Penerapan Data Mining Untuk Prediksi Rating Penjualan Buku Menggunakan Metode Naive Bayes

Sulastri*, Yusuf Sulistyo Nugroho
Program Studi Informatika, Fakultas Komunikasi dan Informatika
Universitas Muhammadiyah Surakarta
sulastri2829@gmail.com*

Abstrak

Lahirnya sebuah buku, dipengaruhi oleh beberapa aspek yaitu pengarang, percetakan, penerbit dan distributor. Untuk mendukung perkembangan ke empat aspek tersebut peran teknologi sangat dibutuhkan. Dewasa ini banyak penerbit memanfaatkan teknologi sebagai sarana untuk mempromosikan buku hasil terbitannya. Namun hal tersebut terkadang tidak mempengaruhi ketertarikan pembaca dalam pemilihan buku bacaannya. Melihat banyaknya buku yang beredar di pasaran dengan segala kelebihan dan kelemahannya maka diperlukan sebuah sistem prediksi untuk mengetahui faktor yang mempengaruhi rating penjualan buku untuk menghasilkan sebuah buku yang memenuhi kriteria pembaca. Teknik data mining yang digunakan dalam prediksi rating penjualan buku ini menggunakan metode Naïve Bayes. Naïve Bayes diterapkan untuk mencari nilai probabilitas terbesar dalam setiap variabel yang ada. Variabel yang digunakan untuk memprediksi rating penjualan buku terdiri dari kategori buku, genre, harga, sampul, penerbit, kategori usia dan rating. Dari penelitian ini diperoleh tingkat akurasi sebesar 74,60 %, recall 83,62 % dan presisi 80,61 % .

Kata Kunci: buku, data mining, naïve bayes, sistem prediksi.

I. PENDAHULUAN

Buku merupakan salah satu sumber informasi dan ilmu yang mudah ditemukan. Lahirnya sebuah buku untuk sampai ke tangan pembaca melalui proses yang panjang dan adanya usaha bersama yang melibatkan minimal empat unsur pendukung yaitu pengarang, percetakan, penerbit dan distributor (Djuhro, 2016). Namun jenis

bacaan yang termuat dalam sebuah buku terkadang tidak mempengaruhi tingkat ketertarikan pembaca dalam pemilihan buku tersebut. Hasil studi *'Most Littered Nation In the World'* yang dilakukan oleh *Central Connecticut State Univesity* menunjukkan bahwa minat membaca di Indonesia menduduki peringkat ke-60 dari 61 negara, artinya minat membaca masyarakat Indonesia sangat rendah (Gewati, 2016).

Dewasa ini banyak penerbit memanfaatkan teknologi sebagai sarana untuk mempromosikan buku hasil terbitannya. Penjualan buku melalui internet dianggap sebagai salah satu cara mudah untuk menyebarluaskan buku hasil terbitannya tanpa memakan banyak waktu dan biaya. Sementara itu, tingkat penjualan buku juga dipengaruhi oleh beberapa faktor, salah satunya adalah genre buku yang sesuai dengan keinginan pembaca tidak menutup kemungkinan akan laris dipasaran (Probo & Afrisa, 2015).

Melihat banyaknya buku yang beredar di pasaran dengan segala kelebihan dan kelemahannya, maka dalam penelitian ini akan dilakukan proses data mining untuk memprediksi rating penjualan buku. Data mining dapat digunakan oleh sebuah organisasi dan suatu lembaga yang memiliki data melimpah untuk memberikan informasi yang dapat mendukung pengambilan keputusan (Bhardwaj et al, 2012). Metode yang dapat digunakan untuk memprediksi suatu hal dengan data mining sangat banyak. Oleh karena itu, penelitian ini akan menggunakan metode *Naïve Bayes*. *Naïve Bayes* merupakan metode yang bersifat sederhana dan efektif dalam pengolahan data yang besar (Chandrasekar & Qian, 2016). Prinsip klasifikasi metode *Naïve Bayes* adalah memanfaatkan teorema menghitung probabilitas posterior dari objek tertentu (Wang et al, 2016).

Berdasarkan latar belakang tersebut, maka akan dilakukan suatu prediksi terhadap rating penjualan buku di pasaran berdasarkan data buku yang telah diperoleh dan telah diketahui ratingnya yang akan digunakan sebagai data pelatihan. Sehingga penerapan data mining dengan menggunakan metode *Naïve Bayes* dapat membantu penerbit buku untuk mempertimbangkan aspek yang mempengaruhi rating penjualan sebuah buku sehingga mampu memenuhi kriteria yang

dibutuhkan pembaca yang nantinya akan meningkatkan minat baca masyarakat.

II. TINJAUAN PUSTAKA

2.1 Visual Basic .NET

Visual Basic .NET adalah sebuah tool development Microsoft yang digunakan untuk membangun aplikasi di lingkungan kerja berbasis sistem operasi windows diatas platform Framework.NET (Rahmat & Suryadi, 2008). Visual Basic .NET merupakan bahasa pemrograman yang menawarkan IDE yang terintegrasi untuk membuat program.

2.2 Sistem Prediksi

Sistem prediksi merupakan sistem yang digunakan untuk kegiatan prediksi atau memperkirakan segala sesuatu yang akan terjadi. Prediksi digunakan sebagai informasi,ulasan dan pengetahuan berdasarkan faktor-faktor yang mempengaruhi maupun berdasarkan fakta yang ada. Prediksi berdasarkan teori probabilitas yaitu dimana prediksi dilakukan berdasarkan data-data yang telah dikumpulkan dalam waktu tertentu yang didukung perhitungan dengan metode statistik (Imadha, 2014).

2.3 Data Mining

Data mining merupakan proses yang mempermudah dalam pengolahan data besar yang digunakan untuk menghasilkan suatu infromasi yang berguna. Informasi yang dihasilkan tersebut nantinya dapat bermanfaat bagi berbagai bidang. Data mining diterapkan untuk menggali berbagai pengetahuan maupun informasi dari data yang kita miliki sehingga data yang kita miliki dapat bermanfaat (Jananto, 2013).

2.4 Naïve Bayes

Naïve Bayes adalah metode data mining yang sederhana dalam pengolahan data yang besar. Dimana metode naïve bayes mempunyai nilai akurasi yang tinggi jika diterapkan ke dalam database yang besar. Naïve bayes merupakan pendekatan statistik dalam pengenalan pola

(pattern recoginition) yang didasarkan pada asumsi penyederhanaan nilai atribut (Ridwan et al, 2013).

III. METODE PENELITIAN

3.1 Analisis Kebutuhan Data

Pada tahap ini dilakukan pengumpulan data buku untuk dianalisis dan diolah berdasarkan variabel yang telah ditentukan untuk mencari probabilitas terbesar yang digunakan untuk memprediksi rating penjualan buku. Kemudian mencari referensi-referensi studi pustaka berupa jurnal, artikel dan literatur yang berkaitan dengan sistem.

3.2 Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data yang diperoleh dari website Gramedia dan Amazon yang digunakan sebagai data pelatihan yang berjumlah 500 data yang dijelaskan pada tabel 1.

Tabel 1. Jumlah data awal

NAMA VARIABEL	JUMLAH DATA
FIKSI	245
NONFIKSI	255
HARGA < 100000	224
HARGA > 100000	276
SAMPUL SOFTCOVER	332
SAMPUL HARDCOVER	168
PENERBIT DALAM NEGERI	39
PENERBIT LUAR NEGERI	461
USIA <6 TAHUN	84
USIA 6-12 TAHUN	271
USIA >12 TAHUN	145
RATING TINGGI	348
RATING RENDAH	152

Dalam kategori buku fiksi jumlah data terdiri dari 5 genre buku yaitu 72 genre drama, 31 genre komik, 92 genre sains-fiksi, 27 genre dongeng dan 23 genre misteri. Sedangkan dalam kategori buku nonfiksi jumlah data terdiri dari 6 genre buku yaitu 35 genre advice, 89 genre sejarah, 38 genre bisnis, 37 genre sains, 21 genre hukum dan 35 genre bisnis. Data uji didapatkan dari hasil survei yang diberikan oleh pembaca.

3.2.1 Penentuan Variabel Data

Tahap ini dilakukan untuk menentukan kebutuhan data yang nantinya akan membantu pengembangan sistem. Berdasarkan tabel 1 maka akan dilakukan tahap penentuan dan konversi data sebagai berikut:

a. Rating

Pembuatan range untuk kategori rating:

Tabel 2. Rating

Rating Awal	Rating Baru
1– 3.5	Rendah
3.5– 5	Tinggi

b. Kategori Buku

Untuk kategori buka, dikarenakan dalam website Gramedia dan Amazon telah ditentukan kategori buku fiksi dan nonfiksi maka tidak dilakukan konversi data.

c. Genre

Untuk kategori genre, dikarenakan dari sumber data awal di website Gramedia dan Amazon telah terdapat genre sesuai dengan masing-masing kategori buku, maka tidak dilakukan konversi.

d. Harga

Berdasarkan hasil wawancara kepada 30 mahasiswa UMS, kategori harga dibagi menjadi 3 yaitu:

Tabel 3. Harga

Harga Awal	Harga Baru
Harga Awal < 100.000	Murah
100.001<= Harga Awal<= 170.000	Sedang
Harga Awal >= 170.001	Mahal

e. Sampul
Untuk kategori sampul, dikarenakan dari sumber data awal di website telah ditentukan masing-masing sampul buku, maka tidak dilakukan konversi.

f. Penerbit
Untuk kategori penerbit, karena data diperolah dari dua website yang berbeda, maka untuk penerbit dari website Gramedia dikategorikan ke dalam negeri dan untuk penerbit dari website Amazon dikategorikan ke luar negeri.

g. Kategori Usia
Pembuatan range untuk kategori usia sebagai berikut:

Tabel 4.Usia

Usia Awal	Usia Baru
1 th – 12 th	Anak
13 th – 17 th	Remaja
Usia >18 th	Dewasa

Dari hasil penentuan dan konversi data maka diperoleh sususan variabel tampak pada tabel 5 yang digunakaan dalam penelitian untuk membantu dalam pengembangan sistem.

Tabel 5. Variabel yang digunakan dalam penelitian

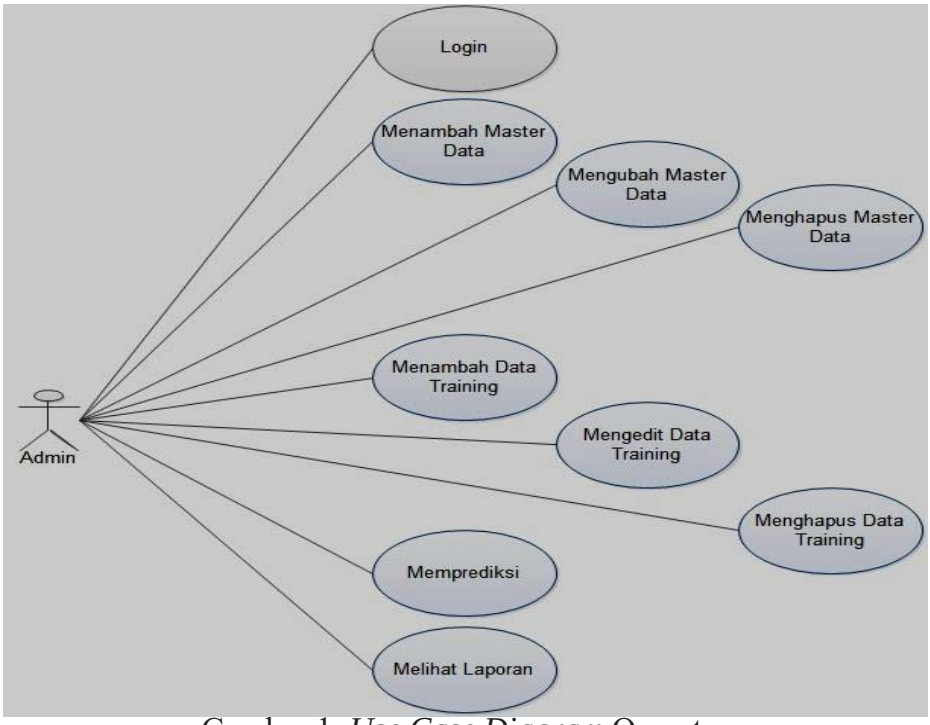
Y	Rating
X1	Kategori Buku
X2	Genre
X3	Harga
X2	Sampul
X4	Penerbit
X5	Kategori Usia

3.3 Pengembangan Sistem

Sistem prediksi ini dibangun berbasis desktop dengan menggunakan Visual Basic .NET dan menggunakan MySQL sebagai pengolahan basis data.

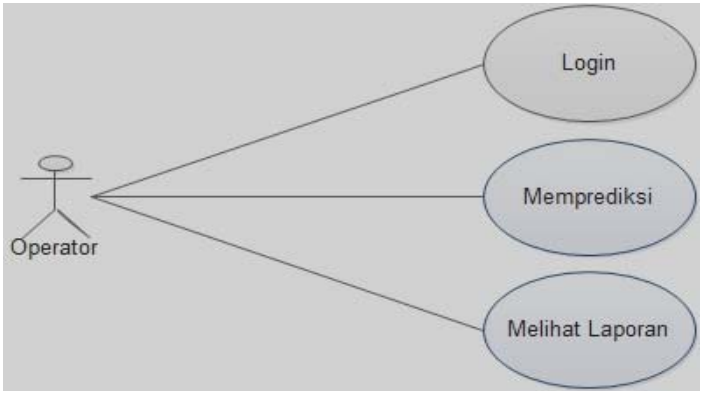
2.3.1 Perancangan Sistem

Pada tahap perancangan sistem ini dibuat *Use Case Diagram*. Ada 2 (dua) *use case diagram* pada sistem prediksi ini, yaitu *use case diagram* untuk admin dan *use case diagram* untuk operator. Pertama, *use case diagram* admin yang memiliki 9 hak akses yaitu login, menambah master data, mengubah master data, menghapus master data, menambah data *training*, mengedit data *training*, menghapus data *training*, memprediksi dan melihat laporan.



Gambar 1. *Use Case Diagram* Operator

Kedua, *use case* diagram operator ini .memiliki 3 hak akses yaitu login, memprediksi dan melihat laporan.



Gambar 2. *Use Case Diagram* Admin

3.4 Metode Naïve Bayes

Menurut Jananto (2013) Metode *Naïve Bayes Classifier* (NBC) merupakan pengklasifikasian statistik yang didasarkan pada teorema *Bayes* yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Perhitungan algoritma *Naïve Bayes* menurut Nugroho dan Haryati (2015) ditunjukkan seperti persamaan (1).

$$P(S|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (1)$$

Keterangan:

X : Data dengan class yang belum diketahui.
S : Hipotesis data X merupakan suatu class spesifik.
P(S|X) : Probabilitas hipotesis S berdasarkan kondisi x (*posteriori probability*)
P(Y) : Probabilitas hipotesis Y (*prior probability*)
P(X|Y) : Probabilitas X berdasarkan kondisi tersebut
P(X) : Probabilitas dari X

IV. HASIL DAN PEMBAHASAN

Pengumpulan data dalam penelitian menghasilkan dua jenis data yaitu data pelatihan dan data uji. Tabel 6 merupakan potongan dari 500 data pelatihan yang digunakan sebagai data pendukung sistem. Isi dari data pelatihan di kategorikan berdasarkan hasil dari penentuan variabel dan konversi data. Jenis data yang kedua adalah data uji yang berjumlah 180 data, ditampilkan pada tabel 7.

Tabel 6. Potongan data pelatihan

ID	Kategori Buku	Kategori Genre	Kategori Harga	Kategori Penerbit	Kategori Sampul	Katgori Usia	Rating
1	FIKSI	Drama	Murah	Dalam Negeri	Softcover	Remaja	TINGGI
2	FIKSI	Drama	Murah	Dalam Negeri	Softcover	Remaja	TINGGI
3	FIKSI	Drama	Murah	Dalam Negeri	Softcover	Remaja	TINGGI
4	FIKSI	Drama	Murah	Dalam Negeri	Softcover	Remaja	TINGGI
5	FIKSI	Drama	Murah	Dalam Negeri	Softcover	Remaja	TINGGI
6	FIKSI	Drama	Murah	Dalam Negeri	Softcover	Remaja	TINGGI
7	FIKSI	Drama	Murah	Dalam Negeri	Softcover	Remaja	TINGGI
8	FIKSI	Drama	Murah	Dalam Negeri	Softcover	Remaja	TINGGI
9	FIKSI	Drama	Murah	Luar Negeri	Softcover	Dewasa	TINGGI
10	FIKSI	Drama	Murah	Luar Negeri	Softcover	Remaja	TINGGI
11	FIKSI	Drama	Murah	Luar Negeri	Softcover	Remaja	TINGGI
12	FIKSI	Drama	Murah	Luar Negeri	Softcover	Remaja	TINGGI
13	FIKSI	Drama	Murah	Luar Negeri	Softcover	Remaja	TINGGI
14	FIKSI	Drama	Murah	Luar Negeri	Softcover	Remaja	TINGGI
15	FIKSI	Drama	Murah	Luar Negeri	Softcover	Remaja	TINGGI
16	FIKSI	Drama	Murah	Luar Negeri	Softcover	Remaja	TINGGI
17	FIKSI	Drama	Murah	Luar Negeri	Softcover	Remaja	TINGGI
18	FIKSI	Drama	Murah	Luar Negeri	Softcover	Dewasa	TINGGI

Tabel 7. Potongan data uji

1	Kategori Buku	Kategori genre	Harga	Kategori Sampul	Kategori Penerbit	Kategori Usia	Rating
2	Fiksi	Drama	150.000	Softcover	Dalam Negeri	Remaja	?

Menggunakan data pelatihan yang ada maka dapat dilakukan prediksi terhadap data uji. Sistem akan melakukan perhitungan dengan menggunakan Naïve Bayes, untuk mencari nilai confidence atau HMAP terbesar pada variabel Y untuk setiap atribut. HMAP adalah nilai yang dapat digunakan sebagai dasar dalam melakukan prediksi (Nugroho & Haryati, 2015).

Perhitungan data uji berdasarkan input dari user seperti potongan pada tabel 7.

1) Menghitung jumlah data *training*

P(Y) = 500

P(Y = Tinggi) = 348 / 500 = 0,696

P(Y = Rendah) = 152 / 500 = 0,304

2) Menghitung nilai probabilitas masing-masing variabel X terhadap variabel Y

P(X1 = Fiksi | Y = Tinggi) = 180 / 348 = 0,5172
P(X2 = Drama | Y = Tinggi) = 55 / 348 = 0,1580
P(X3 = 150.000 | Y = Tinggi) = 89 / 348 = 0,2557
P(X4 = Softcover | Y = Tinggi) = 254 / 348 = 0,7298
P(X5 = Dalam Negeri | Y = Tinggi) = 34 / 348 = 0,0977
P(X6 = Remaja | Y = Tinggi) = 171 / 348 = 0,4913

P(X1 = Fiksi | Y = Rendah) = 65 / 152 = 0,4276
P(X2 = Drama | Y = Rendah) = 17 / 152 = 0,1118
P(X3 = 150.000 | Y = Rendah) = 65 / 152 = 0,4276
P(X4 = Softcover | Y = Rendah) = 78 / 152 = 0,5131
P(X5 = Dalam Negeri | Y = Rendah) = 5 / 152 = 0,0328
P(X6 = Remaja | Y = Rendah) = 88 / 152 = 0,5789

3) Menghitung nilai HMAP

P (PP = Tinggi) = P (X|Y) . P(Y= Tinggi)
= 0,0007319

P (PP = Rendah) = P(X|Y) . P(Y= Rendah)
= 0, 0001992

Hasil perhitungan dengan menggunakan rumus HMAP, menghasilkan nilai probabilitas tertinggi pada P (Rating = Tinggi).

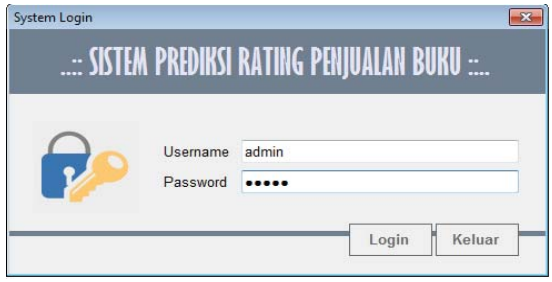
Selanjutnya dilakukan evaluasi dengan mencari nilai akurasi, presisi dan recall dengan bantuan software rapidminer. Hasil dapat dilihat pada tabel 8.

Tabel 8. Hasil pengukuran akhir

Nilai	Hasil
Presisi	80,61 %
Recall	83,62 %
Akurasi	74,60 %

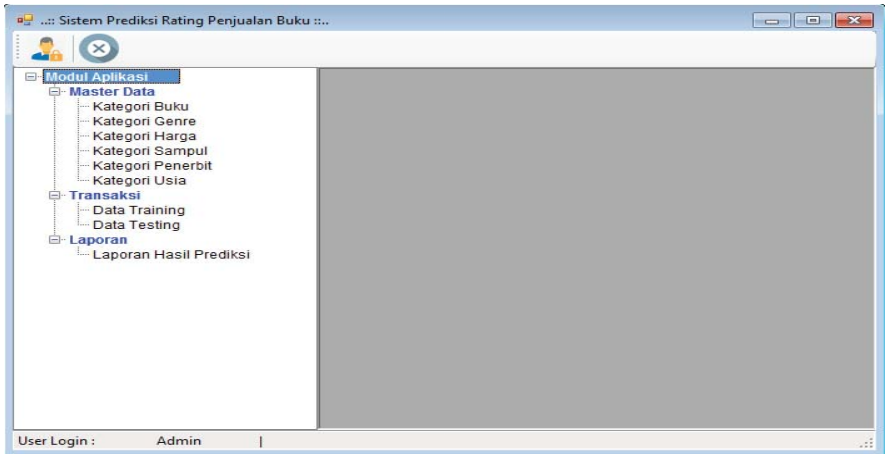
4.1 Tampilan Sistem

Tampilan utama ketika user membuka sistem prediksi ini yaitu berupa menu login untuk masuk sebagai Admin atau Operator. Seperti yang ditampilkan pada gambar 3.



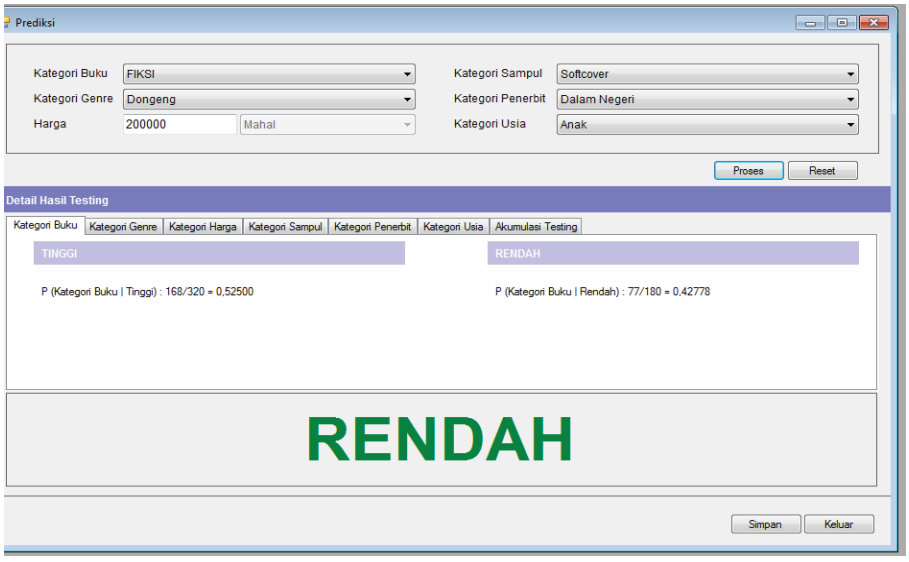
Gambar 3. Halaman Login

Gambar 4 merupakan halaman admin yang terdapat beberapa menu utama untuk mengelola data dan melakukan pengujian pada sistem prediksi ini, yang meliputi master data, transaksi dan laporan. Pada menu master data terdapat 5 submenu kategori sesuai dengan variabel x yang ada. Dimana admin dapat menambah, mengedit dan menghapus master data sesuai variabel yang dibutuhkan. Master data dimaksudkan untuk menambah data utama jika nantinya ada perubahan-perubahan data, seperti penambahan genre buku, perubahan range harga, perubahan range usia dan perubahan range rating.



Gambar 4. Halaman Admin

Menu transaksi pada halaman admin berisi menu data *training* dan data *testing*. Dimana admin dapat menambah, mengedit dan menghapus data *training*. Serta menu data *testing* yang dilakukan untuk melakukan prediksi. Pada menu data *testing*, terdapat beberapa tab menu yang berisi perhitungan dari masing-masing variabel yang dijadikan acuan dalam prediksi beserta total akumulasi dari perhitungan tiap variabel, yang ditampilkan dalam gambar 5.



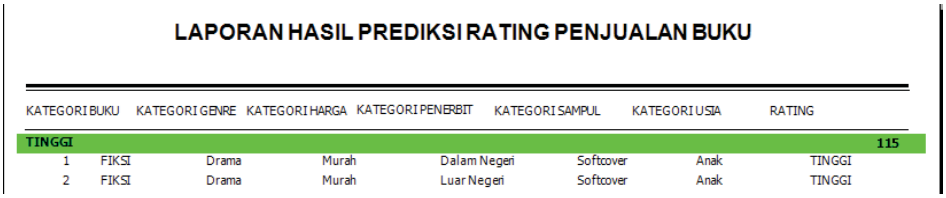
Gambar 5. Halaman Prediksi

Selanjutnya adalah menu laporan, menu ini berisi laporan data yang telah diprediksi dan disimpan. Dari 180 data uji, diperoleh 65 data dengan hasil rendah tampak pada gambar 6. dan 115 data dengan hasil tinggi tampak pada gambar 7.

LAPORAN HASIL PREDIKSI RATING PENJUALAN BUKU

KATEGORI BUKU	KATEGORI GENRE	KATEGORI HARGA	KATEGORI PENERBIT	KATEGORI SAMPUL	KATEGORI USIA	RATING	
RENDAH							**65**
1	FIKSI	Drama	Sedang	Luar Negeri	Softcover	Anak	RENDAH
2	FIKSI	Drama	Sedang	Dalam Negeri	Hardcover	Anak	RENDAH
3	FIKSI	Komik	Sedang	Luar Negeri	Softcover	Anak	RENDAH

Gambar 6. Potongan laporan hasil prediksi rendah

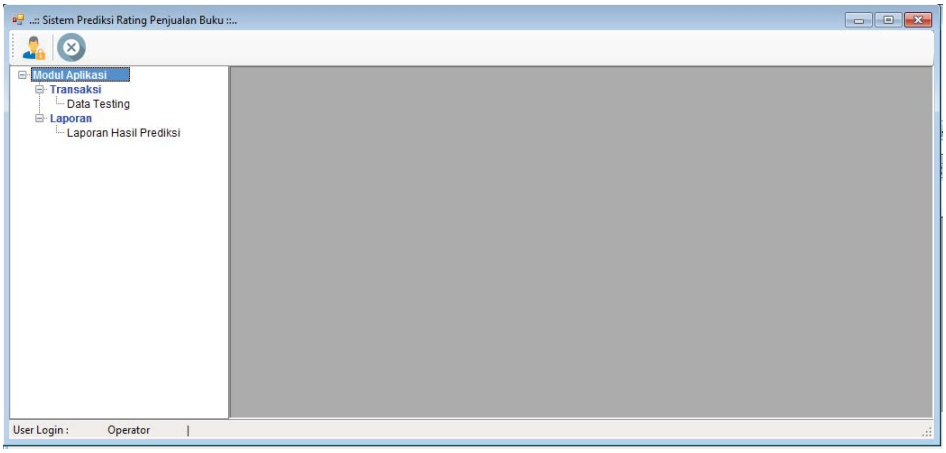


LAPORAN HASIL PREDIKSI RATING PENJUALAN BUKU

KATEGORI BUKU	KATEGORI GENRE	KATEGORI HARGA	KATEGORI PENERBIT	KATEGORI SAMPUL	KATEGORI USIA	RATING	
TINGGI							**115**
1	FIKSI	Drama	Murah	Dalam Negeri	Softcover	Anak	TINGGI
2	FIKSI	Drama	Murah	Luar Negeri	Softcover	Anak	TINGGI

Gambar 7. Potongan laporan hasil prediksi tinggi

Halaman operator berisi menu utama yaitu menu transaksi dan laporan. Dalam menu transaksi terdapat submenu data *testing* yang digunakan untuk melakukan prediksi, yang ditampilkan dalam gambar 8.



Gambar 8. Halaman Admin

V. KESIMPULAN

Dari hasil penelitian yang telah dilakukan, dapat diambil kesimpulan bahwa:

1. Sistem data mining menggunakan metode naïve bayes untuk memprediksi rating penjualan buku memiliki kinerja yang baik dari segi akurasi, presisi dan recall. Dengan nilai akurasi sebesar 74,60%.
2. Berdasarkan pengujian yang dilakukan terhadap 180 data uji dalam memprediksi rating penjualan buku, metode ini dinilai berhasil karena memiliki nilai dengan hasil tinggi sebanyak 115 data, hasil dapat dilihat dalam menu laporan.

DAFTAR PUSTAKA

Bhardwaj, A., Sharma, A., & Shrivastava, V. K. (2012). Data Mining Techniques and Their Implementation in Blood Bank Sector – A Review. *International Journal of Engineering Research and Applications (IJERA)*, *2*(4), 1303–1309.

Chandrasekar, P., & Qian, K. (2016). The Impact of Data Preprocessing on the Performance of a Naive Bayes Classifier. *IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, 618–619. http://doi.org/10.1109/COMPSAC.2016.205

Djuhro, M. (2016). Dunia Penerbitan dan Proses Lahirnya Sebuah Buku. *Al-Maktabah*, *1*(2), 39–55. Retrieved from http://respository.uinjkt.ac.id/dspace/handle/123456789/31046

Gewati, M. (2016). Minat Baca Indonesia Ada di Urutan ke-60 Dunia. Retrieved from http://edukasi.kompas.com/read/2016/08/29/07175131/minat.baca.indonesia.ada.di.urutan.ke-60.dunia.

Imadha, H. (2014). Bahasa Macam-Macam Prediksi. Retrieved from http://abajakarta.wordpress.com/2014/02/15/bahasa-macam-macam-prediksi/.

Jananto, A. (2013). Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa. *Teknologi Informasi DINAMIK*, *18*(1), 9–16.

Nugroho, Y. S., & Haryati, S. N. (2015). Klasifikasi dan Klastering Penjurusan Siswa SMA Negeri 3 Boyolali. *Khazanah Informatika*, *I*(1), 3–8.

Probo, V., & Afrisia, R. S. (2015). Penjualan Buku Agama dan Anak Kalahkan Novel. Retrieved from http://m.cnnindonesia.com/hiburan/20150706063457-241-64528/penjualan-buku-agama-dan-anak-kalahkan-novel/.

Rahmat, A., & Suryadi, Y. (2008). Pemrograman Visual Basic.Net 2005 dan Aplikasi Nyata dengan DBMS SQL Server 2005. Bandung: Yrama Widya.

Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *EECCIS*, *7*(1), 59–64.

Wang, X., Sheng, B., Xue, L., & Xiao, Z. (2016). Classification of Customer Requirements on Map Reduce-based Naive Bayes. *IEEE International Conference on Big Data Analysis (ICBDA)*, 1–4.