, dan IP semester 2. Penelitian tersebut diketahui bahwa dengan menggunakan algoritme C4.5 menghasilkan akurasi sebesar 93.91%. Selain itu juga, Al-Barrak (2016) melakukan penelitian dengan data mahasiswa Fakultas Ilmu Komputer, Universitas King Saud sebanyak 236 mahasiswa yang mempunyai tujuan untuk memprediksi Indeks Prestasi Kumulatif (IPK) dengan melihat nilai sebelumnya. Penelitian tersebut menggunakan algoritme C4.5 dengan membandingkan mata kuliah dari setiap semester.

Penelitian ini melakukan analisis terhadap nilai mata kuliah seperti yang dilakukan oleh Al-Barrak (2016), tetapi penelitian ini menggunakan algoritme C5.0 yang merupakan pengembangan dari algoritme *Iterative Dichotomiser3* (ID3) dan C4.5 (Larose 2005). Algoritme C5.0 memiliki akurasi prediksi yang sebanding dengan algoritme C4.5, tetapi algoritme C5.0 memiliki set aturan yang lebih kecil. Algoritme C5.0 juga lebih cepat dan menggunakan lebih sedikit memori dibandingkan algoritme C4.5 (Rulequest 2017).

Perbedaan lain penelitian ini dengan Al-Barrak (2016) ialah ruang lingkup data yang digunakan. Atribut dalam penelitian ini ialah mata kuliah tingkat 1, dan capaian pada mata kuliah mayor dan interdepartemen di tingkat 2 sebagai kelas yang diprediksi. Penelitian ini juga melihat keterhubungan mata kuliah dengan prasyaratnya menggunakan korelasi Spearman dan algoritme C5.0.

Keterhubungan mata kuliah diukur karena nilai pada mata kuliah prasyarat akan menentukan mata kuliah yang dapat diambil oleh mahasiswa. Korelasi Spearman digunakan untuk keterhubungan nilai 2 mata kuliah karena korelasi Spearman tidak mensyaratkan hubungan yang linear (Hauke dan Kossowski 2011). Keterhubungan mata kuliah yang memiliki lebih dari 2 mata kuliah prasyarat menggunakan pohon keputusan dengan algoritme C5.0.

## Perumusan Masalah

Berdasarkan latar belakang di atas, perumusan masalah dalam penelitian ini adalah:

1. Bagaimana membuat model prediksi tingkat keberhasilan mahasiswa PSIK IPB di tingkat 2 menggunakan algoritme C5.0?
2. Bagaimana keterhubungan mata kuliah dan prasyaratnya pada PSIK IPB ?

## Tujuan Penelitian

Tujuan penelitian ini adalah membuat model prediksi tingkat keberhasilan mahasiswa tingkat 2 pada PSIK IPB menggunakan algoritme C5.0dan mengukur keterhubungan antara mata kuliah dan prasyaratnya.

## Manfaat Penelitian

Model prediksi dari penelitian ini diharapkan dapat membantu PSIK IPB dalam menyiapkan peringatan dini terhadap mahasiswa berpotensi gagal di tingkat 2 berdasarkan capaian akademik di tingkat 1. Selain itu, mahasiswa PSIK IPB dapat mengetahui mata kuliah-mata kuliah di tingkat 1 yang berkontribusi terhadap keberhasilan di tingkat 2. Hasil pengukuran keterhubungan mata kuliah dengan prasyaratnya juga dapat menjadi masukan dalam pengembangan kurikulum PSIK IPB.

## Ruang Lingkup Penelitian

Penelitian ini menggunakan nilai mahasiswa PSIK IPB tahun masuk 2012 hingga 2014 pada mata kuliah wajib.

# METODE

## Data Penelitian

Data yang digunakan pada penelitian ini adalah nilai mata kuliah mahasiswa PSIK tahun masuk 2012 hingga 2014. Ketersediaan data nilai mahasiswa ditunjukkan pada Tabel 1. Mata kuliah dalam kurikulum program sarjana terdiri atas mata kuliah Program Pendidikan Komptensi Umum (PPKU), mata kuliah mayor, mata kuliah interdepartemen, mata kuliah minor, dan mata kuliah penunjang (*supporting courses*). Mata kuliah PPKU diselenggarakan ketika mahasiswa IPB memasuki tingkat 1 (semester 1 dan 2) yang dapat dilihat pada Lampiran 1. Mahasiswa mengambil mata kuliah mayor, interdepartemen, minor, dan penunjang (*supporting courses*) pada semester 3 dan selanjutnya yang dapat dilihat pada Lampiran 2.

Tabel 1 Jumlah data nilai mahasiswa yang tersedia

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tahun Masuk | Jumlah Mahasiswa pada Semester Ke- | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 2012 | 118 | 118 | 115 | 115 | 112 | 111 | 110 | 109 |
| 2013 | 96 | 96 | 96 | 94 | 96 | 96 | - | - |
| 2014 | 91 | 91 | 89 | 88 | - | - | - | - |

## Tahapan Penelitian

Penelitian ini terdiri atas beberapa tahapan berikut: pengumpulan data, praproses data, pembagian data, pemodelan klasifikasi dengan pohon keputusan, model prediksi, pengujian model prediksi, dan evaluasi. Tahapan pada penelitian ini dapat dilihat pada Gambar 1.

## Praproses Data

Tahapan praproses data dilakukan untuk mendapatkan data yang siap untuk digunakan dalam tahapan berikutnya. Langkah pertama pada praproses penelitian ini ialah menyeleksi variabel data nilai mahasiswa semester 1 dan 2 sebagai atribut pengklasifikasi dan menyeleksi variabel data nilai mahasiswa semester 3 dan 4 menjadi kelas sebagai pengklasifikasi.



Gambar 1 Tahapan penelitian

Kurikulum pada tahun masuk 2013 dan 2014 mengalami perubahan sehingga terdapat kode mata kuliah yang berubah. Oleh sebab itu, langkah selanjutnya ialah kode mata kuliah yang baru disamakan dengan kode mata kuliah sebelumnya. Penelitian ini tidak mengikutsertakan mata kuliah minor dan penunjang (*supporting courses*) karena bukan merupakan mata kuliah yang wajib oleh seluruh mahasiswa. Nilai mata kuliah mayor dan interdepartemen pada semester 3 dan 4 diakumulasikan menjadi nilai IPK untuk pembentuk kelas. Perhitungan nilai IPK dapat dilihat pada Persamaan 1.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

dengan *Ni* merupakan nilai mutu kuliah *i*, *n* merupakan jumlah mata kuliah, dan *ki* merupakan bobot sks mata kuliah *i* (IPB 2014).

IPK dikategorikan untuk membentuk kelas target. Penelitian ini melakukan 3 percobaan dengan kelas target yang berbeda. Percobaan ke-1 dilakukan dengan 2 kelas berdasarkan batasan nilai IPK yang dapat dilihat pada Tabel 2. Kelas *low risk* berarti mahasiswa dinilai berhasil di tingkat 2.

Tabel 2 Rentang IPK mata kuliah mayor dan interdepartemen dengan 2 kelas pada tingkat 2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Kelas | | Rentang IPK | Jumlah Mahasiswa pada Tahun Masuk | | | |
| 2012 | 2013 | 2014 | Gabungan |
| *Low risk* | 2.76 ≤ IPK ≤ 4.00 | | 77 | 57 | 62 | 195 |
| *High risk* | 0 ≤ IPK< 2.76 | | 35 | 37 | 26 | 99 |
|  | Total Mahasiswa | | 112 | 94 | 88 | 293 |

Batas IPK 2.76 ditentukan berdasarkan IPB (2014) yang memungkinkan mahasiswa mengambil beban studi maksimum lebih besar dibanding mahasiswa dengan IPK < 2.76. Dengan demikian, mahasiswa dengan IPK 2.76 diharapkan dapat menyelesaikan studi tepat waktu.

Percobaan ke-2 dan 3 dilakukan dengan 3 kelas namun memiliki ketentuan yang berbeda. Penelitian ini menyebut pengelompokan kategori kelas pada percobaan ke-2 menjadi 3 kelas versi 1 dan percobaan ke-3 menjadi 3 kelas versi 2. Kategori kelas pada percobaan ke-2 terdiri atas *low risk, medium risk,* dan *high risk* yang dapat dilihat pada Tabel 3*.* Kelas *medium risk* ditambahkan dengan batas bawah IPK 2.00 yang merupakan IPK kelulusan minimum (IPB 2014).

Tabel 3 Rentang IPK mata kuliah mayor dan interdepartemen dengan 3 kelas versi 1 pada tingkat 2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Kelas | Rentang IPK | | Jumlah Mahasiswa pada Tahun Masuk | | | |
| 2012 | 2013 | 2014 | Gabungan |
| *Low risk* | | 2.76 ≤ IPK ≤ 4.00 | 71 | 57 | 62 | 190 |
| *Medium risk* | | 2.00≤ IPK< 2.76 | 37 | 29 | 22 | 78 |
| *High risk* | | 0 ≤ IPK< 2 | 4 | 8 | 4 | 16 |
|  | | Total Mahasiswa | 112 | 94 | 88 | 294 |

Dari data yang tersedia, percobaan 3 kelas versi 1 menghasilkan ukuran kelas yang sangat tidak imbang. Hal ini dapat menyebabkan model yang dihasilkan tidak akurat. Oleh sebab itu, percobaan ke-3 menggunakan rentang IPK berdasarkan sebaran data yang dapat dilihat pada Tabel 4.

Tabel 4 Rentang IPK mata kuliah mayor dan interdepartemen dengan 3 kelas versi 2 pada tingkat 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Kelas | Rentang IPK | Jumlah mahasiswa pada  Tahun Masuk | | | |
| 2012 | 2013 | 2014 | Gabungan |
| *Low risk* | IPK ≥ 3.23 | 29 | 35 | 27 | 91 |
| *Medium risk* | 2.78 ≤ IPK< 3.23 | 49 | 23 | 37 | 109 |
| *High risk* | 0 ≤ IPK< 2.78 | 34 | 36 | 24 | 94 |
|  | Total Mahasiswa | 112 | 94 | 88 | 294 |

Permasalahan klasifikasi dalam penelitian ini dapat dinyatakan dalam bentuk Persamaan 2.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

dengan *D* merupakan data yang terdiri atas *X* dan *Y*. *X* merupakan vektor atribut klasifikasi yang berisikan data nilai mahasiswa PSIK tingkat 1, *Y* merupakan kelas target berdasarkan data nilai tingkat 2, dan *n* merupakan jumlah data mahasiswa yang tersedia. Pengklasifikasi C memetakan vektor atribut *X*  ke kelas *Y* seperti pada Persamaan 3.

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Percobaan ke-1 dinyatakan dalam bentuk Persamaan 4 dan 5, sedangkan percobaan ke-2 dan ke-3 masing-masing menggunakan Persamaan 6 dan 7 dengan kelas pada Persamaan 8.

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

## Data Latih dan Data Uji

*Dataset* dibagi menjadi data latih dan data uji untuk melakukan klasifikasi. Data latih digunakan untuk membangun pohon keputusan, sedangkan data uji digunakan untuk menghitung akurasi pohon keputusan. Pembagian *dataset* menggunakan teknik *k-fold cross validation*. Metode *k-fold cross validation* adalah metode yang membagi himpunan contoh secara acak menjadi *k* himpunan bagian (*subset*) (Refaeilzadeh *et al*. 2009). Penelitian ini menggunakan *k* sebesar 10 karena menghasilkan akurasi terbesar jika dibanding dengan nilai *k* yang lain (Kohavi 1995).

Sebelum pembagian *dataset*, metode *oversampling* dan *undersampling* digunakan untuk mengatasi masalah jumlah data yang tidak imbang (Longadge *et al.* 2013). Data yang tidak imbang merupakan *dataset* yang memiliki jumlah *instance* jauh lebih kecil bila dibandingkan dengan *subset* lainnya. Metode *oversampling* dilakukan dengan cara menambahkan data di kelas yang berukuran kecil, sedangkan metode *undersampling* dilakukan dengan mengurangi data kelas mayoritas (Liu 2014).

## Pemodelan Klasifikasi Pohon Keputusan dengan Algoritme C5.0

Tahapan ini membangun model prediksi dengan menggunakan algoritme C5.0. Model prediksi yang digunakan ialah pohon keputusan. Pohon keputusan merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon. Setiap *node* pada pohon keputusan merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. *Node* paling atas dari pohon keputusan disebut sebagai *node* akar atau *root* yangdiperoleh dari *information gain* terbesar. Ukuran pemilihan atribut dengan *information gain* didefinisikan pada Persamaan 9 (Han *et.al* 2012).

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

denganmerupakan informasi yang dibutuhkan untuk mengklasifikasi dari semua label kelas, *m* merupakan jumlah kelas target, dan 𝑝𝑖 adalah jumlah sample untuk kelas 𝑖. juga dikenal sebagai entropi.

Atribut *A* dapat digunakan pada partisi *S* ke dalam *v subset*, *{S1, S2, …, Sv},* dengan *Sj* berisi datapada *S* yang bernilai *aj* pada *A*. Jika *A* dipilih sebagai atribut terbaik maka *subset* tersebut akan berhubungan pada cabang dari *node* himpunan *S*. *Sij* adalah jumlah datapada kelas dalam sebuah *subset Sj*. Nilaientropiyang dihasilkan untuk mengklasifikasi *subset* atribut *A* dapat dilihat pada Persamaan 10 (Han *et al*. 2012):

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

dengan adalah jumlah *subset j* yang dibagi dengan jumlah sampel pada *S* dan *m* merupakan jumlah atribut pada *S*, maka untuk mendapatkan nilai *gain* dapat dilihat pada Persamaan 11 sebagai berikut:

|  |  |  |
| --- | --- | --- |
|  |  | (11) |

## Pengujian Model Prediksi

Tahap pengujian model prediksi dilakukan dengan menghitung akurasi, *precision,* dan *recall*. Akurasi dihitung dari model prediksi yang menunjukkan tingkat kesesuaian hasil klasifikasi data terhadap kelas yang sebenarnya. Tingkat akurasi yang baik adalah tingkat akurasi yang mendekati 100%. Semakin tinggi tingkat akurasi maka semakin rendah kesalahan klasifikasi. Akurasi diperoleh dari data uji dengan menggunakan Persamaan 12. Penelitian ini membandingkan rata-rata akurasi yang dihasilkan dari setiap percobaan untuk mengetahui keseluruhan akurasi yang bagus.

|  |  |  |
| --- | --- | --- |
|  |  | (12) |

Selain akurasi, nilai *precision* dan *recall* juga diukur*.* Dalam temu kembali informasi, *precision* adalah rasio jumlah dokumen yang relevan dari jumlah dokumen yang ditemukan oleh sistem, sedangkan *recall* adalah rasio jumlah dokumen relevan yang ditemukan dengan jumlah dokumen relevan dalam koleksi dokumen (Weng *et al*. 2008). Tabel 5 menunjukkan bentuk *confusion matrix* untuk memudahkan dalam melakukan perhitungan *precision* dan *recall* dengan Persamaan 13 dan 14.

Tabel 5 *Confusion matrix*

|  |  |  |  |
| --- | --- | --- | --- |
| Aktual Prediksi | Positif | Negatif | |
| Positif | TP | FP |
| Negatif | FN | TN |

|  |  |
| --- | --- |
|  | (13) |
|  |  |
|  | (14) |

Perhitungan *precision* dan *recall* dapat dijadikan berbentuk kurva yang disebut *Precision Recall Curve* dan *Receiver Operating Characteristic* (ROC). *Precision Recall Curve* dibuat berdasarkan nilai yang telah didapatkan pada perhitungan dengan *confusion matrix*. Kurva ROCadalah grafik antara sensitifitas (*true positive*) pada sumbu Y dengan 1 - spesifisitas pada sumbu X (*false positive*) untuk menentukan karakteristik dari *classifier* (Gorunescu 2011). *Area Under Curve* (AUC)padaROC merupakan area untuk mengukur keakuratan model yang dihasilkan pada kurva (Han *et al.* 2012). Jika nilai AUC mendekati 1, keakuratan model yang dihasilkan dinilai baik.

## Keterhubungan Mata Kuliah dan Prasyaratnya

Mata kuliah memiliki prasyarat yang mempengaruhi pengambilan mata kuliah di semester selanjutnya. Struktur prasyarat mata kuliah tahun masuk 2012 dapat dilihat pada Lampiran 3. Sebagai contoh, mahasiswa yang akan mengambil mata kuliah KOM206 (Organisasi Komputer) di semester 4 harus telah lulus dari mata kuliah KOM203 (Rangkaian Digital) pada semester 3 karena mata kuliah KOM203 merupakan mata kuliah prasyarat untuk mata kuliah KOM206.

Pengukuran keterhubungan dalam penelitian ini berfokus kepada mata kuliah dan prasyaratnya dalam kurikulum mahasiswa tahun masuk 2012 karena data nilai tersedia hingga semester 8. Korelasi Spearman digunakan jika suatu mata kuliah hanya memiliki 1 mata kuliah prasyarat. Jika suatu mata kuliah, *Q*, memiliki lebih dari 2 mata kuliah prasyarat (*P1*, *P2*, …), algoritme C5.0 digunakan dengan nilai *Q*menjadi dasar kelas target dan nilai mata kuliah prasyarat menjadi atribut.

Korelasi Spearman merupakan korelasi yang menunjukkan hubungan di antara peubah dan mengolah data ordinal (Hauke dan Kossowski 2011). Patokan hasil perhitungan korelasi ialah sebagai berikut:

* < 0.20 : hubungan dapat dianggap tidak ada
* 0.20 – 0.40 : hubungan ada tapi rendah
* > 0.40 – 0.70 : hubungan cukup
* > 0.70 – 0.90 : hubungan tinggi
* > 0.90 – 1.00 : hubungan sangat tinggi

Korelasi dapat bernilai positif yang artinya searah: jika variabel pertama besar, variabel kedua semakin besar juga atau bernilai negatif yang artinya berlawanan arah: jika variabel pertama besar, variabel kedua semakin mengecil.

Kelas yang terbentuk pada pengolahan lebih dari 2 mata kuliah prasyarat berdasarkan nilai mutu mata kuliah yang dapat dilihat pada Tabel 6. Nilai mutu A dan AB masuk ke dalam kategori kelas BAIK. Nilai mutu B, BC, C, D dan E masuk ke dalam kategori kelas CUKUP. Nilai E dimasukkan ke dalam kelas CUKUP karena nilai mahasiswa PSIK tahun masuk 2012 yang mendapat nilai E pada mata kuliah sebagai pembentuk kategori kelas sangat jarang. Nilai E hanya ditemukan pada 1 mahasiswa di mata kuliah KOM312 (Komunikasi Data dan Jaringan Komputer).

Tabel 6 Pembagian kelas untuk keterhubungan yang memiliki lebih dari 2 mata kuliah prasyarat

|  |  |  |
| --- | --- | --- |
| Kelas | | Nilai Mutu Mata Kuliah |
| BAIK | A dan AB | |
| CUKUP | B, BC, C, D dan E | |

## Lingkungan Pengembangan

Spesifikasi perangkat keras dan perangkat lunak yang digunakan untuk penelitian ini adalah sebagai berikut:

1. Perangkat keras yang digunakan berupa komputer personal dengan spesifikasi:

* Intel® Core™ i5 CPU @2.50 GHz
* RAM 8 GB
* *Harddisk Internal* 500 GB

1. Perangkat lunak yang digunakan:

* Sistem operasi Windows 8.1 64-bit
* RStudio versi 1.0.136 sebagai korelasi data dan pembentuk model prediksi dengan *library* *package* C50
* Microsoft Excel 2016 untuk pengolahan data mahasiswa
* Python 2.7 dengan *library* matplotlib untuk menghasilkan kurva ROC dan *Precision Recall*.

# HASIL DAN PEMBAHASAN

## Praproses Data

Data yang digunakan merupakan data gabungan nilai mahasiswa PSIK tahun masuk 2012 hingga 2014. Kode mata kuliah yang mengalami perubahan disajikan pada Tabel 7. Data nilai mata kuliah KOM101 (Algoritme) dihapus karena mata kuliah tersebut baru ada dalam kurikulum untuk mahasiswa tahun masuk 2014. Setelah dihitung, nilai IPK dari setiap tahun masuk dikelompokkan berdasarkan kategori kelas seperti pada Tabel 8.

Tabel 7 Penyetaraan mata kuliah dari tahun masuk 2012 hingga 2014

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No. | Kode Mata Kuliah | Mata Kuliah | Semester | Tahun Masuk | | |
| 2012 | 2013 | 2014 |
| 1 | MAT215 | Aljabar Linear | 3 |  |  | ✓ |
| MAT219 | ✓ | ✓ |  |
| 2 | IPB112 | Olahraga dan Seni | 1 |  | ✓ | ✓ |
| IPB109 | ✓ |  |  |
| 3 | KOM421 | Pengolahan Citra Digital | 6 | ✓ | ✓ |  |
| KOM324 |  |  | ✓ |
| 4 | KOM200 | Dasar Pemrograman | 3 |  |  | ✓ |
| KOM202 | Algoritme dan Pemrograman | 3 | ✓ | ✓ |  |

Tabel 8 Contoh bentuk data olahan dengan label kelas

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID Mahasiswa | IPK | Percobaan ke- | | |
| 1 | 2 | 3 |
| 2003 | 2.70 | *High risk* | *Medium risk* | *High risk* |
| 2010 | 3.40 | *Low risk* | *Low risk* | *Low risk* |
| 2019 | 2.41 | *High risk* | *Medium risk* | *High risk* |

Mahasiswa dengan ID 2003 memiliki IPK dari nilai mata kuliah mayor dan interdepartemen di semester 3 dan 4 sebesar 2.70. IPK mahasiswa tersebut memasuki kelas *high risk* pada percobaan ke-1 dan ke-3 yang berarti bahwa mahasiswa tersebut dianggap berisiko tidak berhasil pada tingkat selanjutnya, sedangkan pada percobaan ke-2 mahasiswa tersebut diberi label kelas *medium risk*.

## Data Latih dan Data Uji

Pembagian data latih dan data uji untuk masing-masing percobaan disajikan pada Tabel 9. Percobaan 3 kelas versi 1 menghasilkan jumlah mahasiswa yang sedikit pada kelas *high risk* sehingga tidak bisa menghasilkan pohon keputusan yang akurat. Metode *balancing* menghasilkan jumlah mahasiswa setiap kelas menjadi imbang pada setiap kategori kelas. Metode *undersampling* tidak dilakukan pada data nilai mahasiswa dengan kategori 3 kelas versi 1 karena memiliki jumlah data yang sedikit pada kelas *high risk*.

Tabel 9 Jumlah mahasiswa pada setiap kelas (LR = *Low Risk,* MR = *Medium Risk,* dan HR = *High Risk*)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Jenis Data | Jumlah Mahasiswa | | | | | | | |
| 2 Kelas | | 3 Kelas Versi 1 | | | 3 Kelas Versi 2 | | |
| LR | HR | LR | MR | HR | LR | MR | HR |
| Tidak Imbang | 195 | 98 | 195 | 81 | 17 | 195 | 109 | 100 |
| *Undersampling* | 98 | 98 | - | - | - | 100 | 100 | 100 |
| *Oversampling* | 195 | 195 | 195 | 195 | 195 | 195 | 195 | 195 |

Pembagian data latih dan data uji percobaan ke-1, ke-2, dan ke-3 untuk data tidak imbang dengan 10-*fold* *cross validation* masing-masing disajikan pada Lampiran 4, 5, dan 6. Pembagian serupa dilakukan terhadap hasil *undersampling* dan *oversampling.* Peimbangan data dengan metode *undersampling* pada pembagian data latih dan data uji percobaan ke-1 dan ke-3 masing-masing disajikan pada Lampiran 7 dan 10. Pembagian data latih dan data uji percobaan ke-1, ke-2, dan ke-3 dengan metode *oversampling* masing-masing disajikan pada Lampiran 8, 9, dan 11.

## Pemodelan Klasifikasi Pohon Keputusan dengan Algoritme C5.0

Model prediksi dibangun dalam bentuk pohon keputusan menggunakan algoritme C5.0. Algoritme ini menggunakan ukuran *information gain* dalam membuat pohon keputusan. Pohon keputusan yang ditampilkan merupakan pohon keputusan dengan akurasi terbaik dari model yang terbentuk. Model prediksi dengan algoritme C5.0 dibangun dengan perintah R di *console* berikut:

1. Membaca *dataset* yang akan digunakan

> filedata <- "ds2kall"

> data <- read.csv(file = paste(filedata, ".csv",sep = ""), head = TRUE,sep = ", ", fill = TRUE)

1. Data dibentuk dengan mengacak dari data yang tersedia

> data <- data[sample(nrow(data)),]

> set.seed(123)

1. Pembagian data menggunakan *cross fold validation*

> folds <- cut(seq(1,nrow(dataset)), breaks = 10, labels = FALSE)

> for(i in 1:10){

> testIndexes <- which(folds == i, arr.ind = TRUE)

> testData <- data[testIndexes, ]

> trainData <- data[-testIndexes,]

1. Pembuatan model pohon keputusan menggunakan perintah berikut:

> oneTree <- C5.0(CLASS~., data=trainData)

> summary(oneTree)}

1. Nilai akurasi dari model pohon keputusan dihitung menggunakan fungsi predict. Data yang digunakan adalah data uji dengan perintah berikut:

> oneTreePred <- predict(oneTree, testData)

> conf\_test <- table(oneTreePred, testData$CLASS)

> akurasi <- sum(diag(conf\_test))/sum(conf\_test)

Pohon keputusan yang ditampilkan merupakan pohon keputusan dengan akurasi terbesar dari model yang terbentuk. Pohon keputusan model terbaik pada percobaan ke-1 (2 kelas) dihasilkan dengan metode *oversampling*. Pohon keputusan lengkap disajikan pada Lampiran 12. Model ini menghasilkan akurasi yang tertinggi yaitu sebesar 94.87%. Hasil akurasi semua model pada percobaan ke-1 dapat dilihat pada Tabel 10.

Tabel 10 Akurasi untuk percobaan ke-1

|  |  |  |  |
| --- | --- | --- | --- |
| Model ke- | Data Tidak Imbang(%) | Data Imbang | |
| *Undersampling* (%) | *Oversampling* (%) |
| 1 | 76.67 | 85.00 | 89.74 |
| 2 | 79.31 | 95.00 | 92.31 |
| 3 | 89.66 | 89.47 | 89.74 |
| 4 | 86.21 | 80.00 | **94.87** |
| 5 | 80.00 | 73.68 | 79.49 |
| 6 | 89.66 | 85.00 | 84.62 |
| 7 | 86.21 | 94.74 | 84.62 |
| 8 | 75.86 | 90.00 | 94.87 |
| 9 | 89.66 | 78.95 | 89.74 |
| 10 | 86.67 | 90.00 | 74.36 |
| Rata-rata | 83.99 | 86.18 | 87.44 |

Pohon keputusan model terbaik memiliki 13 *node* dan menampilkan mata kuliah MAT103 (Kalkulus) sebagai atribut yang memperoleh nilai *gain* tertinggi atau *root*. Berikut beberapa aturan yang dihasilkan oleh pohon keputusan tersebut:

1. IF MAT103 in {A, AB, B} AND BIO100 in {A, AB, B, BC} AND EKO100 in {A, AB, B, BC, D, E} THEN CLASS = High Risk (134/4)
2. IF MAT103 in {BC, C, D, E} AND FIS100 in {A, C, D, E} CLASS = High Risk (115/7)

Persentase penggunaan atribut pada model ini dapat dilihat pada Tabel 11.

|  |  |  |
| --- | --- | --- |
| No. | Nama Atribut | Persentase Penggunaan Atribut (%) |
| 6 | IPB100 | 11.68 |
| 7 | KIM101 | 5.41 |
| 8 | KPM130 | 5.13 |
| 9 | KOM201 | 3.13 |

Tabel 11 Penggunaan atribut pada model terbaik di percobaan ke-1 dengan metode *oversampling*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Nama Atribut | | Persentase Penggunaan Atribut (%) | |
| 1 | MAT103 | 100.00 | |
| 2 | BIO100 | 50.43 | |
| 3 | FIS100 | 49.57 | |
| 4 | EKO100 | 38.75 | |
| 5 | IPB108 | 12.25 | |

Tabel 11 menunjukkan bahwa mata kuliah MAT103 (Kalkulus) memiliki persentase sebesar 100% yang berarti bahwa atribut MAT103 selalu muncul pada setiap *rule* yang dihasilkan. Penggunaan atribut MAT103 (Kalkulus), BIO100 (Biologi Dasar), FIS100 (Fisika), dan EKO100 (Ekonomi Umum) menunjukkan peran dalam keberhasilan mahasiswa PSIK tahun masuk 2012 hingga 2014 di tingkat 2. Atribut dengan persentase penggunaan yang rendah dinilai tidak terlalu berperan.

Model terbaik pada percobaan ke-1 memiliki *confusion matrix* dan perhitungan *precision* dan *recall* yang dapat dilihat pada Tabel 12. *Precision* pada kelas *high risk* sebesar 96.00% dan *low risk* sebesar 92.86% dengan 1 data yang prediksinya tidak sesuai dengan kelas aktual. *Recall* pada kelas *high risk* sebesar 96.00% dan *low risk* sebesar 92.86% dengan 1 data yang prediksinya tidak sesuai dengan kelas aktual.

Tabel 12 Nilai *precision* dan *recall* untuk percobaan ke-1

|  |  |  |  |
| --- | --- | --- | --- |
| Aktual Prediksi | *High risk* | *Low risk* | *Class Precision* (%) |
| *High risk* | 24 | 1 | 96.00 |
| *Low risk* | 1 | 13 | 92.86 |
| *Class Recall* (%) | 96.00 | 92.86 |  |

Dalam percobaan ke-2 (3 kelas versi 1), pohon keputusan model terbaik juga dihasilkan oleh metode *oversampling* dengan akurasi sebesar 94.92%. Pohon keputusan lengkap disajikan pada Lampiran 13. Percobaan ini tidak menggunakan metode *undersampling* karena kelas *low risk* berukuran terlalu kecil ada beberapa model yang menghasilkan kelas targetnya kosong. Hasil akurasi semua model pada percobaan ke-2 dapat dilihat pada Tabel 13.

Tabel 13 Akurasi untuk percobaan ke-2

|  |  |  |
| --- | --- | --- |
| Model ke- | Data Tidak Imbang(%) | Data Imbang |
| *Oversampling* (%) |
| 1 | 70.00 | **94.92** |
| 2 | 72.41 | 86.21 |
| 3 | 82.76 | 79.66 |
| 4 | 82.76 | 94.83 |
| 5 | 73.33 | 89.83 |
| 6 | 72.41 | 89.66 |
| 7 | 79.31 | 91.38 |
| 8 | 65.52 | 93.22 |
| 9 | 86.21 | 86.21 |
| 10 | 76.67 | 89.83 |
| Rata-rata | 76.14 | 89.58 |

Pohon keputusan model terbaik pada percobaan ke-2 memiliki 43 *node* dan menampilkan mata kuliah EKO100 (Ekonomi Umum) sebagai *root*. Berikut beberapa aturan yang dihasilkan oleh model tersebut:

1. IF EKO100 in {A, AB} AND FIS100 = C THEN CLASS= Medium Risk (13/1)
2. IF EKO100 in {A, AB} AND FIS100 in {A, AB, B, BC} AND KOM201 in {A, AB, B} AND MAT103 {A, AB, B} AND IPB112 {A, AB, BC, C, D, E} THEN CLASS = Low Risk (94)

Tabel 14 menyajikan persentase penggunaan atribut pada model terbaik di percobaan ke-2.

Tabel 14 Penggunaan atribut pada model terbaik di percobaan ke-2 dengan metode *oversampling*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Nama Atribut | | Persentase Penggunaan Atribut (%) | |
| 1 | EKO100 | 100.00 | |
| 2 | MAT103 | 96.20 | |
| 3 | KOM201 | 75.48 | |
| 4 | BIO100 | 70.15 | |
| 5 | IPB108 | 47.15 | |
| 6 | IPB100 | 45.82 | |
| 7 | KPM130 | 40.30 | |
| 8 | FIS100 | 34.41 | |

|  |  |  |
| --- | --- | --- |
| No. | Nama Atribut | Persentase Penggunaan Atribut (%) |
| 9 | MAT100 | 32.51 |
| 10 | KIM101 | 27.38 |
| 11 | IPB112 | 23.57 |
| 12 | IPB107 | 20.15 |
| 13 | IPB106 | 15.97 |
| 14 | IPB111 | 7.60 |
| 15 | AGB100 | 5.89 |

Tabel 14 menunjukkan bahwa mata kuliah EKO100 (Ekonomi Umum) memiliki persentase sebesar 100% yang berarti bahwa atribut EKO100 selalu muncul pada setiap *rule* yang dihasilkan. Persentase penggunaan atribut EKO100 (Ekonomi Umum), MAT103 (Kalkulus), KOM201 (Penerapan Komputer), dan BIO100 (Biologi Umum) pada percobaan ke-2 terbilang besar yang menunjukkan peran besar terhadap keberhasilan mahasiswa.

*Confusion matrix* dan perhitungan *precision* dan *recall* untuk model terbaik pada percobaan ke-2 dapat dilihat pada Tabel 15. Kesalahan prediksi hanya terjadi pada 10% kelas *medium risk* dan 5.56% kelas *low risk*. Nilai *precision* pada *medium risk* sebesar 90.00% dengan 2 data yang kelas aktualnya *low risk*, dan nilai *recall* sebesar 94.73% dengan 1 data yang salah prediksi sebagai kelas *low risk*. Pada kelas *low risk*, *precision* sebesar 94.44% dengan 1 data yang seharusnya kelas *medium risk*,dan nilai *recall s*ebesar 89.47% dengan 2 data yang diprediksi ke dalam kelas *medium risk*.

Percobaan terakhir dilakukan dengan 3 kelas versi 2. Berbeda dengan percobaan ke-1 dan ke-2, pohon keputusan model terbaik pada percobaan ke-3 dihasilkan dengan metode *undersampling*. Pohon keputusan lengkap disajikan pada Lampiran 14. Model ini menghasilkan akurasi yang tertinggi sebesar 80.77%. Akurasi semua model pada percobaan ini dapat dilihat pada Tabel 16.

Tabel 15 Nilai *precision* dan *recall* untuk percobaan ke-2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Aktual Prediksi | *High risk* | *Medium risk* | | *Low risk* | *Class Precision* (%) |
| *High risk* | 21 | 0 | 0 | | 100.00 |
| *Medium risk* | 0 | 18 | 2 | | 90.00 |
| *Low risk* | 0 | 1 | 17 | | 94.44 |
| *Class Recall* (%) | 100.00 | 94.73 | 89.47 | |  |

Tabel 16 Akurasi untuk percobaan ke-3

|  |  |  |  |
| --- | --- | --- | --- |
| Model ke- | Data Tidak Imbang(%) | Data Imbang | |
| *Undersampling* (%) | *Oversampling* (%) |
| 1 | 50.00 | **80.77** | 51.52 |
| 2 | 55.17 | 68.00 | 72.73 |
| 3 | 68.97 | 56.00 | 62.50 |
| 4 | 58.62 | 76.00 | 60.61 |
| 5 | 66.67 | 68.00 | 45.45 |
| 6 | 65.52 | 40.00 | 65.62 |
| 7 | 68.97 | 72.00 | 63.64 |
| 8 | 48.28 | 64.00 | 56.25 |
| 9 | 58.62 | 76.00 | 72.73 |
| 10 | 70.00 | 65.38 | 63.64 |
| Rata-rata | 61.08 | 66.62 | 61.47 |

Pohon keputusan model terbaik tersebut memiliki 36 *node* dengan mata kuliah MAT103 (Kalkulus) sebagai *root*. Berikut beberapa aturan yang dihasilkan oleh pohon keputusan untuk percobaan ke-3:

1. IF MAT103 in {BC, C, D, E} AND FIS100 in {A, C, D} THEN CLASS= High risk (58/7)
2. IF MAT103 in {A, AB, B} AND EKO100 in {A, AB, B, BC} AND IPB106 in {A, AB, BC} AND FIS100 = AB ANDMAT103 in {A, AB} AND IPB112 in {A, B, C, D, E} THEN CLASS = Low risk (36/8)

Persentase penggunaan atribut pada model terbaik pada percobaan ke-3 dapat dilihat pada Tabel 17.

Berbeda dengan percobaan pada 3 kelas versi 1, persentase penggunaan atribut mata kuliah MAT103 (Kalkulus) sebesar 100% yang berarti bahwa atribut MAT103 selalu muncul pada setiap *rule* yang dihasilkan. Penggunaan atribut MAT103 (Kalkulus), FIS100 (Fisika), EKO100 (Ekonomi Umum), IPB106 (Bahasa Indonesia), dan MAT100 (Pengantar Matematika) berperan dalam pengaruh dengan keberhasilan mahasiswa PSIK tahun masuk 2012 hingga 2014.

Model terbaik yang dihasilkan memiliki *confusion matrix* dengan *precision* dan *recall* yang dapat dilihat pada Tabel 18. Kelas prediksi pada kelas *high risk* memiliki nilai *precision* sebesar 75.00% dengan memiliki 2 data yang salah prediksi karena 1 data seharusnya masuk ke dalam kelas *low risk* dan 1 data seharusnya masuk ke dalam kelas *medium risk*. Nilai *recall* pada *high risk* sebesar 100% yang berarti bahwa kelas prediksi sesuai dengan kelas aktual. Kelas *medium risk* menghasilkan *precision* dan *recall* sebesar 80.00% dengan 2 data yang seharusnya kelas *low risk* dan masing-masing 1 data *medium risk* yang salah prediksi ke kelas *low risk* dan *high risk.* Kelas *low risk* menghasilkan *precision* sebesar 87.50% dengan memiliki 1 data yang seharusnya kelas *low risk. Recall* pada kelas *low risk* sebesar 70.00% dengan memiliki 3 data yang salah prediksi karena 1 data masuk ke dalam kelas *medium risk* dan 2 data masuk ke dalam kelas *high risk*.

Tabel 17 Penggunaan atribut pada model terbaik di percobaan ke-3 dengan metode *undersampling*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Nama Atribut | | Persentase Penggunaan Atribut (%) | |
| 1 | MAT103 | 100.00 | |
| 2 | FIS100 | 92.04 | |
| 3 | EKO100 | 58.41 | |
| 4 | IPB106 | 58.41 | |
| 5 | MAT100 | 21.24 | |
| 6 | IPB112 | 19.03 | |

|  |  |  |
| --- | --- | --- |
| No. | Nama Atribut | Persentase Penggunaan Atribut (%) |
| 7 | BIO100 | 12.83 |
| 8 | IPB111 | 10.18 |
| 9 | KOM201 | 8.85 |
| 10 | IPB107 | 3.10 |
| 11 | IPB100 | 2.65 |
| 12 | KPM130 | 1.77 |

Tabel 18 Nilai *precision* dan *recall* untuk percobaan ke-3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Aktual Prediksi | *High risk* | *Medium risk* | *Low risk* | *Class Precision* (%) |
| *High risk* | 6 | 1 | 1 | 75.00 |
| *Medium risk* | 0 | 8 | 2 | 80.00 |
| *Low risk* | 0 | 1 | 7 | 87.50 |
| *Class Recall* (%) | 100.00 | 80.00 | 70.00 |  |

Rata-rata akurasi dan simpangan baku dari semua percobaan dapat dilihat pada Gambar 2. Rata-rata akurasi tertinggi dicapai oleh percobaan ke-2 (3 kelas versi 1) dengan metode *oversampling*. Percobaan tersebut juga memiliki simpangan baku terkecil yang menunjukkan akurasi model-model dalam percobaan tersebut cukup konsisten.

Hasil akurasi terbaik dari semua percobaan dapat dilihat pada Tabel 19. Dari 3 percobaan (masing-masing dengan 10 model) yang dilakukan, salah satu model pada percobaan ke-2 (3 kelas versi 1) dengan metode *oversampling* memiliki akurasi yang terbesar yaitu 94.92%. Potongan pohon keputusan pada percobaan ke-2 dengan *oversampling* dapat dilihat pada Gambar 3.

Gambar 2 Rata-rata akurasi dari setiap kategori kelas

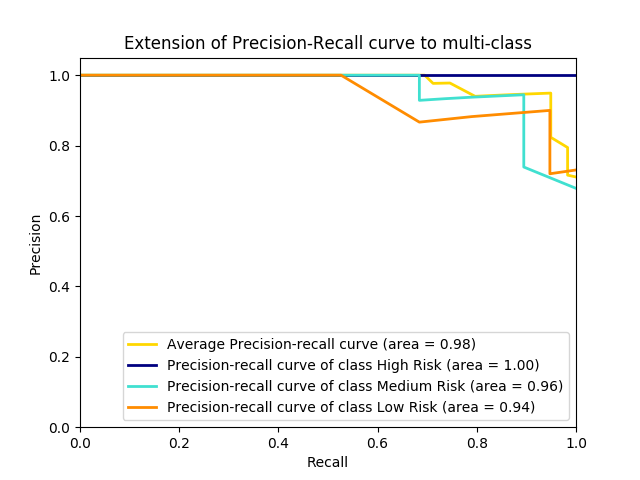
Tabel 19 Hasil akurasi terbaik untuk semua percobaan

|  |  |  |  |
| --- | --- | --- | --- |
| Percobaan ke- | Data Tidak Imbang (%) | Data Imbang | |
| *Undersampling* (%) | *Oversampling* (%) |
| 1 | 89.66 | 94.77 | 94.87 |
| 2 | 86.21 | - | **94.92** |
| 3 | 70.00 | 80.77 | 65.62 |

|  |
| --- |
| EKO100 in {A,AB}:  : FIS100 = C: MediumRisk (13/1)  : FIS100 in {A,AB,B,BC}:  : KOM201 in {BC,D}:  : :...AGB100 in {A,B}: MediumRisk (6)  : : AGB100 = AB: LowRisk (1)  : KOM201 in {A,AB,B}:  : MAT103 in {A,AB,B}:  : :...IPB112 in {A,AB,BC,C,D,E}: LowRisk (94)  EKO100 in {B,BC,C,D,E}:  :...BIO100 in {A,AB,B,BC,E}:  : MAT103 = D:  : :...IPB111 in {A,AB,C}: HighRisk (11)  : : IPB111 in {B,BC}: MediumRisk (10) |

Gambar 3 Potongan pohon keputusan percobaan ke-2 dengan metode *oversampling*

Kurva *precision* dan *recall* dengan model terbaik tersebut dapat dilihat pada Gambar 4. Ketiga kelas memiliki nilai area *precision recall* yang mendekati 1.00 yang berarti model tersebut memiliki kinerja yang baik. Kurva *precision* dan *recall* untuk model terbaik pada percobaan ke-1 dan ke-3 dapat dilihat pada Lampiran 18 dan 19. Rata-rata area *precision recall* semua percobaan mencapai nilai di atas 90% (*lihat* Gambar 5).

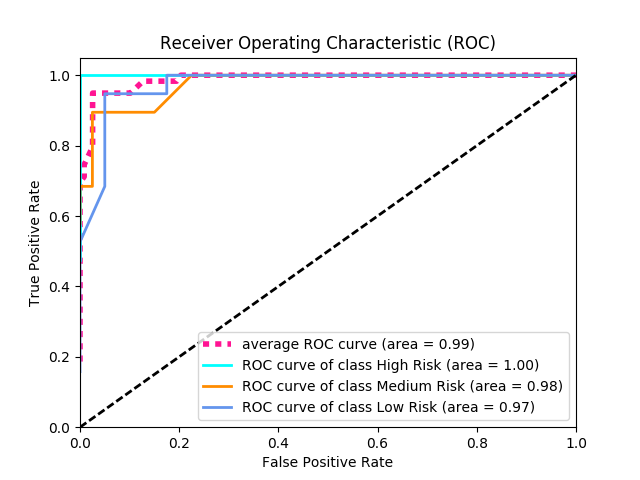


Gambar 4 Kurva *precision recall* pada model terbaik percobaan ke-2 dengan metode *oversampling*

Gambar 5 Rata-rata area *precision recall* setiap kategori kelas

Kurva ROC untuk model terbaik di percobaan ke-2 disajikan pada Gambar 6. Nilai AUC yang dihasilkan pada setiap kelas memiliki nilai yang mendekati 1.00 menunjukkan kinerja yang baik. Kurva ROCpada percobaan ke-1 dan ke-3 dapat dilihat pada Lampiran 20 dan 21.

Kurva ROCdari semua percobaan memiliki rata-rata nilaiAUCyang bagus dengan dapat dilihat pada Gambar 7. Percobaan ke-1 dan ke-2 memiliki kinerja yang konsisten, terlihat dari simpangan baku yang kecil.



Gambar 6 Kurva ROC pada model terbaik percobaan ke-2 dengan metode *oversampling*

Gambar 7 Rata-rata AUC setiap kategori kelas

## Keterhubungan Mata Kuliah dan Prasyaratnya

Keterhubungan mata kuliah dan prasyaratnya pada kurikulum PSIK untuk mahasiswa tahun masuk 2012 dapat dilihat pada Tabel 20. Beberapa mata kuliah dapat membentuk suatu rangkaian prasyarat. Sebagai contoh, MAT100 menjadi prasyarat MAT221 yang kemudian menjadi prasyarat KOM208. Suatu mata kuliah juga dapat memiliki lebih dari 1 prasyarat, misalnya KOM205 dan STK211 menjadi prasyarat untuk KOM332. Penelitian ini melihat keterhubungan mata kuliah dalam suatu rangkaian sehingga KOM332 dianggap sebagai 2 rangkaian yaitu KOM205 ke KOM332 dan STK211 ke KOM332.

Rangkaian Mata Kuliah (RMK) ke-1 hingga ke-8 adalah rangkaian yang terdiri atas lebih dari 2 mata kuliah, sedangkan RMK ke-9 hingga ke-16 adalah rangkaian dengan 2 mata kuliah.

Tabel 20 Prasyarat mata kuliah tahun masuk 2012

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RMK ke- | Faktor | | | Kelas |
| 1 | MAT100 | MAT221 |  | KOM208 |
| 2 | KOM203 | KOM206 | KOM311 | KOM312 |
| 3 | KOM202 | KOM207 | KOM321 | KOM323 |
| 4 | MAT100 | MAT221 | KOM321 | KOM323 |
| 5 | MAT103 | MAT217 |  | MAT321 |
| 6 | KOM202 | KOM331 |  | KOM334 |
| 7 | MAT100 | MAT215 |  | KOM301 |
| 8 | KOM206 | KOM311 |  | KOM312 |
| 9 | MAT103 | MAT217 |  |  |
| 10 | MAT100 | MAT103 |
| 11 | KOM202 | KOM204 |
| 12 | KOM202 | MAT321 |
| 13 | KOM205 | KOM335 |
| 14 | KOM205 | KOM332 |
| 15 | STK202 | KOM322 |
| 16 | STK211 | KOM332 |

RMK ke-1 hingga ke-8 diolah menggunakan algoritme C5.0 dengan data asli serta dengan metode *undersampling* dan *oversampling*. Akurasi terbaik untuk masing-masing RMK disajikan pada Tabel 21. RMK ke-9 hingga ke-16 diolah menggunakan korelasi Spearman dengan hasil yang disajikan pada Tabel 22.

Tabel 21 Akurasi terbaik masing-masing RMK

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| RMK ke- | Kelas | Metode *Balancing* | *Root* | Akurasi (%) | Pohon Keputusan |
| 1 | KOM208 | *Undersampling* | MAT221 | 92.31 | Lampiran 22 |
| 2 | KOM312 | *Oversampling* | KOM206 | 88.24 | Lampiran 23 |
| 3 | KOM323 | *Oversampling* | KOM207 | 80.00 | Lampiran 24 |
| 4 | KOM323 | *Oversampling* | MAT100 | 81.82 | Lampiran 25 |
| 5 | MAT321 | *Oversampling* | MAT103 | 100.00 | Lampiran 26 |
| 6 | KOM334 | *Undersampling* | KOM331 | 81.82 | Lampiran 27 |
| 7 | KOM301 | *Oversampling* | MAT215 | 86.67 | Lampiran 28 |
| 8 | KOM312 | *Oversampling* | KOM206 | 78.95 | Lampiran 29 |

Keseluruhan RMK menghasilkan akurasi lebih dari 80%, kecuali RMK ke-8 yang memiliki akurasi 78.95%. Hasil terbaik ditunjukkan oleh RMK ke-5 yang menggunakan metode *oversampling* dengan akurasi 100%. Hal ini menunjukkan keterhubungan yang sangat erat antara MAT321 (Analisis Numerik) dengan prasyaratnya, terutama MAT103 (Kalkulus). Kedua mata kuliah tersebut diampu oleh Departemen Matematika, FMIPA IPB.

Tabel 22 Korelasi RMK dengan 2 mata kuliah pada kurikulum mahasiswa tahun masuk 2012

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RMK ke- | Kode | Nama Mata Kuliah | *Spearman Correlation* | Tingkat Keterhubungan |
| 9 | MAT100 | Pengantar Matematika | 0.839 | Sangat tinggi |
|  | MAT103 | Kalkulus |  |  |
| 10 | MAT103 | Kalkulus | 0.761 | Tinggi |
| MAT217 | Kalkulus Lanjut |  |  |
| 11 | KOM202 | Algoritme dan Pemrograman | 0.455 | Cukup |
| KOM204 | Bahasa Pemrograman |  |  |
| 12 | KOM202 | Algoritme dan Pemrograman | 0.308 | Rendah |
| MAT321 | Analisis Numerik |  |  |
| 13 | KOM205 | Basis Data | 0.447 | Cukup |
| KOM335 | Sistem Informasi |  |  |
| 14 | KOM205 | Basis Data | 0.492 | Cukup |
| KOM332 | Data Mining |  |  |
| 15 | STK202 | Pengantar Hitung Peluang | 0.399 | Rendah |
| KOM322 | Metode Kuantitatif |  |  |
| 16 | STK211 | Metode Statistika | 0.486 | Cukup |
| KOM332 | Data Mining |  |  |

Untuk RMK dengan mata kuliah kelas yang diampu oleh PSIK, RMK ke-1 dengan metode *undersampling* menghasilkan akurasi tertinggi sebesar 92.31%. Mata kuliah MAT221 (Matematika Diskret) menjadi *root* untuk KOM208 (Teori Bahasa dan Otomata). Keterhubungan ini menunjukkan eratnya teori komputasi dengan matematika. Hal ini juga ditunjukkan oleh RMK ke-4 dan ke-7.

Semua RMK dengan 2 mata kuliah menunjukkan adanya hubungan walaupun tidak semuanya tinggi. Hubungan tertinggi ditunjukkan oleh MAT100 (Pengantar Matematika) dan MAT103 (Kalkulus).

MAT100 dan MAT103 menjadi *root* pada RMK ke-4 dan ke-5. Selain itu, MAT103 memiliki persentase penggunaan atribut yang tinggi (> 96%) di semua model prediksi dengan akurasi tertinggi. Ini menunjukkan pentingnya MAT103 dalam capaian akademik mahasiswa PSIK. Berdasarkan tingginya hubungan MAT100 dan MAT103, MAT100 dapat menjadi penanda awal keberhasilan mahasiswa PSIK karena mata kuliah tersebut diambil di semester 1.

# SIMPULAN DAN SARAN

## Simpulan

Penelitian ini menghasilkan model prediksi capaian akademik mahasiswa PSIK IPB berupa pohon keputusan dengan algoritme C5.0. Model percobaan ke-2 (3 kelas versi 1) dengan menggunakan metode *oversampling* menghasilkan akurasi tertinggi sebesar 94.92%.

EKO100 (Ekonomi Umum), MAT103 (Kalkulus), dan KOM201 (Penerapan Komputer) memiliki persentasi penggunaan atribut tertinggi dalam model tersebut sebesar masing-masing 100%, 96.20%, dan 75.48%. Hal tersebut menunjukkan peran ketiga mata kuliah terhadap capaian akademik mahasiswa PSIK tahun masuk 2012 hingga 2014 di tingkat 2.

Walaupun MAT100 (Pengantar Matematika) memiliki persentase penggunaan atribut relatif kecil (< 33%), mata kuliah tersebut memiliki keterhubungan yang kuat dengan MAT103 dan menjadi *root* untuk mata kuliah KOM323 (Sistem Pakar). Dengan posisinya yang diambil di semester 1, capaian di MAT100 dapat menjadi penanda awal capaian akademik mahasiswa PSIK IPB.

## Saran

Pengolahan data penelitian ini mengandalkan Microsoft Excel sehingga data belum dapat ditambahkan dengan efisien. Oleh sebab itu, penelitian selanjutnya diharapkan dapat membuat sistem untuk pengolahan data agar lebih banyak data yang dapat diolah dengan fokus pada kurikulum untuk mahasiswa tahun masuk 2014 dan setelahnya.

# DAFTAR PUSTAKA

Al-Barrak M A, Al-Razgan M. 2016. Predicting Students Final GPA Using Decision Trees: A Case Study. *International Journal of Information and Education Technology*. 6 (7):528-533.doi:10.7763/IJIET.2016.V6.745.

Galathiya AS, Ganatra AP, Bhensdadia CK. 2012. Improved Decision Tree Induction Algorithm with Feature Selection, Cross Validation, Model Complexity and Reduced Error Pruning. *International Journal of Computer Science and Information Technologies (IJCSIT).* 3(2):3427-3431.

Gorunescu F. 2011. *Data Mining Concept, Models and Techniques*. Verlag (DE): Springer.

Han J, Kamber M, Pei J. 2012. *Data Mining: Concepts and Techniques*. Ed ke-3. Waltham (US): Morgan Kaufman Publisher.

Hauke J, Kossowski T. 2011. Comparison of Values of Pearson’s and Spearman’S Correlation Coefficients on The Same Sets of Data. *Quaestiones Geographicae* .30 (2):87-93.doi:10.2478/v10117-011-0021-1.

[IPB] Institut Pertanian Bogor. 2014. *Panduan Program Pendidikan Sarjana Edisi Tahun 2014*. Bogor (ID): IPB Pr.

Kohavi R. 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*. 14th International Joint Conference on Artificial Intelligence* [internet]; 1995 Agu 20-25; Montreal, Canada. San Fransisco (US): Morgan Kaufmann Publishers, Inc. hlm 1137-1143; [diunduh 2017 April 5]. Tersedia pada: http://ai.stanford.edu/~ronnyk/accEst.pdf*.*

Larose TD. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining.* Hoboken (US):John Wiley & Sons, Inc.

Liu AY. 2014. The Effect of Oversampling and Undersampling on Classifying Imbalanced Text Datasets [tesis]. Austin (US): University of Texas.

Longadge R, Dongre S S, Malik L. 2013. Class Imbalance Problem in Data Mining: Review. *International Journal of Computer Science and Network (IJSN)*. 2(1):83-87.

Refaeilzadeh P, Tang L, Liu H. 2009. Cross-Validation. Di dalam: *Encyclopedia of Database Systems*; Boston (US): Springer. hlm 532-538..

Rulequest Research. 2017. *Data mining tools see5 and C5.0*. [Internet]. [Diunduh 2017 Mar 18]. Tersedia pada: http://rulequest.com/see5-comparison.html.

Swastina L. 2013. Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa. *J GEMA AKTUALITA*. 2(1):93-98.

Weng CG, Poon J. 2008. A New Evaluation Measure for Imbalanced Datasets. Di dalam: Roddick J F, Li J, Christen P, Kennedy P, editor. *7th Australasian Data Mining Conference (AusDM 2008)* [internet]; 2008 Nov 27-28; Glenelg, Australia. Glenelg (AU): Australian Computer Society, Inc. hlm 27-32; [diunduh 2017 Mar 8]. Tersedia pada: http:*//*crpit.com/confpapers/CRPITV87Weng.pdf.

**LAMPIRAN**

Lampiran 1 Atribut *dataset* mahasiswa tingkat 1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Kode Mata Kuliah | Nama Mata Kuliah | Semester | | Jenis Mata Kuliah | Tahun Masuk | | |
| Ganjil | Genap | 2012 | 2013 | 2014 |
| AGB100 | Pengantar Kewirausahaan | 1 |  | PPKU |  |  |  |
| FIS100 | Fisika | 1 |  | PPKU |  |  |  |
| IPB107 | Pengantar Ilmu Pertanian | 1 |  | PPKU |  |  |  |
| IPB108 | Bahasa Inggris | 1 |  | PPKU |  |  |  |
| IPB111 | Pendidikan Pancasila | 1 |  | PPKU |  |  |  |
| IPB112 | Olahraga dan Seni | 1 |  | PPKU |  |  |  |
| KPM130 | Sosiologi Umum | 1 |  | PPKU |  |  |  |
| MAT100 | Pengantar Matematika | 1 |  | PPKU |  |  |  |
| BIO100 | Biologi Dasar |  | 2 | PPKU |  |  |  |
| KIM101 | Kimia Dasar |  | 2 | PPKU |  |  |  |
| KOM201 | Penerapan Komputer |  | 2 | PPKU |  |  |  |
| EKO100 | Ekonomoi Umum |  | 2 | PPKU |  |  |  |
| IPB106 | Bahasa Indonesia |  | 2 | PPKU |  |  |  |
| MAT103 | Kalkulus |  | 2 | PPKU |  |  |  |
| IPB100 | Agama |  | 2 | PPKU |  |  |  |
| KOM101 | Algoritme |  | 2 | PPKU |  |  |  |

Lampiran 2 Mata Kuliah Semester 3 dan selanjutnya di Program Sarjana Ilmu Komputer

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Kode Mata Kuliah | Nama Mata Kuliah | Semester | | Jenis Mata Kuliah | Tahun Masuk | | |
| Ganjil | Genap | 2012 | 2013 | 2014 |
| AGB111 | Dasar-dasar Bisnis |  | 4 | SC |  |  |  |
| AGB221 | Kewirausahaan | 3 | 4 | SC |  |  |  |
| AGB336 | Koperasi dan Kelembagaan Agribisnis |  | 4 | SC |  |  |  |
| FIS254 | Elektronika Digital |  | 4 | SC |  |  |  |

Lampiran 2 Lanjutan

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Kode Mata Kuliah | Nama Mata Kuliah | Semester | | Jenis Mata Kuliah | Tahun Masuk | | | | |
| Ganjil | Genap | 2012 | | 2013 | 2014 | |
| FIS253 | Elektronika Analog |  |  | SC |  |  | | |  | |
| IKK232 | Pengantar Ekonomi Keluarga |  | 4 | SC |  |  | | |  | |
| IKK233 | Perilaku Konsumen | 3 |  | SC |  |  | | |  | |
| IKK335 | Manajemen Keuangan Konsumen |  | 4 | SC |  |  | | |  | |
| KPM210 | Dasar-dasar Komunikasi |  | 2 | SC |  |  | | |  | |
| MAN111 | Pengantar Manajemen |  | 4 | SC |  |  | | |  | |
| MAT232 | Pemrograman Linear |  | 4 | SC |  |  | | |  | |
| KOM200 | Dasar Pemrograman | 3 |  | Mayor |  |  | | |  | |
| KOM201 | Penerapan Komputer | 2 |  | Mayor |  |  | | |  | |
| KOM202 | Algoritme dan Pemrograman | 3 | 4 | Mayor |  |  | | |  | |
| KOM203 | Rangkaian Digital | 3 |  | Mayor |  |  | | |  | |
| KOM204 | Bahasa Pemrograman |  | 4 | Mayor |  |  | | |  | |
| KOM205 | Basis Data |  | 4 | Mayor |  |  | | |  | |
| KOM206 | Organisasi dan Arsitektur Komputer |  | 4 | Mayor |  |  | | |  | |
| KOM207 | Struktur Data |  | 4 | Mayor |  |  | | |  | |
| KOM208 | Teori Bahasa dan Otomata |  | 4 | Mayor |  |  | | |  | |
| KOM209 | Struktur Diskret | 3 |  | Mayor |  |  | | |  | |
| KOM220 | Pengantar Matematika Komputasi | 3 | 4 | Mayor |  |  | | |  | |
| KOM301 | Komputer Grafik | 5 |  | Mayor |  |  | | |  | |
| KOM302 | Etika Komputasi | 7 |  | Mayor |  |  | | |  | |
| KOM311 | Sistem Operasi | 5 |  | Mayor |  |  | | |  | |

Lampiran 2 Lanjutan

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Kode Mata Kuliah | Nama Mata Kuliah | Semester | | Jenis Mata Kuliah | Tahun Masuk | | | | |
| Ganjil | Genap | 2012 | | 2013 | 2014 | |
| KOM312 | Komunikasi Data dan Jaringan Komputer |  | 6 | Mayor |  |  | | |  |
| KOM321 | Kecerdasan Buatan | 5 |  | Mayor |  |  | | |  |
| KOM322 | Metode Kuantitatif | 5 |  | Mayor |  |  | | |  |
| KOM323 | Sistem Pakar |  | 6 | Mayor |  |  | | |  |
| KOM324 | Pengolahan Citra Digital |  | 6 | Mayor |  |  | | |  |
| KOM325 | Komputasi Numerik | 5 | 4 | Mayor |  |  | | |  |
| KOM330 | Manajemen Proyek Perangkat Lunak | 7 |  | Mayor |  |  | | |  |
| KOM331 | Rekasa Perangkat Lunak | 5 | 6 | Mayor |  |  | | |  |
| KOM332 | Data Mining |  | 6 | Mayor |  |  | | |  |
| KOM333 | Interaksi Manusia dan Komputer | 5 | 6 | Mayor |  |  | | |  |
| KOM334 | Pengembangan Sistem Berorientasi Objek |  | 6 | Mayor |  |  | | |  |
| KOM335 | Sistem Informasi |  | 6 | Mayor |  |  | | |  |
| KOM398 | Metode Penelitian dan Telaah Pustaka | 7 |  | Mayor |  |  | | |  |
| KOM399 | Praktik Kerja Lapang | 7 |  | Mayor |  |  | | |  |
| KOM401 | Analisis Algoritme | 7 |  | Mayor |  |  | | |  |
| KOM421 | Pengantar Pengolahan Citra Digital | 7 |  | Mayor |  |  | | |  |
| KOM497 | Kolokium | 7 |  | Mayor |  |  | | |  |
| KOM498 | Seminar |  | 8 | Mayor |  |  | | |  |
| KOM499 | Tugas Akhir |  | 8 | Mayor |  |  | | |  |
| MAT217 | Kalkulus Lanjut | 3 |  | Interdep |  |  | | |  |
| MAT215 | Aljabar Linear | 3 |  | Interdep |  |  | | |  |
| MAT219 |  |  |  | | |  |

Lampiran 2 Lanjutan

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Kode Mata Kuliah | Nama Mata Kuliah | Semester | | Jenis Mata Kuliah | Tahun Masuk | | | | | | |
| Ganjil | Genap | 2012 | | 2013 | | 2014 | | |
| MAT221 | Matematika Diskret | 3 |  | Interdep |  |  | | | |  | |
| MAT321 | Analisis Numerik | 5 |  | Interdep |  |  | | | |  | |
| FMP400 | Pengantar Bioinformatika | 7 |  | Interdep |  |  | | | |  | |
| FMP400 | Pengantar Bioinformatika | 7 |  | Interdep |  | | |  | | |  |
| STK202 | Pengantar Hitung Peluang | 3 |  | Interdep |  | | |  | | |  |
| STK211 | Metode Statistika | 3 |  | Interdep |  | | |  | | |  |

Lampiran 3 Struktur prasyarat mata kuliah tahun masuk 2012

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Mata Kuliah Prasyarat | | Mata Kuliah yang Berhubungan |
|
| 1 | KOM203 | KOM206 | |
| 2 | KOM202 | MAT321, KOM204, KOM207, dan KOM331 | |
| 3 | KOM205 | KOM332 dan KOM335 | |
| 4 | KOM206 | KOM311 | |
| 5 | KOM207 | KOM321 | |
| 6 | KOM311 | KOM312 dan KOM411 | |
| 7 | KOM321 | KOM323 | |
| 8 | KOM331 | KOM334 | |
| 9 | MAT100 | MAT215 dan MAT221 | |
| 10 | MAT103 | MAT217 | |
| 11 | MAT215 | KOM301 dan KOM421 | |
| 12 | MAT217 | MAT321 dan KOM401 | |
| 13 | MAT221 | KOM208 | |
| 14 | STK202 | KOM321, KOM412, dan KOM332 | |
| 15 | STK211 | KOM332 dan KOM431 | |

Lampiran 4 Jumlah pembagian data latih dan data uji untuk percobaan ke-1 dengan data tidak imbang

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Fold* | Data Latih | | | Data Uji | | |
| *Low Risk* | *High Risk* | Jumlah | *Low Risk* | *High Risk* | Jumlah |
| 1 | 176 | 87 | 263 | 19 | 11 | 30 |
| 2 | 179 | 85 | 264 | 16 | 13 | 29 |

Lampiran 4 Lanjutan

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Fold* | Data Latih | | | Data Uji | | |
| *Low Risk* | *High Risk* | Jumlah | *Low Risk* | *High Risk* | Jumlah |
| 3 | 178 | 86 | 264 | 17 | 12 | 29 |
| 4 | 173 | 91 | 264 | 22 | 7 | 29 |
| 5 | 178 | 85 | 263 | 17 | 13 | 30 |
| 6 | 172 | 92 | 264 | 23 | 6 | 29 |
| 7 | 177 | 87 | 264 | 18 | 11 | 29 |
| 8 | 177 | 87 | 264 | 18 | 11 | 29 |
| 9 | 172 | 92 | 264 | 23 | 6 | 29 |
| 10 | 173 | 90 | 263 | 22 | 8 | 30 |

Lampiran 5 Jumlah pembagian data latih dan data uji untuk percobaan ke-2 dengan data tidak imbang

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Fold* |  | Data Latih | | | | | Data Uji | | | |
| *Low Risk* | | *Medium Risk* | *High Risk* | Jumlah | *Low Risk* | | *Medium Risk* | *High Risk* | Jumlah |
| 1 | 176 | | 71 | 16 | 263 | 19 | | 10 | 1 | 30 |
| 2 | 179 | | 71 | 14 | 264 | 16 | | 10 | 3 | 29 |
| 3 | 178 | | 71 | 15 | 264 | 17 | | 10 | 2 | 29 |
| 4 | 173 | | 75 | 16 | 264 | 22 | | 6 | 1 | 29 |
| 5 | 178 | | 71 | 14 | 263 | 17 | | 10 | 3 | 30 |
| 6 | 172 | | 75 | 17 | 264 | 23 | | 6 | 0 | 29 |
| 7 | 177 | | 72 | 15 | 264 | 18 | | 9 | 2 | 29 |
| 8 | 177 | | 71 | 16 | 264 | 18 | | 10 | 1 | 29 |
| 9 | 172 | | 75 | 17 | 264 | 23 | | 6 | 0 | 29 |
| 10 | 173 | | 77 | 13 | 263 | 22 | | 4 | 4 | 30 |

Lampiran 6 Jumlah pembagian data latih dan data uji untuk percobaan ke-3 dengan data tidak imbang

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Fold* |  | Data Latih | | | | | Data Uji | | | |
| *Low Risk* | | *Medium Risk* | *High Risk* | Jumlah | *Low Risk* | | *Medium Risk* | *High Risk* | Jumlah |
| 1 | 76 | | 98 | 89 | 263 | 8 | | 11 | 11 | 30 |
| 2 | 75 | | 102 | 87 | 264 | 9 | | 7 | 13 | 29 |
| 3 | 79 | | 97 | 88 | 264 | 5 | | 12 | 12 | 29 |
| 4 | 77 | | 94 | 93 | 264 | 7 | | 15 | 7 | 29 |
| 5 | 76 | | 100 | 87 | 263 | 8 | | 9 | 13 | 30 |
| 6 | 73 | | 97 | 94 | 264 | 11 | | 12 | 6 | 29 |
| 7 | 78 | | 97 | 89 | 264 | 6 | | 12 | 11 | 29 |
| 8 | 77 | | 99 | 88 | 264 | 7 | | 10 | 12 | 29 |

Lampiran 6 Lanjutan

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Fold* |  | Data Latih | | | | | Data Uji | | | |
| *Low Risk* | | *Medium Risk* | *High Risk* | Jumlah | *Low Risk* | | *Medium Risk* | *High Risk* | Jumlah |
| 9 | 71 | | 100 | 93 | 264 | 13 | | 9 | 7 | 29 |
| 10 | 74 | | 97 | 92 | 263 | 10 | | 12 | 8 | 30 |

Lampiran 7 Jumlah pembagian data latih dan data uji untuk percobaan ke-1 dengan metode *undersampling*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Fold* | Data Latih | | | Data Uji | | |
| *Low Risk* | *High Risk* | Jumlah | *Low Risk* | *High Risk* | Jumlah |
| 1 | 87 | 89 | 176 | 11 | 9 | 20 |
| 2 | 86 | 90 | 176 | 12 | 8 | 20 |
| 3 | 89 | 88 | 177 | 9 | 10 | 19 |
| 4 | 91 | 85 | 176 | 7 | 13 | 20 |
| 5 | 90 | 87 | 177 | 8 | 11 | 19 |
| 6 | 89 | 87 | 176 | 9 | 11 | 20 |
| 7 | 88 | 89 | 177 | 10 | 9 | 19 |
| 8 | 86 | 90 | 176 | 12 | 8 | 20 |
| 9 | 86 | 91 | 177 | 12 | 7 | 19 |
| 10 | 90 | 86 | 176 | 8 | 12 | 20 |

Lampiran 8 Jumlah pembagian data latih dan data uji untuk percobaan ke-1 dengan metode *oversampling*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Fold* | Data Latih | | | Data Uji | | | |
| *Low Risk* | *High Risk* | Jumlah | | *Low Risk* | *High Risk* | Jumlah |
| 1 | 171 | 180 | 351 | | 24 | 15 | 39 |
| 2 | 181 | 170 | 351 | | 14 | 25 | 39 |
| 3 | 175 | 176 | 351 | | 20 | 19 | 39 |
| 4 | 181 | 170 | 351 | | 14 | 25 | 39 |
| 5 | 173 | 178 | 351 | | 22 | 17 | 39 |
| 6 | 170 | 181 | 351 | | 25 | 14 | 39 |
| 7 | 177 | 174 | 351 | | 18 | 21 | 39 |
| 8 | 176 | 175 | 351 | | 19 | 20 | 39 |
| 9 | 173 | 178 | 351 | | 22 | 17 | 39 |
| 10 | 178 | 173 | 351 | | 17 | 22 | 39 |

Lampiran 9 Jumlah pembagian data latih dan data uji untuk percobaan ke-2 dengan metode *oversampling*

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Fold* | Data Latih | | | | | Data Uji | | | | | | |
| *Low Risk* | *Medium Risk* | *High Risk* | Jumlah | *Low Risk* | | *Medium Risk* | *High Risk* | | | Jumlah | |
| 1 | 176 | 176 | 174 | 526 | 19 | | 19 | 21 | | | 59 | |
| 2 | 174 | 175 | 178 | 527 | 21 | | 20 | 17 | | | 58 | |
| 3 | 175 | 178 | 173 | 526 | 20 | | 17 | 22 | | | 59 | |
| 4 | 175 | 174 | 178 | 527 | 20 | | 21 | 17 | | 58 | | |
| 5 | 176 | 174 | 176 | 526 | 19 | | 21 | 19 | | 59 | | |
| 6 | 176 | 177 | 174 | 527 | 19 | | 18 | 21 | | 58 | | |
| 7 | 175 | 176 | 176 | 527 | 20 | | 19 | 19 | 58 | | |
| 8 | 177 | 177 | 172 | 526 | 18 | | 18 | 23 | 59 | | |
| 9 | 176 | 177 | 174 | 527 | 19 | | 18 | 21 | 58 | | |
| 10 | 175 | 171 | 180 | 526 | 20 | | 24 | 15 | 59 | | |

Lampiran 10 Jumlah pembagian data latih dan data uji untuk percobaan ke-3 dengan metode *undersampling*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Fold* | Data Latih | | | | | Data Uji | | | |
| *Low Risk* | *Medium Risk* | *High Risk* | Jumlah | *Low Risk* | | *Medium Risk* | *High Risk* | Jumlah |
| 1 | 74 | 74 | 78 | 226 | 10 | | 10 | 6 | 26 |
| 2 | 76 | 79 | 72 | 227 | 8 | | 5 | 12 | 25 |
| 3 | 74 | 76 | 77 | 227 | 10 | | 8 | 7 | 25 |
| 4 | 78 | 77 | 72 | 227 | 6 | | 7 | 12 | 25 |
| 5 | 73 | 77 | 77 | 227 | 11 | | 7 | 7 | 25 |
| 6 | 80 | 69 | 78 | 227 | 4 | | 15 | 6 | 25 |
| 7 | 78 | 72 | 77 | 227 | 6 | | 12 | 7 | 25 |
| 8 | 74 | 77 | 76 | 227 | 10 | | 7 | 8 | 25 |
| 9 | 72 | 79 | 76 | 227 | 12 | | 5 | 8 | 25 |
| 10 | 77 | 76 | 73 | 226 | 7 | | 8 | 11 | 26 |

Lampiran 11 Jumlah pembagian data latih dan data uji untuk percobaan ke-3 dengan metode *oversampling*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Fold* | Data Latih | | | | Data Uji | | | |
| *Low Risk* | *Medium Risk* | *High Risk* | Jumlah | *Low Risk* | *Medium Risk* | *High Risk* | Jumlah |
| 1 | 96 | 97 | 101 | 294 | 13 | 12 | 8 | 33 |
| 2 | 101 | 100 | 93 | 294 | 8 | 9 | 16 | 33 |
| 3 | 99 | 100 | 96 | 295 | 10 | 9 | 13 | 32 |
| 4 | 97 | 97 | 100 | 294 | 12 | 12 | 9 | 33 |
| 5 | 100 | 95 | 99 | 294 | 9 | 14 | 10 | 33 |

Lampiran 11 Lanjutan

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Fold* | Data Latih | | | | Data Uji | | | |
| *Low Risk* | *Medium Risk* | *High Risk* | Jumlah | *Low Risk* | *Medium Risk* | *High Risk* | Jumlah |
| 6 | 96 | 99 | 100 | 295 | 13 | 10 | 9 | 32 |
| 7 | 100 | 97 | 97 | 294 | 9 | 12 | 12 | 33 |
| 8 | 98 | 99 | 98 | 295 | 11 | 10 | 11 | 32 |
| 9 | 97 | 100 | 97 | 294 | 12 | 9 | 12 | 33 |
| 10 | 97 | 97 | 100 | 294 | 12 | 12 | 9 | 33 |

Lampiran 12 Model prediksi pohon keputusan terbaik dengan *oversampling* untuk percobaan ke-1 dengan algoritme C5.0

|  |
| --- |
| MAT103 in {A,AB,B}:  :...BIO100 = E: LowRisk (0)  : BIO100 in {A,AB,B,BC}:  : :...EKO100 in {A,AB,B,BC,D,E}: LowRisk (134/4)  : : EKO100 = C: HighRisk (2)  : BIO100 in {C,D}:  : :...IPB100 = A: HighRisk (0)  : IPB100 in {AB,B,BC}: LowRisk (22/5)  : IPB100 in {C,D}:  : :...KIM101 in {A,B,BC,C,D,E}: HighRisk (16)  : KIM101 = AB: LowRisk (3)  MAT103 in {BC,C,D,E}:  :...FIS100 in {A,C,D,E}: HighRisk (115/7)  FIS100 in {AB,B,BC}:  :...MAT103 in {D,E}: HighRisk (16/1)  MAT103 in {BC,C}:  :...IPB108 in {A,BC,C,D}: LowRisk (15/2)  IPB108 in {AB,B}:  :...FIS100 = AB: HighRisk (10)  FIS100 in {B,BC}:  :...KPM130 in {A,C,D}: LowRisk (4)  KPM130 in {B,BC}: HighRisk (3)  KPM130 = AB:  :...KOM201 in {A,B}: HighRisk (6/1)  KOM201 in {AB,BC,C,D}: LowRisk (5) |

Lampiran 13 Model prediksi pohon keputusan terbaik dengan *oversampling* untuk percobaan ke-2 dengan algoritme C5.0

|  |
| --- |
| EKO100 in {A,AB}:  :...FIS100 in {D,E}: LowRisk (0)  : FIS100 = C: MediumRisk (13/1)  : FIS100 in {A,AB,B,BC}:  : :...KOM201 = C: LowRisk (0)  : KOM201 in {BC,D}:  : :...AGB100 in {A,B}: MediumRisk (6)  : : AGB100 = AB: LowRisk (1)  : KOM201 in {A,AB,B}:  : :...MAT103 = E: LowRisk (0)  : MAT103 in {A,AB,B}:  : :...IPB112 in {A,AB,BC,C,D,E}: LowRisk (94)  : : IPB112 = B:  : : :...IPB107 in {A,AB,BC,C,D}: LowRisk (12)  : : IPB107 = B: MediumRisk (2)  : MAT103 in {BC,C,D}:  : :...KOM201 = AB: LowRisk (9)  : KOM201 in {A,B}:  : :...FIS100 = A: MediumRisk (0)  : FIS100 = BC: LowRisk (4)  : FIS100 in {AB,B}:  : :...IPB108 in {BC,C,D}: MediumRisk (0)  : IPB108 = A: LowRisk (3)  : IPB108 in {AB,B}:  : :...IPB106 in {A,B,BC,C}: MediumRisk (12)  : IPB106 = AB: LowRisk (1)  EKO100 in {B,BC,C,D,E}:  :...BIO100 in {A,AB,B,BC,E}:  :...MAT103 = E: MediumRisk (0)  : MAT103 = D:  : :...IPB111 in {A,AB,C}: HighRisk (11)  : : IPB111 in {B,BC}: MediumRisk (10)  : MAT103 in {A,AB,B,BC,C}:  : :...KIM101 = E: MediumRisk (0)  : KIM101 in {A,AB}:  : :...IPB107 in {A,AB,BC,C,D}: LowRisk (20)  : : IPB107 = B:  : : :...IPB111 in {A,B,BC,C}: MediumRisk (5)  : : IPB111 = AB: LowRisk (4) |

Lampiran 13 Lanjutan

|  |
| --- |
| : KIM101 in {B,BC,C,D}:  : :...MAT100 = A: LowRisk (2)  : MAT100 in {C,D,E}: MediumRisk (9)  : MAT100 = AB:  : :...FIS100 in {A,AB,B,BC,D,E}: LowRisk (2)  : : FIS100 = C: MediumRisk (2)  : MAT100 = B:  : :...KIM101 in {B,BC,D}: MediumRisk (14/2)  : : KIM101 = C: LowRisk (3)  : MAT100 = BC:  : :...AGB100 in {A,AB}: MediumRisk (23)  : AGB100 = B: LowRisk (1)  BIO100 in {C,D}:  :...MAT103 = AB:  :...IPB111 = A: MediumRisk (3)  : IPB111 in {AB,B,BC,C}: LowRisk (7)  MAT103 in {A,B,BC,C,D,E}:  :...KOM201 = D: HighRisk (0)  KOM201 = A:  :...IPB100 in {A,AB,BC,C,D}: MediumRisk (18)  : IPB100 = B: LowRisk (3/1)  KOM201 in {AB,B,BC,C}:  :...IPB108 in {AB,C,D}: MediumRisk (12/1)  IPB108 in {A,B,BC}:  :...IPB100 = BC:  :...FIS100 in {B,C}: LowRisk (2)  : FIS100 in {A,AB,BC,D,E}: MediumRisk (6)  IPB100 in {A,AB,B,C,D}:  :...KPM130 = AB:  :...IPB112 in  {A,AB,BC,C,D,E}: MediumRisk (12/3)  : IPB112 = B: HighRisk (4)  KPM130 in {A,B,D}:  :...MAT100 in {A,B,C,D,E}: HighRisk (92/2)  : MAT100 in {AB,BC}:  : :...EKO100 in {B,C}: MediumRisk (10)  : EKO100 in {BC,D,E}: HighRisk (13)  KPM130 in {BC,C}: |

Lampiran 13 Lanjutan

|  |
| --- |
| :...EKO100 in {B,E}: HighRisk (0)  EKO100 = BC: MediumRisk (10)  EKO100 in {C,D}:  :...IPB106 in {BC,C}:  :...FIS100 in {A,AB,B,BC,C,E}  : MediumRisk(7)  : FIS100 = D: LowRisk (1)  IPB106 in {A,AB,B}:  :...IPB107 = A: HighRisk (0)  IPB107 in {AB,D}:  :...FIS100 = BC: LowRisk (1)  : FIS100 in {A,AB,B,C,D,E}:  MediumRisk (3)  IPB107 in {B,BC,C}:  :...KIM101 in {AB,BC}:  MediumRisk (2)  KIM101 in {A,B,C,D,E}:  HighRisk (57/1) |

Lampiran 14 Model prediksi pohon keputusan terbaik dengan *undersampling* untuk percobaan 3 kelas versi 2 dengan algoritme C5.0

|  |
| --- |
| MAT103 in {BC,C,D,E}:  :...FIS100 in {A,C,D}: HighRisk (58/7)  : FIS100 in {AB,B,BC}:  : :...MAT100 = A: LowRisk (1)  : MAT100 in {AB,C,D,E}: HighRisk (8/1)  : MAT100 = BC:  : :...MAT103 in {BC,C}: MediumRisk (5)  : : MAT103 in {D,E}: HighRisk (5)  : MAT100 = B:  : :...IPB111 in {BC,C}: HighRisk (1)  : IPB111 = A:  : :...IPB112 in {A,AB,C,D,E}: MediumRisk (4/1)  : : IPB112 = B: LowRisk (1)  : IPB111 = AB:  : :...IPB106 in {A,B,BC,C}: HighRisk (4)  : : IPB106 = AB: MediumRisk (1)  : IPB111 = B:  : :...IPB100 in {A,C}: HighRisk (3/1)  : IPB100 in {AB,B,BC,D}: MediumRisk (3)  MAT103 in {A,AB,B}:  :...EKO100 in {D,E}: LowRisk (0)  EKO100 = C:  :...MAT103 = AB: MediumRisk (1) |

Lampiran 14 Lanjutan

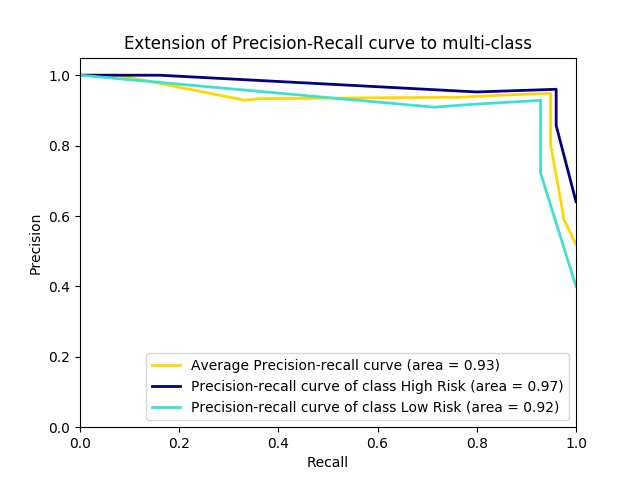
|  |
| --- |
| : MAT103 in {A,B}: HighRisk (4)  EKO100 in {A,AB,B,BC}:  :...IPB106 = C: LowRisk (0)  IPB106 = B: MediumRisk (13/1)  IPB106 in {A,AB,BC}:  :...FIS100 = C: MediumRisk (2)  FIS100 = D: HighRisk (1)  FIS100 = A:  :...KOM201 in {A,B,BC,C,D}: LowRisk (16)  : KOM201 = AB:  : :...EKO100 in {A,AB,BC}: LowRisk (2)  : EKO100 = B: MediumRisk (2)  FIS100 = AB:  :...MAT103 in {A,AB}:  : :...IPB112 in {A,B,C,D,E}: LowRisk (36/8)  : : IPB112 = AB: MediumRisk (2)  : MAT103 = B:  : :...IPB106 in {A,BC}: MediumRisk (5)  : IPB106 = AB: LowRisk (3/1)  FIS100 = BC:  :...IPB106 in {AB,BC}: MediumRisk (4)  : IPB106 = A:  : :...MAT100 in {A,AB,BC,D,E}: LowRisk (7)  : MAT100 in {B,C}: MediumRisk (5)  FIS100 = B:  :...BIO100 in {AB,BC,E}: LowRisk (9/1)  BIO100 = D: MediumRisk (1)  BIO100 = B:  :...MAT103 in {A,B}: MediumRisk (5/1)  : MAT103 = AB: LowRisk (3)  BIO100 = C:  :...KPM130 = A: MediumRisk (2)  : KPM130 in {AB,B,BC,C,D}:HighRisk (2)  BIO100 = A:  :...IPB107 in {AB,BC,C,D}:MediumRisk (0)  IPB107 = B: HighRisk (1)  IPB107 = A:  :...IPB111 in {A,AB,BC,C}: MediumRisk (4)  IPB111 = B: LowRisk (2) |

Lampiran 15 Hasil akurasi pada percobaan ke-1

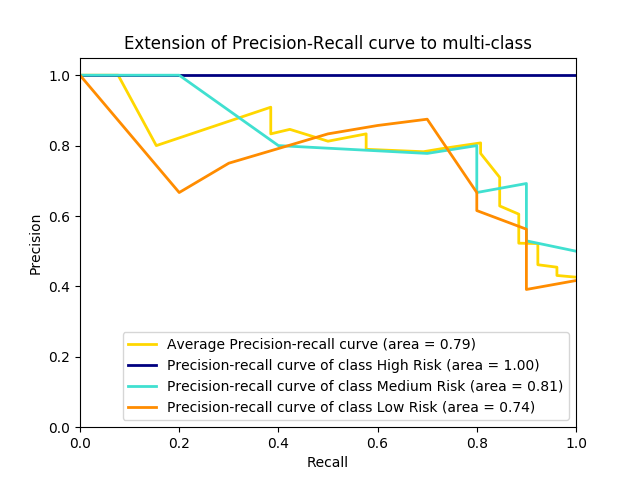
Lampiran 16 Hasil akurasi pada percobaan ke-2

Lampiran 17 Hasil akurasi pada percobaan ke-3

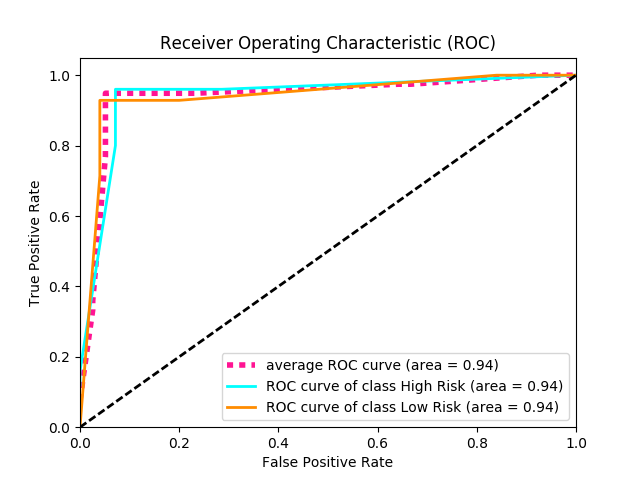
Lampiran 18 Kurva *precision recall* pada percobaan ke-1 dengan metode *oversampling*



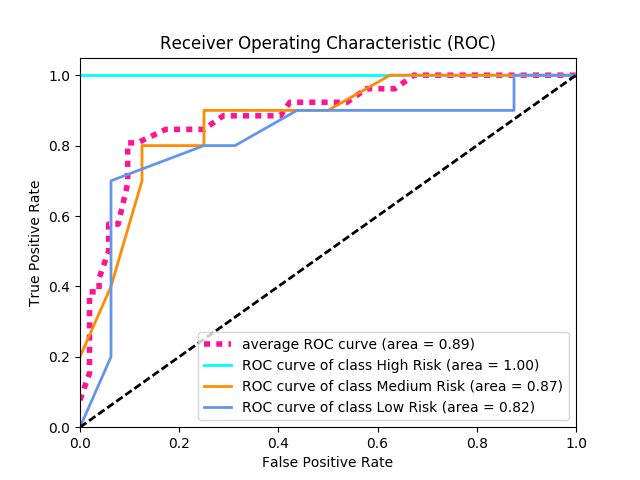
Lampiran 19 Kurva *precision recall* pada model terbaik percobaan ke-3 dengan metode *undersampling*



Lampiran 20 Kurva ROCpada model terbaik percobaan ke-1 dengan metode *oversampling*



Lampiran 21 Kurva ROCpada model terbaik percobaan ke-3 dengan metode *undersampling*



Lampiran 22 Model prediksi pohon keputusan pada korelasi data 1 (MAT100, MAT221, dan KOM208)

|  |
| --- |
| MAT221 in {A,AB,B,BC}: BAIK (78/24)  MAT221 in {C,D}: CUKUP (21/6) |

Lampiran 23 Model prediksi pohon keputusan pada korelasi data 2 (KOM203, KOM206, KOM311, dan KOM312)

|  |
| --- |
| KOM206 in {A,AB}: BAIK (75/20)  KOM206 in {C,D,E}: CUKUP (15/2)  KOM206 = BC:  :...KOM311 in {A,BC}: BAIK (9/2)  : KOM311 in {AB,B,C,D}: CUKUP (16/1)  KOM206 = B:  :...KOM311 in {A,D}: BAIK (12/5)  KOM311 in {AB,BC,C}: CUKUP (6)  KOM311 = B:  :...KOM203 in {A,B}: CUKUP (8/3)  KOM203 in {AB,BC,C}: BAIK (6) |

Lampiran 24 Model prediksi pohon keputusan pada korelasi data 3 (KOM202, KOM207, KOM321, dan KOM323)

|  |
| --- |
| KOM207 in {BC,C}: CUKUP (14)  KOM207 in {A,AB,B}:  :...KOM321 in {A,AB,B,D}: BAIK (41/11)  KOM321 = C: CUKUP (13/4)  KOM321 = BC:  :...KOM202 in {A,AB,C,D}: CUKUP (15/4)  KOM202 in {B,BC}: BAIK (9/2) |

Lampiran 25 Model prediksi pohon keputusan pada korelasi data 4 (MAT100, MAT221, KOM321, dan KOM323)

|  |
| --- |
| MAT100 = A: BAIK (43/14)  MAT100 in {AB,B,BC,C,D}: CUKUP (83/31) |

Lampiran 26 Model prediksi pohon keputusan pada korelasi data 5 (MAT103, MAT217, dan MAT321)

|  |
| --- |
| MAT103 in {BC,D}: CUKUP (23)  MAT103 in {A,AB,B,C}:  :...MAT217 = AB: BAIK (27/9)  MAT217 in {BC,D}: CUKUP (14)  MAT217 = A:  :...MAT103 in {A,AB,C}: BAIK (37/5)  : MAT103 = B: CUKUP (2)  MAT217 = B:  :...MAT103 in {A,AB,C}: BAIK (40/15)  : MAT103 = B: CUKUP (5)  MAT217 = C:  :...MAT103 in {A,AB,C}: CUKUP (9)  MAT103 = B: BAIK (11/3) |

Lampiran 27 Model prediksi pohon keputusan pada korelasi data 6 (KOM202, KOM331, dan KOM334)

|  |
| --- |
| KOM331 in {A,AB,B}: BAIK (36/7)  KOM331 in {BC,C,D}: CUKUP (57/16) |

Lampiran 28 Model prediksi pohon keputusan pada korelasi data 7 (MAT100, MAT215, dan KOM301)

|  |
| --- |
| MAT215 in {C,D}: CUKUP (26/1)  MAT215 in {A,AB,B,BC}:  :...MAT100 in {A,D,E}: BAIK (58/12)  MAT100 in {B,BC,C}: CUKUP (25/8)  MAT100 = AB:  :...MAT215 in {A,AB}: BAIK (4)  MAT215 in {B,BC}: CUKUP (18/7) |

Lampiran 29 Model prediksi pohon keputusan pada korelasi data 8 (KOM206, KOM311, dan KOM312)

|  |
| --- |
| KOM206 in {C,D}: CUKUP (22)  KOM206 in {A,AB,B,BC}:  :...KOM311 in {B,C}: BAIK (101/35)  KOM311 in {BC,D}: CUKUP (14/3)  KOM311 = A:  :...KOM206 in {A,B}: CUKUP (8)  : KOM206 in {AB,BC}: BAIK (15/4)  KOM311 = AB:  :...KOM206 = A: BAIK (6)  KOM206 in {AB,B,BC}: CUKUP (7) |

**RIWAYAT HIDUP**

Penulis dilahirkan di Jakarta pada tanggal 2 Juni 1993. Penulis adalah anak ke-3 dari 3 bersaudara, anak dari pasangan Herry Poerwanto dan Etty Mulhetty.

Penulis menempuh pendidikan sekolah menengah atas di SMA Negeri 17 Bekasi pada tahun 2008 hingga 2011. Penulis melanjutkan pendidikan perguruan tinggi diploma di Program Diploma Institut Pertanian Bogor dengan mengambil Program Keahlian Manajemen Informatika pada tahun 2011 hingga 2014. Penulis melanjutkan pendidikan sarjana di Departemen Ilmu Komputer Institut Pertanian Bogor.