

PENERAPAN DATA MINING ALGORITMA NAIVES BAYES DAN PART UNTUK MENGETAHUI MINAT BACA MAHASISWA DI PERPUSTAKAAN STMIK AMIKOM PURWOKERTO

Mohammad Imron

Teknik Informatika
STMIK Amikom Purwokerto
Email : imron@amikompurwokerto.ac.id

ABSTRAK

Perpustakaan STMIK AMIKOM Purwokerto didirikan dengan tujuan untuk membantu serta memenuhi kebutuhan sivitas akademik yang ada di Perguruan Tinggi, koleksi buku yang ada sekitar 12,518, dengan judul 6,375 eksemplar. Penelitian ini bertujuan untuk mengetahui minat baca mahasiswa dengan membandingkan tingkat akurasi pada *data mining* dengan menggunakan algoritma *naive bayes* dan PART berdasarkan data peminjaman. Pengambilan data dari bulan Januari 2016 sampai bulan Maret 2017 dengan total data 4022. Hasil pengujian yang telah dilakukan terhadap kedua algoritma yaitu algoritma *Naive Bayes* dan Algoritma PART, maka dapat disimpulkan dari evaluasi dengan *confusion matrix* didapat hasil nilai akurasi masing-masing algoritma, nilai akurasi dari algoritma *naive bayes* sebesar 97,01% dan nilai akurasi dari algoritma PART sebesar 97,19%, sehingga algoritma yang terbaik dari pengujian tersebut untuk mengetahui minat baca mahasiswa di Perpustakaan STMIK AMIKOM Purwokerto adalah algoritma PART.

Kata Kunci—Perpustakaan, *data mining*, algoritma *naive bayes*, algoritma PART, *k-fold cross validation*.

ABSTRACT

Library STMIK AMIKOM Purwokerto was established with the aim to help and meet the needs of existing academic civitas in Higher Education, the collection of books that there are about 12,518, with the title of 6.375 copies. This study aims to determine the interest of reading students by comparing the accuracy level in data mining by using naive bayes algorithm and PART based on loan data. Data retrieval from January 2016 until March 2017 with data 4022. The results of testing has been done on both algorithms Naive Bayes algorithm and PART Algorithm, it can be concluded from the evaluation with confusion matrix obtained the results of the accuracy value of each algorithm, the value of accuracy of the naive bayes algorithm of 97.01% and the accuracy of the PART algorithm of 97.19%, so the best algorithm of the test is to know the interest of reading students in STMIK AMIKOM Library Purwokerto is PART algorithm.

Keywords - Library, data mining, naive bayes algorithm, PART algorithm, k-fold cross validation.

PENDAHULUAN

Perpustakaan sendiri merupakan lembaga yang menghimpun, mengelola, melestarikan, dan menyebarluaskan informasi kepada pemakaiannya berupa media informasi baik yang tercetak maupun non cetak. Kemampuan membaca berhubungan dengan minat dan kebiasaan membaca pada setiap mahasiswa yang dituntut untuk memiliki minat dan kemampuan membaca yang baik.

Perpustakaan STMIK AMIKOM Purwokerto didirikan dengan tujuan untuk membantu serta memenuhi kebutuhan sivitas akademik yang ada di Perguruan Tinggi dan dengan adanya perpustakaan diharapkan menjadi bermanfaat bagi mahasiswa ataupun sivitas yang lain yang meminjam dan membaca. Perkembangan teknologi dan penyebaran informasi untuk saat ini sangat pesat, dimana perkembangan tersebut terus tumbuh seiring dengan perkembangan jumlah mahasiswa di STMIK AMIKOM Purwokerto hingga sekarang. Sedangkan visi Perpustakaan sendiri adalah menjadi perpustakaan masa depan dengan mengikuti perkembangan ilmu dan teknologi agar dapat bersaing di era global, dan misi dari Perpustakaan sendiri berupaya senantiasa ikut mencerdaskan anak bangsa dan selalu berusaha menyediakan berbagai sumber ilmu yang dibutuhkan oleh mahasiswa dan Dosen dilingkungan STMIK AMIKOM Purwokerto pada khususnya dan masyarakat luas pada umumnya (Buku Panduan, 2012).

Menurut Abdurrahman (2003) ada lima tahap perkembangan membaca, yaitu kesiapan membaca, membaca permulaan, ketrampilan membaca cepat, membaca luas dan membaca yang sesungguhnya. Perpustakaan merupakan salah satu bagian dari sekolah yang menyediakan bahan bacaan yang diminati siswa. Membaca menjadi menyenangkan apabila materi bacaan memiliki daya tarik bagi siswa, sehingga siswa akan membaca dengan bersungguh-sungguh yang selanjutnya akan menunjang pemahaman bacaan siswa (Masri, 2008).

Perpustakaan STMIK AMIKOM Purwokerto untuk saat ini memiliki koleksi buku 12.518, dengan judul 6.375 Exemplar. Dalam hal ini peneliti melakukan penelitian untuk mengetahui minat baca mahasiswa dengan membandingkan tingkat akurasi pada *data mining* menggunakan algoritma *naive bayes* dan *PART* berdasarkan data peminjaman buku perpustakaan yang nantinya dapat diketahui seberapa besar tingkat akurasi dalam minat baca mahasiswa di Perpustakaan. *Data Mining* adalah sebuah proses dasar dimana metode kecerdasan digunakan untuk mengetahui pola data (Han, Dkk., 2013) . Dan dapat disebutkan juga bahwa *data mining* atau KDD adalah proses untuk menemukan *interesting knowledge* dari sejumlah besar data yang disimpan baik di dalam *databases*, *data warehouses* atau tempat penyimpanan informasi lainnya (Romario, 2013)

Menurut William J.Frawley, dkk. (1992). *Data mining* adalah penyaringan data secara implisit dimana sebelumnya tidak diketahui terdapatnya informasi yang potensial. *Data mining* menganalisis data menggunakan *tool* untuk menentukan pola dan aturan dalam himpunan pola dengan mengidentifikasi aturan

dan fitur pada data. *Tool Data Mining* diharapkan mampu mengenal pola ini dalam data dengan input minimal dari user.

Bedasarkan latarbelakang yang ada maka peneliti merasa perlu melakukan penelitian sejenis dalam rangka menggali informasi yang berharga dari kumpulan informasi minat baca di perpustakaan. Penelitian yang dilakukan dengan menggunakan algoritma *naive bayes*, teknik prediksi berbasis probalitas sederhana yang berdasarkan pada penerapan aturan *bayes* dengan asumsi ketidak tergantungan yang kuat atau bersifat bebas. Selain itu *naive bayes* juga dapat menganalisa *variable* yang paling mempengaruhinya dalam bentuk peluang (Cao, dkk, 2004). Algoritma untuk mengkomparasi adalah algoritma PART yang memiliki kinerja komputasi yang cukup teratur serta membangkitkan *unrestricted decision list* menggunakan prosedur *separate-and-conquer* (Gens, dkk.,2010), (Omar, dkk.,2012).

LANDASAN TEORI

1. Definisi Data Mining

Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data Mining* merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban, 2010). Istilah *data mining* memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. *Data mining*, sering juga disebut sebagai *Knowledge Discovery in Database* (KDD).

KDD merupakan sebuah proses yang terdiri dari serangkaian proses interaksi yang terurut, dan *data mining* merupakan salah satu langkah dalam proses KDD (Thomas, 2004). Urutan langkah dalam KDD diilustrasikan, sebagai berikut:

- a. Pembersihan Data, pembersihan terhadap data dilakukan untuk menghilangkan *noise* dan data yang tidak konsisten.
- b. Integrasi Data, merupakan proses integrasi data dilakukan untuk menggabungkan data dari berbagai sumber.
- c. Seleksi Data, bertujuan untuk mengambil data yang relevan, yang akan digunakan untuk proses analisis dalam *data mining*.
- d. Transformasi Data, merupakan proses untuk mentransformasikan data ke dalam bentuk yang tepat untuk di-*mine*.
- e. *Data Mining*, merupakan proses untuk mengaplikasikan suatu metode untuk mengekstrak pola-pola dalam data.
- f. Evaluasi pola, diperlukan untuk mengidentifikasi beberapa pola yang menarik yang merepresentasikan pengetahuan.
- g. Presentasi pengetahuan, dengan tujuan untuk mempresentasikan pengetahuan yang telah digali kepada pengguna dengan memvisualisasikan pengetahuan tersebut.

2. Naive Bayes

Naive Bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasarkan pada penerapan teorema *bayes* (aturan *bayes*) dengan asumsi independensi (ketidak tergantungan) yang kuat (naif). Dengan kata lain, dalam *naive bayes* model yang digunakan adalah model fitur independen (Prasetyo, 2012).

Naive Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Naive Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Naive Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar (Prasetyo, 2012). Prediksi *bayes* didasarkan pada formula teorema *bayes* dengan formula umum sebagai berikut:

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)}$$

3. PART (*Projective Adaptive Resonance Theory*)

Algoritma PART dikembangkan oleh Cao dan Wu (2004) telah terbukti sangat efektif dalam mengelompokkan data set dalam ruang dimensi tinggi. Algoritma PART didasarkan pada asumsi bahwa model persamaan PART (dengan skala besar dan luar biasa terganggu terhadap sistem persamaan diferensial yang digabungkan dengan mekanisme reset) memiliki kinerja komputasi yang cukup teratur.

4. Confusion Matrix

Confusion Matrix digunakan untuk evaluasi kinerja model klasifikasi yang berdasarkan pada kemampuan akurasi prediktif suatu model. Akurasi prediktif merupakan parameter untuk mengukur ketepatan aturan klasifikasi yang dihasilkan dalam mengklasifikasikan *test set* berdasarkan atribut yang ada ke dalam kelasnya.

Akurasi dinyatakan dalam persentase, sehingga aturan dengan akurasi 100% artinya semua kasus yang tercakup oleh aturan klasifikasi, diklasifikasikan dengan benar ke dalam kelas yang diprediksinya. Untuk mendapatkan nilai akurasi prediktif diperlukan perhitungan jumlah kasus yang diprediksikan dengan benar dan jumlah kasus yang diprediksikan dengan salah. Perhitungan tersebut ditabulasikan ke dalam tabel yang disebut *confusion matrix* (Tabel 1).

Tabel 1. *Confusion Matrix*

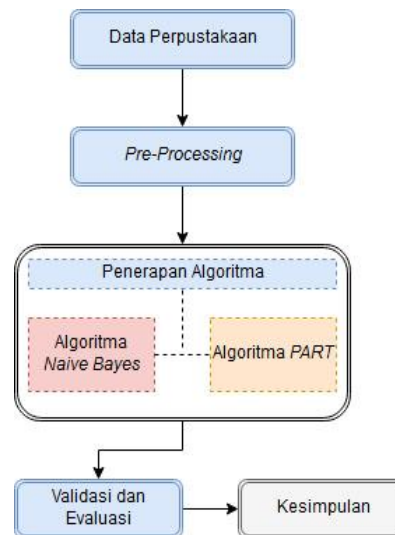
Actual Class	Predicted Class	
	Kelas=Ya	Kelas=Tidak
Kelas = Ya	a (TP)	b (FN)
Kelas = Tidak	c (FP)	d (TN)

METODE PENELITIAN

1. Pengeumpulan Data

Sumber data yang didapat pada penelitian ini dengan dua jenis yaitu data primer dan data sekunder, data primer didapat diperoleh dari objek yang diteliti atau orang yang dijadikan objek adalah pustakawan STMIK AMIKOM Purwokerto yang mengelola Perpustakaan, sedangkan data sekunder dari database Perpustakaan STMIK AMIKOM Purwokerto. Pengambilan data dari bulan Januari 2016 sampai bulan Maret 2017 dengan jumlah data 4022. *Tools* yang digunakan untuk pengujian hasil klasifikasi adalah weka 3.7.10, yang didapatkan gratis *opensource* (Weka, akses 18 Nopember 2016).

2. Diagram Alir Penelitian



Gambar 1. Diagram Alir Penelitian

HASIL DAN PEMBAHASAN

1. Pre-Processing

Pada penelitian ini data yang digunakan adalah data peminjaman buku di Perpustakaan STMIK AMIKOM Purwokerto, yang terdiri dari enam atribut kriteria dan satu atribut keputusan, diantaranya Id anggota, Nama anggota, Kode eksemplar, Judul, Tanggal peminjaman, Tanggal kembali, dan status peminjam. Berikut adalah data peminjaman buku yang didapat dari Perpustakaan yang belum dilakukan penyesuaian pada Tabel 2.

Tabel 2. Data Peminjaman Buku Sebelum Penyesuaian

ID Anggota	Nama Anggota	Kode Eksemplar	Judul	Tanggal Peminjam	Tanggal Harus Kembali	Status Peminjaman
14.11.0311	Nandya Tiara Nugrahaeni	7878	Budaya Organisasi Kepemimpinan & Kinerja; Proses Terbentuk, Tumbuh Kembang, Dinamika, Dan Kinerja Organisasi	2016-01-04	2016-01-18	1
14.11.0311	Nandya Tiara Nugrahaeni	8023	Kepemimpinan; Teori Dan Praktik Edisi Keenam	2016-01-04	2016-01-18	1
15.11.0284	Rilas Agung Pambudi	7341	Fakta Keren	2016-01-04	2016-01-18	1
14.12.0068	Fanindya Bayu Pambudi	5383	Mudah Membuat Aplikasi Android	2016-01-04	2016-01-11	1
14.12.0068	Fanindya Bayu Pambudi	7558	Cara Mudah Membuat Game Di Android	2016-01-04	2016-01-11	1
15.12.0020	Faiz Agil Wirawan	5649	PAS (Panduan Aplikatif & Solusi) Mudah Membuat Portal Berita Online Dengan Php Dan Mysql	2016-01-04	2016-01-18	1
13.12.0286	Dian Rakhma Iryanti	7429	Sistem Teknologi Informasi Edisi III	2017-03-07	2017-03-13	0
13.12.0197	Pipit Andri Himawan	8611	SWOT Balanced Scorecard; Teknik Menyusun Korporat Yang Efektif Plus Cara Mengelola Kinerja Dan Risiko	2017-03-07	2017-03-13	0

13.12.0197	Pipit Andri Himawan	5862	Analisis SWOT Teknik Membedah Kasus Bisnis	2017-03-07	2017-03-13	0
13.12.0286	Dian Rakhma Iryanti	5072	Analisis & Desain Sistem Informasi : Pendekatan Terstruktur Teori Dan Praktik Aplikasi Bisnis	2017-03-07	2017-03-13	0
13.12.0286	Dian Rakhma Iryanti	8441	Rekayasa Perangkat Lunak Terstruktur Dan Berorientasi Objek	2017-03-07	2017-03-13	0

(Sumber:Perpustakaan STMIK AMIKOM Purwokerto)

Dari data diatas akan dilakukan proses diskritisasi atribut penghasilan yang bertujuan mempermudah mengelompokkan nilai berdasarkan kriteria yang ditetapkan. Data yang sudah diolah melalui tahap *preprocessing*, dimana tahap ini untuk menyesuaikan atribut-atribut yang akan digunakan dalam mengolah *dataset*. Atribut-atribut yang dimaksud adalah:

a. Program Studi

Jurusan mahasiswa STMIK AMIKOM Purwokerto yang pernah berkunjung dan meminjam buku di Perpustakaan.

b. Kode Buku

Nomer buku yang telah ditentukan oleh perpustakaan.

c. Kategori Buku

Keseluruhan buku yang telah dipisah menjadi sepuluh kategori.

d. Peminjam Buku

Jumlah peminjam buku yang tercatat pada bulan januari 2016 sampai dengan bulan Maret 2017.

e. Minat Baca

Minat baca digunakan sebagai atribut keputusan dengan menyatakan YA dan TIDAK.

Setelah melalui tahapan *pre-processing* didapat atribut yang semula ada tujuh atribut dan setelah dilakukan *pre-processing* menjadi lima yang dipakai menjadi *dataset* untuk menentukan minat baca mahasiswa di Perpustakaan STMIK AMIKOM Purwokerto, berikut tabel hasil *pre-processing* data.

Tabel 3. Hasil *Pre-processing* data

Prodi	Kode Eksemplar	Kategori Buku	Peminj. Buku	Minat Baca
TI	225	Ilmu Terapan	2016 TI	YA

TI	4361	Ilmu Murni	2016 TI	TIDAK
TI	3982	Ilmu Terapan	2016 TI	YA
TI	4786	Karya Umum	2016 TI	YA
TI	6160	Kesenian	2016 TI	YA
TI	194	Karya Umum	2016 TI	TIDAK
TI	730	Karya Umum	2016 TI	YA
TI	8463	Ilmu Terapan	2016 TI	YA

Tabel 4. Hasil *Pre-processing* data (lanjutan)

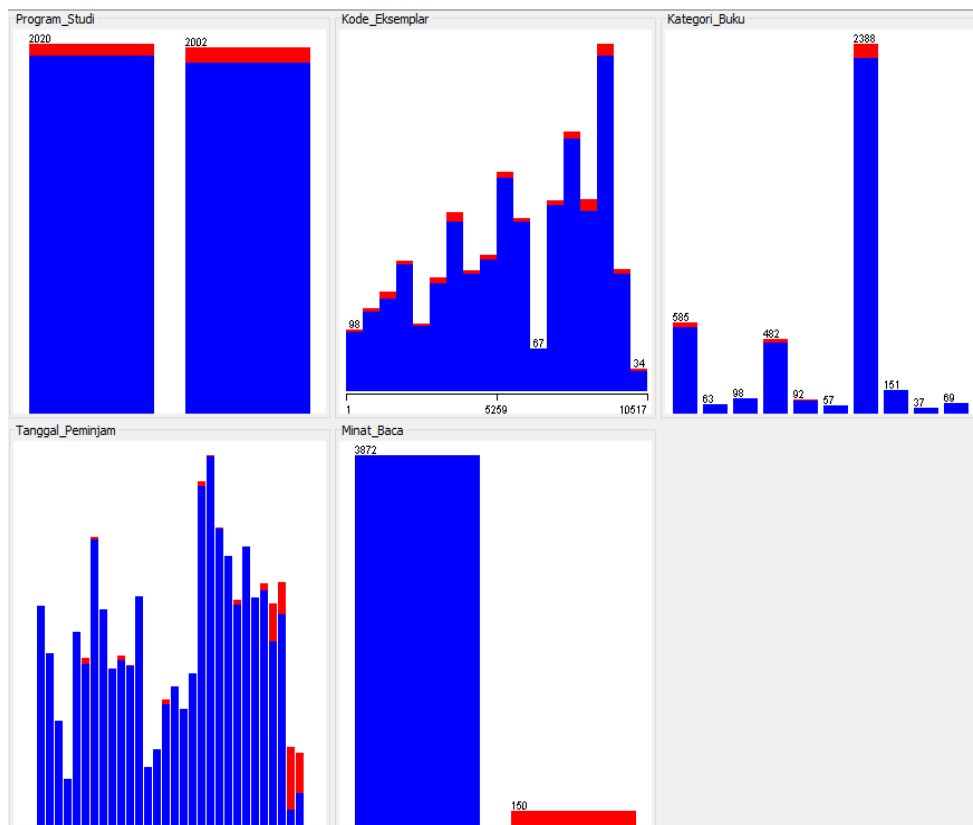
Prodi	Kode Eksemplar	Kategori Buku	Peminj. Buku	Minat Baca
TI	5572	Ilmu Terapan	2016 TI	YA
TI	7878	Ilmu Sosial	2016 TI	YA
TI	8023	Agama	2016 TI	YA
TI	7341	Geografi	2016 TI	TIDAK
TI	5383	Filsafat	2016 TI	TIDAK
TI	7558	Ilmu Terapan	2016 TI	YA
TI	5649	Ilmu Terapan	2016 TI	YA
...
...
...
SI	7429	Ilmu Terapan	2017 SI	YA
SI	8611	Ilmu Sosial	2017 SI	YA
SI	5862	Karya Umum	2017 SI	YA
TI	5072	Ilmu Terapan	2017 TI	TIDAK
TI	8441	Ilmu Terapan	2017 TI	TIDAK
SI	3532	Ilmu Sosial	2017 SI	YA
SI	8948	Ilmu Terapan	2017 SI	YA
SI	1718	Ilmu Terapan	2017 SI	YA
SI	7931	Ilmu Terapan	2017 SI	YA
TI	3062	Ilmu Terapan	2017 TI	TIDAK
TI	4170	Karya Umum	2017 TI	YA
TI	10144	Ilmu Terapan	2017 TI	TIDAK
SI	8954	Ilmu Terapan	2017 SI	YA
SI	9054	Ilmu Terapan	2017 SI	YA
SI	8460	Ilmu Terapan	2017 SI	YA

(Sumber:Perpustakaan STMIK AMIKOM Purwokerto)

Setelah proses penanganan nilai yang tidak lengkap (*missing value*) yang dilakukan dengan aturan diatas, maka didapatkan 4022 data minat baca, yang menyatakan YA sekitar 3872 dan yang menyatakan TIDAK 150 dari total 10446 data asli dan siap diolah lebih lanjut dengan atribut Program Studi, Kode Buku, Kategori Buku, Tanggal Pinjam dan Minat Baca.

2. Penerapan Metode Klasifikasi

Setelah melalui tahap *pre-processing* selesai untuk langkah selanjutnya *dataset* mulai diolah dengan menggunakan aplikasi weka. Pada tahapan ini bertujuan untuk menghasilkan *confusion matrix* berdasarkan berdasarkan evaluasi *10-fold cross validation*, dimana *dataset* tersebut dibagi menjadi 10 *subsets* (9 *subsets* sebagai *training sets* dan satu *subset* sebagai *testing set*) dengan jumlah sepuluh kali iterasi. Adapun *classifier* yang digunakan untuk uji coba pada *dataset* adalah Algoritma *Naive Bayes* dan PART. Gambar dibawah ini menunjukkan visualisasi data secara keseluruhan yang dijabarkan sesuai atributnya setelah melewati tahapan *pre-processing*.



Gambar 2. Visualisasi *dataset* Minat Baca di Perpustakaan STMik AMIKOM Purwokerto

3. Penerapan Algoritma Naive Bayes

Pengujian terhadap hasil analisa ini sangat penting dilakukan untuk menentukan dan memastikan apakah hasil akurasi dari masing-masing algoritma yang diujikan berdasarkan *dataset* yang telah ditentukan memiliki nilai akurasi yang baik atau tidak. Berikut hasil *output classifier* algoritma *Naive Bayes* yang telah dilakukan dengan menggunakan aplikasi weka, dengan hasil sebagai berikut:

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      3902           97.0164 %
Incorrectly Classified Instances    120            2.9836 %
Kappa statistic                    0.4461
Mean absolute error                 0.0489
Root mean squared error             0.1508
Relative absolute error              67.8361 %
Root relative squared error         79.5768 %
Total Number of Instances          4022

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                0.995   0.66    0.975     0.995   0.985     0.925    YA
                0.34    0.005   0.708     0.34    0.459     0.925    TIDAK
Weighted Avg.   0.97    0.636   0.965     0.97    0.965     0.925

=== Confusion Matrix ===

  a    b  <-- classified as
3851  21 |  a = YA
  99   51 |  b = TIDAK
    
```

Gambar 3. Classifier dari Algoritma Naive Bayes

Pada pengujian diatas algoritma *naive bayes* menghasilkan nilai akurasi sebesar 97.0164 %. Dengan nilai akurasi tersebut didapatkan dari hasil perhitungan *precision*, *recal*, dan *f-measure*, dimana proses perhitungan hasil akurasi penjabaran *confusion matrix* dapat dilihat pada tabel 3.3 dibawah ini.

Tabel 5. *Confusion Matrix Tested Negative*

3851 <i>Positive</i>	(<i>True</i>)	21 <i>Negative</i>	(<i>False</i>)
99 <i>Positive</i>	(<i>False</i>)	51 <i>Negative</i>	(<i>True</i>)

Dengan perhitungan atau penjabaran dari *confusion matrix* dari hasil *classifier* diatas dengan kelas *Tested Negative* dapat dilihat dibawah ini:

$$Precision = \frac{TP}{TP+FP} = \frac{3851}{3851+99} = \frac{3851}{3950} = 0,975$$

$$Recall = \frac{TP}{TP+FN} = \frac{3851}{3851+21} = \frac{3851}{3872} = 0,995$$

$$F-Measure = \frac{2 \times precision \times recall}{precision+recall} = \frac{2 \times 0,975 \times 0,995}{0,975+0,995} = \frac{1,940}{1,970} = 0,985$$

Tabel 6. *Confusion Matrix Tested Positive*

51 <i>Positive</i>	(<i>True</i>)	99 <i>Negative</i>	(<i>False</i>)
21 <i>Positive</i>	(<i>False</i>)	3851 <i>Negative</i>	(<i>True</i>)

Sedangkan perhitungan *confusion matrix* diatas dari hasil *classifier Tested Positive*, dibawah ini:

$$Precision = \frac{TP}{TP+FP} = \frac{51}{51+21} = \frac{51}{72} = 0,708$$

$$Recall = \frac{TP}{TP+FN} = \frac{51}{51+99} = \frac{51}{150} = 0,34$$

$$F-Measure = \frac{2 \times precision \times recall}{precision + recall} = \frac{2 \times 0,708 \times 0,34}{0,708 + 0,34} = \frac{0,481}{1,048} = 0,459$$

Untuk perhitungan dari hasil nilai akurasi berdasarkan *confusion matrix* dapat dilihat dibawah ini:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} = \frac{3851+51}{4022} = 0,970164 = 97,0164\%$$

Dari hasil perhitungan diatas dapat dilihat juga di tabel 3.5 dari hasil perhitungan *confusion matrix* dibawah ini:

Tabel 7. Nilai Akurasi berdasarkan *Confusion Matix*

<i>Class</i>	<i>Precisi on</i>	<i>Reca ll</i>	<i>F-Measure</i>
<i>Tested Negative</i>	0,975	0,995	0,985
<i>Tested Positive</i>	0,708	0,34	0,459
<i>Accuracy</i>	97,0164%		

Dari hasil pengujian diatas menunjukkan bahwa algoritma *naive bayes* memiliki nilai yang cukup baik dengan nilai akurasi sebesar 97.0164 %, untuk pengujian berikutnya dengan menggunakan algoritma PART sebagai langkah untuk mengetahui nilai akurasi yang terbaik dari kedua algoritma yang diujikan.

4. Penerapan Algoritma PART (*Projective Adaptive Resonance Theory*)

Dan pengujian berikutnya dengan menggunakan algoritma PART sebagai langkah mengkomparasi atau membandingkan algoritma sebelumnya yaitu *naive bayes*. Berikut hasil *output classifer* algoritma PART yang telah dilakukan dengan menggunakan aplikasi weka, dengan hasil sebagai beriku:

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      3909           97.1905 %
Incorrectly Classified Instances    113           2.8095 %
Kappa statistic                    0.5213
Mean absolute error                 0.0449
Root mean squared error             0.1532
Relative absolute error             62.357 %
Root relative squared error         80.8705 %
Coverage of cases (0.95 level)     99.279 %
Mean rel. region size (0.95 level) 55.3456 %
Total Number of Instances          4022

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                0.993   0.567   0.978     0.993   0.986     0.927   YA
                0.433   0.007   0.699     0.433   0.535     0.927   TIDAK
Weighted Avg.   0.972   0.546   0.968     0.972   0.969     0.927

=== Confusion Matrix ===

  a    b  <-- classified as
3844  28 |    a = YA
  85   65 |    b = TIDAK
    
```

Gambar 4. Classifier dari Algoritma PART

Pada pengujian diatas algoritma PART menghasilkan nilai akurasi sebesar 97.1905 %. Dengan nilai akurasi tersebut didapatkan dari hasil perhitungan *precision*, *recal*, dan *f-measure*, dimana proses perhitungan hasil akurasi penjabaran *confusion matrix* dapat dilihat pada tabel 3.6 dibawah ini.

Tabel 8. Confusion Matrix Tested Negative

3844 <i>(True Positive)</i>	28 <i>(False Negative)</i>
85 <i>(False Positive)</i>	65 <i>(True Negative)</i>

Dengan perhitungan atau penjabaran dari *confusion matrix* dari hasil *classifier* diatas dengan kelas *Tested Negative* dapat dilihat dibawah ini:

$$Precision = \frac{TP}{TP+FP} = \frac{3844}{3844+85} = \frac{3844}{3929} = 0,978$$

$$Recall = \frac{TP}{TP+FN} = \frac{3844}{3844+28} = \frac{3844}{3872} = 0,992$$

$$F-Measure = \frac{2 \times precision \times recall}{precision+recall} = \frac{2 \times 0,978 \times 0,992}{0,978+0,992} = \frac{1,940}{1,97} = 0,945$$

Tabel 9. Confusion Matrix Tested Positive

65 <i>(True Positive)</i>	85 <i>(False Negative)</i>
28 <i>(False Positive)</i>	3844 <i>(True Negative)</i>

Sedangkan perhitungan *confusion matrix* diatas dari hasil *classifier Tested Positive*, dibawah ini:

$$Precision = \frac{TP}{TP+FP} = \frac{65}{65+28} = \frac{65}{93} = 0,698$$

$$Recall = \frac{TP}{TP+FN} = \frac{65}{65+85} = \frac{65}{150} = 0,43$$

$$F-Measure = \frac{2 \times precision \times recall}{precision + recall} = \frac{2 \times 0,698 \times 0,43}{0,698 + 0,43} = \frac{0,600}{1,128} = 0,459$$

Untuk perhitungan dari hasil nilai akurasi berdasarkan *confusion matrix* dapat dilihat dibawah ini:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} = \frac{3851+51}{4022} = 0,971904 = 97,1905\%$$

Dari hasil perhitungan diatas dapat dilihat juga di tabel 3.5 dari hasil perhitungan *confusion matrix* dibawah ini:

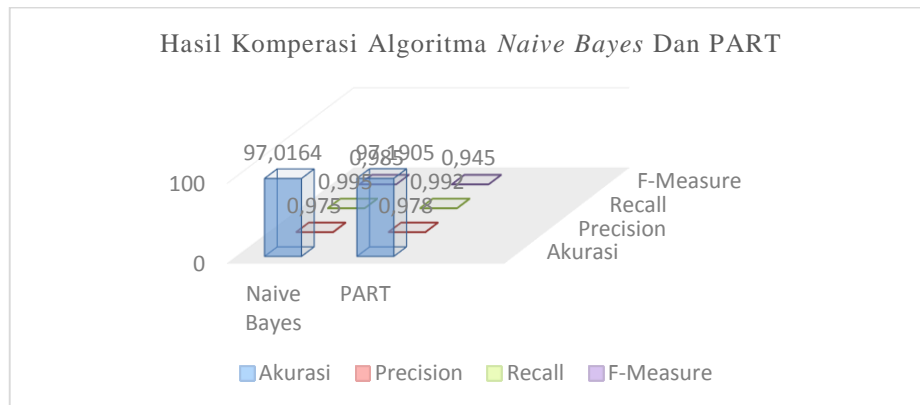
Tabel 10. Nilai Akurasi berdasarkan *Confusion Matix*

<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
<i>Tested Negative</i>	0,978	0,992	0,945
<i>Tested Positive</i>	0,698	0,43	0,459
<i>Accuracy</i>	97,1905%		

Dari hasil pengujian dengan menggunakan algoritma PART dari perhitungan *confusion matrix* bahwa nilai akurasi dari nilai tersebut sebesar 97,1905%.

5. Hasil Pengujian

Dari hasil pengujian kedua algoritma maka dapat disimpulkan dari perhitungan *confusion matrix* dari masing-masing pengujian yang telah dilakukan, pengujian dengan algoritma *naive bayes* didapat perhitungan dari kelas *tested negative* hasil *classifier* tersebut menunjukkan nilai *precision* sebesar 0,975, *recall* 0,995, *f-measure* 0,985, sedangkan untuk kelas *tested positive* nilai *precision* sebesar 0,708, *recall* 0,34, dan *f-measure* 0,459 dengan nilai akurasi 97,0164%. Pengujian dengan menggunakan algoritma PART dengan perhitungan kelas *tested negative* hasil *classifier* tersebut didapat dari *precision* sebesar 0,978, *recall* 0,992, dan *f-measure* 0,945, sedangkan perhitungan dengan kelas *tested positive* nilai *precision* sebesar 0,698, *recall* 0,43, dan *f-measure* 0,459 dengan nilai akurasi sebesar 97,1905%. Sehingga dari hasil pengujian dapat dilihat pada grafik dari masing-masing algoritma yang dapat dilihat pada gambar 3.3 dibawah ini.



Gambar 5. Hasil Komparasi Algoritma *Naive Bayes* dan PART

Dari grafik diatas dapat dilihat perhitungan diatas bahwa perbandingan kinerja Algoritma *Naive Bayes* dan PART untuk mengetahui minat baca mahasiswa di Perpustakaan STMIK AMIKOM Purwokerto lebih baik menggunakan algoritma PART karena tingkat akurasinya lebih baik, walaupun hampir sama persentasenya.

KESIMPULAN DAN SARAN

1. Kesimpulan

Dari hasil pengujian yang telah dilakukan terhadap kedua algoritma yaitu algoritma *Naive Bayes* dan Algoritma PART, maka dapat disimpulkan dari evaluasi dengan *confusion matrix* didapat hasil nilai akurasi masing-masing algoritma, nilai akurasi dari algoritma *naive bayes* sebesar 97,01% dan nilai akurasi dari algoritma PART sebesar 97,19%, sehinggann algoritma yang terbaik dari pengujian tersebut untuk mengetahui minat baca mahasiswa di Perpustakaan STMIK AMIKOM Purwokerto adalah algoritma PART.

2. Saran

Dari analisis yang sudah dilakukan oleh peneliti untuk menentukan minat baca mahasiswa di Perpustakaan STMIK AMIKOM Purwokerto didapat algoritma terbaik dari algoritma PART, selanjutnya untuk mendapatkan hasil akurasi yang lebih baik maka peneliti memberikan saran berharap agar dapat diteliti kembali menggunakan algoritma yang lain.

DAFTAR PUSTAKA

STMIK AMIKOM Purwokerto, Buku Panduan Akademik Mahasiswa Tahun Ajaran 2012-2013. Purwokerto, Jawa Tengah: STMIK AMIKOM Purwokerto, 2012.

Mulyono Abdurrahman. (2003). *Pendidikan Bagi Anak Berkesulitan Belajar*. Jakarta: PT Rineka Cipta.

- R. Masri Sareb Putra. (2008). *Menumbuhkan Minat Baca: Panduan Praktis bagi Pendidik, Orang Tua, dan Penerbit*. Indonesia: PT. Macanan Jaya Cemerlang.
- Han, Jiawei. Kamber, Micheline. Pei, Jian. 2012. *Data Mining Concepts and Techniques 3rd Edition*. Morgan Kaufmann, USA.
- Romario. (2013). *Penerapan Data Mining Pada Rsup Dr.Moh Hoesin Sumatera Selatan Untuk Pengelompokan Hasil Diagnosa Pasien Pengguna Asuransi Kesehatan Miskin (askin)*. Palembang: Universitas Bina Darm Palembang.
- William J. Frawley, Gregory Piatetsky-Shapiro, and Christopher J. Matheus, "Knowledge Discovery in Databases : An Overview", *AI Magazine*, Volume 13, Number 3, 1992.
- Y. Cao and J. Wu, "Dynamics of Projective Adaptive Resonance Theory Model: The Foundation of PART Algorithm", *IEEE Transactions On Neural Networks*, vol. XX, no. XX, 2004.
- S. Genç, F. E. Boran, D. Akay, and Z. Xu, "Interval multiplicative transitivity for consistency, missing values and priority weights of interval fuzzy preference relations," *Inf. Sci.*, vol. 180, no. 24, pp. 4877–4891, Dec. 2010.
- W. Nor Haizan W. Mohamed, Mohd Najib Mohd Salleh, Abdul Halim Omar, "A Comparative Study of Reduced Error Pruning Method in Decision Tree Algorithms", *IEEE International Conference on Control System, Computing and Engineering*, 23 -Penang, Malaysia , 25 Nov. 2012.
- Turban, E., J.E. Aronson dan T.P. Liang. 2005. *Decision Support System and Intelligent Systems - 7th ed*. Pearson Education, Inc. Pearson Education, Inc. Dwi Prabantini (penterjemah). 2005. *Sistem Pendukung Keputusan dan Sistem Cerdas*. Penerbit ANDI. Yogyakarta.
- Thomas E. 2004. *Data Mining : Definition and Decision Tree Examples*, e-book.
- J Prasetyo, E., 2012, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*, Penerbit Andi, Yogyakarta.
- Eko Prasetyo, *Data Mining : Konsep dan Aplikasi menggunakan MATLAB*, 1st ed. Yogyakarta, Indonesia: Andi, 2012.
- <http://www.cs.waikato.ac.nz/ml/weka/documentation.html> (diakses terakhir 18 Nopember 2016) .