

PREDIKSI PENCAPAIAN TARGET KELULUSAN TAHFIDZ MENGUNAKAN PENDEKATAN DATA MINING MBA

Langgeng Listiyoko¹, Rosalia Wati²

*STMIK Muhammadiyah Banten
langgeng.listiyoko@stmikmbanten.ac.id¹
rosaliawati@stmikmbanten.ac.id²*

ABSTRACT

Evaluating graduation achievement of school will deliver a meaningful information regarding to next achievement prediction. It's possible by implementing market basket analyse (MBA) methodology to find out a pattern in the historical data. Every single entity attribute value related each other to create a pattern which used to predict achievement. The research gives prediction of each student whether will pass the target or not. Finally the report will be used by a decision maker in the managerial level to set a good improvement in order to achieve optimal result as mentioned in their vision.

Keywords: *graduation achievement, prediction, market basket analyse, pattern*

ABSTRAK

Dengan mempelajari pencapaian kelulusan sekolah tahun lalu dapat memberikan informasi yang berarti dalam rangka memprediksi pencapaian yang akan datang. Hal ini dapat dilakukan dengan mengimplementasikan metodologi *market basket analyse (MBA)* untuk mencari pola tertentu yang terbentuk dari data latih atau data histori. Setiap atribut berkaitan erat satu sama lain yang akan digunakan untuk membentuk pola yang dimaksud. Penelitian ini memberikan prediksi terhadap setiap siswa sekolah apakah ia akan lulus sesuai target atau tidak. Laporan prediksi ini nantinya akan membantu manajer tingkat menengah untuk menyusun langkah perbaikan agar dapat mencapai visi sekolah.

Kata Kunci: *pencapaian kelulusan, prediksi, market basket analyse, pola*

PENDAHULUAN

Ada banyak cara dilakukan institusi penyelenggara pendidikan (sekolah) untuk dapat meningkatkan mutu lulusannya. Usaha ini tentu terkait dengan visi yang dicanangkan sebagai tujuan jangka panjang dan dituangkan dalam misi sekolah. Perbaikan terus menerus harus dilakukan demi menjaga konsistensi kualitas sekolah dari tahun ke tahun. Pencapaian di tahun yang telah lalu bermanfaat untuk evaluasi guna mencapai hasil yang lebih baik di tahun berikutnya.

Salah satu cara mempercepat implementasi perbaikan sistem adalah dengan mengetahui lebih dini kemungkinan yang akan diperoleh pada akhir periode tahun berjalan. Usaha untuk mengetahui lebih dini dilakukan dengan prediksi, yaitu memperhatikan keterkaitan variabel siswa dengan prestasi yang telah dicapai oleh siswa yang telah lulus. Analisis yang dapat mendukung penelitian ini adalah analisa data *roll up*, yaitu mengetahui fakta terlebih dahulu kemudian menentukan hipotesis yang sesuai. Analisis yang dilakukan dikenal sebagai *knowledge discovery in database* (Maimoon, 2013). Di dalam KDD ditemui pula metodologi analisis dalam rangka menemukan informasi yang terbentuk dari relasi antar tabel.

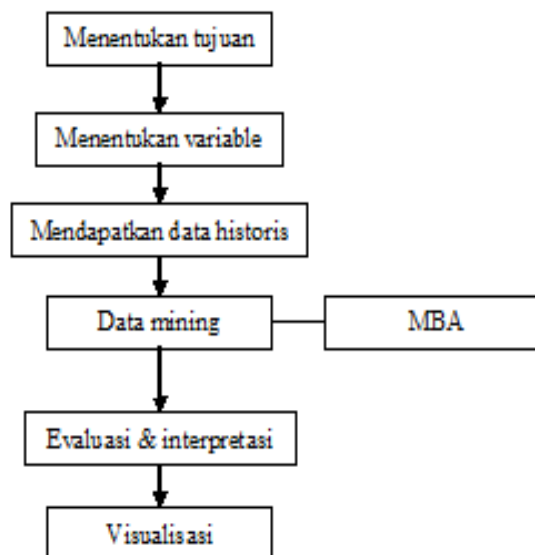
Proses pengolahan database dalam rangka mendapatkan informasi yang berguna dari gudang basis data yang berupa ilmu pengetahuan selanjutnya dikenal sebagai *data mining* (Yanto, 2015). Teknik *data mining* digunakan untuk mencari informasi penting berupa pola pada sekumpulan data kompleks sehingga dapat diterapkan pada data latih untuk mendapatkan informasi yang mendukung pengambilan keputusan (Gorunescu, 2011). Kemampuan ini sangat membantu manajer dalam menentukan sikap jika keputusan yang harus diambil melibatkan banyak variabel untuk dipertimbangkan mengingat keterbatasan yang dimiliki otak manusia untuk mengkalkulasi pertimbangan (variabel) yang kompleks secara akurat.

Data mining dapat bekerja sama dengan *datawarehouse*, kecerdasan buatan, dan statistik dalam menghasilkan informasi yang akan disajikan (Listiyoko, 2015). Integrasi yang baik dari ketiga unsur tersebut selanjutnya dapat dipandang sebagai *Business Intelligence* (Vercellis, 2009). BI sangat membantu dalam pengelolaan database dalam ukuran yang sangat besar.

METODE PENELITIAN

Terdapat beberapa metodologi *data mining* yang dapat mendukung pengambilan keputusan. Metodologi *market basket analyse (MBA)* dikembangkan untuk fungsi prediksi dengan memperhatikan pola pada event-event tunggal yang terbentuk sebagai akibat dari transaksi (Susanto, 2010). Transaksi yang terjadi pada penelitian ini adalah penyimpanan data perolehan nilai siswa sejak semester awal. Selain itu untuk menambah densitas pengambilan keputusan ditambahkan variabel yang bersifat statis seperti kegiatan eksternal.

Metode penelitian yang dianut adalah seperti diilustrasikan Gambar 1 sebagai berikut :



Gambar 3. Metode Penelitian

Penelitian diawali dengan pengumpulan data yang secara alami telah dimiliki institusi. Oleh karenanya data yang diperlukan bersifat historis, bukan merupakan data baru. Data dimaksud diperoleh dengan memperhatikan kebutuhan atribut yang akan digunakan. Adapun atribut yang akan diteliti adalah nilai prestasi siswa secara akumulatif, nilai mata pelajaran unggulan, kegiatan eksternal pendukung prestasi, serta keaktifan ekstrakurikuler (ekskul).

Seluruh data tersebut didapatkan dari master data nilai siswa yang dikenal sebagai buku besar, master data kegiatan sekolah, dan master data kesiswaan. Adapun atribut keputusan dalam prediksi kelulusan ini adalah pencapaian target hafalan pada akhir tahun keenam. Prediksi dengan *value* YA atau TIDAK dengan tingkat kepercayaan dapat dievaluasi pada tahun berikutnya untuk angkatan siswa yang lebih muda. Tabel 1 berikut ini adalah koleksi kebutuhan variabel yang akan digunakan dalam penelitian.

Tabel 1. Tabel Kebutuhan Data

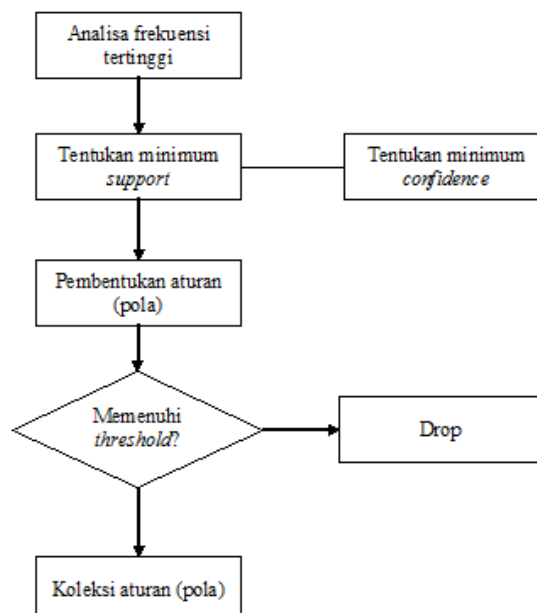
Atribut	Tabel Data	Sumber Data
Nilai akumulasi	Tabel nilai	Rapor
Nilai pelajaran unggulan	Tabel nilai	Buku besar
Ekskul	Tabel kegiatan siswa	Kesiswaan
Prestasi eksternal	Tabel prestasi sekolah	Umum
Nilai tahfidz akhir	Tabel nilai	Hasil ujian

Setelah data historis yang mengandung variabel mandatori didapatkan, maka langkah selanjutnya adalah melakukan data mining dengan algoritma MBA. Algoritma MBA adalah usaha memperoleh pola dengan memperhatikan kombinasi item (Kusrini, 2009). Dalam eksperimen yang dikerjakan kombinasi yang dimaksud dimodifikasi menjadi kombinasi nilai antar atribut. Perbedaan yang cukup signifikan dibanding penelitian menggunakan metodologi MBA lain terletak pada populasi itemset yang diteliti.

Pada penelitian MBA umumnya meneliti item pada sekumpulan transaksi pada waktu tertentu yang memuat nilai dari satu variabel saja, misalnya item barang yang dibeli pada satu struk belanja (Gunadi, 2012). Modifikasi yang dilakukan peneliti adalah mengganti item barang belanja dengan variabel-variabel yang mempengaruhi keberhasilan siswa layaknya metodologi pohon keputusan. Perbedaan kedua adalah jumlah item pembentuk kombinasi dalam penelitian ini adalah sama, sedangkan dalam penelitian lainnya jumlah item pembentuk kombinasi tergantung transaksi yang terjadi pada setiap *event* belanja (Hermawaty, 2013). Secara umum algoritma MBA dapat dijelaskan pada Gambar 2 berikut.

1. Analisis frekuensi tertinggi

Frekuensi yang dimaksud adalah jumlah kombinasi yang muncul dalam database. Oleh karena variabel yang terlibat dalam setiap transaksi jumlahnya tetap, maka hasil yang didapat adalah 100%. Kondisi ini tidak patut untuk diteliti, sehingga dilakukanlah langkah *pre processing* yang dalam hal ini adalah mengusahakan nilai dalam masing-masing variabel menjadi satu dalam sebuah variabel baru. Dengan demikian maka nilai variabel dianggap sebagai item uji independent.



Gambar 4. Algoritma MBA

Berikut ini adalah proses pengubahan bentuk nilai variabel menjadi item independen.

a. Menentukan kriteria nilai

Ada 3 kelompok nilai variabel yang akan ditentukan kriterianya, yaitu nilai akumulatif (jangkauan 50,00-100,00), nilai pelajaran unggulan (jangkauan 5,00-10,00), dan prestasi eksternal (jangkauan 0-3). Adapun nilai tahfidz akhir dan ekskul yang sifatnya biner tidak dilakukan perubahan.

Tabel 2. Tabel Koleksi Data

Data ke-i	Var 1	Var 2	Var 3	Var 4	Var 5
1	80.00	8.15	3.00	1.00	A
2	68.00	8.90	0.00	0.00	B
3	73.20	8.01	2.00	1.00	B
4	80.00	7.41	2.00	1.00	B
5	75.80	7.44	1.00	1.00	B
...
n	value 1	value 2	value 3	value 4	value 5

Tabel 3. Tabel Konversi Kriteria

Var. 1		Var. 2		Var. 3	
Kriteria	Value	Kriteria	Value	Kriteria	Value
60.00-69.99	D	6.00-6.99	D	0	ALPHA
70.00-79.99	C	7.00-7.99	C	1	BETA
80.00-89.99	B	8.00-8.99	B	2	SIERA
90.00-100.00	A	9.00-10.00	A	3	SIGMA

b. Konversi nilai variabel ke bentuk kriteria

Setelah kriteria yang terbentuk diaplikasikan pada item set, maka didapat hasil sebagai berikut :

Tabel 4. Tabel Hasil Konversi Data

Var 1	Var 2	Var 3	Var 4	Var 5
B	B	SIGMA	1.00	SEMPURNA
D	B	ALPHA	0.00	BAIK
C	B	SIERA	1.00	BAIK
B	C	SIERA	1.00	BAIK
C	C	BETA	1.00	BAIK
...
value 1	value 2	value 3	value 4	value 5

c. Inventarisasi kriteria sebagai item transaksi

Selanjutnya merangkum nilai masing-masing variabel menjadi bentuk transaksi seperti berikut :

Tabel 5. Tabel Kombinasi

Data ke-i	Kombinasi
1	A,B,1,ALPHA,SEMPURNA
2	A,C,0,ALPHA,SEMPURNA
3	A,C,1,ALPHA,SEMPURNA
4	B,0,ALPHA,SEMPURNA
5	B,0,BETA,SEMPURNA
6	B,0,SIERA,SEMPURNA

2. Menentukan minimum support dan minimum confidence

Untuk dapat menghitung *Support* maka harus diketahui bentuk tabular data transaksi terlebih dahulu. Dalam Tabel 6 (*membership*) di bawah ini angka 1

menandakan adanya *membership* nilai variabel (item transaksi) di dalam data ke-i, sedangkan 0 bernilai sebaliknya.

Tabel 6. Tabel Distribusi Nilai Data Konversi

Data ke-i	A	B	C	D	1	0	ALPHA	BETA	SIERA	SIGMA	SEMPURNA	BAIK
1	0	2	0	0	1	0	0	0	0	1	1	0
2	0	1	0	1	0	1	1	0	0	0	0	1
3	0	1	1	0	1	0	0	0	1	0	0	1
4	0	1	1	0	1	0	0	0	1	0	0	1
5	0	0	2	0	1	0	0	1	0	0	0	1
6	0	1	1	0	0	1	1	0	0	0	1	0

Dari tabel selanjutnya akan terbentuk kombinasi item transaksi. Banyaknya kombinasi yang terbentuk dapat dihitung menggunakan formula kombinasi sebagai berikut :

$$Jumlah\ kombinasi = \sum_{j=1}^n jml\ kriteria\ variable(j) \dots\dots\dots (1)$$

Tabel 7 berikut adalah tabel kombinasi yang terbentuk :

Tabel 7. Tabel Frekuensi Kombinasi

Kombinasi	Frekuensi
B,C,1,BETA,BAIK	1
B,C,1,BETA,SEMPURNA	1
B,C,1,SIERA,BAIK	4

Tampak bahwa dari 4 kombinasi yang tersaji diketahui 3 elemen selalu muncul, yakni B,C, dan 1. Hal ini menandakan adanya *support* yang merupakan peluang munculnya elemen-elemen pembentuk kombinasi terhadap total transaksi (dalam hal ini total kombinasi). *Support* dapat dihitung menggunakan Rumus (2) berikut:

$$Support\ (B, C, 1) = \frac{\sum\ transaksi\ mengandung\ B,C\ dan\ 1}{\sum\ transaksi} \dots\dots\dots (2)$$

3. Pembentukan aturan

Aturan dapat dibentuk dengan menghitung nilai *confidence* pada sejumlah transaksi kombinasi terhadap frekuensi kemunculan item, dalam hal ini kemunculan nilai tahfidz (sempurna atau baik). *Confidence* adalah peluang kombinasi yang mengandung dua atau lebih elemen dalam seluruh kombinasi yang mengandung salah satu elemen yang bersangkutan. Secara umum perhitungan *confidence* didapat menggunakan Rumus (3) (Gunadi, 2012).

$$Confidence = P(A|B) = \frac{\sum\ transaksi\ mengandung\ A\ dan\ B}{\sum\ transaksi\ mengandung\ A} \dots\dots\dots (3)$$

4. Seleksi aturan

Aturan yang berhasil dibentuk kemudian diseleksi menurut batas kepercayaan yang ditetapkan.

5. Koleksi aturan final

Aturan final yang akan dihasilkan adalah aturan yang memenuhi aturan nilai *Support x Confidence*. Dengan metode pemilihan aturan seperti ini diharapkan dapat diperoleh akurasi yang optimal.

HASIL DAN PEMBAHASAN

Penelitian yang dilakukan adalah penelitian relasional, yang ditujukan untuk mengetahui hubungan implikasi variabel hipotesa dengan variabel antiseden. Permasalahan yang ingin diketahui adalah hubungan antara perolehan nilai akumulatif siswa, nilai rata-rata mata pelajaran unggulan, keaktifan internal (ekstra kurikuler) dan eksternal terhadap pencapaian target tahfidz di akhir tahun ke enam. Oleh karenanya kombinasi yang akan diteliti meliputi 5 variabel tersebut di atas.

1. Support

Kombinasi yang mungkin dapat terbentuk dari kelima variabel menurut Rumus (1) adalah 64. Jumlah ini tentunya tidak akan pernah tercapai mengingat sample penelitian yang digunakan hanya sejumlah 33 *record*. Adapun dari data historis yang dimiliki hanya 27 kombinasi yang terbentuk atau 81,82% dari potensi 33 kombinasi. Berikut adalah kombinasi dengan frekuensi kemunculan ≥ 2 ($\Phi=2$). Dengan batasan ini maka hanya akan ada 4 aturan yang diteliti lebih lanjut seperti tertera dalam Tabel 8. Adapun nilai *support* diperoleh menggunakan perhitungan Rumus (2).

Tabel 8. Tabel Support

ID	Kombinasi	Frekuensi	Support
9	B ₁ BETA_SEMPURNA	2	6.06%
11	B ₂ C ₀ ALPHA_BAIK	2	6.06%
14	B ₂ C ₁ ALPHA_BAIK	2	6.06%
17	B ₂ C ₁ SIERA_BAIK	4	12.12%

2. Pembentukan aturan dan *confidence*

Menggunakan Rumus (3) maka dapat dibentuk aturan inisial dengan nilai *confidence* seperti dijelaskan dalam Tabel 9 berikut :

Tabel 9. Tabel *Support* Aturan Inisial

Aturan	<i>Confidence</i>	
Jika nilai akumulatif, unggulan B dan mengikuti ekskul dan kelompok prestasi eksternal BETA maka tahfidz SEMPUANA	2/2	100%
Jika nilai akumulatif B dan nilai unggulan C dan tidak mengikuti ekskul dan kelompok prestasi eksternal ALPHA maka tahfidz BAIK	2/8	25%
Jika nilai akumulatif B dan nilai unggulan C dan mengikuti ekskul dan kelompok prestasi eksternal ALPHA maka tahfidz BAIK	2/8	25%
Jika nilai akumulatif B dan nilai unggulan C dan mengikuti ekskul dan kelompok prestasi eksternal SIERA maka tahfidz BAIK	4/8	50%

3. Aturan final

Dari Tabel 8 dan 9 dapat dibentuk tabel aturan final seperti Tabel 10 berikut :

Tabel 10. Tabel Pembentukan Aturan Final

No aturan	ID kombinasi	Support	Confidence	Support x confidence
1	9	6,06%	100%	6,06%
2	11	6,06%	25%	1,52%
3	14	6,06%	25%	1,52%
4	17	12,12%	50%	6,06%

Aturan final yang dapat diterima dengan batasan nilai *support x confidence* 5% adalah aturan 1 dan 4. Nilai *support* tampak sangat kecil mengingat peluang frekuensi dari total populasi juga kecil. Jika dari 33 populasi menghasilkan 33 kombinasi maka masing-masing hanya berpeluang sebesar 3%. Oleh karenanya penetapan nilai ambang batas 5% dirasa masih wajar.

KESIMPULAN

Dengan memperhatikan variabel nilai akumulatif, nilai mata pelajaran unggulan, keaktifan kegiatan internal dan eksternal didapat 2 aturan kuat untuk memprediksi keberhasilan tahfidz siswa. Dapat disimpulkan pula bahwa untuk siswa dengan nilai kumulatif dan mata pelajaran unggulan minimal B, serta aktif dalam kegiatan internal maka diprediksi akan memperoleh nilai tahfidz SEMPURNA.

Dari 33 sample penelitian dan 4 variabel hipotesa terbentuk 27 kombinasi pembentuk aturan, yaitu 81,82% dari potensial (maksimal 33), atau 42,19% dari seluruh potensial kombinasi (64 variasi) dan diperoleh 4 aturan MBA dengan frekuensi tertinggi.

DAFTAR PUSTAKA

- Gorunescu, Florin, 2011, *Data Mining : Concepts, Models and Techniques*, Springer-Verlag Berlin Heidelberg.
- Gunadi, G., Senses, Dana I., 2012, *Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku dengan Menggunakan Algoritma Apriori Dan frequent Pattern Growth (Fp-Growth): Studi Kasus. Percetakan PT. GRAMEDIA.*
- Hermawaty, F. A., 2013, *Data Mining*, Andi Publisher, Yogyakarta.
- Kusrini., Luthfi, E. T., 2009, *Algoritma Data Mining*, Penerbit Andi, Yogyakarta
- Listiyoko, L., Roihan, A., Karwandi, 2015 *Implementasi Data Mining Dalam Rangka Optimalisasi Iklan Perusahaan Ritel*, NS-CCIT
- Oded Maimon · Lior Rokach, 2013, *Data Mining and Knowledge : Discovery Handbook*, Springer-Verlag Berlin Heidelberg.

Prosiding SNRT (Seminar Nasional Riset Terapan)
Politeknik Negeri Banjarmasin, 9 November 2017

ISSN 2341-5662 (Cetak)
ISSN 2341-5670 (Online)

- Susanto, Sani, PhD; Suryadi, Dedy, S.T.,M.S, 2010, *Pengantar Data Mining – Menggali Pengetahuan Dari Bongkahan Data*, Andi Publisher, Yogyakarta.
- Vercellis; Carlo, 2009, *Business Intelligence: Data Mining and Optimization for Decision Making*, Springer-Verlag Berlin Heidelberg.
- Yanto, R., Khoiriah, R., 2015, *Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat*, Citec Journal, Vol. 2, No. 2, Februari 2015 – April 2015 102-113