

Prediksi Penjualan Produk Rokok Pada PT. Indomarco Prismatama Menggunakan Algoritma C4.5

Meliana Leonardi¹, Riska Emilda², Irena Katrin³, Agus Yulianto⁴

¹Universitas Nusa Mandiri
e-mail: melianaleonardi@gmail.com

²Universitas Nusa Mandiri
e-mail: riskaemilda24@gmail.com

³Universitas Nusa Mandiri
e-mail: irenakatrin18@gmail.com

⁴Universitas Nusa Mandiri
e-mail: agus.aag@nusamandiri.ac.id

Abstrak - Pandemi Covid-19 setidaknya membawa pengaruh buruk pada ekonomi nasional sepanjang tahun 2020 yang dampaknya dapat dirasakan tak terkecuali oleh retail. Tujuan dari penelitian ini adalah untuk menginformasikan prediksi penjualan produk rokok pada Indomaret Plus Eco RSUD Cengkareng agar perusahaan dapat melakukan pengadaan barang sesuai dengan permintaan. Adanya prediksi yang tepat diharapkan dapat mengatasi beberapa masalah dalam penjualan produk rokok pada Indomaret Plus Eco RSUD Cengkareng seperti menumpuknya dan ketidakteraturan produk membuat toko harus terus melakukan FEFO (First Expired First Out) yang dalam praktiknya cukup memakan waktu ataupun overproduction dan under production. Metode yang digunakan dalam penelitian ini adalah algoritma C4.5. Algoritma C4.5 merupakan algoritma klasifikasi data dengan tipe pohon keputusan. Nilai akurasi tertinggi diperoleh dari pengujian dengan 58 data training dan 37 data testing, menunjukkan tingkat akurasi 92,11%, tingkat kesalahan 7,894% dan nilai AUC 87,8% dengan akurasi klasifikasi baik. Ukuran merupakan atribut yang paling mempengaruhi penjualan produk tembakau, tergolong bestseller atau underseller.

Kata Kunci: *Data Mining, Decision Tree, Prediksi Penjualan.*

Abstract - The Covid-19 pandemic has had a bad influence on the national economy throughout 2020, the impact of which can be felt, including by retail. The purpose of this study is to inform the prediction of sales of cigarette products at Indomaret Plus Eco RSUD Cengkareng so that companies can procure goods according to demand. The correct prediction is expected to overcome several problems in selling cigarette products at Indomaret Plus Eco RSUD Cengkareng, such as piling up and product irregularities, making the store have to continue to carry out FEFO (First Expired First Out) which in practice is quite a time consuming or overproduction and under production. The method in this research is the C4.5 Algorithm. The C4.5 algorithm is a data classification algorithm with the type of decision tree (decision tree). The highest accuracy value was obtained from experiments with 58 training data and 37 testing data, which showed an accuracy rate of 92.11%, an error rate of 7.894%, and an AUC value of 87.8% with good classification accuracy. Size is the attribute that most influences the sale of cigarette products, which are classified as selling or not selling well.

Keywords: *Data Mining, Decision Tree, Sales Prediction.*

PENDAHULUAN

Perekonomian industri di Indonesia terus berkembang yang menyebabkan persaingan yang ketat di seluruh sektor industri, khususnya pada

industri rokok. Salah satu peran penting bagi perusahaan untuk menentukan strategi penjualan adalah prediksi penjualan. Info berupa pola mampu digunakan sebagai suatu solusi untuk pengambilan keputusan pada dunia perbisnisan serta jua untuk



suatu pengembangan bisnis lainnya (Thabit, Alhomdy, Al-Ahdal, & Jagtap, 2020). Informasi yang ada dapat digunakan sebagai tolak ukur membuat prediksi penjualan. Selain itu, hasil prediksi penjualan dapat memberikan gambaran mengenai persediaan bahan baku yang dibutuhkan perusahaan dalam menjalankan kegiatan produksi. Gambaran yang tepat dapat mencegah kerugian perusahaan akibat *overproduction* maupun *under production* pada biaya penyimpanan.

Perusahaan PT. Indomarco Prismatama pada Indomaret RSUD Cengkareng merupakan toko pada perusahaan retail yang menjual berbagai jenis rokok dari berbagai *supplier*. *Supplier* yang telah bekerja sama dengan PT. Indomarco Prismatama antara lain : PT. Surya Mustika Nusantara, PT. Niaga Nusa Abadi, PT. Korea Tomorrow dan Global Indonesia, PT. Sumber Cipta Multiniaga, PT. Gudang Garam, PT. Gelora Djaja, PT. Bentoel Distribusi Utama, dan PT. HM Sampurna. Akumulasi dari beberapa *supplier* yang bekerjasama mengadakan 3 jenis kategori rokok yaitu SKT (Sigaret Kretek Tangan), SKM (Sigaret Kretek Mesin), dan SPM (Sigaret Putih Mesin). Pada toko Indomaret RSUD Cengkareng dalam kurun waktu dua bulan yang lalu terhitung dari tanggal 22 April 2021 mampu menjual 2.990 bungkus yang mana PKM (Permintaan Kuantiti Maksimum) yaitu sebanyak 3.281 bungkus.

Pada bulan Maret – April tahun 2021, PT. Indomarco Prismatama melakukan penarikan bea cukai rokok tahun 2020 dari semua mitra *supplier*. Tercatat terdapat penarikan sebanyak 163 bungkus pada toko Indomaret Plus Eco RSUD Cengkareng. Permasalahan terjadi akibat dari tidak berjalannya FEFO (*First Expired First Out*) dari DC (*Distribution Center*) ataupun bersumber dari toko karena produk yang tidak teratur. Dampak lanjut dari permasalahan ini yaitu menumpuknya persediaan rokok produksi lama dan produk rokok yang tidak sesuai dengan minat pembeli yang berakibat terjadi penarikan pada saat pergantian bea cukai pertahun. Terlebih jika ada penumpukan persediaan barang yang berpotensi menambah kerugian pelaku usaha akibat pandemi Covid-19 (Pusat Data dan Analisa Tempo, 2021).

Berdasarkan uraian di atas, dibutuhkan metode yang bisa membantu prediksi produk yang laku dan kurang laku sesuai dengan minat pembeli di pasaran sehingga toko Indomaret RSUD Cengkareng menentukan jumlah PKM (Permintaan Kuantiti Maksimal) rokok pada periode selanjutnya. Ada suatu tahapan atau langkah untuk menemukan *pattern* atau pola informasi pada sekumpulan data dengan memakai algoritma tertentu di pada data mining. algoritma yang akan digunakan pada perkara ini

merupakan algoritma C4.5. Algoritma C4.5 yaitu suatu algoritma yang bisa dipergunakan untuk membentuk pohon keputusan. Kelebihan berasal algoritma C4.5 diantaranya dapat menghasilkan pohon keputusan yang dipresentasikan dengan mudah, mempunyai tingkat akurasi yang bisa diterima, efisien pada menangani atribut bertipe diskret serta *numeric* (Muhammad Rizal, 2019).

Penelitian terkait menggunakan algoritma C4.5 untuk dengan ruang lingkup lain yaitu jurnal berjudul Penerapan Algoritma C4.5 Untuk Penentuan Ketersediaan Barang *E-Commers* untuk mengklasifikasikan stok barang yang harus ditambah dan yang tidak. Hasil akurasi yang didapatkan dari analisis menggunakan pohon keputusan dalam penelitian ini sebesar 98.9% dari 5.000 sampel data dan 23 atribut (Lukhayu Pritalia, 2018). Selain itu, penelitian terkait lainnya adalah jurnal berjudul Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada PT. Capela Dinamik Nusantara Cabang Muka Kuning. Pada penelitian ini menghasilkan nilai *gain* tertinggi yaitu 0.6965 dan menghasilkan suatu pohon keputusan guna melihat pola prediksi perilaku konsumen pembeli motor (Azwanti, 2018)

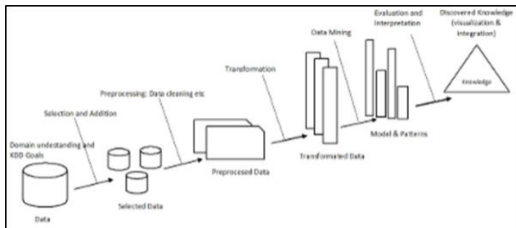
Pada penelitian ini penulis akan menerapkan salah satu algoritma *decision tree* yaitu algoritma C4.5 untuk membentuk pohon keputusan. Algoritma C4.5 ialah suatu algoritma yang termasuk ke dalam fungsi pengelompokan data yang menghasilkan suatu pohon keputusan (*decision tree*) dihasilkan dari perhitungan *entropy* serta *gain ratio* guna pengolahan atribut menjadi node (Izzulhaq & Sulastri, 2020). Pada algoritma C4.5 Jumlah data berbanding lurus dengan tingkat keakuratan, sehingga semakin banyak data maka tingkat keakuratan akan semakin tinggi (Lukhayu Pritalia, 2018). metode algoritma C4.5 bertujuan untuk membangun pohon keputusan diantaranya (Eska, 2018):

- Tentukan atribut yang nantinya akan berperan sebagai akar pohon
- Kemudian, setiap nilai dibentuk sebuah cabang.
- Bagi kasus pada sebuah cabang
- Ulangi proses pada tiap cabang hingga mempunyai kelas yang sama.

Dari penelitian terkait menunjukkan penggunaan algoritma C4.5 untuk memprediksi penjualan memperoleh nilai akurasi yang baik diharapkan dengan penggunaan algoritma C4.5 bisa mencari tahu bahwa kinerja juga tingkat akurasi terbaik dari *dataset* dalam penelitian ini.

METODOLOGI PENELITIAN

Penelitian ini menggunakan metode *Knowledge Discovery in Database* (KDD). Metode yang digunakan untuk mendapatkan pengetahuan asal dari sebuah *database* merupakan definisi dari KDD. Proses kerja KDD terlihat pada gambar di bawah ini: Sumber: (Nikmatun & Waspada, 2019)



Gambar 1. Proses *Knowledge Discovery in Database*

Data Selection (Data Seleksi)

Tahapan ini merupakan tahapan dimana data-data dikumpulkan dan kemudian diolah dengan metode Decision Tree algoritma C4.5. Berikut ini adalah data penjualan rokok pada Indomaret Plus Eco RSUD Cengkareng yang akan digunakan untuk pemrosesan data yang diolah dengan metode decision tree. Jumlah dataset asli pada data produk rokok sebanyak 95 record. Dari semua atribut yang ada terdapat 6 atribut yang akan digunakan dalam proses KDD. Atribut tersebut yaitu:

Tabel 1. Pemilihan Atribut

| Atribut | Detail Penggunaan |
|-----------------|-------------------|
| Id Produk | X Tidak |
| Nama Produk | ✓ ID |
| Jenis | ✓ Nilai Model |
| Kategori Produk | ✓ Nilai Model |
| Ukuran | ✓ Nilai Model |
| Minat | ✓ Nilai Model |

Sumber: (Leonardi, Emilda, & Katrin, 2021)

Tabel di atas menjelaskan atribut-atribut yang akan digunakan dalam survei, indikator ya (✓) berarti atribut-atribut tersebut akan digunakan dalam survei. Meskipun indikator tidak (X) menunjukkan bahwa atribut tersebut tidak digunakan dalam tahap data mining investigasi.

Processing (Cleaning)

Setelah dilakukan *data selection*, atribut atau variabel yang digunakan terdiri dari 5 atribut yang terdapat dalam penentuan minat. Dengan 3 atribut *predictor*, 1 atribut ID dan 1 atribut target. Pada atribut jenis, kategori produk, dan ukuran memiliki *value* yang

dapat dilihat pada tabel dibawah ini:

Tabel 2. Atribut dan *Value*

| Atribut | <i>Value</i> |
|-----------------|--------------------------|
| Jenis | Filter Kretek Mild |
| Kategori Produk | SKT SKM SPM |
| Ukuran | Besar Sedang kecil |

Sumber: (Leonardi et al., 2021)

Data Transformation

Dalam tahap *data transformation* tidak ada data yang dibersihkan, pembagian data set dengan menggunakan *process split data* untuk mnghasilkan data *training* dan data *testing* dengan rasio 60:40. Data *training* berjumlah 58 data dan data *testing* 37 data dari total keseluruhan yaitu 95 data. Tabel data tersebut dapat dilihat pada gambar berikut:

| nama produk | jenis | katego | ukuran | minat |
|----------------------|--------|--------|--------|-------------|
| Lucky Strike Purple | filter | SPM | Besar | Laku |
| Djisamsoe Kretek | kretek | SKT | Sedang | Laku |
| Djisamsoe Kretek | kretek | SKT | Kecil | Laku |
| Envig Kretek | kretek | SKT | Kecil | Laku |
| GG Surya Filter | filter | SKM | Sedang | Laku |
| GG Surya Pro | filter | SKM | Sedang | Laku |
| Djarum Super Filter | filter | SKM | Kecil | Kurang Laku |
| Djarum Super Filter | filter | SKM | Sedang | Laku |
| LA Filter Light | filter | SKM | Sedang | Laku |
| Sampoerna Mild | mild | SKM | Kecil | Kurang Laku |
| Sampoerna Mild | mild | SKM | Sedang | Laku |
| Djarum Black Filter | filter | SKM | Sedang | Laku |
| Clas Mild Filter | filter | SKM | Sedang | Laku |
| U Mild Filter | filter | SKM | Sedang | Laku |
| Dii Samsoe Magnum | filter | SKM | Kecil | Kurang Laku |
| Clas Mild Filter | filter | SKM | Kecil | Kurang Laku |
| Dunhil F/C Mild | mild | SKM | Besar | Kurang Laku |
| Esse Mild | mild | SKM | Besar | Kurang Laku |
| GG Mild Shiver | mild | SKM | Sedang | Laku |
| Dunhil FC Filter | filter | SKM | Sedang | Laku |
| Djarum Coklat Filter | filter | SKM | Kecil | Kurang Laku |
| LA Bold | filter | SKM | Besar | Kurang Laku |
| Sampoerna Awoltn | mild | SKM | Besar | Kurang Laku |
| Sampoerna Awoltn | mild | SKM | Besar | Laku |
| Marlboro Hpack | filter | SPM | Besar | Laku |
| Marlboro Light | filter | SPM | Besar | Laku |
| Marlboro Black | filter | SKM | Besar | Kurang Laku |
| Djarum Mild Black | mild | SKM | Sedang | Laku |
| Djarum Mild Black | mild | SKM | Kecil | Laku |
| Esse Shuffle Pop | mild | SKM | Sedang | Laku |
| Sampoerna Menthol | mild | SKM | Sedang | Laku |
| Dunhill Merah | mild | SPM | Besar | Laku |
| Marlboro Burst | filter | SPM | Besar | Laku |
| Djarum Coklat TIN | kretek | SKM | Kecil | Kurang Laku |
| GG Surya Kretek | kretek | SKM | Kecil | Kurang Laku |
| Insta Rokok | filter | SKM | Sedang | Laku |
| Dii Samsoe Mild | mild | SKM | Besar | Kurang Laku |
| Sampoerna Tropical | mild | SKM | Sedang | Laku |
| Sampoerna Sunnu | mild | SKM | Sedang | Laku |
| Camel Active Purple | mild | SPM | Besar | Laku |
| Class Mild Silver | mild | SKM | Sedang | Laku |
| Dunhil FC Filter | filter | SKM | Kecil | Kurang Laku |
| Ziga Rokok Filter | filter | SKM | Kecil | Kurang Laku |
| Camel Purple | filter | SKM | Kecil | Kurang Laku |
| Phlomis Magnum | filter | SKM | Kecil | Kurang Laku |
| Sampoerna Mild | mild | SKM | Sedang | Laku |
| Esse Cafe | mild | SKM | Besar | Kurang Laku |
| Camel White | filter | SPM | Besar | Laku |
| Marlboro IC Burst SE | mild | SPM | Besar | Laku |
| Djarum Kretek Ckt | kretek | SKM | Kecil | Kurang Laku |
| LA Bold | mild | SKM | Besar | Kurang Laku |

Sumber: (Leonardi et al., 2021)

Gambar 2. Data *Training*

| nama produk | jenis | kategori pr | ukuran | minat |
|---------------------|--------|-------------|--------|-------------|
| Djarum Kretek | kretek | SKT | Kecil | Laku |
| 234 Rokok | kretek | SKT | Kecil | Laku |
| Sampoerna Kretek | kretek | SKT | Kecil | Laku |
| Marlboro Ice Brust | filter | SPM | Besar | Laku |
| Sampoerna Hiaju | kretek | SKT | Kecil | Laku |
| GG Filter Merah | filter | SKM | Kecil | Kurang Laku |
| Wismilak Diplomat | kretek | SKM | Kecil | Kurang Laku |
| LA Light Menthol | mild | SKM | Sedang | Laku |
| GG Surya Pro Mild | mild | SKM | Sedang | Laku |
| Envio Mild | mild | SKM | Sedang | Kurang Laku |
| Djarum Super Mild | mild | SKM | Sedang | Kurang Laku |
| GG Surya | kretek | SKM | Sedang | Laku |
| Djarum Super Mild | mild | SKM | Besar | Kurang Laku |
| GG Signature | filter | SKM | Kecil | Laku |
| LA Ice | mild | SKM | Sedang | Laku |
| Esse Chng Aplmnt | mild | SKM | Besar | Kurang Laku |
| LA Bold | filter | SKM | Kecil | Kurang Laku |
| Dji Samsoe | kretek | SKM | Sedang | Laku |
| Lucky Strike Filter | filter | SPM | Besar | Laku |
| GG Surya Filter | filter | SKM | Kecil | Kurang Laku |
| Esse Change GRP | mild | SKM | Besar | Kurang Laku |
| LA Bold | filter | SKM | Sedang | Laku |
| Esse Change Juicy | mild | SKM | Besar | Kurang Laku |
| Esse Change Doub | mild | SKM | Sedang | Laku |
| Diplomat Evo | kretek | SKM | Sedang | Laku |
| Djarum Wave | mild | SKM | Kecil | Laku |
| Marlboro Hpack | filter | SPM | Besar | Laku |
| Marlboro Advance | filter | SKM | Kecil | Kurang Laku |
| Esse Punch Pop | mild | SKM | Sedang | Laku |
| Forte Original | kretek | SPM | Besar | Laku |
| Forte Menthol | kretek | SPM | Besar | Laku |
| Esse Berry Pop | mild | SPM | Sedang | Laku |
| Djisamsoe EDS | kretek | SKT | Kecil | Laku |
| DG Kretek Spack | kretek | SKT | Kecil | Laku |
| Win Bold | Mild | SKM | Besar | Kurang Laku |
| Esse Golden leaf | filter | SPM | Besar | Laku |
| Forte Brz Mntl | kretek | SPM | Besar | Laku |
| Esse Filter Blue | mild | SPM | Besar | Laku |

Sumber: (Leonardi et al., 2021)
Gambar 3. Data Testing

Data Mining

Metode yang penulis gunakan dalam mencari pola menarik adalah metode *Decision Tree*, Algoritma C4.5 yang merupakan salah satu algoritma dalam fungsi klasifikasi. Algoritma C4.5 memberikan hasil akhir berupa pohon keputusan (*decision tree*) dari hasil perhitungan *entropy* dan *gain ratio* untuk pengolahan atribut menjadi *node*.

Interpretation / Evalaution

Pada tahap evaluasi melibatkan hasil perhitungan *accuracy*, *precision*, *error rate* dan *recall* dari data *testing* dengan bantuan *Confusion Matrix*. Sedangkan, dalam tahap interpretasi merupakan visualisasi dari hasil evaluasi yang sudah dilakukan.

HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan perhitungan secara manual dan dengan bantuan tools RapidMiner untuk mendapatkan sebuah pohon keputusan (*decision tree*) yang dapat memprediksi penjualan produk rokok pada PT Indomarco Prismatama Cabang Indomaret Plus Eco RSUD Cengkareng.

1. Perhitungan Secara Manual

Sesuai dengan langkah pembuatan pohon keputusan (*decision tree*) setelah melakukan analisis data yang menghasilkan data *training* dan data *testing* maka dilanjutkan dengan menentukan nilai *entropy* dan *gain* pada data *training*. Hasil perhitungan akan digunakan untuk menentukan *root* atribut atau atribut yang akan menjadi *the best classifier*.

Rumus menghitung nilai *entropy*:

$$Entropy (S) = \sum_{i=1}^n -p_i * \log_2 p_i \dots\dots\dots(1)$$

Keterangan :

- S = himpunan kasus
- n = jumlah partisi S
- p_i = proporsi S_i terhadap S

Menghitung nilai *gain* menggunakan rumus:

$$Gain (S,A) = Entropy (S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy (S) \dots\dots\dots(2)$$

Keterangan :

- S = Himpunan Kasus
- A= Atribut
- n = Jumlah Partisi Atribut A
- $|S_i|$ = Jumlah Kasus pada partisi ke-i
- $|S|$ = Jumlah Kasus dalam S

Didapatkan hasil perhitungan seperti pada gambar di bawah ini:

| atribut | value | jumlah kasus (s) | laku (S ₁) | kurang laku | entropy | gain |
|-----------------|--------|------------------|------------------------|-------------|-------------|-------------|
| total | | 57 | 36 | 21 | 0.949452015 | |
| Kategori Produk | | | | | | 0.195360384 |
| | SKT | 3 | 3 | 0 | 0 | |
| | SKM | 43 | 22 | 21 | 0.999609836 | |
| | SPM | 11 | 11 | 0 | 0 | |
| Ukuran | | | | | | 0.374010544 |
| | Kecil | 17 | 4 | 13 | 0.787126586 | |
| | Sedang | 20 | 20 | 0 | 0 | |
| | Besar | 20 | 12 | 8 | 0.970950594 | |
| Jenis | | | | | | |
| | Kretek | 6 | 3 | 3 | 1 | 0.011631625 |
| | Filter | 28 | 17 | 11 | 0.966618633 | |
| | Mild | 23 | 16 | 7 | 0.886540893 | |

Sumber: (Leonardi et al., 2021)
Gambar 4. Hasil Perhitungan Entropy dan Gain

| Atribut | Value | jumlah kasus (s) | laku (S _l) | kurang laku | entropy | gain |
|-----------------|---------------|------------------|------------------------|-------------|-------------|-------------|
| Ukuran | | | | | | |
| | Kecil & Besar | 37 | 16 | 21 | 0.98678672 | |
| Kategori Produk | | | | | | 0.634204379 |
| | SKT | 2 | 2 | 0 | 0 | |
| | SKM | 24 | 3 | 21 | 0.543564443 | |
| | SPM | 11 | 11 | 0 | 0 | |
| Jenis | | | | | | 0.008187511 |
| | Kretek | 5 | 2 | 3 | 0.970950594 | |
| | Filter | 18 | 7 | 11 | 0.964078765 | |
| | Mild | 14 | 7 | 7 | 0.910768985 | |

Sumber: (Leonardi et al., 2021)

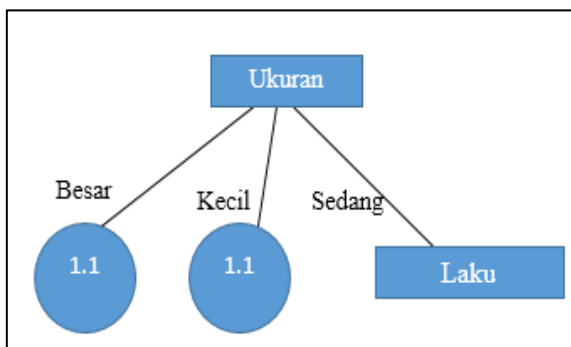
Gambar 5. Lanjutan Hasil Perhitungan *Entropy* dan *Gain*

| atribut | value | jumlah kasus (s) | laku (S _l) | kurang laku | entropy | gain |
|-----------------|--------|------------------|------------------------|-------------|-------------|-------------|
| Kategori Produk | | | | | | |
| | SKM | 24 | 3 | 21 | 0.543564443 | |
| Jenis | | | | | | 0.910768985 |
| | Kretek | 3 | 0 | 3 | 0 | |
| | Filter | 11 | 0 | 11 | 0 | |
| | Mild | 10 | 3 | 7 | 0.881290899 | |

Sumber: (Leonardi et al., 2021)

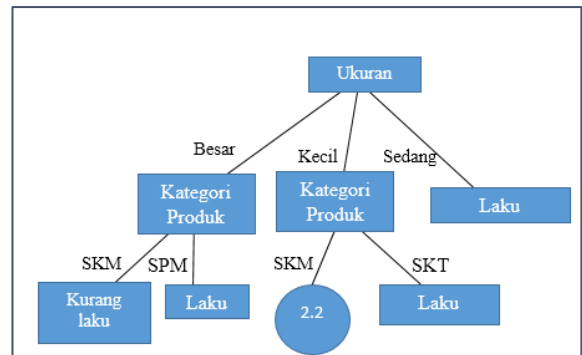
Gambar 6. Lanjutan Hasil Perhitungan *Entropy* dan *Gain*

Didapatkan pohon keputusan secara manual sebagai berikut.



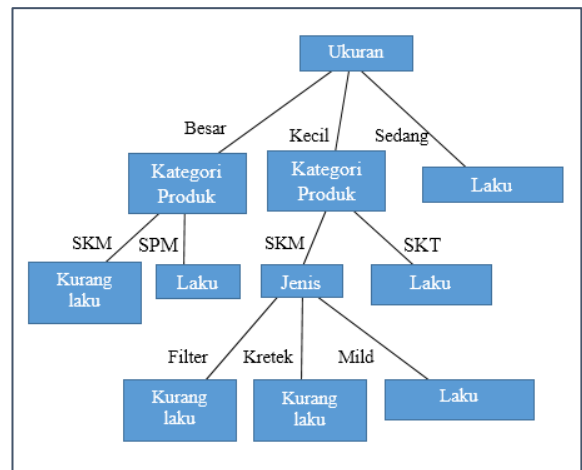
Sumber: (Leonardi et al., 2021)

Gambar 7. Pohon Hasil Perhitungan Menentukan *Root*



Sumber: (Leonardi et al., 2021)

Gambar 8. Pohon Hasil Perhitungan Menentukan *Subtree* Pertama



Gambar 9. Pohon Hasil Perhitungan Menentukan *Subtree* Kedua

2. Evaluasi Hasil Manual

Setelah selesai menghitung *entropy* dan *gain*, selanjutnya yang akan dilakukan adalah menghitung *accuracy*, *precision*, *recall* dan *error rate* pada data dengan bantuan *confusion matrix*. *Confusion matrix* memiliki empat istilah hasil proses klasifikasi, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, *False Negative (FN)*.

| Actual | Predicted | |
|--------|-----------|----|
| | - | + |
| - | 10 | 2 |
| + | 1 | 25 |

Sumber: (Leonardi et al., 2021)

Gambar 10. Hasil *Confusion Matrix*

a. Accuracy

Accuracy dapat dihitung menggunakan rumus:

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \times 100\% \dots\dots\dots(3)$$

b. *Error Rate*

Nilai *error rate* dapat dihitung dengan menggunakan rumus:

$$Error Rate = \frac{FP+FN}{TN+TP+FN+FP} \times 100\% \dots\dots\dots(4)$$

c. *Precision*

Precision dapat dihitung menggunakan rumus:

$$Precision = \frac{TP}{FP+TP} \times 100\% \dots\dots\dots(5)$$

d. *Recall*

Nilai *recall* dapat dihitung dengan menggunakan rumus:

$$Recall = \frac{TP}{FN+TP} \times 100\% \dots\dots\dots(6)$$

e. *Spesificity*

Nilai *specificity* dapat dihitung dengan menggunakan rumus:

$$Spesificity = \frac{TN}{TN+FP} \times 100\% \dots\dots\dots(7)$$

Didapatkan hasil perhitungan seperti pada gambar di bawah ini.

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \times 100\% = \frac{10+25}{10+25+1+2} \times 100\% = 92,105\%$$

$$Error Rate = \frac{FP+FN}{TN+TP+FN+FP} \times 100\% = \frac{1+2}{10+25+1+2} \times 100\% = 7,894\%$$

$$Precision = \frac{TP}{FP+TP} \times 100\% = \frac{25}{2+25} = 92,595\%$$

$$Recall = \frac{TP}{FN+TP} \times 100\% = \frac{25}{1+25} \times 100\% = 96,153\%$$

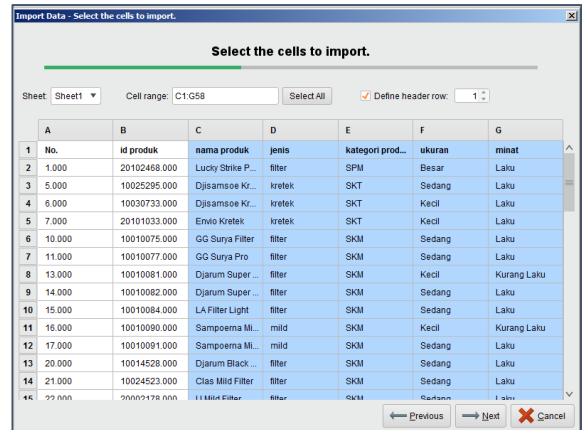
$$Spesificity = \frac{TN}{TN+FP} \times 100\% = \frac{10}{10+2} \times 100\% = 83,333\%$$

Sumber: (Leonardi et al., 2021)

Gambar 11. Hasil Perhitungan Evaluasi

3. Perhitungan Menggunakan RapidMiner

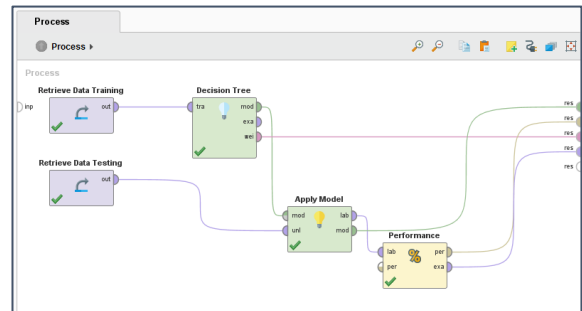
Dengan menggunakan algoritma C4.5, akan dibentuk pohon keputusan (decision tree) menggunakan bantuan tools RapidMiner. Langkah pertama adalah *import data*



Sumber: (Leonardi et al., 2021)

Gambar 12. *Import Data*

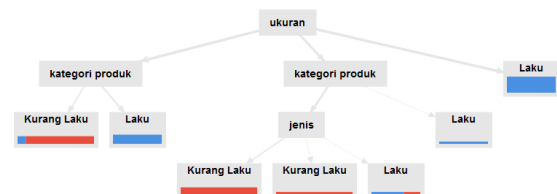
Dilanjutkan dengan menambahkan *operators* yang dibutuhkan.



Sumber: (Leonardi et al., 2021)

Gambar 13. Menambahkan *Operators*

Lakukan *running proses* untuk mendapatkan hasil berupa pohon keputusan dari data *training* atau *training samples* dan mendapatkan keakuratan dari data *testing* atau *testing samples*. Berikut adalah pohon keputusan yang didapatkan dengan bantuan tools RapidMiner.



Sumber: (Leonardi et al., 2021)

Gambar 14. Pohon Keputusan

Rule yang dihasilkan adalah sebagai berikut.

Tree

```

ukuran = Besar
| kategori produk = SKM: Kurang Laku {Laku=1, Kurang Laku=8}
| kategori produk = SPM: Laku {Laku=11, Kurang Laku=0}
ukuran = Kecil
| kategori produk = SKM
| | jenis = filter: Kurang Laku {Laku=0, Kurang Laku=9}
| | jenis = kretek: Kurang Laku {Laku=0, Kurang Laku=3}
| | jenis = mild: Laku {Laku=2, Kurang Laku=1}
| kategori produk = SKT: Laku {Laku=2, Kurang Laku=0}
ukuran = Sedang: Laku {Laku=20, Kurang Laku=0}
    
```

Sumber: (Leonardi et al., 2021)

Gambar 15. *Rule* Pohon Keputusan

1. Jika produk berukuran Besar dengan kategori SKM, maka produk tersebut Kurang Laku. Perbandingan Laku dan Kurang Laku adalah 1:8.
2. Jika produk berukuran Besar dengan kategori SPM, maka produk tersebut Laku. Perbandingan Laku dan Kurang Laku adalah 11:0.
3. Untuk produk berukuran Kecil dengan kategori SKM dan berjenis filter maka produk Kurang Laku. Perbandingan Laku dan Kurang Laku adalah 0:9.
4. Untuk produk berukuran Kecil dengan kategori SKM dan berjenis kretek maka produk Kurang Laku. Perbandingan Laku dan Kurang Laku adalah 0:3.
5. Untuk produk berukuran Kecil dengan kategori SKM dan berjenis mild maka produk Laku. Perbandingan Laku dan Kurang Laku adalah 2:1.
6. Jika produk berukuran Kecil dan berkategori SKT maka produk tersebut Laku dengan perbandingan jumlah Laku dan Kurang Laku 2:0.
7. Untuk produk berukuran sedang mendapat jumlah Laku sebanyak 20 dan Kurang Laku 0 sehingga dapat disimpulkan produk berukuran sedang, Laku.

Dari *rule* di atas didapatkan prediksi produk rokok Laku adalah seperti gambar di bawah ini.

| No. | Rule Laku | Jumlah Produk Laku | Nama Produk |
|-----|---|--------------------|---|
| 1 | Ukuran: Besar Kategori: SPM | 11 | 1. Lucky Strike Purple Boster 2. Marlboro Hpack 3. Marlboro Light 4. Dunhill Merah 5. Marlboro Burst 6. Camel Active Purple Mint 7. Camel White 8. Marlboro IC Burst SE 9. Dunhill Biru Light 10. Dunhill Hijau Menthol 11. Esse Blue Change |
| 2 | Ukuran: Kecil Kategori: SKM Jenis: Mild | 2 | 1. Sampoerna Mild 2. GG Sigt Mild |
| 3 | Ukuran: Kecil Kategori: SKT | 2 | 1. Djsamsoe Kretek 2. Envio Kretek |
| 4 | Ukuran: Sedang | 20 | 1. Djsamsoe Kretek 2. GG Surya Filter 3. GG Surya Pro 4. Djarum Super Filter 5. LA Filter Light 6. Sampoerna Mild 7. Djarum Black Filter 8. Clas Mild Filter 9. U Mild Filter 10. GG Mild Shiver 11. Dunhill FC Filter 12. Djarum Mild Black 13. Esse Shuffie Pop 14. Sampoerna Menthol 15. Insta Rokok 16. Sampoerna Tropical 17. Sampoerna Sunny 18. Class Mild Silver 19. Sampoerna Mild 20. Marlboro Black |

Sumber: (Leonardi et al., 2021)

Gambar 16. Prediksi Produk Rokok Laku

4. Hasil *Performance Testing Samples* pada RapidMiner
Performance meliputi hasil perhitungan *accuracy* dan ROC/AUC.

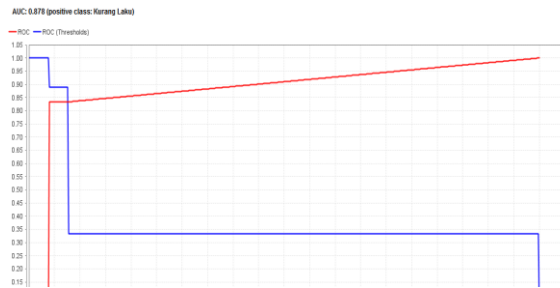
| | True Laku | True Kurang Laku | class precision |
|------------------|-----------|------------------|-----------------|
| accuracy: 92.11% | | | |
| AUC (optimized) | 25 | 2 | 92.59% |
| AUC | | | |
| AUC (desametric) | 1 | 10 | 90.91% |
| class recall | 96.15% | 93.33% | |

Sumber: (Leonardi et al., 2021)

Gambar 17. *Accuracy Data Testing* pada RapidMiner

Performance keakurasian AUC (Gorunescu, 2010) dapat diklasifikasikan menjadi lima kelompok yaitu (Tusarwenda, 2018):

- a. 0.90 – 1.00 = *Excellent Classification*.
- b. 0.80 – 0.90 = *Good Classification*.
- c. 0.70 – 0.80 = *Fair Classification*.
- d. 0.60 – 0.70 = *Poor Classification*.
- e. 0.50 – 0.60 = *Failure*.



Sumber: (Leonardi et al., 2021)

Gambar 18. AUC Data *Testing* pada RapidMiner

Dari gambar di atas dapat dilihat hasil dari pengolahan AUC data *testing* menggunakan Algoritma C4.5 sebesar 0,878 dengan tingkat akurasi *Good Classification*.

| | Cross Validation |
|----------|------------------|
| Accuracy | 92,11% |
| AUC | 0,878% |

Sumber: (Leonardi et al., 2021)

Dari tabel di atas untuk mengetahui prediksi penjualan pada toko Indomaret Plus Eco RSUD Cengkareng, menggunakan algoritma *decision tree*, model *cross validation* menghasilkan nilai AUC 0,878% akurasi yang tinggi sebesar 92,11%.

5. Potensi Hasil

Penelitian ini sendiri menggunakan salah satu algoritma pada metode *Decision Tree* yaitu Algoritma C4.5 untuk membuat pohon keputusan (*Decision Tree*) dan mendapatkan nilai *accuracy* sebesar 92.11% serta nilai AUC 0,878 atau 87,8% dengan tingkat akurasi *Good Classification*. Sehingga dapat disimpulkan bahwa penggunaan metode *Decision Tree* dengan Algoritma C4.5 ini cukup tepat untuk menentukan prediksi penjualan rokok pada PT. Indomarc Prismatama Cabang Indomaret Plus Eco RSUD Cengkareng.

KESIMPULAN

Dari hasil perhitungan secara manual maupun dengan bantuan *tools* RapidMiner dari data training sebanyak 58 didapatkan pohon keputusan yang menciptakan rule untuk produk rokok yang termasuk ke dalam minat Laku adalah produk rokok dengan kondisi: berukuran Sedang; berukuran Kecil dan berkategori SKT; berukuran Besar dan berkategori SPM; berukuran Kecil, berkategori SKM dan bejenis mild. Berdasarkan evaluasi hasil menggunakan data testing sebanyak 37 data pada penelitian prediksi penjualan produk rokok menggunakan Algoritma C4.5

didapatkan nilai *accuracy* sebesar 92,11% dan AUC 0,878 sehingga mendapatkan tingkat akurasi *Good Classification*.

REFERENSI

- Azwanti, N. (2018). Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 13(1), 33. <https://doi.org/10.30872/jim.v13i1.629>
- Eska, J. (2018). Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5. 2. <https://doi.org/10.31227/osf.io/x6svc>
- Izzulhaq, A. F., & Sulastri. (2020). Klasifikasi Penjualan Aplikasi Android. *Proceeding SENDIU*, 72(6), 978–979.
- Leonardi, M., Emilda, R., & Katrin, I. (2021). PREDIKSI PENJUALAN PRODUK ROKOK PADA PT. INDOMARCO PRISMATAMA CABANG INDOMARET PLUS ECO RSUD CENKARENG MENGGUNAKAN ALGORITMA C4.5.
- Lukhayu Pritalia, G. (2018). Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce. *Indonesian Journal of Information Systems*, 1(1), 47–56. <https://doi.org/10.24002/ijis.v1i1.1727>
- Muhammad Rizal. (2019). Analisa Prediksi Penjualan Produk Dengan Menggunakan Metode C4.5 (Studi Kasus : PT. Kawan Lama Ace Hardware). *Jurnal Riset Komputer*, 6(5), 545–549. Retrieved from <https://ejournal.stmik-budidarma.ac.id/index.php/jurikom/article/view/1656/1253>
- Nikmatun, I. A., & Waspada, I. (2019). Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, 10(2), 421–432.
- Pusat Data dan Analisa Tempo. (2021). Berharap Lonjakan Omzet di Masa Pandemi. In *PENGUSAHA RETAIL MENYIASATI BISNIS DI ERA PANDEMI* (p. 12).
- Thabit, F., Alhomdy, A. P. S., Al-Ahdal, A. H. A., & Jagtap, P. D. S. (2020). Jurnal Comasie. *Global Transitions Proceedings*, 3(3), 21–30.
- Tusarwenda, T. R. I. B. (2018). Penerapan data mining dengan algoritma c4.5 dalam prediksi penjualan botol pada cv. seribukilo.

PROFIL PENULIS

Meliana Leonardi

Tahun 2020 lulus dari Program Ahli Madya (A,Md.)
Program Studi Sistem Informasi Universitas Bina
Sarana Informatika dan sekarang sedang melanjutkan
Program Strata Satu (S1) Universitas Nusa Mandiri.

Riska Emilda

Tahun 2020 lulus dari Program Ahli Madya (A,Md.)
Program Studi Sistem Informasi Universitas Bina
Sarana Informatika dan sekarang sedang melanjutkan
Program Strata Satu (S1) Universitas Nusa Mandiri.

Irena Katrin

Tahun 2020 lulus dari Program Ahli Madya (A,Md.)
Program Studi Sistem Informasi Universitas Bina
Sarana Informatika dan sekarang sedang melanjutkan
Program Strata Satu (S1) Universitas Nusa Mandiri.