

Penerapan Data Mining Classification Untuk Prediksi Perilaku Pola Pembelian Terhadap Waktu Transaksi Menggunakan Metode Naïve Bayes

Susanto, S.Kom, M.Kom¹), Evi Dewi Sri Mulyani, S.Kom, M.Kom²), Irma Ratnasari Nurhasanah³)
STMIK TASIKMALAYA

Jl. RE. Martadinata No. 272 Indihiang Kota Tasikmalaya, Jawa Barat-Indonesia Tlp. (0265) 310830
susantoprata@yahoo.co.id, eviajadech@gmail.com, irmaratna01@gmail.com,

Abstrak

Data Mining adalah cabang ilmu baru untuk mengatasi masalah penggalian informasi atau pola yang penting dari data dalam jumlah yang besar. Jumlah pembeli memiliki pengaruh yang besar terhadap transaksi penjualan. Semakin banyak pembeli, semakin besar kemungkinan transaksi yang dapat diperoleh. Jumlah pembeli setiap harinya berbeda-beda dan memiliki persentase yang tidak sama dalam melakukan transaksi penjualan. Salah satu cara dalam meningkatkan pendapatan penjualan adalah dengan memprediksi penjualan berdasarkan rata-rata jumlah pembeli. Pada penelitian ini, penulis berusaha menambang data (data mining) transaksi penjualan untuk memprediksi perilaku pola pembeli pada tahun berikutnya (2015) dengan menggunakan data tahun sebelumnya (2014). Data yang ada dianalisis menggunakan perhitungan Rapid Miner 5.1 dengan algoritma Naïve Bayes. Naïve Bayes merupakan salah satu metode pada Probabilistic Reasoning. Algoritma Naïve Bayes bertujuan untuk melakukan klasifikasi data pada kelas tertentu, kemudian pola tersebut dapat digunakan untuk memperkirakan data transaksi penjualan dalam memprediksi perilaku pola pembeli.

Kata kunci: Data Mining, Klasifikasi, Algoritma Naïve Bayes, Rapid Miner 5.1

1. Pendahuluan

Dengan kemajuan teknologi informasi diharapkan dapat menjadi media yang paling efektif untuk mencari dan menyebarkan informasi yang sangat dibutuhkan dalam kehidupan saat ini, sehingga informasi yang tinggi kadang tidak diimbangi dengan penyajian informasi yang memadai, sering kali informasi tersebut masih harus digali ulang dari data yang jumlahnya sangat besar. Perkembangan teknologi saat ini mengalami kemajuan yang sangat besar, sehingga pemanfaatan teknologi informasi dapat mempermudah kinerja manusia. Pemanfaatan data yang ada didalam sistem informasi untuk menunjang kegiatan pengambilan kesimpulan, tidak hanya mengandalkan data operasional yang tersedia saja, diperlukan suatu analisis data untuk menggali potensi-potensi informasi yang ada.

Pengambilan keputusan berusaha untuk memanfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna membantu mengambil kesimpulan, hal ini menjadikan munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi atau pola yang penting atau menarik dari data dalam jumlah yang besar, yang disebut dengan data mining. Penggunaan teknik data mining diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya.

2. Metode Penelitian

Metode yang digunakan pada penelitian skripsi ini adalah metode kualitatif dengan pendekatan deskriptif.

Metode kualitatif sering disebut metode penelitian naturalistik karena penelitiannya dilakukan pada kondisi yang alamiah, disebut juga metode etnographi, karena pada awalnya metode ini lebih banyak digunakan untuk penelitian.

maka penelitian kualitatif dapat diartikan sebagai usaha untuk menyelidiki keadaan yang sebenarnya, dalam memprediksi perilaku pola pembelian berdasarkan waktu transaksi penjualan sehingga

perencanaan strategi penjualan dapat tercapai. pustaka guna mendukung data-data sekunder yang diperoleh dari objek penelitian serta referensi lainnya yang berkaitan dengan penelitian ini.

3. Hasil dan Pembahasan

Dalam memprediksi pola perilaku pembelian terhadap waktu transaksi berdasarkan rumus *naïve bayes*, maka H merepresentasikan jumlah pembeli. $P(H)$ adalah prior probability dimana kasus ini merupakan probabilitas pembeli terhadap transaksi dengan waktu tertentu. $P(H|X)$ merefleksikan probabilitas transaksi dengan X berupa waktu pembelian. $P(X|H)$ adalah *posterior* probability yang menunjukkan kemungkinan pola perilaku pembeli berdasarkan prediktor X . $P(X)$ adalah prior probability yang merupakan probabilitas transaksi dengan kriteria X . Dari data transaksi penjualan yang diperoleh diambil sampel data sebanyak 1768.

No Nota	Tanggal	Nama Barang	Merk Barang	Harga Barang	Jumlah	Harga Jual	Jam	Hari
#010000	1/1/2014	Sepatu	Toha	245000	1	245000	08.00-10.00	Rabu
#010001	1/1/2014	Sepatu	Toha	245000	1	245000	08.00-10.00	Rabu
#010002	1/1/2014	Sepatu	Glorpy	245000	1	245000	10.00-12.00	Rabu
#010003	1/1/2014	Sepatu	Toha	245000	1	245000	12.00-15.00	Rabu
#010004	1/1/2014	Kerudung	Kaila La Fauza	75000	1	75000	15.00-17.00	Rabu
#010005	1/1/2014	Kerudung	Kaila La Fauza	75000	1	75000	15.00-17.00	Rabu
#010006	1/1/2014	Sepatu	Davy Nelo	315000	1	315000	08.00-10.00	Kamis
#010007	2/1/2014	Baju Muslim Wanita	Dress Zoya Iner Alba	99000	2	198000	08.00-10.00	Kamis
#010008	2/1/2014	Kerudung	Kaila La Fauza	75000	1	75000	08.00-10.00	Kamis
#010009	2/1/2014	Sepatu	Toha	245000	1	245000	10.00-12.00	Kamis
#010010	2/1/2014	Sepatu	Big Pro	315000	1	315000	10.00-12.00	Kamis
#010011	2/1/2014	Kerudung	Zoya Zaida Algarve	109000	1	109000	12.00-15.00	Kamis
#010012	2/1/2014	Kerudung	Kaila La Fauza	75000	2	150000	15.00-17.00	Kamis
#010013	2/1/2014	Baju Muslim Pria	ZeeFora	100000	1	100000	15.00-17.00	Kamis
#010014	2/1/2014	Sepatu	Toha	245000	2	490000	08.00-10.00	Jumat

Gambar 3.1. Data Training

2.1. Proses Data Mining

2.1.1. Pemrosesan Awal

Tahap pertama dalam proses data mining adalah memahami domain aplikasi untuk mengetahui dan menggali pengetahuan awal serta apa sasaran pengguna. Dalam pemrosesan ini yang menjadi sasarannya adalah data transaksi penjualan pada tahun 2014-2015, yang akan menghasilkan grafik atau pola untuk memperlihatkan prediksi pada tahun 2016.

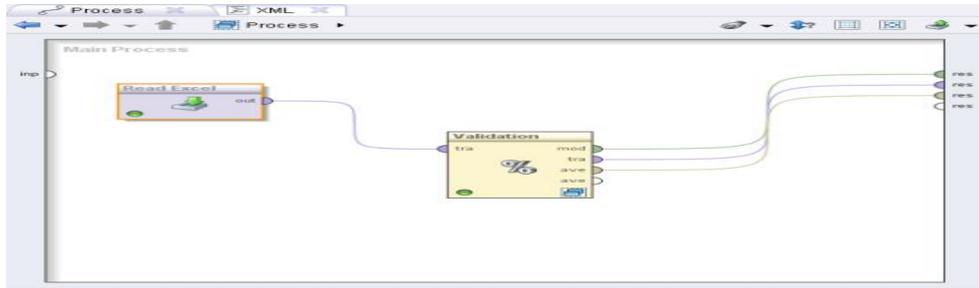
Pembuatan model diawali dengan pembacaan file data (read excel). Data training dan data testing disimpan dalam satu file excel 2010. Data tersebut kemudian divalidasi. Untuk data testing terdapat 2 klasifikasi yang menjadi label yaitu Pagi dan Siang.

Tanggal	Nama Barang	Harga Barang	Jam	Hari	Klasifikasi
01/01/2014	Sepatu	245000	08.00-10.00	Rabu	Pagi
01/01/2014	Sepatu	245000	08.00-10.00	Rabu	Pagi
01/01/2014	Sepatu	245000	10.00-12.00	Rabu	Pagi
01/01/2014	Sepatu	245000	12.00-15.00	Rabu	Siang
01/01/2014	Kerudung	75000	15.00-17.00	Rabu	Siang
01/01/2014	Kerudung	75000	15.00-17.00	Rabu	Siang
01/01/2014	Kerudung	75000	15.00-17.00	Rabu	Siang
01/01/2014	Sepatu	315000	08.00-10.00	Kamis	Pagi
01/02/2014	Baju Muslim Wanita	99000	08.00-10.00	Kamis	Pagi
01/02/2014	Kerudung	75000	08.00-10.00	Kamis	Pagi
01/02/2014	Sepatu	245000	10.00-12.00	Kamis	Pagi
01/02/2014	Sepatu	315000	10.00-12.00	Kamis	Pagi
01/02/2014	Kerudung	109000	12.00-15.00	Kamis	Siang
01/02/2014	Kerudung	75000	15.00-17.00	Kamis	Siang
01/02/2014	Baju Muslim Pria	100000	15.00-17.00	Kamis	Siang
01/02/2014	Sepatu	245000	08.00-10.00	Jumat	Pagi
01/03/2014	Sandal	185000	10.00-12.00	Jumat	Pagi

Gambar 3.2. Data Testing

3.2. Validation

Melakukan validation yaitu melakukan analisis berbagai model dan memilih model dengan kinerja prediksi yang baik. Pada gambar 3.3. merupakan proses validasi, setelah pembacaan file data, blok read excel dihubungkan dengan blok validation. Proses data training dan testing didalam blok proses (validation) dapat ditampilkan dengan melakukan klik 2x menggunakan mouse sehingga muncul proses training dan testing.

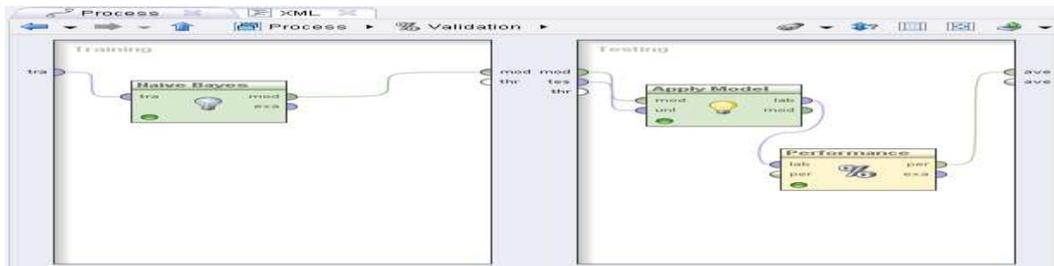


Gambar 3.3. Validation

3.3. Proses Training dan Testing

Proses training yaitu melakukan proses pelatihan data pada model (naïve bayes). Sedangkan proses testing yaitu melakukan pengujian data yang akan menghasilkan grafik atau pola.

Pada gambar 3.4 dijelaskan bahwa proses training digunakan untuk blok model naive bayes, dihubungkan dengan garis penghubung pada blok apply model dan blok performance dengan bagian testing sebagai penampil informasi hasil dari pengujian data, hasil pengujian ini akan menghasilkan arsitektur naïve bayes. Dengan menggunakan Rapid Miner 5.1 dapat menganalisis dataset transaksi penjualan menggunakan metode naïve bayes.



Gambar 3.4. Training dan Testing

3.4. Hasil Performance Vector

Proses klasifikasi dengan Rapid Miner 5.1 dengan metode naïve bayes yang digunakan untuk mengklasifikasikan data transaksi pada penelitian ini sehingga diperoleh nilai Accuracy, precision, dan recall.

3.4.1. Accuracy

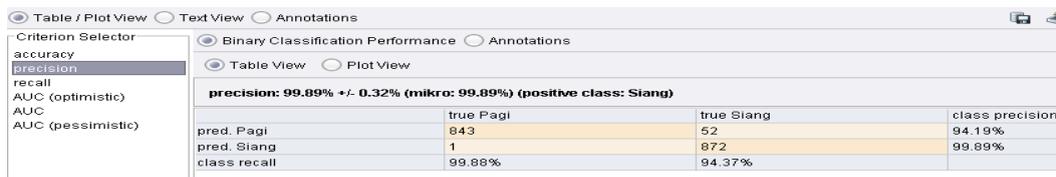
Dengan mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi hasil prediksi yaitu 97.00% dari hasil data testing.

Criterion Selector		Multiclass Classification Performance	
accuracy		accuracy: 97.00% +/- 1.13% (mikro: 97.00%)	
precision		pred. Pagi	true Siang
recall		843	52
AUC (optimistic)		pred. Siang	872
AUC		1	
AUC (pessimistic)		class recall	94.37%
		99.88%	

Gambar 3.5. Accuracy

3.4.2. Precision

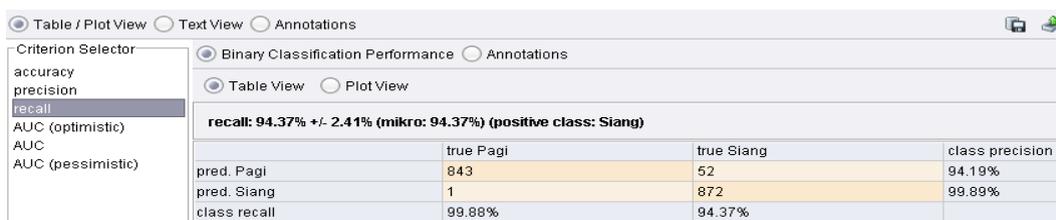
Precision adalah jumlah data yang true positive (jumlah data positif yang dikenali secara benar sebagai positif) dibagi dengan jumlah data yang dikenali sebagai positif. Dari hasil pengujian nilai precision yaitu 99.89% untuk class siang dan 94.19% untuk class pagi.



Gambar 3.6. Precision

3.4.3. Recall

Recall adalah jumlah data yang true positive dibagi dengan jumlah data yang sebenarnya positif (true positive + true negative). Untuk nilai recall yaitu 94.37% pada class siang dan nilai class pagi yaitu 99.88%.



Gambar 3.7. Recall

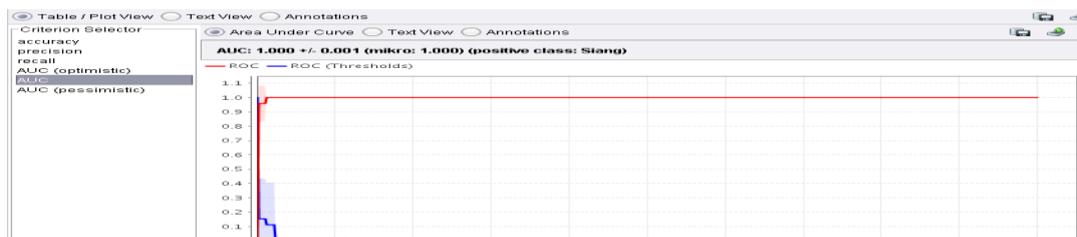
Dari hasil pengujian untuk nilai accuracy 97,00%, untuk nilai precision yaitu 99.89% pada class siang dan 94.19% pada class pagi, sedangkan nilai recall 94.37% pada class siang dan 99.88% untuk nilai class pagi. Hasil class accuracy, precision dan recall untuk memprediksi perilaku pola pembelian terhadap waktu transaksi penjualan di toko Toha tasikmalaya dapat di lihat pada Tabel 3.1.

Tabel 3.1. Hasil Class Accuracy, Precision, dan Recall

Penerapan Data Mining <i>Classification</i> Untuk Prediksi Prilaku Pola Pembelian Terhadap Waktu Transaksi Penjualan Menggunakan Metode <i>Naïve Bayes</i> Di Toko Toha Tasikmalaya				
Penulis	Class	Accuracy	Precision	Recall
Irma Ratnasari Nurhasanah	Pagi	97.00%	94.19%	99.88%
	Siang		99.89%	94.37%

3.5. AUC (Area Under Curve)

Selain confusion matrix, kurva Receiver Operating Characteristic (ROC) dihasilkan oleh rapid miner 5.1. kurva tersebut dapat dilihat pada gambar 3.8.



Gambar 3.8. AUC (Area Under Curve)

Kurva Receiver Operating Characteristic (ROC) digunakan untuk mengekspresikan data confusion matrix. Garis horizontal mewakili nilai false positives (FP) dan garis vertikal mewakili nilai true positives (TP). Dari gambar 4.8. dapat diketahui bahwa nilai Area Under Curve (AUC) model

algoritma naïve bayes adalah 1.000. hal ini menunjukkan bahwa model algoritma naïve bayes mencapai klasifikasi sempurna.

3.6. Hasil Klasifikasi Class dengan Metode Naïve Bayes

3.6.1. Simple Distribution Model

Dengan menggunakan Rapid Miner 5.1. untuk menganalisis tabel data transaksi penjualan dalam memprediksi perilaku pola pembelian dengan menggunakan metode naïve bayes dapat menghasilkan 4 (empat) kelas utama pembagian seperti pada gambar 3.9



Gambar 3.9 Simple Distribution Model

Hasil klasifikasi dari data transaksi penjualan dengan metode naïve bayes membagi 2 kelas klasifikasi yaitu class pagi dan class siang. Untuk nilai class pagi yaitu 0.477 dan nilai class siang yaitu 0.523.

Tabel 3.2. Simple Distribution Model

Class	Nilai
Pagi	0.477
Siang	0.523

3.6.2. Distribution Tabel

Dimana dapat dilihat tabel distribusi hasil analisa dengan metode naïve bayes terhadap tabel data transaksi penjualan, dalam memprediksi perilaku pola pembelian di Toko Toha Tasikmalaya.



Gambar 3.10. Distribution Table

4. Simpulan

Dengan demikian dari hasil penelitian, pembahasan dan pengujian model dapat disimpulkan bahwa :

1. Pengujian menggunakan 1768 data dengan perhitungan Rapid Miner 5.1 dihasilkan nilai Accuracy, Precision, dan Recall dapat disimpulkan untuk nilai accuracy 97,00%, untuk nilai precision yaitu 99,89% pada class siang dan nilai precision pada class pagi yaitu 94,19%, sedangkan nilai recall yaitu 94,37% pada class siang dan nilai recall pada class pagi yaitu 99,88%.
2. Hasil klasifikasi dari data transaksi penjualan dengan metode naïve bayes membagi 2 kelas klasifikasi yaitu class pagi dengan nilai 0.477, dan class siang dengan nilai 0.523

-
3. Hasil model plot view analisa data transaksi penjualan untuk melihat perilaku pola pembeli dapat diprediksi dengan density tertinggi pada class siang mencapai 16,8% pada hari minggu.

Daftar Pustaka

- [1] Sandi Fajar Rodiyansyah dan Edi Winarko, "Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayes Classification," *IJCCS*, vol. 7, no. ISSN, pp. 13-22, Jan. 2013.
- [2] J. F. Ulysses, "Data Mining Classification Untuk Prediksi Lama Masa Studi Mahasiswa Berdasarkan Jalur Penerimaan Dengan Metode Naive Bayes".
- [3] Y. Salim, "Penerapan Algoritma Naive Bayes Untuk Penentuan Turn Over Pegawai," *Media Sains*, vol. 4, no. ISSN, 2012.
- [4] Bustami, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi".
- [5] I. B. Sangadji, "Prediksi Prilaku Pola Pengunjung Terhadap Transaksi Pada Toko Buku Gramedia menggunakan jaringan syaraf tiruan metode Back Propagation," *Junral Informatika*, vol. 5, pp. 135-150, 2009.
- [6] E. Prasetyo, *DATA MINIG Mengolah Data Informasi Menggunakan Matlab*, A. Sahala, Ed. Yogyakarta: ANDI, 2014.
- [7] F. A. Hermawati, *Data Mining*. Yogyakarta: Andi, 2013.
- [8] R. S. Wahono, *Data Mining*. 2012.
- [9] (2010) Sejarah dan Pengertian Microsoft Excel. [Online].
<http://mairatnasari.blogspot.com/2013/01/sejarah-pengertian-microsoft-excel-2010.html>
- [10] *Mengoptimalkan Microsoft Excel 2010 Untuk Analisis Data*. Yogyakarta: ANDI, 2011.
- [11] I. ., M. ., B. Hariyanto, *Rekayasa Sistem Berorientasi Objek*. Bandung, 2004.
- [12] P. D. Sugiyono, *Memahami Penelitian Kualitatif*. Bandung: ALFABETA, 2009.
- [13] Z. F. & D. M. A. H. Nurhadi, *Metodologi Penelitian Kualitatif: Teori dan Paradigma*. Bandung: CV Alfabeta, 2012.