

PREDIKSI PENJUALAN KARTON DUS SUSU CHIL MIL DENGAN PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA METODE C4.5

U. Darmanto Soer¹, Mustijah²

Program Studi Teknik Informatika Fakultas Teknik
Universitas Pelita Bangsa
darmanto@pelitabangsa.ac.id

Disetujui, 30 Desember 2019

Abstraksi

Penjualan merupakan hal yang sangat penting di dalam perusahaan, namun tidak mudah bagi marketing untuk mengetahui seberapa laku tidaknya produk dalam perusahaan setiap periode dengan begitu banyak data yang ada. Selama ini banyak data yang besar di sebuah perusahaan khususnya dalam bidang penjualan yang belum dapat diolah dimaksimalkan dengan baik, Jika data tersebut dapat dimanfaatkan maka dapat digunakan untuk membantu marketing dalam penjualan produk. Panggilan data dalam jumlah yang besar biasa disebut data mining. Data ini diuji dengan dua algoritma yaitu decision tree dan prediksi. Penelitian ini bertujuan mengetahui hasil penjualan setiap tahun produk apa yang laku dipasaran dan memprediksi penjualan. Dari pengujian ini memperoleh dengan hasil akurasi 90,29% precision mendapatkan nilai 91,67%, Recall 88%. Hasil dari perhitungan rapidminer adalah dengan prediksi lebih dari 19.500 lembar laku dan kurang dari 1.500 lembar tidak laku.

Kata Kunci : Data Mining, Prediksi, dan Algoritma C4.5.

Abstract

Sales are very important in the company, but it is not easy for marketing to find out how well the products in the company sell every year with so much data. During this time a lot of big data in a company, especially in the field of sales that have not been able to be maximized properly if the data can be utilized then it can be used to assist marketing in product sales. Large amounts of data calls are usually called data mining. This data was tested with two algorithms, namely decision tree and prediction. This study aims to determine the results of sales every year what products sell in the market and predict sales. From this test the accuracy of 90,29 % results obtained 91,67 % precision and 88 % recall . with predictions of more than 19.500 pieces of cardboard and less than 1.500 pieces of cardboard unsold. The results of the rapidminer calculation are with predictions of more than 19.500 salable sheets and less than 1.500 unsold sheets.

Keywords: Data Mining, Prediction, and Algorithm C4.5.

1. Pendahuluan

Perkembangan Teknologi Informasi dan Komunikasi di era sekarang ini semakin canggih dan modern, dan kebutuhan akan sebuah informasi semakin meluas dan bersifat komputerisasi yang bertujuan untuk mempermudah kinerja para pemakainya. Kemampuan untuk diakses kapan saja dan dimana saja merupakan salah satu alasan sistem ini berkembang dengan pesat. Penggunaan teknologi informasi dalam sebuah instansi atau perusahaan saat ini sangat penting karena dapat memudahkan sebuah instansi atau perusahaan untuk melakukan pengolahan data untuk menjadi sebuah informasi yang dapat digunakan oleh penggunanya. Dalam dunia bisnis yang dinamis dan penuh persaingan seiring dengan pertumbuhan bisnis, perusahaan tidak lagi bisa unggul hanya dengan memanfaatkan teknologi yang ada. Karena teknologi tersebut digunakan hanya untuk mendukung kegiatan sehari-hari perusahaan.

Persaingan yang terjadi dalam dunia bisnis memaksa para pelakunya untuk selalu memikirkan strategi-strategi dan terobosan yang dapat menjamin kelangsungan dari bisnis yang di jalankannya. Database dalam jumlah yang besar merupakan salah satu aset berharga yang dimiliki sebuah perusahaan. Sebagai salah satu perusahaan yang bergerak dibidang percetakan, pihak *marketing* PT. SERUNIGRAF haruslah memikirkan strategi dalam pemasaran untuk mampu menghasilkan suatu informasi yang siap digunakan untuk membantu pihak *marketing* dalam mengambil keputusan strategi pemasaran. Pihak

marketing perusahaan tersebut ingin mengetahui produk apa yang harus di kembangkan, Seberapa jauh pencapaian yang telah dicapai oleh perusahaan, dan pasar apa yang harus di fokuskan. Untuk memenuhi kebutuhan-kebutuhan diatas, banyak cara yang dapat ditempuh salah satunya adalah dengan melakukan pemanfaatan data base perusahaan menggunakan teknik Data Mining. Data Mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Ketersediaan data yang banyak dan dibutuhkan akan informasi atau pengetahuan sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis dan dukungan infrastruktur dibidang teknik informatika merupakan cikal bakal dari lahirnya teknologi data mining. Sehingga informasi tersebut bisa digunakan sebagai solusi pengambilan keputusan didunia bisnis, untuk pengembangan bisnis.

Pada PT.SERUNIGRAF terdapat beberapa permasalahan yang sering muncul mengenai penjualan produk. Perusahaan sulit mendapatkan informasi-informasi strategis seperti tingkat penjualan per periode laku tidaknya produk, dan penjualan produk yang dihasilkan. Ketersediaan data penjualan yang besar di database *server* tidak digunakan semaksimal mungkin, sehingga data penjualan tersebut hanya digunakan untuk kegiatan sehari-hari. Dari permasalahan diatas peneliti akan melakukan penelitian dalam bentuk laporan penelitian dengan judul. **“PREDIKSI PENJUALAN KARTON DUS SUSU CHIL MIL DENGAN PENERAPAN DATA MINING MENGGUNAKAN METODE ALGORITMA C.4.5”**.

2. Tinjauan Pustaka

2.1 Data mining

Data mining adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. Istilah lain yang sering digunakan diantaranya *knowledge discovery (mining) in databases (KDD)*, *knowledge extraction*, *data /pattern analysis*, *data archeology*, *data dredging*, *information harvesting*, dan *business intelligence* [1].

Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. Data mining merupakan proses besar. Teknik data mining digunakan untuk dapatkan korelasi antar produk yang sebenarnya memeriksa basis data berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna. Tidak semua pekerjaan pencarian informasi dinyatakan sebagai data mining. Sebagai contoh, pencarian *record individual* menggunakan database *management system* atau pencarian halaman web tertentu melalui *kueri* ke semua *search engine* adalah pekerjaan pencarian informasi yang erat kaitannya dengan *information retrieval*.

2.2 Tahapan Data Mining

Tahapan data mining dibagi menjadi lima bagian yaitu :

1. Pembersihan data (*data cleaning*) Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang 4 duplikasi data, memeriksa data yang *konsisten*, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*). Juga dilakukan proses *enrichment*, yaitu proses —memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD seperti data atau informasi *eksternal* [2].
2. Integrasi data (*data integration*) Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu data base baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada *atribut- atribut* yang mengidentifikasi *entitas -entitas* yang unik seperti *atribut* nama, jenis produk, nomor pelanggan dan lainnya. *Integrasi* data perlu dilakukan secara cermat karena kesalahan pada *integrasi* data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila *integrasi* data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan di tidak ada [2].
3. Proses mining adalah sebuah proses yang paling utama pada saat metode diterapkan untuk mencari pengetahuan tersembunyi dan berharga dari data [2].
4. Evaluasi pola (*pattern evaluation*), untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah *hipotesa* yang ada memang tercapai [2].
5. Presentasi pengetahuan (*knowledge presentation*), merupakan penyajian dan visualisasi pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau lima aksi dari hasil analisis yang di dapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data

mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining.[2].

3. Metode Penelitian

Pada penelitian ini data yang digunakan adalah data penjualan karton dus susu chil mil Perbulan pada PT.SERUNIGRAF. Data tersebut akan diolah menggunakan data mining sehingga diperoleh satu metode yang akurat yang digunakan dalam prediksi.

1. Observasi

Yaitu teknik pengumpulan data yang dilakukan dengan mengadakan pengamatan langsung terhadap aktifitas sumber daya manusia yang ada pada perusahaan. Atau dengan kata lain suatu teknik mencari atau mengumpulkan data dengan jalan mengamati dan terjun langsung kelapangan dengan segala aspek yang berhubungan langsung dengan penelitian.

2. Wawancara

Yaitu teknik pengumpulan data dengan mengadakan tanya jawab dengan responden dan pihak-pihak yang ada kaitannya dengan masalah yang diteliti.

3. Daftar Pustaka

Sumber dan data sekunder sebagai pelengkap data primer yaitu mengkaji, mengutip, dan mempelajari berbagai jenis buku, jurnal, dan artikel dari internet yang berhubungan dengan permasalahan yang diteliti, dimana teori-teori yang digunakan dijadikan sebagai referensi dalam penyusunan penelitian tersebut.

4. Browsing

Peneliti juga mengutip dan mencari artikel-artikel dengan browsing internet sebagai membantu dan referensi dalam penyusunan penelitian.

3.1 Tahapan Penelitian

Untuk melakukan sebuah penelitian, diperlukan adanya tahapan-tahapan yang tersusun dengan baik dan sistematis agar pelaksanaan penelitian tepat tercapai tujuan yang diharapkan. Maka dari itu disusunlah tahapan-tahapan metodologi dalam penelitian ini yaitu :

1. Data Selection

Pemilihan atau seleksi data dari sekumpulan data operasional. Data hasil seleksi akan digunakan untuk proses *data mining*, dan disimpan dalam suatu berkas dan terpisah dari basis data operasional. Data yang digunakan dalam penelitian ini adalah data set penjualan karton dus susu chil mil yang diperoleh dari tarikan data yang dikeluarkan melalui aplikasi penjualan pada PT. SERUNIGRAF.

Data tersebut kemudian disalin dalam sebuah file dengan tipe *xls*,diperoleh jumlah tarikan data sebanyak kurang lebih 700 jenis data penjualan. Data yang diperoleh terdiri dari, normalisasi penjualan, nama tipe karton, jumlah Orderan *customer*,jumlah Orderan yang dikirim ke *customer*, keterangan target orderan.

2. Cleaning Data

Proses cleaning dilakukan dengan menghilangkan noise, membuang duplikasi data, memeriksa data yang kesalahan pada data, seperti kesalahan cetak (tipografi). Dan juga pada data set penjualan yang diperoleh dari tarikan data yang dikeluarkan melalui aplikasi penjualan, terdapat beberapa data yang bernilai kosong null. Data kosong ini semua statusnya bernilai — NO —, karena keterbatasan data yang bernilai — YESI maka peneliti memutuskan untuk memakai data yang — *missing value*l.

3. Transformasi Data

Pada tahap ini dilakukan peringkasan data atau proses pengubahan data mentah menjadi data yang mudah dikelola. Data mentah akan dimasukan pada kategorinya masingmasing kemudian data juga dipilih, Sehingga data tersebut sesuai untuk proses *data mining* dan mempermudah jalannya sistem aplikasi RapidMiner sebagai berikut:

4. Hasil Penelitian dan Pembahasan

Metode penelitian yang peneliti lakukan menghasilkan informasi dan datayang digunakan sebagai acuan peneliti sebagai dasar penelitian. Data penjualan PT.SERUNIGRAF merupakan hasil pembukuan dari produk yang sudah terjual selama 1 tahun terakhir.Data penjualan ini nantinya akan di prediksi yang bertujuan untuk mempermudah proses perhitungan dengan algoritma decision tree. Dan dari perhitungan tersebut akan menghasilkan aturan pohon keputusan yang akan dipergunakan untuk memprediksi penjualan sebuah produk pada PT.SERUNIGRAF. Penelitian ini menggunakan semua data dari data penjualan yang ada di PT.SERUNIGRAF.

Tujuan dari prediksi adalah untuk menemukan model dari training set yang membedakan atribut ke dalam kategori atau kelas yang sesuai model tersebut kemudian digunakan untuk memprediksi atribut yang kelasnya belum diketahui sebelumnya. Hasil dari dataset penjualan yang telah didapat, diolah dan dilakukan klasifikasi guna untuk melakukan perhitungan.

4.1 Data Pemasaran Karton Dus Susu Chil Mil

Berikut ini adalah data testing yang diambil dari data pemasaran karton dus susu chil mil dengan data testing 700 data dan data training 350 data. Pengujian sistem yang dilakukan dengan data pemasaran menggunakan algoritma C4.5.

Tabel 4.1 Hasil Perhitung Entropy dan Gain

JENIS BARANG	JUMLAH KASUS	LAKU	TIDAK LAKU	KEY	ENTROPY
CHIL MIL RS. COKLAT 200 GR	6	3	3	TIDAK LAKU	1
CHIL MIL RS. VANILA 200 GR	6	5	1	LAKU	0,650022
CHIL MIL RS. COKLAT 400 GR	3	1	2	TIDAK LAKU	0,918296
CHIL MIL RS. VANILA 400 GR	3	1	2	TIDAK LAKU	0,918296
CHIL MIL RS. COKLAT 800 GR	3	1	2	TIDAK LAKU	0,918296
CHIL MIL RS. VANILA 800 GR	6	6	0	LAKU	0
CHIL MIL RS. JERUK 200 GR	5	4	1	LAKU	0,918296
CHIL MIL RS. JERUK 400 GR	8	3	5	TIDAK LAKU	0,503358
CHIL MIL RS. JERUK 800 GR	20	15	5	LAKU	0
CHIL MIL RS. MELON 200 GR	9	7	2	LAKU	0
CHIL MIL RS. MELON 400 GR	10	4	6	TIDAK LAKU	0,970951
CHIL MIL RS. MELON 800 GR	8	6	2	LAKU	0
CHIL MIL RS. MANGGA 400 GR	10	7	3	LAKU	0
CHIL MIL RS. MANGGA 200 GR	11	7	4	LAKU	0
DUPLIX COATED	10	4	6	TIDAK LAKU	0
ART CARTON	17	16	1	LAKU	0,3227557
FULL MERAH	3	0	3	TIDAK LAKU	0
PUTH MERAH	3	1	2	TIDAK LAKU	0
COKLAT HIJAU	3	2	1	LAKU	0
KECIL	18	11	7	LAKU	0
BESAR	9	8	1	LAKU	0

4.2 Perhitungan Menggunakan *Rapidminer*

1. Proses Validation

Proses *validation* yaitu melakukan analisis berbagai model dan memilih model dengan kinerja prediksi yang baik, Setelah pembacaan file data, Blok data testing yang akan diuji kemudian dihubungkan dengan validation. Dari proses validation tersebut dilakukan pengolahan data pada RapidMiner .

2. Proses Training dan Testing

Proses training yaitu proses melakukan penelitian data pada model C4.5, Sedangkan proses testing yaitu melakukan pengujian data yang akan menghasilkan grafik dan pola. Proses training digunakan untuk blok model decision tree, dihubungkan dengan garis penghubung pada blok apply model dan blok performance dengan bagian testing sebagai tampilan informasi hasil dari pengujian data, hasil dari pengujian ini akan menghasilkan arsitektur algoritma C4.5.

3. Accuracy *RapidMiner*

Accuracy adalah Sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. Nilai minimum data yang dihitung accuracy adalah 60%. Nilai *accuracy* dihitung dengan cara membagi jumlah data benar yang bernilai positif (*true positive*) ditambah dengan jumlah benar yang bernilai negatif (*true negative*) dibagi dengan jumlah data benar yang bernilai positif (*true positive*), ditambah dengan jumlah data yang benar yang bernilai negatif (*true negative*), ditambah data yang salah yang bernilai positif (*false positive*) dan data salah yang bernilai negatif (*false negative*).

Accuracy yang dihasilkan dari data pengujian pada rapidminer adalah 90,29%.

4. **Precision RapidMiner**

Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan sistem. Nilai yang minimum precision adalah 60 %. Nilai *precision* dihitung dengan cara membagi jumlah data benar yang bernilai positif (*true positive*) dibagi dengan jumlah data benar yang bernilai positif (*true positive*) dan data salah bernilai negatif (*false negative*). Precision yang dihasilkan dari pengujian pada *rapidminer* sebesar 91.67 %.

5. **Recall RapidMiner**

Recall adalah keberhasilan sistem dalam menemukan kembali sebuah informasi. Nilai minimum data recall adalah 60% . Nilai *Recall* dihitung dengan cara membagi jumlah data benar yang bernilai positif (*true positive*) dibagi dengan jumlah data benar yang bernilai positif (*true positive*) dan data salah bernilai positif (*false positive*). *Recall* yang dihasilkan dari pengujian dari *rapidminer* sebesar 88 %.

5. **Kesimpulan**

Dari Hasil Pengolahan data pemasaran karton dus susu chil mil, dapat diambil kesimpulan bahwa hasil mining dan pemasaran dapat di implementasikan menggunakan algoritma C4.5 dengan tool RapidMiner untuk memprediksi produk karton dus susu chil mil yang bahwa karton Chil Mil Art Carton merupakan produk yang laku terjual. Dengan besar tingkat *Accuracy* 90.29% , *Precision* 91.67 % , *Recall* 88 % . Hasil dari perhitungan rapidminer adalah dengan prediksi lebih dari 19.500 lembar laku dan kurang dari 1.500 lembar tidak laku.

Daftar Pustaka

- [1] J.Han,"05_dbdm2007_Data Mining,"2007. Sri Wahyuni., Suherman.Implementasi Data Mining Mining Memprediksi Stok Barang Menggunakan Algoritma C4.5 Jurnal Informatika, Volume Nomor 2 Juli 2019.
- [2] Muchlisin, Riadi , Kamis 2 September 2007: <https://www.kajianpustaka.com>
- [3] H. Widya, S. Nasution, N. Silalahi" Pengertian Karton.vol. 2, no. 1, pp. 93-98,2013.
- [4] Sabna, Eka, 10 November 2018: <https://jurnalmanajemen.com>
- [5] Fae-zya.blogspot.com, 2017: Jenis-Jenis Kertas Karton Dan Penjualan.
- [6] L. Swastina, "Penerapan Algoritma C4.5 untuk Penentuan Prediksi Penjualan ," J. Gema Aktual ., vol. 2, no. 1, pp. 93-98, 2013.
- [7] A. Andriani, " Sistem Prediksi Penjualan Berbasis Decision Tree," Bianglala Inform., vol. 1, no. 1, pp.1-10,2013.
- [8] Indriyani, Novi, 2009, Penerapan Metode C4.5 Menggunakan Aplikasi RapidMiner.
- [9] David Hartanto Kamagi, Seng Hansun, 05 Juni 2014." Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa.
- [10] Erlin Elisa, Join. Vol 2 no.1 Juni 2017, " Analisa Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT. Arupadhatu Adisesanti.
- [11] Nurul Azwanti, vol. 13, No.1 Februari 2018: Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada PT. Capella Dinamik Nusantara Cabang Muka Kuning.