

PREDIKSI BESARNYA PENGELUARAN BIAYA RUMAH TANGGA DENGAN METODE KLASIFIKASI NAÏVE BAYES

Yuniar Andi Astuti

Manajemen Informatika, Politeknik Unggul LP3M Medan

Abstrak

Pengeluaran biaya rumah tangga sangat penting bagi setiap lapisan masyarakat agar kehidupan manusia dapat berjalan dengan baik dan lancar. Setiap keluarga tentunya memiliki kebutuhan dan pengeluaran yang berbeda dan tentunya harus disesuaikan dengan pendapatan.. Oleh karena itu setiap rumah tangga haruslah paham dalam mengelolah pendapatan dan pengeluaran yang efektif. Penerapan metode naïve bayes diharapkan mampu untuk memprediksi besarnya pengeluaran biaya tiap rumah tangga agar lebih mudah mengatur pengeluaran biaya rumah tangga. dari 60 data rumah tangga yang diuji dengan metode naïve bayes, maka diperoleh hasil persentase 85% untuk keakuratan prediksi, di mana dari 60 data pengeluaran biaya rumah tangga yang diuji terdapat 51 data pengeluaran biaya rumah tangga yang berhasil diklasifikasikan dengan benar.

Kata Kunci — *Data Mining, Naïve Bayes, Pengeluaran biaya rumah tangga*

I. PENDAHULUAN

Rumah tangga tentu memiliki sejumlah kebutuhan yang kompleks, di mana harus dapat memenuhi semua kebutuhan seluruh anggota keluarga dengan sangat tepat. Bukan hanya diri sendiri saja, namun anggaran ini juga harus memenuhi semua kebutuhan pasangan serta anak-anak dengan tepat, sehingga rumah tangga juga bisa berjalan dengan baik. Kotler (1996) bahwa kebutuhan rumah tangga adalah suatu keadaan yang timbul oleh rasa kekurangan terhadap suatu hal, bila tidak dipenuhi akan menimbulkan deprivasi yaitu kualitas hidup yang dibawah kewajaran, secara umum kebutuhan rumah tangga meliputi: a) Kebutuhan fisik yaitu makanan, pakaian dan keamanan, b) Kebutuhan sosial yaitu memiliki seseorang dan kasih sayang, c) Kebutuhan individual yaitu pengetahuan, dan kemampuan mengekspresikan diri. Ketika kebutuhan ini muncul, maka rumah tangga mencari objek yang dapat memuaskan kebutuhan, namun jika hal itu tidak ditemukan maka rumah tangga akan berusaha mengurangi tingkat kebutuhannya dengan menggunakan apa yang dalam jangkauannya. Ernest Engel (1857 dalam BPS, 2014) bahwa persentase pengeluaran untuk makan akan menurun sejalan dengan meningkatnya pendapatan. Oleh karena itu komposisi pengeluaran rumah tangga dapat dijadikan sebagai indikator untuk kesejahteraan penduduk. Semakin rendah persentase pengeluaran untuk makanan terhadap total pengeluaran, maka semakin baik tingkat perekonomian penduduk. Perbedaan tingkat pendapatan akan mengakibatkan perbedaan pola distribusi pendapatan termasuk pola konsumsi rumah tangga. Dalam kondisi terbatas (pendapatan kecil), maka seseorang akan mendahulukan pemenuhan kebutuhan makanan dan sebagian besar pendapatan tersebut dibelanjakan untuk konsumsi makanan. Semakin rendah pangsa pengeluaran pangan, berarti tingkat kesejahteraan masyarakat semakin baik (Ariani et al., 2007).

Untuk memenuhi kebutuhan, rumah tangga keluarga harus mengeluarkan dana sebagai pengorbanan untuk memperoleh barang atau jasa. Besar kecil pengeluaran tergantung pada macam, banyak, dan tingkat harga barang atau jasa yang dibutuhkan. Pengeluaran rumah tangga antara lain untuk: membeli makanan, minuman, dan pakaian; membayar rekening listrik dan telepon; membayar langganan surat kabar dan majalah; membayar iuran televisi; biaya transportasi; kepenluan sekolah; dan sumbangan sosial. Rumah tangga keluarga yang rasional (pemenuhan kebutuhan berdasarkan akal sehat) harus selektif dalam menentukan barang atau jasa yang akan dibeli. Dalam memilih barang atau jasa, keluarga perlu selalu memperhatikan kepentingan seluruh anggota keluarga dan dana yang tersedia.

Data Mining

Data Mining merupakan proses pengekstraksian informasi dari sekumpulan data yang sangat besar melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data. *Data mining* adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Definisi lain mengatakan *Data Mining* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk

menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Dari beberapa definisi di atas dapat ditarik kesimpulan bahwa *Data Mining* merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi – informasi yang nantinya dapat digunakan.

Tahap-tahap Data Mining

Sebagai suatu rangkaian proses, *Data Mining* dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge base*.

Tahap-tahap *Data Mining* adalah sebagai berikut:

- a. Pembersihan data (Data Cleaning)
Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.
- b. Integrasi data (Data Integration)
Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.
- c. Seleksi data (Data Selection)
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.
- d. Transformasi data (Data Transformation)
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *Data Mining*.
- e. Proses Mining
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan *Data Mining*.
- f. Evaluasi pola (Pattern Evaluation)
Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.
- g. Presentasi pengetahuan (Knowledge Presentation)
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

Metode Naive Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya.

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. *Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.*

II. METODE PENELITIAN

Adapun metode penelitian yang diterapkan pada penelitian ini adalah sebagai berikut:

1. Analisis Masalah dan Studi Literatur

Tahap ini adalah langkah awal untuk menentukan rumusan masalah dari penelitian. Dalam hal ini mengamati permasalahan yang berhubungan dengan faktor-faktor penyebab besarnya pemakaian listrik rumah tangga. Permasalahan-permasalahan yang ada, selanjutnya dianalisa untuk mengetahui bagaimana cara penyelesaian terhadap masalah tersebut dan menentukan ruang lingkup permasalahan yang akan diteliti. Mempelajari dasar teori dari berbagai literature mengenai penerapan metode Naive Bayes, konsep dan teori data mining dan pengeluaran biaya rumah

tangga, melalui jurnal-jurnal dan agar mendapatkan dasar pengetahuan untuk melakukan penelitian selanjutnya.

2. Mengumpulkan Data

Prosedur sistematis yang digunakan untuk mengumpulkan data yaitu dengan mengajukan pertanyaan-pertanyaan kepada responden. Hasil kuesioner yang didapatkan akan digunakan untuk bahan penganalisisan data terhadap metode Naïve Bayes. Setelah data dikumpulkan dilakukan analisa data untuk menyesuaikan proses data yang akan diolah pada metode Naïve Bayes.

3. Implementasi dan Pengujian

Sesuai dengan pengolahan data maka pada tahap implementasi adalah tentang bagaimana pengolahan datanya diterapkan dalam sebuah tools. Tools yang akan digunakan dalam implementasi penelitian ini adalah dengan menggunakan Software Weka. Selanjutnya Pengujian dilakukan untuk mengetahui apakah penelitian yang dilakukan telah sesuai dengan tujuan yang diharapkan.

III. HASIL DAN PEMBAHASAN

Penerapan Metode Naïve Bayes

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. *Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Dalam metode Naive Bayes data String yang bersifat konstan dibedakan dengan data numerik yang bersifat kontinyu, perbedaan ini akan terlihat pada saat menentukan nilai probabilitas setiap kriteria baik itu kriteria dengan nilai data string maupun kriteria dengan nilai data numerik. Adapun penerapan metode Naive Bayes sebagai berikut.*

Baca Data Training

Untuk menentukan data yang nantinya akan dianalisis dengan metode *Naive Bayes* maka langkah pertama yang dilakukan adalah membaca data latih. Adapun data latih yang digunakan dapat dilihat pada tabel berikut:

Tabel 1. Data Training

No	Jumlah Pendapatan	Jml Tanggungan	Biaya RT	Biaya Makan	Biaya Sekolah
1	Tinggi	Banyak	Tinggi	Besar	Tinggi
2	Tinggi	Banyak	Tinggi	Besar	Sedang
3	Tinggi	Banyak	Tinggi	Standar	Tinggi
4	Tinggi	Banyak	Sedang	Standar	Tinggi
5	Sedang	Kecil	Kecil	Standar	Sedang
6	Sedang	Kecil	Kecil	Standar	Sedang
7	Sedang	Kecil	Sedang	Standar	Sedang
8	Sedang	Kecil	Sedang	Sedang	Sedang
9	Sedang	Kecil	Sedang	Sedang	Rendah
10	Sedang	Kecil	Sedang	Standar	Tinggi
11	Sedang	Kecil	Kecil	Standar	Tinggi
12	Sedang	Sedang	Kecil	Sedang	Sedang
13	Rendah	Kecil	Kecil	Kecil	Sedang
14	Sedang	Sedang	Sedang	Sedang	Sedang
15	Tinggi	Banyak	Sedang	Sedang	Tinggi
60	Sedang	Kecil	Kecil	Kecil	Sedang

Kriteria dan Probabilitas

Adapun nilai probabilitas setiap kriteria didapatkan dari data latih pada tabel 1. Adapun nilai probabilitas setiap kriteria sebagai berikut.

Probabilitas Kriteria Jumlah Pendapatan

Berdasarkan data training pada tabel 1, diketahui jumlah data latih sebanyak 60 data, dimana dari 60 data tersebut terdapat 5 keluarga dengan jumlah pendapatan tinggi dan jumlah tanggungan kecil, 2 keluarga dengan jumlah pendapatan tinggi dan tanggungan sedang, 5 keluarga dengan jumlah pendapatan tinggi dan tanggungan banyak kemudian 30 keluarga dengan jumlah pendapatan sedang

dan tanggungan kecil, 9 keluarga dengan jumlah pendapatan dan tanggungan sedang, 1 keluarga dengan jumlah pendapatan sedang dan tanggungan banyak, dan 8 keluarga dengan jumlah pendapatan rendah dan jumlah tanggungan kecil.

Tabel 2. Probabilitas Kriteria Jumlah Pendapatan

Jumlah Pendapatan	Jumlah Tanggungan			Probabilitas		
	Kecil	Sedang	Banyak	Kecil	Sedang	Banyak
Tinggi	5	2	5	0.11	0.18	0.83
Sedang	30	9	1	0.70	0.80	0.17
Rendah	8	0	0	0.18	0.00	0.00
Jumlah	43	11	6	0.72	0.18	0.10

Probabilitas Kriteria Biaya Rumah Tangga

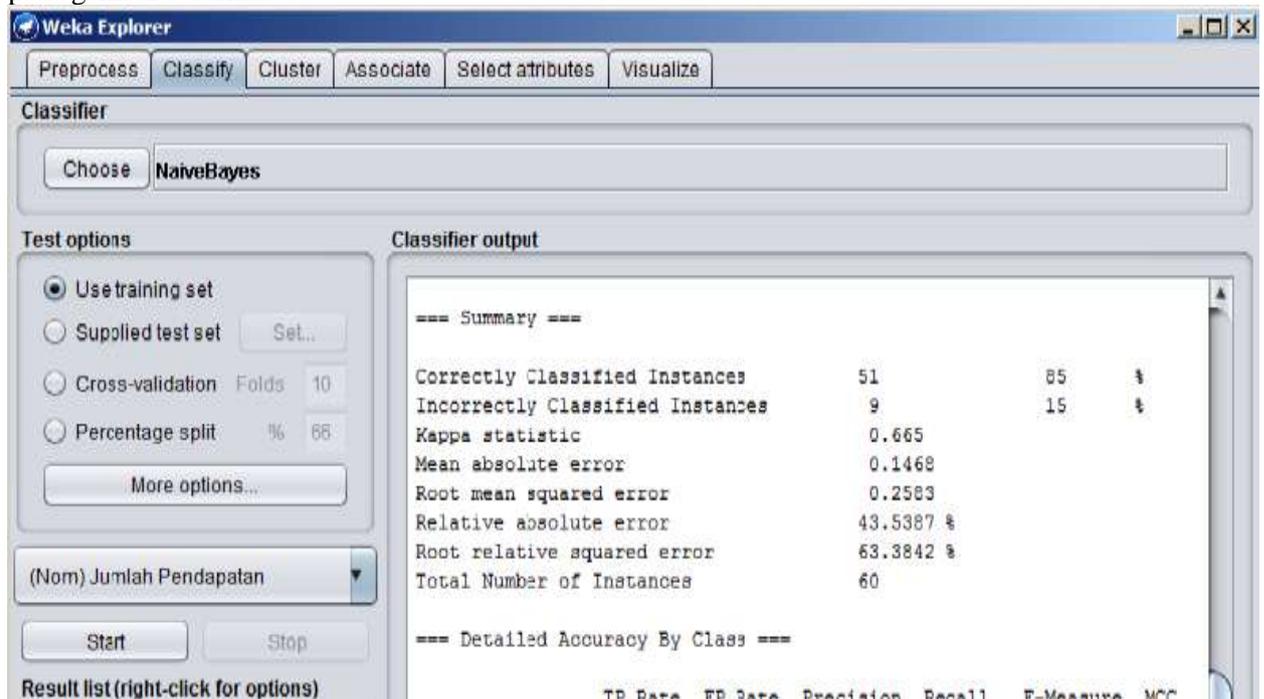
Berdasarkan data training pada tabel 1, diketahui jumlah data latih sebanyak 60 data, dimana dari 60 data tersebut terdapat 3 keluarga dengan jumlah pendapatan rendah dan biaya rumah tangga tinggi, 6 keluarga dengan jumlah pendapatan dan biaya rumah tangga tinggi, 1 keluarga dengan jumlah pendapatan rendah dan biaya rumah tangga sedang, 23 keluarga dengan jumlah pendapatan dan biaya rumah tangga sedang, 6 keluarga dengan jumlah pendapatan tinggi dan biaya rumah tangga sedang, 4 keluarga dengan jumlah pendapatan rendah dan biaya rumah tangga kecil dan 17 keluarga dengan jumlah pendapatan sedang dan biaya rumah tangga kecil.

Tabel 2. Probabilitas Kriteria Biaya Rumah Tangga

Biaya Rumah T.	Jumlah Pendapatan			Probabilitas		
	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
Tinggi	3	0	6	0.37	0.00	0.50
Sedang	1	23	6	0.13	0.57	0.50
Kecil	4	17	0	0.50	0.43	0.00
Jumlah	8	40	12	0.14	0.66	0.20

Pengujian Metode Naïve Bayes

Dari nilai probabilitas di atas akan diuji sebanyak 60 data dan diselesaikan dengan menggunakan tools weka sehingga dihasilkan hasil klasifikasi biaya pengeluaran rumah tangga seperti pada gambar 3 berikut:



Gambar 3. Hasil Klasifikasi Metode Naïve Bayes

Berdasarkan gambar 3 di atas dapat dilihat persentase untuk *Correctly Classified Instance* adalah sebesar 85% sementara persentase untuk *Incorrectly Classified Instance* adalah sebesar 15%.

Di mana dari 60 data pengeluaran biaya rumah tangga, ada sebanyak 51 data pengeluaran biaya rumah tangga berhasil diklasifikasikan dengan benar dan sebanyak 9 data tidak berhasil diklasifikasikan dengan benar.

IV. KESIMPULAN

Berdasarkan penelitian tentang memprediksi biaya pengeluaran rumah tangga dapat ditarik beberapa kesimpulan sebagai berikut :

1. Metode *Naive Bayes* memanfaatkan data *training* untuk menghasilkan probabilitas setiap kriteria untuk *class* yang berbeda, sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk memprediksi biaya pengeluaran rumah tangga berdasarkan proses klasifikasi yang dilakukan oleh metode *Naive Bayes* itu sendiri.
2. Berdasarkan data rumah tangga yang dijadikan data *training*, metode *Naive Bayes* berhasil mengklasifikasikan 51 data dari 60 data yang diuji. Sehingga metode *Naive Bayes* berhasil memprediksi besarnya pengeluaran biaya rumah tangga dengan persentase keakuratan sebesar 85%.

V. DAFTAR PUSTAKA

- Manjusha K. K., Sankaranarayanan, K., Seena P., 2014, Prediction of Different Dermatological Conditions Using Naive Bayesian Classification, *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol 4, No 1, Hal 864- 868.
- Ting, S. L., Ip, W. H., Tsang, A. H.C., 2011, Is Naive Bayes a Good Classifier for Document Classification?, *International Journal of Software Engineering and Its Applications*, Vol. 5, No. 3, Hal 37-46.
- Taruna R., S., Hiranwal, S., 2013, Enhanced Naive Bayes Algorithm for Intrusion Detection in Data Mining, *International Journal of Computer Science and Information Technologies*, Vol.6, No. 4, Hal 960-962.
- Mabrur, A. G., Lubis, R., 2012, Penerapan Data Mining untuk Memprediksi Kriteria Nasabah Kredit, *Jurnal Komputer dan Informatika (KOMPUTA)*, Vol.1, No. 1, Hal 53-57.
- Mujiasih, S., 2011, Pemanfaatan Data Mining Untuk Prakiraan Cuaca, *Jurnal Meteorologi dan Geofisika*, Vol. 12, No. 2, Hal 189 - 195.
- Ridwan, M., Suyono, H., Sarosa, M., 2013, Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier, *Jurnal EECCIS*, Vol 1, No. 7, Hal. 59-64.
- Patil, T. R., Sherekar, M. S., 2013, Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification, *International Journal of Computer Science and Applications*, Vol. 6, No. 2, Hal 256-261.
- Bustami., 2013, Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi, *TECHSI : Jurnal Penelitian Teknik Informatika*, Vol. 3, No.2, Hal. 127-146.
- Pattekari, S. A., Parveen, A., 2012, Prediction System for Heart Disease Using Naive Bayes, *International Journal of Advanced Computer and Mathematical Sciences*, ISSN 2230-9624, Vol. 3, No 3, Hal 290-294.