

# Prediksi Potensi Pemasaran Produk Baru dengan Metode Naïve Bayes Classifier dan Regresi Linear

Taghsya Izmi Andini\*, Wina Witanti, Faiza Renaldi

Jurusan Informatika, Fakultas MIPA

Universitas Jenderal Achmad Yani

Jl. Terusan Sudirman, Cimahi

\*izmiandini@gmail.com

**Abstrak**—CV. Lampegan Jaya merupakan salah satu perusahaan yang bergerak di bidang pendistribusian produk minuman dan makanan. Produk tersebut didistribusikan ke outlet-outlet yang ada di Jawa Barat. Penjualan produk di CV. Lampegan Jaya terkadang mengalami peningkatan dan penurunan penjualan, hal tersebut disebabkan kurangnya daya beli masyarakat terhadap salah satu produk yang ditawarkan, untuk meningkatkan penjualan maka perusahaan melakukan alternatif dengan cara menjual produk baru. Produk baru tersebut memiliki kesamaan spesifikasi dengan produk sebelumnya, salah satu cara untuk mengetahui tingkat kedekatan spesifikasi antar produk yaitu dengan Metode Naïve Bayes Classifier dan Regresi Linear. Produk yang memiliki tingkat kedekatan terbesar akan dilakukan pengambilan data penjualan produk serta daerah pemasarannya untuk di tampilkan sebagai prediksi bagi produk baru. Beberapa penelitian terdahulu dalam melakukan pencarian nilai kedekatan antar produk dengan melakukan prediksi nilai mahasiswa, prediksi persediaan alat-alat kesehatan, prediksi besarnya penggunaan listrik rumah tangga, dan beberapa penelitian menggunakan Nearest Neighbors, Regresi Linear, Metode Clustering dan Algoritma Apriori yang termasuk kedalam Metode Data Mining. Pada penelitian ini menggunakan Metode Naïve Bayes Classifier dan Regresi Linear yang dapat diterapkan pada kasus dengan class dan data yang banyak, maka diharapkan dapat memudahkan perusahaan dalam mendapatkan informasi mengenai prediksi penjualan produk baru dan daerah pemasarannya

**Kata Kunci**—penjualan; Naïve bayes classifier; regresi linear; data mining

## I. PENDAHULUAN

CV. Lampegan Jaya merupakan perusahaan yang bergerak di bidang pendistribusian produk makanan dan minuman Meses Tulip Coklat, Vita Zone, Teh Bintang Sobo, Teh Preso, Okky Jelly Drink, Fruit Tea, Kuku Bima Energi, Kopikap Cappuccino, Mizon dan My Tea. Produk-produk tersebut didistribusikan kepada outlet-outlet yang tersebar di daerah Jawa Barat. Distribusi produk yang dilakukan berdasarkan permintaan outlet terhadap produk yang ditawarkan sales ataupun permintaan outlet terhadap produk baru.

CV. Lampegan Jaya melakukan pendistribusian produk terhadap 7000 outlet yang mengakibatkan permintaan outlet terhadap produk baru ataupun produk yang sudah ada menjadi beragam. Perusahaan menjadikan permintaan outlet sebagai dasar dalam pengambilan keputusan untuk melakukan prediksi

penjualan produk baru dan daerah pemasarannya. Permintaan outlet didapatkan dari sales yang memberikan daftar permintaan kepada trader untuk selanjutnya diberikan ke manager pemasaran. Manager pemasaran dan direktur mengambil peran penting dalam hal menentukan prediksi penjualan.

Permintaan produk oleh outlet selanjutnya dilakukan penambahan sesuai keinginan direktur dalam hal prediksi penjualan, hal tersebut menyebabkan sering terjadi kelebihan ataupun kekurangan produk. Produk yang ditawarkan perusahaan termasuk kedalam jenis food, maka penyimpanan produk di gudang tidak dapat terlalu lama. Produk yang mengalami pemesanan lebih dari permintaan sering terjadi cacat atau rusak. Perusahaan juga sering mengalami kerugian akibat tidak terpenuhinya permintaan outlet terhadap produk yang diinginkan. Berdasarkan pada permasalahan tersebut, diperlukan sebuah pengelolaan data yang memungkinkan perusahaan untuk melakukan prediksi penjualan produk baru. Data penjualan dapat dimanfaatkan dengan Data Mining, seperti pada penelitian sebelumnya terhadap data di dunia pendidikan menggunakan Metode Spectral Clustering dan Bagging Regresi Linear [1][2], data penjualan mengenai prediksi penjualan [3] dan penggunaan listrik rumah tangga [7].

Representasi Data Mining diyakini dapat berguna dalam melakukan pengelolaan, memberikan informasi dan meningkatkan nilai akurasi mengenai prediksi penjualan produk baru dan daerah pemasarannya, seperti pada penelitian sebelumnya dalam pengolahan data mining, hal tersebut didasarkan pada penelitian sebelumnya menggunakan Metode Regresi Linear [1] dan Naïve Bayes Classifier yang dapat melakukan klasifikasi terhadap data baru [9].

## II. LANDASAN TEORI

### A. Data Mining

Data mining adalah proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Teknik dalam Data Mining yaitu bagaimana menelusuri data yang ada untuk membangun sebuah model. Model tersebut digunakan untuk mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan.

Penelitian terdahulu telah menggunakan data jenis item alat-alat kesehatan sebanyak 30 data untuk melakukan prediksi persediaan barang dengan pengujian pertama yaitu menghitung frekuensi *itemset* dengan *parameters* adalah minimal *support* sama dengan 16% dan maksimal *support* sama dengan 100%, *support* adalah nilai persentasi banyak jumlah jenis *items* dan jumlah persentasi banyak *items* dan pola kombinasi dua *items* didalam transaksi menggunakan Metode Apriori [4][5], lalu terdapat penelitian terdahulu menggunakan Metode NBC dengan data sampel sebanyak 60 data, setelah dilakukan pengujian, terdapat 47 data penggunaan listrik rumah tangga berhasil diklasifikasi dengan benar dan sebanyak 13 data penggunaan listrik rumah tangga tidak berhasil diklasifikasikan dengan benar [7], selain itu penelitian lain menggunakan Metode *Regresi Linear* dengan menggunakan data sebanyak 6 bulan yaitu dari bulan July hingga Desember 2008 untuk memprediksi bursa yang ada di Negara Nigeria [12][13] dan penelitian lainnya menggunakan data dari Badan *Meteorology* [8] serta data dari dunia pendidikan [1] yang dapat membantu dalam pengambilan keputusan

### B. Naïve Bayes Classifier

*Naïve Bayes Classifier* merupakan sebuah proses klasifikasi probabilistik sederhana yang berdasarkan pada penerapan Teorema *Bayes* (atau Aturan *Bayes*) dengan asumsi independensi (ketidaktergantungan) yang kuat, dengan kata lain, dalam *Naïve Bayes*, model yang digunakan adalah model fitur independen. Dalam *Bayes* (terutama *Naïve Bayes*), maksud independensi yang kuat pada fitur adalah sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama. Keuntungan dari klasifikasi adalah metode ini hanya membutuhkan sejumlah kecil data pelatihan untuk memperkirakan parameter (sarana dan varians dari variabel) yang diperlukan untuk klasifikasi.

Beberapa penelitian menggunakan Metode *Naïve Bayes Classifier* menghasilkan probabilitas setiap kriteria untuk *class* yang berbeda, sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk memprediksi penggunaan listrik dengan data latih sebanyak 60 data, setelah dilakukan pengujian menggunakan Metode *Naïve Bayes Classifier* dari 60 data penggunaan listrik rumah tangga, ada sebanyak 47 data penggunaan listrik rumah tangga yang diklasifikasikan dengan benar dan 13 lainnya tidak berhasil diklasifikasikan dengan benar [7]. Selanjutnya, *Naïve Bayes* merupakan model yang efektif dalam memprediksi pasien dengan penyakit jantung, ginjal karena model tersebut dapat menjawab pertanyaan kompleks dengan kemudahan dalam melakukan penafsiran akses ke informasi [9], seperti pada penelitian sebelumnya dengan menggunakan 14 data dengan 4 atribut yaitu *age*, *income*, *student* dan *credit\_rating* dalam melakukan prediksi penyakit jantung [10] selain itu Metode *Naïve Bayes* juga dapat digunakan dalam melakukan prediksi penjualan mobil di perusahaan manufaktur [14][15][16] dan prediksi kinerja akademik mahasiswa dengan faktor yang berpengaruh meliputi Indeks Prestasi Kumulatif (IPK), Indeks Prestasi (IP) semester 1, IP semester 4 dan jenis kelamin dengan menghasilkan nilai akurasi sebesar 70% [11].

Perhitungan *Naïve Bayes Classifier* untuk klasifikasi dapat dilakukan dengan rumus berikut:

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^n P(x_i|Y)}{P(X)} \dots\dots\dots(1)$$

Dimana,  
 $P(Y|X)$  = Probabilitas data dengan vector X pada kelas Y  
 $P(Y)$  = Probabilitas awal kelas Y  
 $P(x_i|Y)$  = Probabilitas X berdasarkan kondisi hipotesis H  
 $P(X)$  = Probabilitas X

### C. Regresi Linear

Regresi Linear merupakan Metode Regresi dimana persamaan yang dihasilkan berupa persamaan linear. Berdasarkan pada persamaan yang dihasilkan, dapat dihitung prediksi dengan memasukkan nilai-nilai variabel prediktor pada persamaan tersebut. Berdasarkan pada proses tersebut dapat dihasilkan nilai prediksi variabel respon. Variabel adalah besaran yang berubah-ubah nilainya. Dalam penelitian ini, variabel atau atribut yang digunakan yaitu harga, rasa, jenis, jumlah per dus, kemasan dan berat bersih.

Beberapa penelitian menggunakan Metode Regresi Linear memiliki nilai akurasi yang tinggi, hal tersebut berdasarkan pada nilai *error* yang kecil terhadap penelitian tersebut dalam melakukan peramalan penjualan [6] dan dari hasil perhitungan menggunakan enam metode *forecasting* maka didapat hasil MAD sebesar 46,7289 dan MSE sebesar 3.436,6822 sehingga dalam melakukan perhitungan selanjutnya dapat menggunakan hasil dari Metode Regresi Linear dengan peramalan siswa tahun ajaran 2011/2012 sebesar 603 siswa [2] serta penelitian sebelumnya dengan variabel *predictor* berupa nilai evaluasi dan variabel *respon* berupa nilai akhir atau UAS yang menghasilkan nilai *error* sebesar 0.05-0.08 dari dataset yang digunakan [1].

Perhitungan Regresi Linear dapat dilakukan dengan persamaan berikut:

$$\beta_1 = \frac{\sum_{i=1}^n y_i x_i - \frac{(\sum_{i=1}^n y_i)(\sum_{i=1}^n x_i)}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} \dots\dots\dots(2)$$

Dimana,  
 $y$  = data uji  
 $x$  = persentase produktif (P.E) rasio  
 $n$  = Total jumlah terjadinya variabel

## III. METODOLOGI PENELITIAN

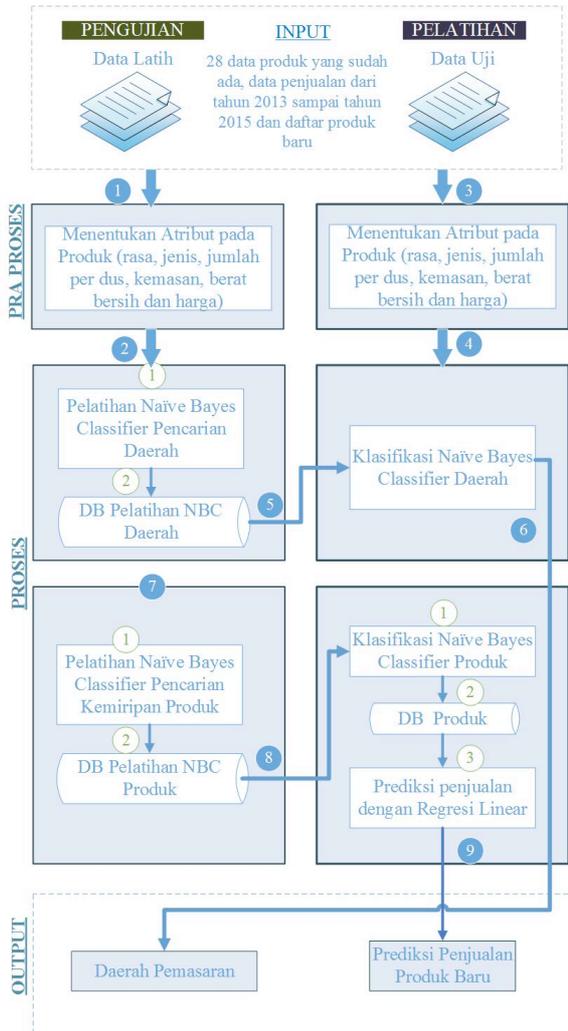
### A. Analisis Sistem Berjalan

Analisis Sistem berjalan digunakan untuk mengidentifikasi permasalahan yang terjadi di perusahaan untuk selanjutnya dapat menghasilkan perancangan sistem yang sesuai dengan kebutuhan CV. Lampegan Jaya. Aktivitas pengajuan permintaan produk baru dilakukan dari konsumen sampai kepada *manager* pemasaran atau bersifat *button up*.

### B. Gambaran Umum Sistem

Perancangan sistem untuk prediksi penjualan produk baru dan daerah pemasarannya berdasarkan data produk sebelumnya dan data produk baru, menghasilkan data latih untuk pelatihan di *Naïve Bayes Classifier* sebanyak 28 data produk. Sebelum dilakukan pelatihan menggunakan NBC, sebelumnya harus dilakukan proses penentuan atribut dan perhitungan mean serta

varian, setelah data produk baru dan daerah pemasarannya telah ditemukan maka selanjutnya dilakukan pencarian data penjualan produk baru menggunakan metode *Regresi Linear*. Sistem prediksi penjualan produk baru dapat dilihat pada Gambar 1.

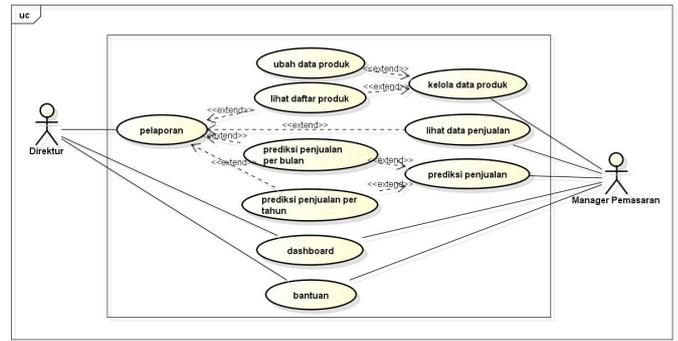


Gambar 1. Gambaran umum sistem prediksi penjualan produk baru

### C. Perancangan Sistem

#### 1) Use Case Diagram

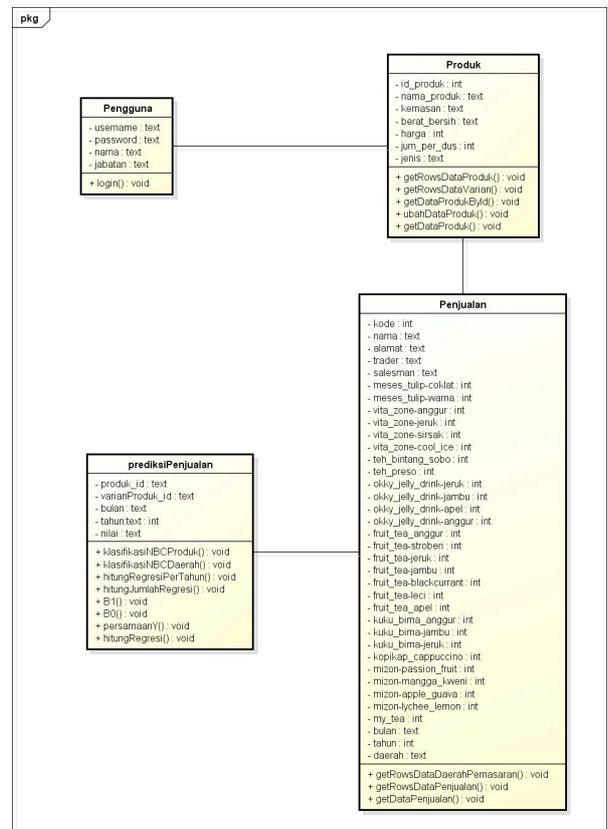
*Use Case Diagram* adalah teknik untuk merekam persyaratan fungsional sebuah sistem. Sistem mendeskripsikan interaksi tipikal antara pengguna sistem dengan sistem itu sendiri. *Use Case Diagram* digambarkan dengan aktor dan *use case*. Aktor menggambarkan siapa saja yang terlibat dalam menggunakan sistem, *use case* pada sistem prediksi penjualan produk baru digambarkan seperti pada Gambar 2.



Gambar 2. Use case diagram prediksi penjualan produk baru

#### 2) Class Diagram

*Class Diagram* adalah diagram yang menggambarkan *class-class* yang bekerja pada sistem. Terdapat beberapa *class* yang saling terhubung dan berkaitan pada Sistem Prediksi Penjualan Produk Baru, seperti pada Gambar 3.



Gambar 3. Class diagram sistem prediksi penjualan produk baru

#### 3) Data Masukan

Data masukan untuk sistem prediksi penjualan produk baru menggunakan data produk sebelumnya sebanyak 28 data, 6 data produk baru dan data penjualan produk sebelumnya selama tiga tahun dari tahun 2013 sampai tahun 2015.

#### 4) Perhitungan Mean dan Varian

Perhitungan *mean* dan *varian* dilakukan setelah atribut pada produk telah ditentukan, atribut yang terdapat pada sistem

prediksi penjualan produk baru yaitu nama, nama produk, kemasan, berat bersih, harga, jumlah per dus, jenis dan rasa, seperti pada Tabel I.

TABEL I. DATA LATIH KLASIFIKASI PRODUK

Nama	Nama Produk	.....	Daerah
Warnas/Diah	Vita Zone Anggur	.....	Bandung
Bp. Edi	Vita Zone Anggur	.....	Bandung
PB. Riksa	Teh Bintang Sobro	.....	Bandung
Ibu Ai	Teh Bintang Sobro	.....	Bandung
.....	.....	.....	.....
Ratna, Ibu	Okky Jelly Drink Apel	.....	Cimahi

Setelah atribut pada produk telah ditentukan, maka dilakukan perhitungan *mean* dan varian untuk masing-masing kelas yang memiliki atribut yang menggunakan tipe numerik, pada Tabel I terdapat atribut harga dan jumlah per dus yang menggunakan tipe numerik, maka selanjutnya dilakukan perhitungan sebagai berikut:

Perhitungan mean dan varian untuk atribut harga sebagai berikut:

$$\bar{x}_{Bandung} = 3227,78$$

$$\bar{x}_{Cilenyi} = 4841,67$$

$$\bar{x}_{Cimahi} = 6455,56$$

$$S^2_{Bandung} = 86845828,90$$

$$S_{Bandung} = \sqrt{86845828,90} = 9319,11$$

$$S^2_{Cilenyi} = 119535589,70$$

$$S_{Cilenyi} = \sqrt{119535589,70} = 10933,23$$

$$S^2_{Cimahi} = 145477804,72$$

$$S_{Cimahi} = \sqrt{145477804,72} = 12061,41$$

Perhitungan mean dan varian untuk atribut jumlah per dus sebagai berikut:

$$\bar{x}_{Bandung} = 2$$

$$\bar{x}_{Cilenyi} = 3$$

$$\bar{x}_{Cimahi} = 4$$

$$S^2_{Bandung} = 34,35$$

$$S_{Bandung} = \sqrt{34,35} = 5,86$$

$$S^2_{Cilenyi} = 47,45$$

$$S_{Cilenyi} = \sqrt{47,45} = 6,88$$

$$S^2_{Cimahi} = 58$$

$$S_{Cimahi} = \sqrt{58} = 7,6$$

### 5) Perhitungan Nilai Probabilitas

Tahap selanjutnya setelah melakukan perhitungan *mean* dan varian terhadap data produk sebelumnya maka dilakukan perhitungan nilai probabilitas untuk atribut dengan tipe numerik, yaitu harga dan jumlah per dus pada data produk baru, seperti pada Tabel II.

TABEL II. DATA UJI

Nama Produk	Kemasan	Berat Bersih	Harga	.....	Rasa
TehKita	Cup	240 ml	20600	.....	Teh Alami

Perhitungan nilai probabilitas untuk atribut dengan tipe numerik dilakukan menggunakan rumus:

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(x_i - \mu_i)^2}{2\sigma^2} \right)$$

$$P(\text{Harga} = 20600 | \text{Bandung}) = 0,9988$$

$$P(\text{Harga} = 20600 | \text{Cilenyi}) = 0,9996$$

$$P(\text{Harga} = 20600 | \text{Cimahi}) = 0,9994$$

$$P(\text{Jumlah per Dus} = 24 | \text{Bandung}) = 0,3151$$

$$P(\text{Jumlah per Dus} = 24 | \text{Cilenyi}) = 0,0031$$

$$P(\text{Jumlah per Dus} = 24 | \text{Cimahi}) = 0,5135$$

### 6) Pelatihan Naïve Bayes Classifier

Tahap selanjutnya setelah nilai probabilitas setiap atribut dengan tipe numerik yaitu harga dan jumlah per dus didapatkan, kemudian melakukan pencarian probabilitas atribut dan kelas, seperti pada Tabel III.

TABEL III. PROBABILITAS FITUR DAN KELAS

Kemasan		.....	Rasa
Bandung	.....	Bandung	.....
botol plastic PET = 6	.....	Anggur = 3	.....
Cup = 12	.....	Jeruk = 3	.....
Harga		Jumlah Per Dus	
Bandung		Bandung	
Bandung = 3227,78		Bandung = 2	
Bandung = 86845828,90		Bandung = 34,35	
S <sub>Bandung</sub> = 9319,11		S <sub>Bandung</sub> = 5,86	
		Cimahi	
		Anggur = 1	
		Jeruk = 1	
		Kelas	
		Bandung	
		Bandung = 18	
		P(Bandung) = 18/39 = 0,46	

### 7) Klasifikasi Naïve Bayes Classifier

Setelah nilai probabilitas atribut dan kelas sudah didapatkan, maka selanjutnya melakukan perhitungan probabilitas akhir untuk setiap kelas terhadap data produk baru atau data uji

$$P(X | Bandung) = P(\text{Kemasan} = \text{Cup} | \text{Bandung}) \times P(\text{Berat} = 240 \text{ ml} | \text{Bandung}) \times P(\text{Jenis} = \text{Teh} | \text{Bandung}) \times P(\text{Rasa} = \text{Teh Alami} | \text{Bandung}) \times P(\text{Harga} = 20600 | \text{Bandung}) \times P(\text{Jumlah per Dus} = 24 | \text{Bandung})$$

$$= 12/18 \times 4/18 \times 9/18 \times 9/18 \times 0,9988 \times 0,3151$$

$$= 0,67 \times 0,22 \times 0,5 \times 0,5 \times 0,9988 \times 0,3151 = 0,01159$$

$$P(X | Cilenyi) = P(\text{Kemasan} = \text{Cup} | \text{Cilenyi}) \times P(\text{Berat} = 240 \text{ ml} | \text{Cilenyi}) \times P(\text{Jenis} = \text{Teh} | \text{Cilenyi}) \times P(\text{Rasa} = \text{Teh Alami} | \text{Cilenyi}) \times P(\text{Harga} = 20600 | \text{Cilenyi}) \times P(\text{Jumlah per Dus} = 24 | \text{Cilenyi})$$

$$= 7/12 \times 1/12 \times 5/12 \times 5/12 \times 0,9996 \times 0,0031$$

$$= 0,58 \times 0,08 \times 0,41 \times 0,41 \times 0,9996 \times 0,0031 = 0,00002$$

$$P(X | Cimahi) = P(\text{Kemasan} = \text{Cup} | \text{Cimahi}) \times P(\text{Berat} = 240 \text{ ml} | \text{Cimahi}) \times P(\text{Jenis} = \text{Teh} | \text{Cimahi}) \times P(\text{Rasa} = \text{Teh Alami} | \text{Cimahi}) \times P(\text{Harga} = 20600 | \text{Cimahi}) \times P(\text{Jumlah per Dus} = 24 | \text{Cimahi})$$

$$= 7/9 \times 2/9 \times 6/9 \times 6/9 \times 0,9994 \times 0,5135$$

$$= 0,78 \times 0,22 \times 0,67 \times 0,67 \times 0,9994 \times 0,5135 = 0,03953$$

Selanjutnya, nilai tersebut dimasukkan untuk mendapatkan probabilitas akhir.

$$P(\text{Bandung} | X) = \alpha \times 0,46 \times 0,01159 = 0,005331\alpha$$

$$P(\text{Cilenyi} | X) = \alpha \times 0,30 \times 0,00002 = 0,000006\alpha$$

$$P(\text{Cimahi} | X) = \alpha \times 0,23 \times 0,03953 = 0,009091\alpha$$

$$\alpha = 1/P(\text{Bandung} | X) = 1/0,005331 = 187,58$$

$$\alpha = 1/P(\text{Cilenyi} | X) = 1/0,000006 = 166666,67$$

$$\alpha = 1/P(\text{Cimahi} | X) = 1/0,009091 = 109,99$$

Nilai akhir terbesar ada di kelas Cimahi, data uji TehKita diprediksi sebagai kelas Cimahi dan produk TehKita memiliki kesamaan spesifikasi dengan produk Teh Preso.

#### 8) Prediksi Penjualan dengan Regresi Linear

Setelah diketahui produk yang memiliki kesamaan spesifikasi dengan TehKita yaitu Teh Preso, selanjutnya dilakukan prediksi penjualan TehKita berdasarkan data penjualan yang dimiliki oleh Teh Preso selama tiga tahun. Pertama melakukan prediksi penjualan per bulan selama satu tahun dari bulan Januari sampai bulan Desember 2015 dengan melakukan perincian perhitungan untuk regresi linear sederhana, seperti pada Tabel IV.

TABEL IV. PERINCIAN PERHITUNGAN UNTUK REGRESI LINIER SEDERHANA

No i	Bulan Sebelumnya X	Bulan Selanjutnya Y	YiXi	X <sub>i</sub> <sup>2</sup>
1	200	196	39200	40000
2	196	201	39396	38416
3	201	196,5	39496,5	40401
4	196,5	198,5	39005,3	38612,3
5	198,5	201,5	39997,8	39402,3
6	201,5	197,5	39796,3	40602,3
7	197,5	197,5	39006,3	39006,3
8	197,5	199,5	39401,3	39006,3
9	199,5	133	26533,5	39800,3
10	133	196,5	26134,5	17689
11	196,5	234	45981	38612,3
Jumlah	2117,5	2151,5	413948	411548
Rata-rata	192,5	195,591		

Dengan menggunakan hasil perhitungan pada tabel IV, selanjutnya dilakukan perhitungan gradien  $\beta_1$  dan konstanta  $\beta_0$ , seperti berikut:

$$\beta_1 = -0,054$$

$$\beta_0 = 195,59 - (-0,054)(192,5) = 205,98$$

Selanjutnya dilakukan pencarian persamaan garis regresi. Berdasarkan hasil diatas, maka diperoleh persamaan  $Y = 205,98 + (-0,054)(12) = 205,332$ . Jadi prediksi penjualan produk TehKotak untuk bulan Januari 2016 yaitu sebesar 205,332 dus produk.

Setelah pencarian prediksi penjualan per bulan, selanjutnya dilakukan prediksi penjualan per tahun yaitu dari tahun 2013 sampai tahun 2015, dengan melakukan perincian perhitungan untuk regresi linear sederhana, seperti pada Tabel V.

TABEL V. PERINCIAN PERHITUNGAN UNTUK REGRESI LINIER SEDERHANA

No i	Bulan Sebelumnya X	Bulan Selanjutnya Y	YiXi	X <sub>i</sub> <sup>2</sup>
1	203	201	40803	41209
2	201	200	40200	40401
Jumlah	404	401	81003	81610
Rata-rata	202	200,5		

Dengan menggunakan hasil perhitungan pada tabel V, selanjutnya dilakukan perhitungan gradien  $\beta_1$  dan konstanta  $\beta_0$ , seperti berikut:

$$\beta_1 0,5$$

$$\beta_0 = 200,5 - 0,5(202) = 99,5$$

Selanjutnya dilakukan pencarian persamaan garis regresi. Berdasarkan hasil diatas, maka diperoleh persamaan  $Y=99,5+0,5(3) = 101$ . Jadi prediksi penjualan per tahun produk TehKotak untuk bulan Januari 2016 yaitu sebesar 101 dus produk.

#### IV. KESIMPULAN DAN SARAN

##### A. Kesimpulan

Berdasarkan penelitian dan pembahasan yang dilakukan dapat disimpulkan bahwa:

1. Untuk menentukan daerah pemasaran penjualan produk baru, dapat dilakukan dengan Metode Naïve Bayes Classifier, seperti pada produk TehKita yang termasuk kedalam kelas Cimahi sebagai daerah pemasarannya.
2. Produk baru yaitu TehKita memiliki kesamaan spesifikasi dengan produk Teh Preso, maka data penjualan Teh Preso dijadikan data latih untuk melakukan prediksi penjualan bagi produk baru pada bulan Januari untuk prediksi penjualan per bulan sebesar 205,332 dan prediksi penjualan per tahun sebesar 101 dus produk.

##### B. Saran

Rancangan Sistem Prediksi Penjualan Produk Baru dengan metode Naïve Bayes Classifier dan Regresi Linear diharapkan dapat membantu perusahaan dalam melakukan prediksi penjualan berdasarkan kesamaan spesifikasi dengan produk sebelumnya. Rancangan sistem Prediksi Penjualan ini, menurut penulis masih jauh dari kesempurnaan, untuk itu penulis berharap penelitian ini dikembangkan dengan metode yang dapat mempelajari data latih dengan cepat dan dapat membaca data latih dengan atribut yang memiliki arti yang sama, supaya dapat meningkatkan nilai akurasinya.

#### DAFTAR PUSTAKA

- [1] I. Ariesanti, H. Ginardi and A. Yusuf, "Pengembangan Perangkat Lunak Prediktor Nilai Mahasiswa Menggunakan Metode Spectral Clustering dan Bagging Regresi Linear," *Jurnal Teknik ITS*, vol. 1, no. ISSN: 2301-9271, pp. 246-250, 2012.
- [2] H. Sarjono and L. Sanny, "Peramalan Jumlah Sisw A/I Sekolah Menengah Atas Swasta Menggunakan Enam Metode Forecasting," *Forum Ilmiah*, vol. 10, no. No. 2, pp. 198-208, 2013.
- [3] H. Sarjono, "Peramalan Penjualan Dengan Pendekatan Sebelas Metode Forecasting Secara Manual," *Buletin Ekonomi*, vol. 12, no. No. 1, pp. 1-14, 2014.
- [4] A. Saleh, "Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citec Journal*, vol. 2, no. No. 3, ISSN: 2354-5771, pp. 207-217, September 2015.
- [5] K. Dahal, A. Hossain, C. M. Rahman and M. F. Kabir, "Enhanced Classification Accuracy on Naive Bayes Data Mining Models," *International Journal of Computer Applications*, vol. 28, no. No. 3, pp. 9-16, 2011.
- [6] B. Reza, H. Saragih and K. Tampubolon, "Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan," *Informasi dan Teknologi Ilmiah (INTI)*, vol. I, no. No. I, ISSN: 2339-210X, pp. 93-106, 2013.

- [7] R. Panigrahi and N. Padhy, "Data Mining: A Prediction Technique for The Workers In The PR Department of Orissa (Block and Panchayat)," *International Journal of Computer Science, Engineering and Information Technology (IJCEIT)*, vol. 2, no. No. 5, pp. 19-36, 2012.
- [8] P. S. Dutta and H. Tahbilder, "Prediction Of Rainfall Using Datamining Technique Over Assam," *Indian Journal of Computer Science and Engineering (IJCSE)*, vol. 5, no. No. 2, ISSN: 0976-5166, pp. 85-90, 2014.
- [9] G. Subbalakshmi, K. Ramesh and M. C. Rao, "Decision Support in Heart Disease Prediction System Using Naive Bayes," *Indian Journal of Computer Science and Engineering*, vol. 2, no. No. 2, ISSN: 0976-5166, pp. 170-176, April-June 2011.
- [10] K. Jyoti and S. Kaur, "Predicting The Future of Car Manufacturing Industry Using Naive Bayes Classifier," *International Journal for Science and Emerging Technologies with Latest Trends*, no. ISSN No.(Online): 2250-3641, ISSN No.(Print): 2277-8136, pp. 25-34.
- [11] M. Sarosa, H. Suryono and M. Ridwan, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Jurnal EECCIS*, vol. 7, no. No. 1, pp. 59-64, 2013.
- [12] I. S. J. Djie, "Analisis Peramalan Penjualan dan Penggunaan Metode Linear Programming dan Decision TREE Guna Mengoptimalkan Keuntungan pada PT Primajaya Pantes Garment," *Jurnal The Winners*, vol. 14, no. No. 2, pp. 113-119, 2013.
- [13] J. R. G. A. K. S and S. A. S. Olaniyi, "Stock Trend Prediction Using Regression Analysis - A Data Mining Approach," *ARPN Journal of Systems and Software*, vol. 1, no. No. 4, ISSN: 2222-9833, pp. 154-157, 2011.
- [14] R. R. Patil, "Heart Disease Prediction System using Naive Bayes and Jelinek-Mercer Smoothing," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 3, no. 5, ISSN (Online): 2278-1021, ISSN (Print): 2319-5940, pp. 6787-6789, 2014.
- [15] M. S. Dhayanand and D. S. Vijayarani, "Data Mining Classification Algorithms for Kidney Disease Prediction," *International Journal on Cybernetics & Informatics (IJCI)*, vol. 4, no. No. 4, pp. 13-25, 2015.
- [16] S. Soni, U. Ansari, J. Soni and D. Sharma, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction," *International Journal of Computer Applications*, vol. 17, no. No. 8, pp. 43-48, 2011.