

Prediksi Peringkat Kelulusan Mahasiswa Untuk Menentukan Strategi Pemasaran Kampus Menggunakan Pohon Keputusan

Achmad Fahrudin¹, Langgeng Listiyoko², Panji Surya³, Ali Maksum⁴

^{1,2,3,4}Sistem Informasi STMIK Muhammadiyah Banten

Jl. Syech Nawawi KM 4 No. 13 Matagara Tigaraksa Tangerang

E-mail: achmad.fahrudin@stmikmbanten.ac.id¹, langgeng.listiyoko@stmikmbanten.ac.id²,
panjisurya@stmikmbanten.ac.id³, alimaksum@stmikmbanten.ac.id⁴

ABSTRAKSI

Kepuasan pelanggan merupakan kualifikasi yang harus dipertimbangkan masyarakat dalam rangka menentukan pilihan untuk belajar di perguruan tinggi. Di lain pihak proses bisnis yang dilaksanakan oleh penyelenggara pendidikan juga harus dibangun dengan sangat teliti untuk mencapai efisiensi. Pemilihan waktu yang tepat untuk mengimplementasikan strategi pemasaran dapat menarik outcome dalam hal jumlah pendaftar atau peminat di satu perguruan tinggi. Dalam makalah ini ditunjukkan prediksi apakah program-program yang diselenggarakan kampus akan efektif atau tidak, serta memberi solusi berdasarkan prediksi dimaksud bagian mana saja yang memerlukan perhatian khusus demi meningkatkan predikat kelulusan. Untuk dapat melaksanakannya, digunakan satu teknik data mining pohon keputusan dengan memanfaatkan data-data yang sudah dimiliki untuk ditemukan pola tertentu.. Kemudian sebagai alat bantu digunakan program Rapidminer. Penelitian ini menyimpulkan bahwa nilai awal semester berpengaruh terhadap pencapaian akhir tentang predikat kelulusan (memuaskan, sangat memuaskan, dengan pujian), dan disarankan juga bahwa pengelolaan perpustakaan sebagai penunjang harus diperhatikan. Dengan meningkatnya indeks kelulusan perguruan tinggi maka daya jual semakin tinggi.

Kata Kunci : Pohon keputusan, prediksi, efisiensi, Rapidminer

1. PENDAHULUAN

Beragam upaya dilakukan manajemen kampus dalam rangka menjaring calon mahasiswa sebanyak-banyaknya. Keberhasilan kegiatan tersebut didasari pada seberapa tinggi tingkat kepuasan pelanggan terhadap pelayanan yang diberikan, di samping usaha promosi yang baik. Kegiatan internal seperti seminar, kemahasiswaan, fasilitas pendukung perkuliahan, jajaran dosen dan staff juga turut mempengaruhi keberhasilan kampus secara umum.

Penurunan angka peminat dapat disebabkan oleh rendahnya kualitas kelulusan mahasiswa. Persentase kelulusan mahasiswa dengan predikat “Dengan Pujian” mempengaruhi minat calon mahasiswa, sehingga kampus akan bekerja keras dalam meningkatkan predikat kelulusan mahasiswanya.

Di lain sisi masalah efisiensi menjadi isu penting agar proses bisnis dapat tetap berjalan dengan baik. Artinya dalam setiap kesempatan pelaksanaan program-program penunjang kuliah harus dipertimbangkan pula *outcome* yang ditimbulkan. Pertimbangan tentang apakah sudah perlu melaksanakan program dapat dijawab dengan memperhatikan prediksi kelulusan di masa yang akan datang. Jika diprediksi akan terjadi penurunan pencapaian target, maka beberapa program penting akan menjadi kebutuhan mendesak.

Dengan mengambil data kelulusan pada 3 tahun sebelumnya sebagai data latih, penelitian ini akan memberikan prediksi tentang banyaknya mahasiswa

yang akan lulus tahun depan dengan predikat “Dengan Pujian”. Hasil penelitian ini akan berguna bagi manajemen untuk menentukan sikap berkaitan dengan peningkatan mutu kelulusan mahasiswa dengan tetap mempertimbangkan efisiensi sumber daya.

2. Data Mining

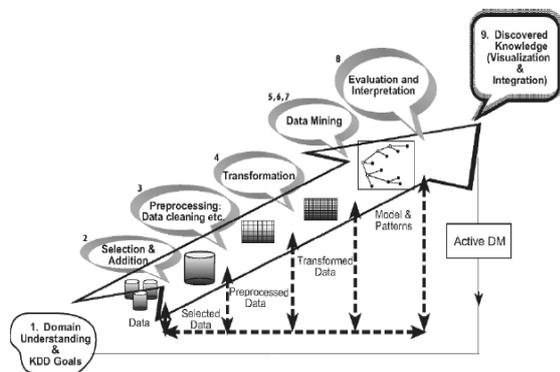
Data mining hanya dapat bekerja berdasarkan hasil penambangan pola pada data history. Pola yang diperoleh selanjutnya diterapkan pada data uji untuk mendapatkan informasi terselubung yang berguna bagi manajemen secara umum. Keluaran dari *data mining* umumnya dipakai sebagai masukan bagi sistem yang lain, meskipun secara alamiah dapat langsung digunakan oleh karena formatnya yang berupa informasi.

Salah satu teknik data mining yang juga digunakan dalam penelitian ini adalah pohon keputusan. Teknik tersebut menghasilkan keluaran berupa data biner, yaitu kesimpulan “Ya” atau “Tidak”. Atribut yang dikenai kesimpulan tersebut selanjutnya dikenal sebagai atribut tujuan. Gambar 1 berikut menjelaskan secara sederhana konsep metode data mining dalam rangka menggali informasi.



Gambar 1 Metode penelitian data mining

Maimon dalam penjelasannya mengenai KDD (*Knowledge Discovering in Database*) menyebutkan bahwa data mining merupakan salah satu tahapan penting sebagai sub sistem. Kedudukan data mining dalam KDD diilustrasikan seperti dalam gambar 2.



Gambar 2 Proses KDD [7]

Dilihat dari fungsinya, data mining dapat dikelompokkan sebagai fungsi klasifikasi, asosiasi, dan prediksi (Susanto, S., & Suryadi, D., 2010). Pohon keputusan termasuk dalam kategori fungsi prediksi, sehingga sesuai untuk menyelesaikan masalah dalam penelitian ini.

Pohon keputusan merupakan salah satu metode data mining yang menganut teknik learning (Suyanto, 2014). Pohon keputusan mengambil salah satu atribut sebagai atribut tujuan (variabel keputusan) yang akan menjadi output. Secara umum pohon keputusan akan membentuk rule "if...then..." pada akhir prosesnya. Rule tersebut akan memberikan nilai pada atribut tujuan sehingga diketahui kategori masing-masing item dataset.

Secara garis besar proses data mining meliputi proses persiapan, pengolahan data, dan interpretasi. Persiapan data mining dimulai dari penentuan objek data record yang akan diteliti yang diikuti oleh atribut-atribut yang berpengaruh. Atribut data record tidak seluruhnya berguna bagi proses pengolahan data, sehingga dilakukan *screening* terhadap atribut/variabel-variabel yang melekat.

Tabel 1 Atribut Data Histori (X1-X5)

NIM	NAMA	ASAL (X1)	JENIS KELAMIN (X2)	AGAMA (X3)	JENJANG (X4)	JURISAN (X5)
0611457081	Phi Siewanto	Tangerang	Laki-Laki	Islam	Sarjana	Sistem Informasi
0613457373	Ahmad Nur Bhsan	Tangerang	Laki-Laki	Islam	Sarjana	Sistem Informasi
0622407790	Hamzah Hartono	Simalungun	Laki-Laki	Islam	Sarjana	Teknik Informatika
071338285	Dewi Yunawati	Jakarta	Perempuan	Islam	Diploma 3	Manajemen Informatika
071349192	Helmi Iskandar	Tangerang	Laki-Laki	Islam	Sarjana	Sistem Informasi
0713460137	Syafidawati	Medan	Perempuan	Islam	Sarjana	Sistem Informatika
0721460054	Reza Anugerah Hermansyah	Jakarta	Laki-Laki	Islam	Sarjana	Teknik Informatika
072248767	Maria Luisa	Jakarta	Perempuan	Protestan	Sarjana	Teknik Informatika
072249226	Latif Hariyandi Suntoro	Jakarta	Laki-Laki	Islam	Sarjana	Teknik Informatika
073149150	Reza Adhi Saputra Somantri	Tangerang	Laki-Laki	Islam	Sarjana	Sistem Komputer
0811461463	Anisa Nugraha	Muara Bungo	Laki-Laki	Islam	Sarjana	Sistem Informasi
0811461598	Ahmad Zamrud	Tangerang	Laki-Laki	Islam	Sarjana	Sistem Informasi
0812461120	Irfan Firdaus	Tangerang	Laki-Laki	Islam	Sarjana	Sistem Informasi
0812462008	Michael Vardad Alexander Chandra	Tangerang	Laki-Laki	Islam	Sarjana	Sistem Informatika

Tabel 2 Atribut Data Histori (X6-X10)

NILAI SEMESTER 1 (X7)	IPK (X8)	ANGGOTA PERPUSTAKAAN (X9)	MK (X10)	PREDIKAT (Y)
1.81	2.49	TIDAK ADA		Memuaskan
1.67	2.68	ADA		Memuaskan
1.91	2.87	TIDAK ADA		Sangat Memuaskan
2.62	2.73	TIDAK ADA		Memuaskan
2.20	2.92	TIDAK ADA		Memuaskan
3.61	3.58	TIDAK ADA		Dengan Pujian
3.44	3.40	ADA		Sangat Memuaskan
3.23	3.16	ADA		Sangat Memuaskan
2.22	2.55	TIDAK ADA		Memuaskan
2.30	2.57	ADA		Memuaskan
2.38	2.73	TIDAK ADA		Memuaskan
1.92	2.53	TIDAK ADA		Memuaskan
2.40	2.90	ADA		Sangat Memuaskan
3.23	2.97	TIDAK ADA		Sangat Memuaskan

Data mart

Data mart adalah sumber data dari luar system yang mampu memberikan informasi berkenaan dengan item set yang sedang diteliti. Data mart merupakan database eksternal yang saling melengkapi demi membentuk data yang representatif.

Oleh karena kemampuan data mining dalam menangani data kompleks, maka besar kemungkinan persiapan data melibatkan data mart demi mendapatkan variabel yang lebih lengkap, sehingga menimbulkan penambahan variabel. Penambahan variabel terhadap data record tidak hanya terjadi oleh tindakan konsolidasi data mart, tetapi juga dapat berupa variabel turunan dari integrasi variabel lain. Sebagai contoh data asli menyajikan variabel tanggal lahir, namun data mining menuntut adanya data usia. Maka dalam hal ini variabel usia ditambahkan sebagai turunan dari variabel tanggal lahir. Data mart yang dibutuhkan adalah sebagai berikut :

Tabel 3 Data Mart List

Data Mart List	
DM1	Akademik
DM2	Kemahasiswaan
DM3	Perpustakaan

Dalam penelitian ini kontribusi masing-masing data mart dalam pemenuhan kebutuhan variabel adalah sebagai berikut :

Tabel 4 Sumber Data Masing-Masing Atribut

Data Sources List	
Variabel	Sources
Jurusan	DM1
Nilai Sem. 1	DM2
IMK	DM2
Keanggotaan Perpust.	DM3
Pekerjaan	DM1

Masih dalam persiapan, dijumpai tahapan preprocessing terhadap data. Data mining hanya bekerja pada data lengkap. Dapat dijumpai sebuah record tidak memenuhi kriteria data yang valid oleh karena hilangnya nilai pada variabel tertentu. Dalam kasus ini, preprocessing dapat dilakukan dengan alternatif mengisi variabel dengan nilai default, nilai modus, nilai mean, atau bahkan didrop / dibuang.

Tabel 5 Inventarisasi Atribut dari Datamart

Variabel X	
X1	Kota Asal
X2	Jenis Kelamin
X3	Agama
X4	Jenjang
X5	Jurusan
X6	Konsentrasi
X7	Nilai Sem. 1
X8	IMK
X9	Keanggotaan Perpust.
X10	Pekerjaan

Dalam tabel didapat informasi mengenai jumlah atribut yang berhasil dihimpun dari data mart sebanyak 10. Beberapa atribut dianggap tidak terlalu kuat untuk mempengaruhi variabel Y sebagai tujuan penelitian, maka diperoleh atribut sebagai berikut :

Tabel 6 Daftar Atribut Terpilih

Variabel X	
X1	Jurusan
X2	Nilai Sem. 1
X3	IMK
X4	Keanggotaan Perpust.
X5	Pekerjaan

Variabel keputusan dalam tabel data histori adalah predikat kelulusan mahasiswa yang bernilai :

memuaskan, sangat memuaskan, dan dengan pujian. Penelitian fokus terhadap predikat “Dengan Pujian” maka, variabel Y (predikat kelulusan) perlu diubah bentuk dari polynominal menjadi binominal. Atribut data kemudian mengalami transformasi dari “Predikat” (memuaskan, sangat memuaskan, dengan pujian) menjadi “Dengan Pujian” (yes, no).

Tabel 7 Variabel Y Sebelum Transformasi

NIM	NAMA	PEKERJAAN	JURUSAN (XS)	PERPUST.	NILAI SEM.1	IMK	PREDIKAT (Y)
0922363330	Kusendang Fitriyanti	BEKERJA	AK	TIDAK ADA	Sangat Memuaskan	Memuaskan	Sangat Memuaskan
1022364529	Narulita Nova Irlani	BEKERJA	AK	TIDAK ADA	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan
1022365497	Meri Mayang Sari	BEKERJA	AK	ADA	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan
1123366544	Siti Rahma Sita	BEKERJA	AK	TIDAK ADA	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan
1123368830	Puspa Widanti	BEKERJA	AK	TIDAK ADA	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan
1123368854	Siti Ropiah	MAHASISWA	AK	ADA	Dengan Pujian	Sangat Memuaskan	Dengan Pujian
1123369112	Dewi Sekar Tajji	MAHASISWA	AK	ADA	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan

Tabel 8 Variabel Y Setelah Transformasi

NIM	NAMA	PEKERJAAN	JURUSAN (XS)	PERPUST.	NILAI SEM.1	IMK	DP? (Y2)
0922363330	Kusendang Fitriyanti	BEKERJA	AK	TIDAK ADA	Sangat Memuaskan	Memuaskan	No
1022364529	Narulita Nova Irlani	BEKERJA	AK	TIDAK ADA	Sangat Memuaskan	Sangat Memuaskan	No
1022365497	Meri Mayang Sari	BEKERJA	AK	ADA	Sangat Memuaskan	Sangat Memuaskan	No
1123366544	Siti Rahma Sita	BEKERJA	AK	TIDAK ADA	Sangat Memuaskan	Sangat Memuaskan	No
1123368830	Puspa Widanti	BEKERJA	AK	TIDAK ADA	Sangat Memuaskan	Sangat Memuaskan	No
1123368854	Siti Ropiah	MAHASISWA	AK	ADA	Dengan Pujian	Sangat Memuaskan	Yes

Transformasi juga dialami value dari atribut IMK dan nilai semester 1 di mana terjadi perubahan dari bentuk nominal menjadi kategori. Hal ini akan meningkatkan kinerja system dan sangat membantu untuk perhitungan manual untuk menghindari kesalahan pembacaan.

Gain / Entropy

Keputusan yang terbentuk dari metode pohon keputusan diperoleh dari serangkaian pengujian terhadap atribut tertentu, yang kemudian dikenal sebagai node akar. Penentuan atribut sebagai node akar dilakukan dengan penghitungan gain dan entropy. Entropy merupakan jumlah perkalian log2 dan ratio keputusan terhadap total data yang dimiliki. Pada penelitian ini keputusan yang dimaksud adalah “ya” dan “tidak”. Selisih antara entropy total dan jumlah total entropy pada setiap value atribut menghasilkan gain. Atribut dengan nilai gain tertinggi kemudian dipilih sebagai node akar.

Algoritma Pohon Keputusan

Selanjutnya proses data mining berlangsung dengan algoritma tertentu. Dalam penelitian yang menggunakan teknik pohon keputusan, terdapat algoritma sebagai berikut :

1. Pilih atribut/ variabel sebagai akar. Akar didapat dari nilai *Gain* yang tertinggi, dari atribut yang ada. Untuk menghitung *Gain* digunakan formula (1) di bawah ini.

$$Gain(S, A) = Entropy(S) - \left(\sum_{i=1}^n \frac{A_i}{S} * Entropy(A_i) \right) \quad (1)$$

Keterangan

- S : himpunan kasus
- A : atribut
- n : jumlah partisi atribut A
- |Si| : jumlah kasus pada partisi ke-i
- |S| : jumlah kasus dalam S

Sedangkan formula untuk menghitung Entropy (2) adalah seperti berikut :

$$Entropy (S) = \sum_{i=1}^n -pi * \log_2 pi$$

(2)

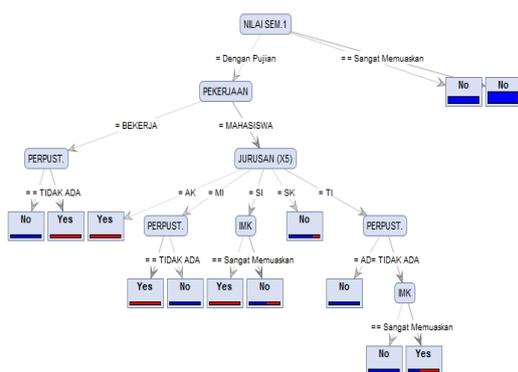
Keterangan :

- S : himpunan kasus
- A : fitur
- n : jumlah partisi S
- Pi : proporsi dari S, terhadap S

2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk setiap masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Variasi penyelesaian dengan pruning bertujuan untuk meningkatkan akurasi dengan membuang cabang yang mengindikasikan adanya noise (Hand, D., Mannila, H., & Smyth, P., 2001).

Pengembangan lebih lanjut pada teknik pohon keputusan dapat memberikan informasi pada variabel keputusan berupa multivalued, tidak sekedar data biner. Namun demikian pada dasarnya penelitian ini dapat menghasilkan kesimpulan berupa kategori, yakni “Memuaskan”, “Sangat Memuaskan”, dan “Dengan Pujian”.



Gambar 3 Hasil Pengolahan Data Histori dengan Rapidminer

Menurut perhitungan Rapidminer diperoleh hasil rule sebanyak 13. Hasil ini mengacu pada aturan pruning, yaitu batas minimal jumlah item pada setiap node. Penelitian mengaplikasikan 3 pruning, artinya proses pada masing-masing node akan

dihentikan ketika diperoleh jumlah item 3 (keputusan “ya” maupun “tidak”), sehingga akan menghasilkan nilai confidence.

Nilai confidence adalah probabilitas terbentuknya keputusan. Di bawah ini adalah tabel rule yang dihasilkan Rapidminer beserta nilai confidence-nya

```

Tree
NILAI SEM.1 = Dengan Pujian
| PEKERJAAN = BEKERJA
| | PERPUST. = ADA: No {No=2, Yes=0}
| | PERPUST. = TIDAK ADA: Yes {No=0, Yes=4}
| PEKERJAAN = MAHASISWA
| | JURUSAN (X5) = AK: Yes {No=0, Yes=1}
| | JURUSAN (X5) = MI
| | | PERPUST. = ADA: Yes {No=0, Yes=2}
| | | PERPUST. = TIDAK ADA: No {No=1, Yes=0}
| | JURUSAN (X5) = SI
| | | IMK = Memuaskan: Yes {No=0, Yes=2}
| | | IMK = Sangat Memuaskan: No {No=6, Yes=4}
| | JURUSAN (X5) = SK: No {No=4, Yes=1}
| | JURUSAN (X5) = TI
| | | PERPUST. = ADA: No {No=3, Yes=0}
| | | PERPUST. = TIDAK ADA
| | | | IMK = Memuaskan: No {No=1, Yes=0}
| | | | IMK = Sangat Memuaskan: Yes {No=2, Yes=3}
NILAI SEM.1 = Memuaskan: No {No=101, Yes=0}
NILAI SEM.1 = Sangat Memuaskan: No {No=209, Yes=4}
    
```

Gambar 4 Interpretasi Rapidminer Mode Text

Proses interpretasi pada data mining akan menghasilkan sajian data yang mudah dicerna bagi pengguna. Dalam proses ini dilakukan pula pengolahan data statistika sehingga mendapatkan butir kesimpulan yang informatif dan mempengaruhi sikap pengambil keputusan. Dari hasil pengolahan Rapidminer dapat diinterpretasikan rule sebagai berikut :

1. If Nilai Sem. 1 = Sangat Memuaskan then Predikat Dengan Pujian = No (confidence = 101/101 = 100%)
2. If Nilai Sem. 1 = Memuaskan then Predikat Dengan Pujian = No (confidence = 209/213 = 98.12%)
3. Dst

Sejalan dengan tujuan penelitian, maka dalam interpretasi akan menyajikan kesimpulan mengenai : peningkatan mutu kelulusan dari tahun yang telah lalu, pencapaian predikat kelulusan secara umum, dan rekomendasi kepada manajemen.

3. HASIL DAN PEMBAHASAN

Data uji adalah data saat ini yang akan dikenai proses prediksi berdasarkan pola yang telah diperoleh dari data histori. Untuk menguji data yang dimiliki dalam penelitian ini akan menggunakan rule yang telah dibentuk oleh Rapidminer. Diperoleh hasil prediksi terhadap sejumlah mahasiswa aktif adalah sebagai berikut :

Tabel 9 Hasil Prediksi Data test

	Jumlah	%
Sampling	395	100%
Dengan Pujian	23	5.82%
Bukan Dengan Pujian	372	94.18%

Tabel 10 Evaluasi Predikat Kelulusan Berdasarkan Prediksi

	2015	2016	Mean	2017 (Prediksi)	Kesimpul
Dengan Pujian	6.67%	6.01%	6.34%	5.82%	Turun 0.5
Bukan Dengan Pujian	93.33%	93.99%	93.66%	94.18%	Naik 0.5

4. KESIMPULAN

Berisi berbagai kesimpulan yang diambil berdasarkan penelitian yang telah dilakukan. Berisi pernyataan singkat tentang hasil yang disarikan dari pembahasan. Saran dapat dituliskan pada bagian paling akhir.

1. Berdasarkan hasil prediksi maka dapat disimpulkan bahwa Nilai Semester awal sangat berpengaruh bagi predikat kelulusan.
2. Dapat direkomendasikan untuk memaksimalkan kegiatan perkuliahan awal dan pemanfaatan perpustakaan sebagai penunjang kuliah..

PUSTAKA

- Gorunescu, F., 2011, *Data Mining Concepts, Models and Techniques*, Springer, Berlin.
- Han, J., & Kamber, M., 2006, *Data Mining : Concepts and Techniques*, Elsevier, San Fransisco.
- Hand, D., Mannila, H., & Smyth, P., 2001, *Principles of Data mining*, MIT Press, London.
- Hermawati, F. A., 2013, *Data Mining*, Andi Publisher, Yogyakarta.
- Kusrini, Luthfi, E.T., 2009, *Algoritma Data Mining*, Andi Publisher.
- Larose, D. T., 2005, *Discovering Knowledge in Data : An Introduction to Data Mining*, John Wiley & Sons, Inc., New jersey.
- Maimon, O., & Rokach, L., 2010, *Data Mining And Knowledge Discovery Handbook* (Second Edition), Springer, New York.
- Maurina, D., Fanani, A.Z., 2015, Penerapan Data Mining Untuk Rekomendasi Beasiswa Pada SMA Muhammadiyah Gubug Menggunakan Algoritma C4.5, *Jurnal Udinus*.
- Meilani, B.D., Slamet, A.F., 2013, Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree, *Jurnal ITATS*.
- Pratiwi, L.D.B., Wibowo, W., Zain, I., 2015, Klasifikasi Nilai Peminat SBMPTN (Seleksi Bersama Masuk Perguruan Tinggi Negeri) ITS dengan Pendekatan CART, *Jurnal Sains dan Seni ITS* Vol.4 No. 2 D193-198.

- Susanti, 2013, Klasifikasi Kredit Menggunakan Metode Decision Tree Pada Nasabah PD BPR BKK Gabus, *Jurnal Udinus*.
- Susanto, S., & Suryadi, D., 2010, *Pengantar Data Mining Menggali Pengetahuan Dari Bongkahan Data*, Andi Publisher, Yogyakarta.
- Suyanto, 2014, *Artificial Intelligence*, Informatika, Bandung.
- Vercellis, C., 2009, *Business Intelligence : Data Mining And Optimization For Decision Making*, John Wiley & Son Ltd, Padstow.