

Prediksi Indeks Prestasi Mahasiswa Yang Berkuliah Sambil Bekerja Di Universitas Advent Indonesia Dengan Menggunakan Metode *Decision Tree* C4.5 Dan SMOTE

Yusran Timur Samuel *¹, Chrystle Beatrix Allbright Nahuway ²

¹Fakultas Teknologi Informasi, Universitas Advent Indonesia

e-mail: *¹y.tarihoran@unai.edu, ²chrystle.beatrix@gmail.com

Abstrak

Pendidikan tinggi adalah salah satu cara agar mendapat pekerjaan lebih mudah, hal tersebut disebabkan karena melalui pendidikan individu tersebut mampu meningkatkan kualitas sumber daya manusia pada zaman ini. Namun biaya pendidikan yang tinggi sangat mahal sehingga individu yang ingin berkuliah harus juga bekerja di saat yang bersamaan, maka penelitian ini bertujuan untuk memprediksi indeks prestasi mahasiswa yang berkuliah sambil bekerja di Universitas Advent Indonesia. Dari hasil penelitian ini terdapat 8 atribut yang berpengaruh dalam memprediksi indeks prestasi mahasiswa di Universitas Advent Indonesia yaitu Departemen Pekerjaan, Jam Kerja, Jurusan, Jenis Kelamin, Tempat Tinggal, Usia, Jumlah SKS dan Indeks Prestasi. Metode yang digunakan dalam penelitian ini adalah *Decision Tree* C4.5 yang diimplementasikan pada program WEKA dengan algoritma J48. Penelitian ini juga menggunakan algoritma SMOTE (*Synthetic Minority Oversampling Technique*) untuk menyeimbangkan jumlah data pada kelas minor. *Root* teratas dari penelitian ini adalah Jenis Kelamin yang mempengaruhi indeks prestasi mahasiswa di Universitas Advent Indonesia. Algoritma SMOTE pada penelitian ini berguna untuk membantu menaikkan hasil dari penelitian ini sebesar 7-8% bisa dilihat dari hasil akurasi pengujian *cross validation 10 folds* adalah 63.6672%, kemudian rata-rata hasil dari *precision* dan *recall* adalah 0.621 dan 0.637. Sementara untuk hasil akurasi dari *split test 70:30* adalah 62.7955%, kemudian rata-rata hasil dari *precision* dan *recall* adalah 0.621 dan 0.628. Jika dibandingkan dengan penggunaan algoritma *decision tree* C4.5 saja maka, akurasi dari pengujian *cross validation 10 folds* adalah 55.5044%, dengan rata-rata hasil dari *precision* dan *recall* adalah 0.545 dan 0.555. Sementara hasil akurasi dari *split test 70:30* adalah 55.2995% dengan rata-rata hasil dari *precision* dan *recall* adalah 0.554 dan 0.553. Hasil analisa menggunakan *confusion matrix* serta kurva ROC dengan hasil dari 0.688 menjadi 0.756, yang berada dalam rentang 0.70 – 0.80 yang masuk kedalam tingkat diagnosa *fair classification*. Dapat disimpulkan bawa terdapat pengaruh berkuliah sambil bekerja yang cukup kuat terhadap indeks prestasi mahasiswa. Dengan urutan atribut dari yang paling atas adalah Jenis Kelamin, Jumlah SKS, Jurusan, Umur, Departemen Kerja, Jam Kerja dan Tempat Tinggal.

Kata Kunci: *Data Mining, Decision Tree C4.5, SMOTE, Prediksi Indeks Prestasi, Berkuliah Sambil Bekerja.*

Predicting Student Grade Point Average Who Is Studying While Working At Adventist University Of Indonesia Using Decision Tree C4.5 Method And SMOTE

Abstract

Higher education is one way to get job easier, this thing happens because through education the individual is able to increase the level of human resources in this era. However, the high cost of education is very expensive so individuals who wants to study must also work at the same time, so this research aims

to predict the student GPA who is studying while working at the same time at Adventist University of Indonesia. From the results of this research there are 8 attributes that have an effect on predicting student GPA at Adventist University of Indonesia, namely the Department of Work, Working Hours, Course, Gender, Residence, Age and Number of Credits. The method that has been used in this research is Decision Tree C4.5 implemented on the WEKA program with the J48 algorithm. This research also uses the SMOTE (Synthetic Minority Oversampling Technique) algorithm to balancing the amount of data in the minor class. The top root of this research is Gender which affects the student GPA at University of Indonesia. The SMOTE algorithm in this research is useful to help raising the result of this research by 7-8% can be seen from the results of the accuracy of the cross validation 10 folds test is 63.6672%, the average result of precision and recall are 0.621 and 0.637. While the accuracy of the split test 70:30 is 62.7955%, then result of precision and recall are 0.621 and 0.628. When compared with the use of the Decision Tree C4.5 algorithm only, the accuracy of the cross validation 10 fold test is 55.5044%, with the average result of precision and recall is -.545 and 0.555. While the accuracy of the split test 70:30 is 55.2995% with the results of precision and recall is 0.554 and 0.553. The analysis results using confusion matrix and ROC curve with results from 0.688 to 0.756, which are in the range of 0.70 - 0.80 which is included in the level of fair classification diagnosis. It can be concluded that there is a strong effect while working on the student GPA. With the order of attributes from the top most are Gender, Total Credit, Department, Age, Department of Work, Working Hours and Residence.

Keywords: Data Mining, Decision Tree C4.5, SMOTE, Predicting Student GPA, studying while working.

1. Pendahuluan

Kesempatan mendapatkan kerja akan lebih mudah didapat jika seorang pencari kerja mempunyai latar belakang pendidikan di Perguruan tinggi. Namun biaya untuk bersekolah di Perguruan tinggi tidak murah yang menyebabkan seorang mahasiswa yang berkuliah sambil bekerja. Motivasi mahasiswa untuk memilih bekerja paruh waktu disebabkan oleh banyak faktor antara lain: guna memenuhi kebutuhan keuangan yang mendesak, atau sekadar mengisi waktu luang, ingin mendapatkan pengalaman bekerja, ingin bertemu dengan banyak orang (meningkatkan ketrampilan 5 komunikasi), ingin meningkatkan ketrampilan dan keahlian dalam bekerja sehingga dapat meningkatkan karir di masa datang. Maka dari itu penulis ingin membuat penelitian ini untuk memprediksi IPK mahasiswa dan mengevaluasi proses pendidikan yang ada di Universitas Advent Indonesia berdasarkan data yang telah ada mengenai mahasiswa yang berkuliah sambil bekerja dan juga indeks prestasi mereka. Atribut yang akan dipakai di penelitian ini antara lain adalah jenis kelamin, SKS yang diambil, dan waktu bekerja yang diikuti akan dijelaskan lebih jelas di bab metode penelitian.

2. Metode

Data penelitian yang digunakan oleh penulis adalah data yang telah didata dan tersusun rapi oleh Pembantu Rektor II dan Biro Administrasi Akademik. Data yang telah diambil oleh penulis tentunya dapat di kelolah oleh penulis untuk mendapatkan hasil yang akurat saat melakukan prediksi terhadap mahasiswa yang berkuliah sambil bekerja di Universitas Advent Indonesia.

Dalam tahapan penelitian ini penulis menggunakan data-data dari *database* UNAI, yang dikumpulkan menjadi 8 atribut. Subyek yang diteliti oleh penulis adalah data-data mahasiswa dari tahun 2013-2017. Kemudian data-data yang telah dikumpulkan akan diolah dengan aplikasi WEKA menggunakan metode *Decision Tree* C4.5 atau algoritma J48 pada program WEKA.

Tahapan Penelitian

1. Mengumpulkan atribut yang akan digunakan pada saat pengelolaan data.
2. Melakukan pengumpulan data dari *database* yang dibutuhkan sesuai dengan atribut yang sudah ditetapkan di *database* Universitas Advent Indonesia.

3. *Cleaning data* yaitu merapikan data-data yang tidak konsisten maupun data-data yang memiliki kekacauan yang harus dirapikan kembali agar data-data tersebut bias diproses. Dalam proses ini penulis menggunakan Microsoft excel dan diubah formatnya menjadi *csv (comma separated values)*.
4. Dalam penelitian ini penulis menggunakan metode *Decision Tree C.45* atau dalam aplikasi WEKA disebut juga *Decision Tree J48*.
5. Kemudian setelah itu data-data tersebut diproses menggunakan aplikasi WEKA.
6. Setelah melewati proses tersebut maka akan mendapatkan hasil dengan menggunakan metode *Decision Tree C4.5* yang diuji dalam aplikasi WEKA.

Fungsi utama dari *preprocessing data* atau pemrosesan awal data terbagi menjadi beberapa bagian yaitu untuk membersihkan data yang tidak valid atau *null* maupun data-data yang bersifat tidak konsisten. Kemudian integrasi data langkah ini berguna untuk menggabungkan data-data dari berbagai tabel, *file*, kubus data atau *file text* lainnya ke dalam satu tabel. Selanjutnya adalah seleksi data yang berguna untuk menyeleksi data-data yang relevan dari tabel, *file*, kubus data maupun *file text* yang diperlukan. Langkah yang terakhir adalah transformasi data yaitu proses pengubahan data dan penggabungan data ke dalam bentuk yang dibutuhkan untuk melakukan operasi agregasi atau mengubah nilai.

[1] SMOTE atau (*Synthetic Minority Oversampling Technique*) adalah solusi mengatasi ketidakseimbangan data. Jika metode *oversampling* lainnya memperbanyak data secara acak maka metode SMOTE ini menambah jumlah data kelas minoritas agar setara dengan kelas mayoritas, data yang ditambah adalah data buatan menggunakan rumus penghitungan jarak Euclidean untuk menghitung data numerik, sementara untuk data kategori dihitung dengan rumus *Value Difference Metric (VDM)* yaitu:

$$\Delta(X, Y) = w_x w_y \sum_{i=1}^N \delta(x_i, y_i)^r \quad (1)$$

Dengan:

$\Delta(X, Y)$: jarak antara amatan X dengan Y

$w_x w_y$: bobot amatan (dapat diabaikan)

N : banyaknya perubah penjelas

R : bernilai 1 (jarak Manhattan) atau 2 (jarak Euclidean)

$\delta(x_i, y_i)^r$: jarak antar kategori, dengan rumus:

$$\delta(V_1, V_2) = \sum_{i=1}^n \left| \frac{C_{1i}}{C_1} - \frac{C_{2i}}{C_2} \right|^k \quad (2)$$

$\delta(V_1, V_2)^r$: jarak antara nilai V1 dan V2

C_{1i} : banyaknya V1 yang termasuk kelas I

C_{2i} : banyaknya V2 yang termasuk kelas I

I : banyaknya kelas; $I = 1, 2, \dots, m$

C_1 : banyaknya nilai 1 terjadi

C_2 : banyaknya nilai 2 terjadi

N : banyaknya kategori

K : konstanta (biasanya 1)

Dengan data yang akan dikumpulkan, penulis mencoba untuk menyelesaikan dengan menggunakan metode *Decision Tree-J48* yang juga bisa disebut *Decision Tree C4.5* dengan rumus di bawah ini.

$$Entropy(S) = -p1log2p1-p2log2p2 (2)$$

Keterangan:

S : Himpunan kasus

pl : Proporsi dari sv terhadap S

Rumus *Gain* digunakan untuk mencari nilai keseluruhan dari *Entropy*, di bawah ini merupakan rumus untuk mencari *Gain*:

$$Gain(S,A) = Entropy(s) - \sum_{v \in values(A)} \frac{sv}{s} * Entropy(Sv)$$

Keterangan:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

sv : Jumlah kasus pada partisi ke v

s : Jumlah kasus dalam S

3. Hasil

Tabel 1 Perbandingan Hasil *Cross Validation*

Algoritma	Akurasi	Precision	Recall	ROC Area
<i>Decision Tree C4.5</i>	55.965%	0.549	0.560	0.676
SMOTE	64.3143%	0.629	0.643	0.770

Dari perbandingan di atas algoritma *Decision Tree C4.5* memiliki hasil akurasi, *precision*, *recall* dan *ROC Area* yang lebih kecil dibandingkan dengan hasil *cross validation* menggunakan algoritma *Decision Tree C4.5* dan algoritma SMOTE secara bersamaan. Algoritma SMOTE berguna untuk meningkatkan hasil akurasi dari 55.965% menjadi 64.3143%, *precision* dari 0.549 menjadi 0.629, *recall* dari 0.560 menjadi 0,643 dan *ROC Area* dari 0.676 menjadi 0.770.

Tabel 2 Perbandingan Hasil *Percentage Split*

Algoritma	Akurasi	Precision	Recall	ROC Area
<i>Decision Tree C4.5</i>	53.7634%	0.521	0.538	0.664
SMOTE	61.5622 %	0.595	0.616	0.750

Dari Perbandingan *percentage split* di atas algoritma *Decision Tree C4.5* memiliki hasil akurasi, *precision*, *recall* dan *ROC Area* yang lebih kecil dibandingkan dengan hasil *percentage split* menggunakan algoritma *Decision Tree C4.5* dan algoritma SMOTE secara bersamaan. Algoritma SMOTE berguna untuk meningkatkan hasil akurasi dari 53.7634% menjadi 61.5622%, *precision* dari 0.521menjadi 0.595, *recall* dari 0.538 menjadi 0.616 dan *ROC Area* dari 0.664 menjadi 0.750.

Urutan atribut ditetapkan oleh penulis melalui hasil penghitungan *entropy* dari hasil yang paling besar sampai kepada hasil yang penghitungan *entropy* yang paling kecil. Urutan atribut di bawah ini juga menunjukkan atribut mana yang paling berpengaruh terhadap indeks prestasi mahasiswa yang berkuliah sambil bekerja di Universitas Advent Indonesia:

1. Jenis Kelamin dengan hasil penghitungan yaitu *entropy* 0.8
2. Jumlah SKS dengan hasil penghitungan yaitu *entropy* 0.1
3. Jurusan dengan hasil perhitungan *entropy* yaitu 0.06

5. Departemen Kerja dengan hasil perhitungan *entropy* yaitu 0.006
6. Umur dengan hasil perhitungan *entropy* yaitu 0.001
7. Jam Kerja dengan hasil perhitungan *entropy* yaitu 0.0017
8. Tempat Tinggal dengan hasil perhitungan *entropy* yaitu 0.0011

4. Pembahasan

Pengujian menggunakan *decision tree* C4.5 tanpa menggunakan SMOTE

(1) Accuracy

Correctly Classified Instances	1215	55.965 %
Incorrectly Classified Instances	956	44.035 %

Gambar 1 Accuracy Cross Validation

Dari gambar di atas mengartikan bahwa data yang diklasifikasikan dengan benar adalah 55.965% sedangkan data yang diklasifikasikan secara tidak benar adalah 44.035%.

(2) Confusion Matrix

```

=== Confusion Matrix ===
      a  b  c  <-- classified as
150  63 324 |  a = HIGH
 52 255 273 |  b = UNDER
111 133 810 |  c = AVERAGE
    
```

Gambar 2 Confusion Matrix Cross Validation

Dari gambar *confusion matrix* 2 dapat dilihat bahwa:

- a. 150 data *High* yang diklasifikasikan menjadi data *High*, 63 data *High* yang diklasifikasikan menjadi data *Under* dan 324 data *High* yang diklasifikasikan menjadi data *Average*.
- b. 52 data *Under* yang diklasifikasikan menjadi data *High*, 255 data *Under* yang diklasifikasikan menjadi data *Under* dan 273 data *Under* yang diklasifikasikan menjadi data *Average*.
- c. 111 data *Average* yang diklasifikasikan menjadi data *High*, 133 data *Average* yang diklasifikasikan menjadi data *Under* dan 810 data *Average* yang diklasifikasikan menjadi data *Average*.

(3) Precision Recall

Precision	Recall
0.479	0.279
0.565	0.440
0.576	0.769
0.549	0.560

Gambar 3 Precision Recall Cross Validation

Dari Gambar 3 dapat dijelaskan bahwa:

- a. *Precision* dari data *High* adalah 0.479 yang dikalkulasikan dari 150 data *High* dibagi dengan 313 data data yang diklasifikasikan menjadi data *High*. Hasil *Recall* adalah 0.279 yang dikalkulasikan dari 150 data *High* dibagi dengan 537 data keseluruhan *High*.

- b. *Precision* dari data *Under* adalah 0.565 yang dikalkulasikan dari 255 data *Under* dibagi dengan 451 data yang diklasifikasikan menjadi data *Under*. Hasil *Recall* adalah 0.440 yang dikalkulasikan dari 255 data *Under* dibagi dengan 580 data keseluruhan *Under*.
- c. *Precision* dari data *Average* adalah 0.576 yang dikalkulasikan dari 810 data *Average* dibagi dengan 1407 data yang diklasifikasikan menjadi data *Average*. Hasil *Recall* adalah 0.769 yang dikalkulasikan dari 810 data *Average* dibagi dengan 1054 data keseluruhan *Average*.
- d. *Average Precision* dari seluruh data adalah 0.549 dan *Average Recall* dari seluruh data adalah 0.560.

Pengujian *Split Test* 70:30 Tanpa SMOTE

(1) *Accuracy*

Correctly Classified Instances	350	53.7634 %
Incorrectly Classified Instances	301	46.2366 %

Gambar 4 *Accuracy Split Test*

Dari gambar di atas mengartikan bahwa data yang diklasifikasikan dengan benar adalah 53.7634% sedangkan data yang diklasifikasikan secara tidak benar adalah 46.2366%.

(2) *Confusion Matrix*

```

=== Confusion Matrix ===
      a  b  c  <-- classified as
39  21  92 |  a = HIGH
15  67  95 |  b = UNDER
36  42 244 |  c = AVERAGE
    
```

Gambar 5 *Confusion Matrix Split Test*

Dari gambar *confusion matrix* 5 dapat dilihat bahwa:

- a. 39 data *High* yang diklasifikasikan menjadi data *High*, 21 data *High* yang diklasifikasikan menjadi data *Under* dan 92 data *High* yang diklasifikasikan menjadi data *Average*.
- b. 15 data *Under* yang diklasifikasikan menjadi data *High*, 67 data *Under* yang diklasifikasikan menjadi data *Under* dan 95 data *Under* yang diklasifikasikan menjadi data *Average*.
- c. 36 data *Average* yang diklasifikasikan menjadi data *High*, 42 data *Average* yang diklasifikasikan menjadi data *Under* dan 244 data *Average* yang diklasifikasikan menjadi data *Average*.

(3) *Precision Recall*

Precision	Recall
0.433	0.257
0.515	0.379
0.566	0.758
0.521	0.538

Gambar 6 *Precision Recall Split Test*

Dari gambar 6 dapat dijelaskan bahwa:

- d. *Precision* dari data *High* adalah 0.433 yang dikalkulasikan dari 39 data *High* dibagi dengan 90 data yang diklasifikasikan menjadi data *High*. Hasil *Recall* adalah 0.257 yang dikalkulasikan dari 39 data *High* dibagi dengan 152 data keseluruhan *High*.

- e. *Precision* dari data *Under* adalah 0.515 yang dikalkulasikan dari 67 data *Under* dibagi dengan 130 data yang diklasifikasikan menjadi data *Under*. Hasil *Recall* adalah 0.440 yang dikalkulasikan dari 67 data *Under* dibagi dengan 177 data keseluruhan *Under*.
- f. *Precision* dari data *Average* adalah 0.566 yang dikalkulasikan dari 244 data *Average* dibagi dengan 431 data yang diklasifikasikan menjadi data *Average*. Hasil *Recall* adalah 0.758 yang dikalkulasikan dari 322 data *Average* dibagi dengan 1054 data keseluruhan *Average*.
- g. *Average Precision* dari seluruh data adalah 0.521 dan *Average Recall* dari seluruh data adalah 0.538.

Pengujian *Cross Validation 10 Folds* Menggunakan SMOTE

(1) *Accuracy*

Correctly Classified Instances	2087	64.3143 %
Incorrectly Classified Instances	1158	35.6857 %

Gambar 7 *Accuracy Cross Validation*

Dari gambar di atas mengartikan bahwa data yang diklasifikasikan dengan benar adalah 64.3143% sedangkan data yang diklasifikasikan secara tidak benar adalah 35.6857%.

(2) *Confusion Matrix*

```

=== Confusion Matrix ===
      a   b   c  <-- classified as
1346   75  190 |   a = HIGH
  176  227  177 |   b = UNDER
  395  145  514 |   c = AVERAGE
    
```

Gambar 8 *Confusion Matrix Cross Validation*

Dari Gambar *confusion matrix* 8 dapat dilihat bahwa:

- a. 1346 data *High* yang diklasifikasikan menjadi data *High*, 75 data *High* yang diklasifikasikan menjadi data *Under* dan 190 data *High* yang diklasifikasikan menjadi data *Average*.
- b. 176 data *Under* yang diklasifikasikan menjadi data *High*, 227 data *Under* yang diklasifikasikan menjadi data *Under* dan 177 data *Under* yang diklasifikasikan menjadi data *Average*.
- c. 395 data *Average* yang diklasifikasikan menjadi data *High*, 145 data *Average* yang diklasifikasikan menjadi data *Under* dan 514 data *Average* yang diklasifikasikan menjadi data *Average*.

(3) *Precision Recall*

<i>Precision</i>	<i>Recall</i>
0.702	0.836
0.508	0.391
0.583	0.488
0.629	0.643

Gambar 9 *Precision Recall Cross Validation*

Dari gambar 9 dapat dijelaskan bahwa:

- Precision* dari data *High* adalah 0.602 yang dikalkulasikan dari 1346 data *High* dibagi dengan 1917 data yang diklasifikasikan menjadi data *High*. Hasil *Recall* adalah 0.836 yang dikalkulasikan dari 1346 data *High* dibagi dengan 1611 data keseluruhan *High*.
- Precision* dari data *Under* adalah 0.508 yang dikalkulasikan dari 227 data *Under* dibagi dengan 447 data yang diklasifikasikan menjadi data *Under*. Hasil *Recall* adalah 0.391 yang dikalkulasikan dari 227 data *Under* dibagi dengan 580 data keseluruhan *Under*.
- Precision* dari data *Average* adalah 0.583 yang dikalkulasikan dari 514 data *Average* dibagi dengan 881 data yang diklasifikasikan menjadi data *Average*. Hasil *Recall* adalah 0.488 yang dikalkulasikan dari 514 data *Average* dibagi dengan 1054 data keseluruhan *Average*.
- Average Precision* dari seluruh data adalah 0.629 dan *Average Recall* dari seluruh data adalah 0.643.

Pengujian *Split Test* 70:30 Menggunakan SMOTE

(1) *Accuracy*

Correctly Classified Instances	599	61.5622 %
Incorrectly Classified Instances	374	38.4378 %

Gambar 10 *Accuracy Split Test*

Dari gambar di atas mengartikan bahwa data yang diklasifikasikan dengan benar adalah 61.5622 % sedangkan data yang diklasifikasikan secara tidak benar adalah 38.4378%.

(2) *Confusion Matrix*

```

=== Confusion Matrix ===
      a  b  c  <-- classified as
416  33  40 |  a = HIGH
  58  62  59 |  b = UNDER
147  37 121 |  c = AVERAGE
    
```

Gambar 11 *Confusion Matrix Split Test*

Dari Gambar *confusion matrix* 4.28 dapat dilihat bahwa:

- 416 data *High* yang diklasifikasikan menjadi data *High*, 33 data *High* yang diklasifikasikan menjadi data *Under* dan 40 data *High* yang diklasifikasikan menjadi data *Average*.
- 58 data *Under* yang diklasifikasikan menjadi data *High*, 62 data *Under* yang diklasifikasikan menjadi data *Under* dan 59 data *Under* yang diklasifikasikan menjadi data *Average*.
- 147 data *Average* yang diklasifikasikan menjadi data *High*, 37 data *Average* yang diklasifikasikan menjadi data *Under* dan 121 data *Average* yang diklasifikasikan menjadi data *Average*.

(3) *Precision Recall*

Precision	Recall
0.702	0.836
0.508	0.391
0.583	0.488
0.629	0.643

Gambar 12 Precision Recall Cross Validation

Dari gambar 12 dapat dijelaskan bahwa:

- a. *Precision* dari data *High* adalah 0.702 yang dikalkulasikan dari 416 data *High* dibagi dengan 612 data yang diklasifikasikan menjadi data *High*. Hasil *Recall* adalah 0.836 yang dikalkulasikan dari 416 data *High* dibagi dengan 489 data keseluruhan *High*.
- b. *Precision* dari data *Under* adalah 0.470 yang dikalkulasikan dari 62 data *Under* dibagi dengan 132 data yang diklasifikasikan menjadi data *Under*. Hasil *Recall* adalah 0.346 yang dikalkulasikan dari 62 data *Under* dibagi dengan 179 data keseluruhan *Under*.
- c. *Precision* dari data *Average* adalah 0.550 yang dikalkulasikan dari 121 data *Average* dibagi dengan 220 data yang diklasifikasikan menjadi data *Average*. Hasil *Recall* adalah 0.397 yang dikalkulasikan dari 121 data *Average* dibagi dengan 305 data keseluruhan *Average*.
- d. *Average Precision* dari seluruh data adalah 0.595 dan *Average Recall* dari seluruh data adalah 0.616.

5. Kesimpulan

Hasil analisa kurva ROC dengan hasil dari 0.676 menjadi 0.770, yang berada dalam rentang 0.70 – 0.80 yang masuk kedalam tingkat diagnose *fair classification*. Dapat disimpulkan bawa terdapat pengaruh berkuliah sambil bekerja yang cukup kuat terhadap indeks prestasi mahasiswa. Dengan atribut yang paling mempengaruhi yang diurutkan dari hasil penghitungan *entropy* adalah Jenis Kelamin dengan hasil *entropy* 0.8, Jumlah SKS hasil *entropy* 0.1, Jurusan hasil *entropy* 0.06, Departemen Kerja hasil *entropy* 0.006, Umur dengan hasil *entropy* 0.001, Jam Kerja hasil *entropy* 0.0017 dan Tempat Tinggal hasil *entropy* 0.0011.

6. Daftar Pustaka

- [1] Andriyan, David. (2016). *Indeks Prestasi Kumulatif Mahasiswa Ditinjau dari Strategi Belajar dan Keaktifan Berorganisasi pada Mahasiswa Pendidikan Akuntansi Universitas Muhammadiyah Surakarta tahun 2014*. [Online]. Available: <http://davidandriyan.blogspot.co.id/2016/07/proposal-penelitian-dengan-judul-indeks.html> [5 April 2018]
- [2] B. Rossi, S. Itasia, and A, Farit, "Penerapan *Synthetic Minority Oversampling Technique (SMOTE)* Terhadap Data Tidak Seimbang Pada Pembuatan Model Komposisi Jamu," vol. 1, no. 1, 2013. Diakses pada: Maret, 16, 2019. [Online]. Tersedia di: <http://dx.doi.org/10.29244/xplore.v1i1.12424>
- [3] Berry, Michael J.A dan Linoff, Gordon S (2004). *Data Mining Techniques For Marketing, Sales, Customer Relationship Management Second Edition*. United States of America: Wiley Publishing, Inc.
- [4] Defiyanti, Sofi (2014). *Perbandingan: Prediksi Prestasi Belajar Mahasiswa Menggunakan Teknik Data Mining (Study Kasus Fasilkom UNSIKA)*. Makasar: KNSI.
- [5] Larose, D. T. (2005). *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- [6] Lestari, Fenti (2016). *Pengaruh Lingkungan Keluarga Dan Fasilitas Belajar Terhadap Motivasi Belajar Dan Hasil Belajar Siswa Kelas XI IPS Pada Mata Pelajaran Ekonomi DI SMAN 2 Kebumen Tahun Pelajaran 2015/2016*. Yogyakarta: Universitas Negeri Yogyakarta.
- [7] Putri, Ratna P.S dan Waspada, Indra (2018). *Penerapan Algoritma C4.5 Pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika*. Semarang: Jurnal Ilmu Komputer dan Informatika