

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN METODE *DECISION TREE* DENGAN PENERAPAN ALGORITMA *C4.5*

F. F. Anwar¹, A. I. Jaya² dan M. Abu³

^{1,2,3}Program Studi Matematika Jurusan Matematika

Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tadulako

Jalan Soekarno-Hatta Km. 09 Tondo, Palu 94118, Indonesia.

¹fatma.f.anwar003@gmail.com, ²jayaindraagus@gmail.com, ³maulidyanisho7992@gmail.com

ABSTRACT

One of the higher education quality's determinants is the percentage of students' ability to complete their study period on time. However, in practice only a few students that are able to complete their studies on time. One thing that can be done to increase student graduation is by using graduation prediction, so that the prevention or treatment can be done early on students who have the opportunity to not graduate on time. Therefore, this study was conducted to determine student graduation on time by applying the C4.5 algorithm with 16 input attributes and 1 target attribute. The data used in this study is the data of alumni of the Mathematics Study Program, FMIPA UNTAD class 2004 - 2015 which collected 255 data. The decision tree was built with 150 training data and processed by using the C4.5 algorithm to produce 40 rules, which were the tested using 75 testing data with the accuracy value is 92% and the AUC value is 0,7138.

Keywords : C4.5 Algorithm, Decision Tree, Graduating on Time, Testing Data, Training Data.

ABSTRAK

Salah satu penentu kualitas perguruan tinggi adalah persentase kemampuan mahasiswa untuk menyelesaikan masa studisecara tepat waktu. Namun, dalam praktiknya hanya sedikit mahasiswa yang mampu menyelesaikan studinya di perguruan tinggi secara tepat waktu. Salah satu hal yang dapat dilakukan untuk meningkatkan kelulusan mahasiswa adalah prediksi kelulusan, sehingga dapat dilakukan pencegahan atau penanganan secara dini terhadap mahasiswa yangberpeluang tidak lulus tepat waktu. Oleh karena itu, penelitian ini dilakukan untuk memprediksi kelulusan mahasiswa tepat waktu dengan penerapan algoritma *C4.5* dengan 16 atribut *input* dan 1 atribut target. Data yang digunakan adalahdata alumni mahasiswa prodi matematika FMIPA UNTAD angkatan 2004 – 2015 sebanyak 225 data. Pohon keputusan dibangun dengan 150 data *training* dan diolah menggunakan algoritma *C4.5* menghasilkan 40 *rules*, yang kemudian diuji menggunakan 75 data *testing* dengan hasil nilai *accuracy* 92% dan nilai AUC sebesar 0,7138.

Kata Kunci : Algoritma C4.5, Data Training, Data Testing, Decision Tree, Lulus Tepat Waktu

I. PENDAHULUAN

Persaingan dalam dunia pendidikan antar perguruan tinggi sudah sangat kompetitif, sehingga untuk setiap perguruan tinggi kini dituntut untuk memiliki keunggulan dan kualitas yang baik dalam bersaing. Salah satu penentu kualitas perguruan tinggi adalah persentase kemampuan mahasiswa untuk menyelesaikan masa studi secara tepat waktu [9], atau yang biasa dikenal dengan "lulus tepat waktu". Mahasiswa dikatakan lulus tepat waktu apabila menyelesaikan studinya di perguruan tinggi selama kurang dari atau sama dengan empat tahun. Namun, dalam praktiknya hanya sedikit mahasiswa yang mampu menyelesaikan pendidikan sarjana dalam kurun waktu empat tahun. Hal ini bisa disebabkan oleh beberapa faktor, diantaranya adalah faktor internal dan faktor eksternal.

Faktor internal merupakan faktor yang berasal dari dalam diri mahasiswa itu sendiri, misalnya tingkat kecerdasan. Sebaliknya faktor eksternal merupakan faktor yang berasal dari luar diri mahasiswa, misalnya faktor lingkungan di sekitarnya. Selain itu, faktor kondisi ekonomi keluarga juga dapat mempengaruhi mahasiswa lulus tepat waktu atau tidak [10].

Untuk meningkatkan jumlah mahasiswa yang lulus tepat waktu, maka perlu diprediksi mana mahasiswa yang bisa lulus tepat waktu dan yang tidak tepat waktu, sehingga dapat dilakukan pencegahan atau penanganan secara dini terhadap mahasiswa yang berpotensi tidak lulus tepat waktu. Bagi mahasiswa yang diprediksikan tidak dapat lulus tepat waktu, bisa diberikan bimbingan atau arahan serta dorongan dan juga motivasi belajar agar mahasiswa tersebut bisa lulus sesuai dengan waktu yang diharapkan.

Salah satu teknik yang dapat digunakan untuk melakukan prediksi yaitu teknik penggalian data / *data mining*. *Data mining* merupakan serangkaian proses untuk mendapatkan pengetahuan atau pola dari kumpulan data [5]. *Data mining* sendiri memiliki beberapa teknik salah satunya klasifikasi. Klasifikasi adalah proses untuk menemukan model atau fungsi yang akan mengelompokkan kelas data, kemudian model tersebut dapat digunakan untuk memprediksi kelas obyek yang tidak diketahui. *Decision tree* (pohon keputusan) merupakan salah satu metode klasifikasi yang paling populer digunakan karena mudah diinterpretasikan oleh manusia [4]. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain *ID3*, *CART*, dan *C4.5* [7].

Penelitian mengenai kelulusan mahasiswa telah banyak dilakukan, salah satu diantaranya yaitu penelitian yang dilakukan oleh Kamagi dan Hansun [6]. Penelitian tersebut bertujuan untuk memprediksi tingkat kelulusan mahasiswa dengan menggunakan algoritma *C4.5* berdasarkan nilai IP semester I-VI, jenis kelamin, asal SMA, dan jumlah SKS pada semester VI. Dari penelitian tersebut, diperoleh hasil bahwa faktor yang paling berpengaruh pada tingkat kelulusan mahasiswa adalah nilai IP semester VI dengan tingkat akurasi sebesar 87,5%. Hal ini berarti algoritma *C4.5* memiliki kinerja yang cukup baik dalam memprediksi tingkat kelulusan mahasiswa.

Penelitian ini bertujuan untuk memprediksi kelulusan mahasiswa tepat waktu pada mahasiswa Prodi Matematika Fakultas MIPA Universitas Tadulako angkatan 2004 – 2015 menggunakan metode *decision tree* dengan penerapan algoritma *C4.5*. *Output* dari *decision tree* dibagi menjadi dua kelas yaitu “Tepat Waktu” dan “Tidak Tepat Waktu” dengan menggunakan 16 atribut, yaitu : jenis kelamin, asal SLTA, nilai IP semester I – VI, jumlah SKS yang dilulusi semester I – VI, status domisili serta penghasilan orang tua/wali per bulan. Dengan menerapkan algoritma *C4.5* akan menghasilkan pola tingkat kelulusan mahasiswa, yaitu berupa *rules* yang dapat digunakan untuk memprediksi kelulusan mahasiswa Prodi Matematika Fakultas MIPA Universitas Tadulako.

II. METODE PENELITIAN

1. Memulai penelitian.
2. Mengkaji literatur – literatur yang berkaitan dengan kelulusan mahasiswa tepat waktu.
3. Mengidentifikasi masalah, yaitu dengan merumuskan masalah dari kelulusan mahasiswa, kemudian mencari solusi dan metode yang dapat digunakan. Dalam penelitian ini akan dilakukan prediksi kelulusan mahasiswa dengan menerapkan algoritma *C4.5*.
4. Mengumpulkan data alumni mahasiswa Prodi Matematika FMIPA UNTAD.
5. Melakukan *preprocessing*, yaitu menghapus data yang tidak lengkap (*missing value*) serta menentukan atribut yang akan digunakan. Selanjutnya melakukan transformasi data, yaitu melakukan peringkasan data atau mengubah data mentah menjadi data yang mudah dikelola.
6. Pembagian data sebesar 150 data untuk *training* dan 75 data untuk *testing*.
7. Menerapkan algoritma *C4.5* untuk membuat model pohon keputusan.
8. Melakukan pengujian terhadap model pohon keputusan yang telah dibangun sebelumnya.
9. Pemaparan hasil dan kesimpulan.

III. HASIL DAN PEMBAHASAN

3.1. *Preprocessing Data*

Data yang digunakan adalah data alumni mahasiswa Prodi Matematika FMIPA UNTAD angkatan 2004 – 2015. Data diperoleh penulis dari berkas kliring alumni yang berada di Prodi Matematika FMIPA UNTAD, dari alumni yang bersangkutan (responden) dan dari berbagai pihak yang berhubungan erat dengan responden. Data yang diperoleh berjumlah 244 data.

Setelah dilakukan *preprocessing data* yaitu membersihkan data – data yang tidak lengkap, maka jumlah data menjadi 225 data yang terdiri atas 21 data mahasiswa yang lulus tepat waktu (masa studi ≤ 4 tahun) dan 204 data mahasiswa yang lulus tidak tepat waktu (masa studi > 4 tahun). Adapun beberapa atribut atau parameter yang dianggap dapat mempengaruhi terhadap tingkat kelulusan mahasiswa, yaitu : jenis kelamin, asal SLTA,

status domisili mahasiswa saat kuliah di UNTAD, penghasilan orang tua / wali mahasiswa, nilai IP semester (I, II, III, IV, V, dan VI), dan jumlah SKS yang dilulusi pada semester (I, II, III, IV, V, dan VI).

3.2. Transformasi Data

Pada tahap ini, akan dilakukan peringkasan data atau mengubah data menjadi data yang mudah dikelola dalam algoritma *C4.5*. Dalam kasus ini akan digunakan data yang bersifat diskrit, oleh karena itu data yang bersifat kontinu akan diubah menjadi data diskrit. Setelah itu, data – data tersebut diubah formatnya menjadi format *Attribute Relation File Format (ARFF)* agar bisa terbaca oleh *WEKA*. Proses transformasi dilakukan dengan cara membuat klasifikasi pada masing-masing atribut *input*, seperti yang terlihat pada Tabel 1.

Tabel 1 : Transformasi Data

No.	Atribut	Skala	Nilai Atribut	
1	Jenis Kelamin (JK)	Nominal	Laki – laki	= L
			Perempuan	= P
2	Asal SLTA Mahasiswa (AS)	Nominal	Sekolah Menengah Atas	= SMA
			Sekolah Menengah Kejuruan	= SMK
			Madrasah Aliyah	= MA
3	Status Domisili (SD)	Nominal	Bersama orang tua	= ORTU
			Kos/kontrakan/sewa	= KOS/ SEWA
			Lainnya (milik pribadi, asrama, bersama wali, dll)	= Lainnya
4	Penghasilan Orang Tua /Wali Perbulan (PO)	Ordinal	≤ Rp. 1.500.000	= Rendah
			Rp. 1.500.001 – Rp. 2.500.000	= Sedang
			Rp. 2.500.001 – Rp. 3.500.000	= Tinggi
			> Rp. 3.500.000	= Sangat Tinggi
5	Indeks Prestasi (IP)	Ordinal	0,00 - 0,99	= 1
			1,00 - 1,99	= 2
			2,00 - 2,99	= 3
			3,00 - 4,00	= 4
6	Jumlah SKS yang dilulusi (SKS)	Ordinal	≤ 15 SKS	= 1
			16 - 18 SKS	= 2
			19 - 21 SKS	= 3
			21 - 24 SKS	= 4

Menurut Badan Pusat Statistik (BPS), golongan pendapatan penduduk dibedakan menjadi 4, yaitu golongan pendapatan sangat tinggi dengan rata – rata lebih dari Rp. 3.500.000 per bulan, golongan pendapatan tinggi dengan rata-rata antara Rp. 2.500.000 – Rp. 3.500.000 per bulan, golongan pendapatan sedang dengan rata-rata antara Rp. 1.500.000 – Rp. 2.500.000 per bulan, dan golongan pendapatan rendah dengan rata-rata kurang dari atau sama dengan Rp. 1.500.000 per bulan [1]. Adapun tujuan penulis menambahkan angka 1 pada nilai batas bawah baru seperti yang terlihat pada Tabel 1 di atas adalah untuk membedakan dengan kriteria nilai batas atas sebelumnya, sehingga tidak terdapat keanggotaan yang sama pada masing – masing golongan pendapatan orang tua/wali mahasiswa per bulannya. Di Prodi Matematika FMIPA UNTAD, nilai IP yang mungkin diperoleh mahasiswa per semesternya adalah antara 0,00 - 4,00, oleh karena itu pada penelitian ini penulis membagi/mentrasformasikan atribut nilai IP menjadi empat bagian seperti yang terlihat pada tabel di atas. Sedangkan untuk transformasi atribut jumlah SKS yang dilulusi merupakan ide dari penulis sendiri yang didasarkan dengan aturan pembelian Kartu Rencana Studi dan disesuaikan dengan data alumni yang diperoleh.

Selanjutnya atribut – atribut *input* tersebut akan dianalisa menggunakan *algoritma C4.5* untuk menghasilkan atribut target. Atribut target merupakan *class output* untuk menentukan mahasiswa lulus tepat waktu atau tidak tepat waktu, sesuai dengan atribut target yang dapat dilihat pada Tabel 2.

Tabel 2 : Atribut Target

Atribut Target		Keterangan
Tepat Waktu	= T	Masa studi \leq 4 tahun
Tidak Tepat Waktu	= TT	Masa studi $>$ 4 tahun

Pada Tabel 2 dapat terlihat bahwa terdapat dua *class output* dari atribut target yang diinginkan, yaitu lulus “Tepat Waktu” dan “Tidak Tepat Waktu” yang ditentukan berdasarkan masa studi mahasiswa. Setelah dilakukan transformasi data dari data yang bersifat kontinu ke diskrit, selanjutnya data tersebut diubah dari excel ke *notepad* seperti yang terlihat pada Gambar 1.

```

@relation prediksi_kelulusan_mahasiswa
@attribute JK {P,L}
@attribute AS {SMA,SMK,MA}
@attribute SD {KOS/SEWA,ORTU,LAINNYA}
@attribute PO {Rendah,Sedang,Tinggi,Sangat_Tinggi}
@attribute IPS1 {1,2,3,4}
@attribute IPS2 {1,2,3,4}
@attribute IPS3 {1,2,3,4}
@attribute IPS4 {1,2,3,4}
@attribute IPS5 {1,2,3,4}
@attribute IPS6 {1,2,3,4}
@attribute SKS1 {1,2,3,4}
@attribute SKS2 {1,2,3,4}
@attribute SKS3 {1,2,3,4}
@attribute SKS4 {1,2,3,4}
@attribute SKS5 {1,2,3,4}
@attribute SKS6 {1,2,3,4}
@attribute KLASIFIKASI {T,TT}

@data
P,MA,ORTU,Sedang,4,4,3,3,4,3,4,4,4,4,3,3,TT
P,SMA,ORTU,Tinggi,4,4,4,4,4,4,4,4,4,4,4,3,T
P,SMA,ORTU,Sangat_Tinggi,4,4,3,4,4,4,4,4,4,4,4,4,T
L,SMA,ORTU,Sedang,4,4,4,4,4,4,4,4,4,4,4,4,T
P,SMA,KOS/SEWA,Rendah,4,4,4,4,4,4,4,4,4,3,4,4,4,TT
P,SMA,ORTU,Sedang,4,4,4,3,4,3,4,4,4,4,3,4,T
P,SMA,KOS/SEWA,Sedang,3,4,3,3,4,3,4,3,3,3,4,TT
L,SMA,KOS/SEWA,Sedang,4,3,3,4,2,4,4,3,3,4,TT
P,SMA,KOS/SEWA,Sedang,4,4,4,3,4,4,4,4,3,3,4,TT
P,SMK,KOS/SEWA,Sedang,4,4,3,4,4,4,4,4,3,4,4,3,TT
L,SMA,KOS/SEWA,Sangat_Tinggi,4,4,4,4,4,4,4,4,4,4,4,T
.
.
.
dst

```

Gambar 1 : Tampilan Data yang Diubah ke *Notepad*

Gambar 1 merupakan tampilan data yang diubah dari excel ke notepad yang kemudian disimpan dalam format *arff* agar bisa terbaca oleh *WEKA*.

3.3. Pembagian Data

Pada tahap ini, data akan dibagi menjadi dua yaitu data *training* dan *testing*. Data *training* berjumlah 150 data yang terdiri dari 12 data mahasiswa yang lulus tepat waktu dan 138 data mahasiswa yang lulus tidak tepat waktu. Sedangkan data *testing* berjumlah 75 data yang terdiri dari 9 data mahasiswa yang lulus tepat waktu dan 66 data mahasiswa yang lulus tidak tepat waktu.

3.4. Pembentukan *Rules Decision Tree*

Berdasarkan model pohon keputusan yang telah dibangun, maka diperoleh hasil bahwa dari 16 atribut yang digunakan, hanya ada 11 atribut yang dapat mempengaruhi mahasiswa lulus tepat waktu dan tidak tepat waktu, di antaranya adalah : IP semester (I,II,III dan V), jumlah SKS yang dilulusi pada semester (II,III,IV dan V), asal SLTA, jenis kelamin dan penghasilan orang tua/ wali. Berikut adalah tampilan *rules* yang dihasilkan pada *WEKA*.

```

IPS5 = 1: TT (1.0)
IPS5 = 2: TT (15.0)
IPS5 = 3: TT (70.0)
IPS5 = 4
| SKS4 = 1: TT (1.0)
| SKS4 = 2: TT (3.0)
| SKS4 = 3: TT (20.0)
| SKS4 = 4
| | IPS1 = 1: TT (0.0)
| | IPS1 = 2: TT (0.0)
| | IPS1 = 3: TT (6.0)
| | IPS1 = 4
| | | AS = SMA
| | | | SKS3 = 1: TT (0.0)
| | | | SKS3 = 2: TT (0.0)
| | | | SKS3 = 3: TT (2.0)
| | | | SKS3 = 4
| | | | | IPS2 = 1: TT (0.0)
| | | | | IPS2 = 2: TT (0.0)
| | | | | IPS2 = 3: TT (1.0)
| | | | | IPS2 = 4
| | | | | | SKS6 = 1: TT (2.0)
| | | | | | SKS6 = 2
| | | | | | | PO = Rendah: T (2.0)
| | | | | | | PO = Sedang
| | | | | | | JK = P: T (1.0)
| | | | | | | JK = L: TT (1.0)
| | | | | | | PO = Tinggi: TT (1.0)
| | | | | | | PO = Sangat_Tinggi: TT (1.0)
| | | | | | | SKS6 = 3

```

Gambar 2 : *Rules Decision Tree* pada WEKA

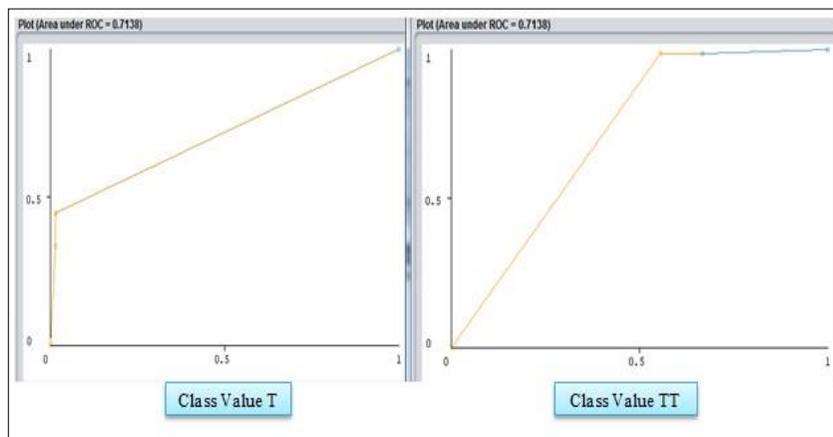
3.5. Pengujian Model

Pada tahap ini, dari 40 *rules* yang telah diperoleh dari proses *mining* kemudian akan dijadikan acuan dalam memprediksi kelas target (lulus tepat waktu dan tidak tepat waktu) pada data *testing* dengan menggunakan metode *Confusion Matrix*. Data *testing* yang digunakan berjumlah 75 data yang terdiri dari 9 data mahasiswa yang lulus tepat waktu dan 66 data mahasiswa yang lulus tidak tepat waktu. Adapun hasil dari pengujian model pohon keputusan tersebut menggunakan *WEKA* dapat dilihat pada Gambar 3.

```
=== Evaluation on test set ===  
  
Time taken to test model on supplied test set: 0.02 seconds  
  
=== Summary ===  
  
Correctly Classified Instances      69      92 %  
Incorrectly Classified Instances    6        8 %
```

Gambar 3 : Hasil Pengujian Model *Decision Tree*

Berdasarkan Gambar 3 di atas, maka dapat diperoleh hasil bahwa dari 75 data, jumlah nilai prediksi yang benar yaitu 69 data atau 92% dan jumlah nilai prediksi yang salah yaitu 6 data atau 8%. Adapun gambar kurva *ROC* dapat dilihat pada Gambar 4.



Gambar 4 : Kurva *ROC*

Untuk klasifikasi *data mining*, nilai *AUC* dapat dibagi menjadi beberapa [3] :

- 0,90 – 1,00 = Klasifikasi sangat baik
- 0,80 – 0,90 = Klasifikasi baik
- 0,70 – 0,80 = Klasifikasi cukup
- 0,60 – 0,70 = Klasifikasi buruk
- 0,50 – 0,60 = Klasifikasi salah/gagal.

Dari Gambar 4 dapat dilihat bahwa nilai *area under ROC* atau nilai *AUC* adalah sebesar 0,7138. Hal ini berarti bahwa model yang diperoleh memiliki kinerja yang cukup dalam memprediksi kelulusan mahasiswa tepat waktu.

IV. KESIMPULAN

1. Dari 150 data *training* yang diolah menggunakan algoritma *C4.5* menghasilkan 40 *rules* dan setelah diterapkan pada 75 data *training* (75 data mahasiswa yang dianggap belum lulus), maka diperoleh hasil bahwa terdapat 69 data mahasiswa yang diprediksikan dengan benar dan 6 data mahasiswa yang diprediksikan dengan salah.
2. Berdasarkan nilai *AUC* sebesar 0,7138, maka algoritma *C4.5* memiliki kinerja yang cukup dalam memprediksi kelulusan mahasiswa tersebut. Adapun tingkat keakuratan algoritma *C4.5* dalam memprediksi kelulusan mahasiswa Prodi Matematika FMIPA UNTAD adalah *Accuracy* sebesar 92%, *Error rate* sebesar 8%.

DAFTAR PUSTAKA

- [1]. Afriani, R., Yulastri, A., dan Syarif, W., Hubungan Pendapatan Orang Tua dengan Hasil Belajar Siswa Mata Pelajaran Praktik Di SMK Negeri 1 Sintuk Toboh Gadang Padang Pariaman, *Journal of Home Economics And Tourism*.8, 2015.
- [2]. Asroni, Respati, B.M., dan Riyadi, Penerapan Algoritma *C4.5* untuk Klasifikasi Jenis Pekerjaan Alumni Di Universitas Muhammadiyah Yogyakarta, *Semesta Teknika*. 21, 2018, 158 – 165.
- [3]. Gorunescu, F., *Data Mining : Concepts, Model and Technique*, Berlin, 2011, Springer.
- [4]. Haryati, S., Sudarsono, A., dan Suryana, E., Implementasi *Data Mining* untuk Memprediksikan Masa Studi Mahasiswa Menggunakan Algoritma *C.45* (Studi Kasus : Universitas Dahasen Bengkulu), *Media Infotama*. 11, 2015, 130 –138.
- [5]. Hastuti, K., Analisis Komparasi Algoritma *Klasifikasi Data Mining* untuk Prediksi Mahasiswa Non Aktif, *Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012 (Semantik 2012)*, 2012, 241 – 249.
- [6]. Kamagi, D.H., dan Hansun, S., Implementasi *Data Mining* dengan Algoritma *C4.5* untuk Memprediksikan Tingkat Kelulusan Mahasiswa, *Ultimatics*. 6, 2014, 15 – 20.
- [7]. Larose, D.T., *Discovering Knowledge in Data: An Introduction to Data Mining*, Hoboken , 2005, Jhon Wiley & Sonsinc.
- [8]. Novianti, B., Rismawan, T., dan Bahri, S., Implementasi *Data Mining* dengan Algoritma *C4.5* untuk Penjurusan Siswa (Studi kasus : SMA Negeri 1 Pontianak), *Coding*. 4, 2016, 75

– 84.

- [9]. Rismayanti, *Decision Tree* Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika (Studi kasus : Fakultas Teknikdan Komputer Universitas Harapan Medan), *Jurnal Sistem Informasi. 2*, 2018, 16 – 24.
- [10]. Soeparman, S., Faktor – Faktor yang Mempengaruhi Keberhasilan Studi Mahasiswa Penyandang Disabilitas, *Indonesian Journal of Disability Studies. 1*, 2014, 12 – 19.