

## **IMPLEMENTATION OF DATA MINING WITH CLASSIFICATION AND FORECASTING METHOD USE MODEL GAUSSIAN NAÏVE BAYES FOR BUILDING STORE (STUDI CASE: TB SINAR JAYA)**

**Fernando Valentinus<sup>1</sup>, Fabian Sujono<sup>2</sup>, Ilham Ariansyah<sup>3</sup>, Dwi Ade Handayani Capah<sup>\*4</sup>**

<sup>1,2,3,4</sup>Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana, Indonesia  
Email: <sup>1</sup>[41819210001@student.mercubuana.ac.id](mailto:41819210001@student.mercubuana.ac.id), <sup>2</sup>[41819210003@student.mercubuana.ac.id](mailto:41819210003@student.mercubuana.ac.id),  
<sup>3</sup>[41818210025@student.mercubuana.ac.id](mailto:41818210025@student.mercubuana.ac.id), <sup>4</sup>[dwi.ade@mercubuana.ac.id](mailto:dwi.ade@mercubuana.ac.id)

(Naskah masuk: 23 November 2022, Revisi: 7 Desember 2022, Diterbitkan: 23 Maret 2023)

### **Abstract**

*In the construction industry, the building materials business is a necessity, one of the businesses that is currently growing rapidly is the property business. The public's need for a place to live becomes a business opportunity sought by the public. The Industry 4.0 transformation brought building materials stores to the online market, one of which is Tokopedia. TB. Sinar Jaya has a dataset of inventory and sales with a total of approximately 15,000 data rows which are updated each month. With large amounts of data, data mining and machine learning methods are needed in data management. 5 years rapid development of TB. Sinar Jaya has not been without problems, such as competition with online stores that offer lower prices than offline stores and a lack of strong marketing strategy. In this case, TB. Sinar Jaya wants help in making marketing strategy decisions by utilizing information system technology and minimizing existing problems. Based on these problems, it is necessary to implement data mining and machine learning gaussian algorithms naïve bayes to find out the average prices available on Tokopedia to increase sales and carry out classification, forecasting and TSA (Time Series Analysis) at TB Sinar Jaya. Based on the results of testing/research, the gaussian naïve Bayes algorithm has good accuracy results with an accuracy level of 0.71 and gains insight, that is, for potential buyers they can do Wishlist efficiency, and for segments that generate profits below 50% of the total profit the researcher recommends carrying out a campaign program according to customer profile in order to improve the resulting profitability.*

**Keywords:** Classification, Data Mining, Forecasting, Gaussian Naïve Bayes, Machine learning.

## **PENERAPAN DATA MINING DENGAN METODE CLASSIFICATION, FORECASTING DENGAN MODEL GAUSSIAN NAÏVE BAYES PADA TOKO BANGUNAN (STUDI KASUS: TB SINAR JAYA)**

### **Abstrak**

Dalam pembangunan, bisnis bahan bangunan adalah sebuah kebutuhan, salah satu bisnis yang saat ini bertumbuh dengan pesat yaitu bisnis properti, sehingga kebutuhan masyarakat dalam tempat tinggal menjadi peluang usaha yang dicari oleh masyarakat. Transformasi 4.0 menghadirkan toko bahan bangunan melalui toko online salah satunya adalah Tokopedia. TB. Sinar Jaya mempunyai dataset stok barang dan penjualan dengan jumlah baris data kurang lebih 15000 baris data yang diperbarui setiap satu bulan sekali. Dengan jumlah data yang besar membutuhkan metode data mining dan machine learning dalam pengelolaan data. Perkembangan pesat TB. Sinar jaya selama 5 tahun tidak lepas dari masalah, seperti persaingan dengan toko online yang menawarkan harga lebih murah dibanding toko offline dan kurang kuatnya strategi marketing. Dalam masalah ini pihak TB Sinar Jaya menginginkan bantuan dalam membuat keputusan strategi marketing dengan pemanfaatan teknologi sistem informasi dan meminimalisir masalah yang ada. Berdasarkan masalah tersebut diperlukan implementasi data mining dan machine learning algoritma gaussian naïve bayes untuk mengetahui harga rata - rata yang tersedia di Tokopedia untuk meningkatkan penjualan dan melakukan classification, forecasting dan TSA (Time Series Analysis) pada TB Sinar Jaya. Berdasarkan hasil pengujian/penelitian, algoritma gaussian naïve bayes memiliki hasil akurasi yang baik dengan tingkat akurasi 0.71 dan mendapatkan insight yaitu untuk potential pembeli bisa melakukan wishlist efficiency, dan untuk segmen yang menghasilkan profit dibawah 50% dari total profit peneliti merekomendasikan untuk melakukan program champaign sesuai customer profile agar memperbaiki profitabilitas yang dihasilkan.

**Kata kunci:** Classification, Data Mining, Forecasting, Gaussian Naïve Bayes, Machine Learning.

## 1. PENDAHULUAN

Bisnis bahan bangunan adalah tempat yang menyediakan bahan untuk tujuan konstruksi seperti semen, batu, pasir, bata merah, batako, hebel, cat, baja ringan, paralon, triplek, keramik, dan kayu. Pada sebuah toko bangunan banyak sekali jumlah data stok barang dan penjualan dengan jumlah baris data kurang lebih 15000 baris data yang diperbarui setiap satu bulan sekali. Dengan jumlah data yang besar membutuhkan metode *data mining* dan *machine learning* dalam pengelolaan data.

Metode *data mining* adalah teknik pencarian informasi yang dilakukan pada data berjumlah besar yang tersimpan pada *database*, *data warehouse*, atau media penyimpanan lainnya[1]. Informasi yang telah diambil dari data besar kemudian dapat dikelola dengan melakukan *machine learning*[2]. *Machine learning* adalah proses pengelolaan data untuk mendeteksi *outlier data*[3], dan *classification data* ke suatu model algoritma untuk melakukan *forecasting* dan mendapatkan *insights* yang berguna untuk meramalkan kejadian pada masa depan berdasarkan data pada masa lalu atau sekarang[4]. Algoritma *naïves bayes* merupakan proses *classification data* untuk menentukan pola yang dapat digunakan untuk *forecasting* sehingga bermanfaat untuk mengambil keputusan atau *insights* yang berguna berdasarkan data masa lalu atau sekarang[5]. Kelebihan dari algoritma *naïves Bayes* yaitu mudah diimplementasikan dan cepat, juga hanya membutuhkan jumlah training data yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian[6]. Karena yang diasumsikan sebagai variabel independen, maka hanya varians di suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan di matriks kovarians[7].

TB Sinar Jaya adalah toko bangunan yang menjual barang-barang yang dibutuhkan untuk melakukan pembangunan seperti besi, paku, kayu, dan sebagainya. TB Sinar Jaya melakukan pelayanan berupa penjualan dari skala kecil yang seperti perorangan dan ahli bangunan serta perusahaan besar. Dalam bisnisnya, TB Sinar Jaya hingga saat ini cukup berkembang pesat dalam 5 tahun terakhir, akan tetapi tidak lepas dari masalah yang ada seperti pencatatan transaksi penjualan dan pembelian masih dilakukan dengan cara konvensional, persaingan dengan toko online yang menawarkan harga lebih murah dibanding toko offline, ditambah dengan kurang kuatnya strategi marketing. Dalam masalah ini pihak TB Sinar Jaya menginginkan bantuan dalam membuat keputusan strategi marketing, pemanfaatan teknologi sistem informasi dan meminimalisir masalah yang ada.

Berdasarkan penelitian terdahulu, penerapan model *gaussian naïve bayes* dengan metode *classification* dan *forecasting* terbukti mudah diimplementasikan dan membutuhkan jumlah training data yang kecil untuk menentukan estimasi dalam *classification* dan *forecasting*. Penelitian ini, peneliti menggunakan pendekatan algoritma algoritma *gaussian naïve bayes* dengan metode *classification* kemudian hasil *classification* diterapkan pada metode *forecasting* dan TSA (*Time Series Analysis*) dengan *XGBoost* untuk mengetahui stok barang, harga, dan profit TB. Sinar Jaya berdasarkan dataset pada tahun 2019 sampai dengan tahun 2022.

## 2. METODE PENELITIAN

*Naïve bayes* merupakan pengklasifikasian sederhana yang digunakan untuk menghitung probabilitas dengan cara menjumlahkan frekuensi dan nilai dari data yang digunakan. *Naïve bayes* berdasar kepada teorema bayes yang digunakan untuk menghitung probabilitas dari tiap kelas dengan mengasumsikan kelas satu dengan yang lain independen (tidak ada keterkaitan)[8]. Definisi lain yaitu *naïve bayes* merupakan metode yang digunakan untuk memprediksi peluang yang akan datang berdasarkan pengalaman yang sudah terjadi sebelumnya [9]. *Naïve bayes* memiliki tingkat akurasi dan kecepatan yang lebih tinggi saat di aplikasikan ke dalam suatu *database* dengan nilai yang besar[10]. Berikut adalah rumus dari Bayes Theorem:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Penjelasan:

X :Data yang kelasnya tidak diketahui.

H :Data hipotesis X pada sebuah kelas.

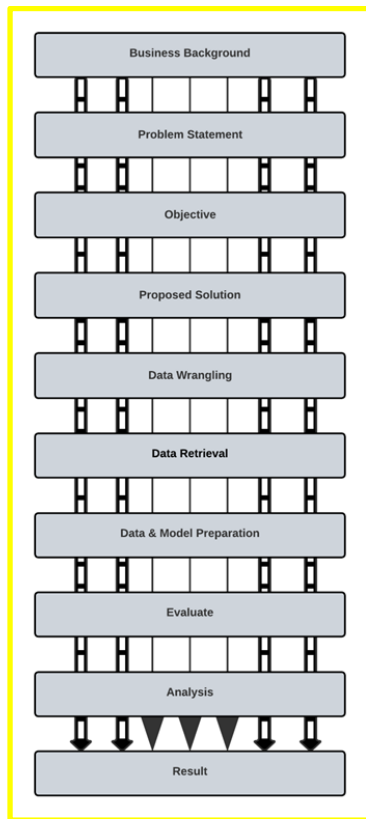
P(H|X) :Probabilitas dari hipotesis berdasarkan dari kondisi (*posteriori probabilitas*).

P(H) : Probabilitas dari hipotesis H (sebelum probabilitas).

P(X|H) :Probabilitas dari X berdasarkan dari kondisi dalam hipotesis.

H P(X) :Probabilitas X ([11]).

Tahapan pada penelitian ini terdapat 10 tahapan, *Business background*, *Proposed Solution*, *Objective*, *Problem Statement*, *Data Wrangling*, *Data Retrieval*, *Data & Model Preparation*, *Evaluate*, *Analysis*, Dan *Result*. Alur tahap penelitian terdapat pada gambar 1



Gambar 1. Diagram Alir Penelitian

### 2.1. Business Background

Bisnis bahan bangunan adalah bisnis yang menjual bahan untuk tujuan konstruksi seperti batu, semen, pasir, cat, dan bahan - bahan bangunan lainnya. TB. Sinar Jaya mempunyai dataset stok barang dan penjualan dengan jumlah baris data kurang lebih 15000 baris data yang diperbarui setiap satu bulan sekali. Dengan jumlah data yang besar membutuhkan metode *data mining* dan *machine learning* dalam pengelolaan data. TB Sinar Jaya cukup berkembang pesat dalam 5 tahun terakhir, akan tetapi tidak lepas dari masalah yang ada seperti persaingan dengan toko online yang menawarkan harga lebih murah dibanding toko offline, ditambah dengan kurang kuatnya strategi marketing.

### 2.2. Problem Statement

TB. Sinar Jaya bisa mencapai *taker rate* lebih tinggi untuk meningkatkan penghasilan B2B & B2C.

### 2.3. Objective

Penerapan predictive model untuk machine learning. Mendefinisikan *taker* dan *non taker* untuk menaikan efektivitas campaign. Merekomendasikan berdasarkan hasil *classification* dan *forecasting*[12].

### 2.4. Proposed Solution

Inisiasi strategi *campaign* untuk prediksi *takers*:

1. Produk yang menghasilkan *profit* diatas 50% akan masuk ke list potential *takers*.
2. Produk yang menghasilkan *profit* dibawah 50% akan masuk ke segment *Prosperous Data Eyeless* dengan memberikan extra diskon untuk meningkatkan produk *attractiveness*.

3. Meningkatkan *taker* dari penjualan produk untuk membantu dalam peningkatan *customer lifetime value*.
4. Meningkatkan *whitelist efficiency* dengan target potential *takers*.
5. Evaluasi produk berdasarkan program *campaign* yang terdapat pada *customer profile*.

### 2.5. Data Wrangling

Proses merubah data mentah pada data TB. Sinar Jaya dan hasil *data mining* tokopedia menjadi format yang lebih rapih yang berisi bermacam-macam tipe data supaya data tersebut lebih mudah untuk dianalisis[13].

### 2.6. Data Retrieval

Peneliti mendapatkan sumber data dari TB. Sinar Jaya yang berlokasi di Jl. Raya Leuwinanggung No.44, Leuwinanggung, Kec. Tapos, Kota Depok, Jawa Barat 16456 pada tanggal 12 february 2022 dengan menggunakan metode observasi, wawancara, dengan datang langsung ke TB Sinar Jaya, dan dokumentasi digunakan untuk memahami model bisnis yang berjalan pada TB. Sinar Jaya dan untuk menentukan system requirement yang dibutuhkan dalam membuat aplikasi. Jumlah responden dari TB. Sinar Jaya adalah tiga yaitu pemilik, *general manager*, dan *head of marketing* TB. Sinar Jaya.

### 2.7. Data & Model Preparation

1. *Feature Engineering*: pembuatan data pada variabel baru, yang berada diluar dari data *training*[14].
2. *Feature Selection*: mengurangi variabel yang masuk dengan cara menggunakan data - data yang relevan dan mengurangi *noise* pada data[15].
3. *Model Selection*: pemilihan model machine learning yang tepat untuk digunakan pada dataset[16].

### 2.8. Evaluate

Evaluasi hasil model *machine learning* berdasarkan:

1. *Model accuracy*>80%  
Tujuan *model accuracy* untuk mempelajari pola dataset bertujuan untuk dataset dapat bekerja dengan baik dengan model yang sudah ditentukan.
2. *Model precision*>80%  
Tujuan *model precision* menggunakan statistik untuk *forecasting* dengan *prediktif modelling* berdasarkan hasil *data mining* untuk menganalisis data historis dan menghasil *model prediksi* yang berguna untuk dimasa depan.
3. *Model Sensitivity*>80%  
Tujuan *model sensitivity* untuk mendeteksi *True Positive Rate* (TPR) atau *recall* dan digunakan untuk mengevaluasi hasil modelling dengan melihat TPR yang dapat diidentifikasi model dengan benar.

4. *F1 Score* > 80%  
Tujuan *F1 Score* untuk *harmonic mean* hasil *precision* dan *recall*.

2.9. Analysis

1. *Exploratory Data Analysis*: profil berdasarkan *customer taker* and *nontaker*. Evaluasi produk berdasarkan tingkat penjualan dan *profitabilitas*.
2. *Predictive Model*: data yang dapat dibaca oleh *Machine Learning*. Memberikan label pada data sebagai ‘*Taker*’ dan ‘*Nontaker*’. *Customer profile* berdasarkan hasil *classification*.
3. *Campaign Strategy*: memberikan strategi *campaign* efektif dan detail.

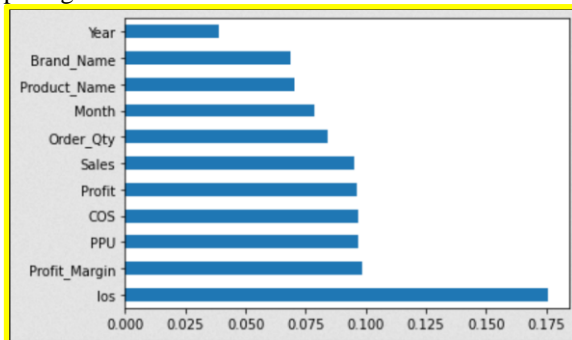
2.10. Result

Rekomendasi berdasarkan hasil akhir dari *classification*, *forecasting*, dan *TSA (time series analysis)* yang telah dilakukan terhadap produk - produk pilihan TB. Sinar Jaya.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Feature Importance

Pada tahap ini penentuan *feature of importance* dari dataset TB. Sinar Jaya untuk menentukan *feature - feature* yang memiliki *importance value* tinggi pada dataset. Hasil dari *feature importance* dapat dilihat pada gambar 1.



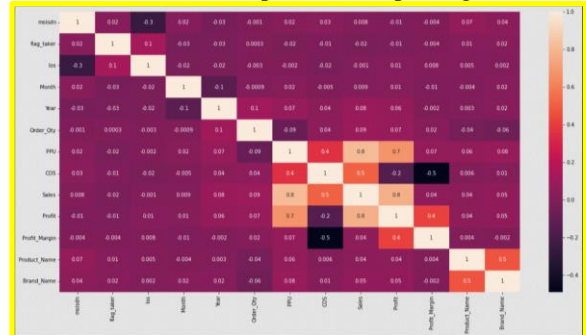
Gambar 1. Feature Importance for better visualization

Pada tahap selanjutnya digunakan untuk melihat feature dataframe yang berguna untuk modelling. Hasil dari *feature importance* dapat dilihat pada gambar 2.

	feature	feature_importance
0	los	0.17584148882539075
1	Profit_Margin	0.09842445763515673
2	PPU	0.0966371963036189
3	COS	0.09659549758036375
4	Profit	0.09618727637696879
5	Sales	0.09531332501816661
6	Order_Qty	0.08421117071703468
7	Month	0.07880809336427846
8	Product_Name	0.07017177872425664
9	Brand_Name	0.06876171480762998
10	Year	0.03904800647134776

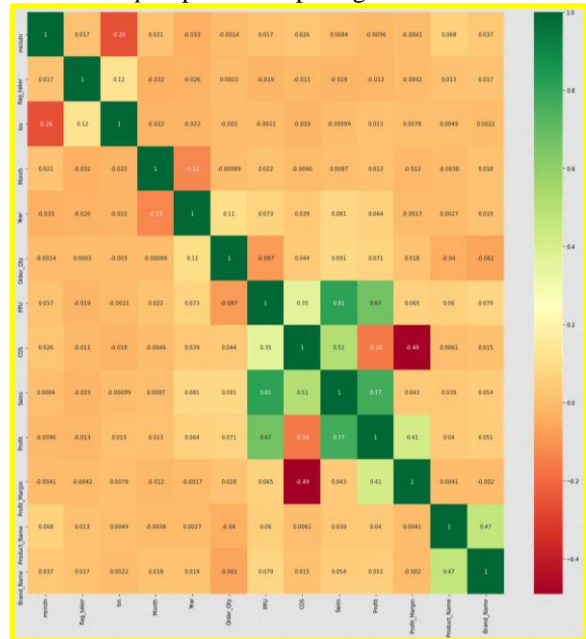
Gambar 2. Feature Importance dataframe

Setelah penentuan feature of importance dilanjutkan dengan *Correlation matrix* pada setiap kolom dataset. Seperti korelasi kolom PPU (*Price Per Unit*) dengan COS (*Cost Of Sale*), *sales* dan *profit*. Hasil *Correlation matrix* dapat dilihat pada gambar 3.



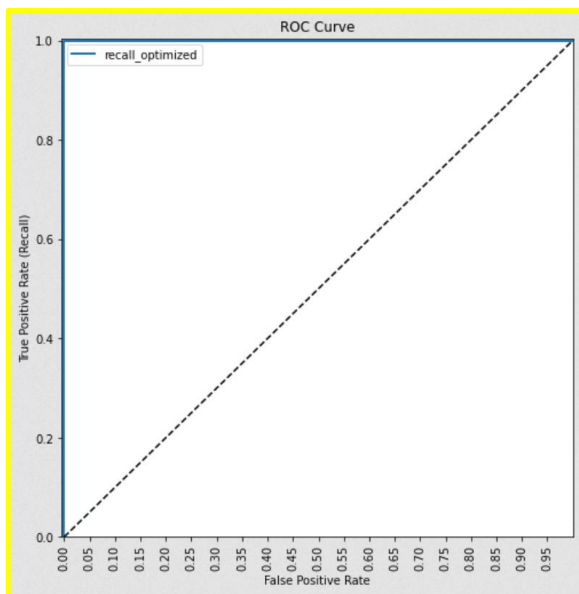
Gambar 3. Correlation matrix

Selanjutnya *heatmap* untuk melihat korelasi antar dataframe dengan menggunakan *heatmap*. tinggi pada korelasi dibedakan per angka dan warna. Hasil dari *heatmap* dapat dilihat pada gambar 4.



Gambar 4. Heatmap

Tahapan selanjutnya *ROC (Receiver Operator Curve)* berfungsi sebagai alat ukur *performance* untuk *classification problem* dalam menentukan *threshold* dan hasil *ROC* memberikan hasil *good precision* berdasarkan taker dengan mendekati 0.79 ditunjukkan oleh gambar 5.



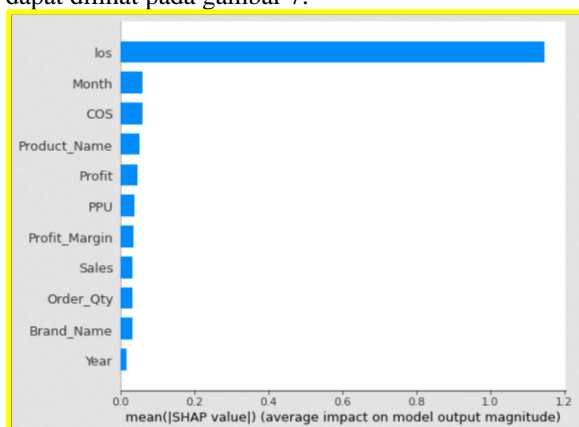
Gambar 5. ROC Curve

Tahapan selanjutnya yaitu *confusion matrix* yang digunakan untuk pengukuran performa machine learning dengan menunjukkan hasil accuracy 0.71, Specificity 0.96, dan Sensitivity 0.039. Hasil dari *confusion matrix* dapat dilihat pada gambar 6.

Confusion Matrix :  
 [[705 22]  
 [267 11]]  
 Accuracy : 0.7124378109452736  
 Specificity : 0.9697386519944979  
 Sensitivity : 0.039568345323741004

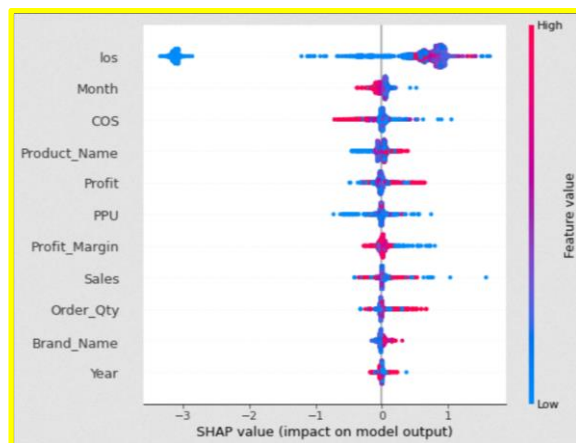
Gambar 6. Confusion Matrix.

Tahap berikutnya adalah visualisasi hasil rata rata *impact* berdasarkan model *output magnitude* dengan menggunakan *bar chart*. Hasil dari visualisasi rata rata *impact* berdasarkan model *output magnitude* dapat dilihat pada gambar 7.



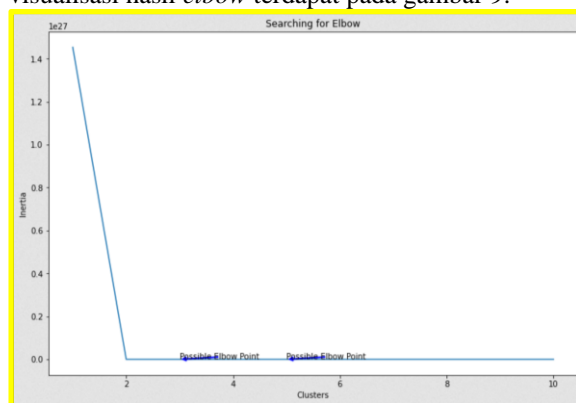
Gambar 7. Mean SHAP Value

Tahap selanjutnya adalah visualisasi hasil *feature value* berdasarkan *impact* pada *model output* dengan indikator semakin ke kanan dan berwarna merah mempunyai *impact* semakin tinggi. Hasil dari visualisasi hasil *feature value* berdasarkan *impact* pada *model output* terdapat pada gambar 8.



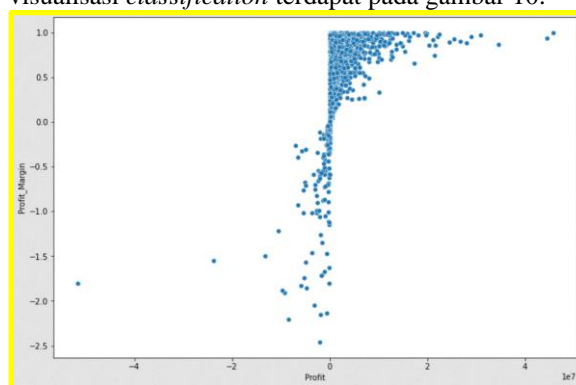
Gambar 8. SHAP Value

Tahapan berikutnya adalah visualisasi hasil *elbow* yang akan digunakan pada *classification*. Hasil dari visualisasi hasil *elbow* terdapat pada gambar 9.



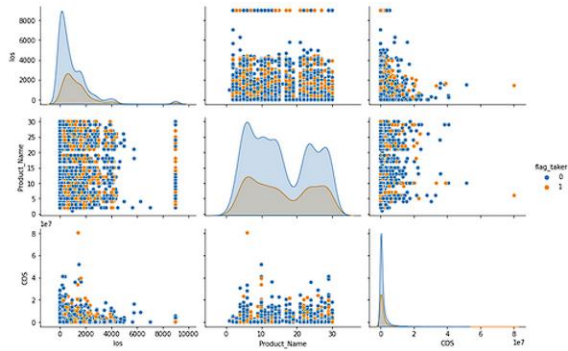
Gambar 9. Elbow

Tahapan selanjutnya adalah visualisasi dari *classification* yang akan digunakan antara kolom *profit margin* dengan kolom *profit*. Hasil dari visualisasi *classification* terdapat pada gambar 10:



Gambar 10. Classification

Tahap berikutnya adalah visualisasi dari *taker* dan *non taker* pada setiap kolom yang akan digunakan sebagai *insight* untuk Tb. Sinar Jaya. Hasil dari visualisasi *taker* dan *non taker* terdapat pada gambar 11.



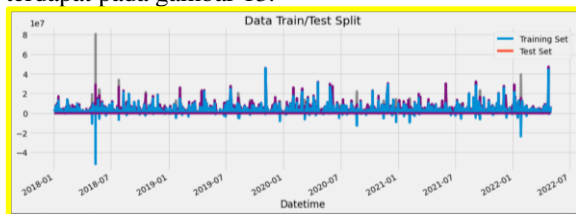
Gambar 11. Flag Taker

Tahap berikutnya adalah hasil RSME yaitu hasil evaluasi dan tuning algoritma yang terdiri dari *accuracy*, *macro avg*, dan *weighted avg*. Hasil RSME dapat dilihat pada gambar 12.

	RMSE			
	Precision	Recall	f1-score	support
Accuracy	-		0.71	1005
Macro Avg	0.52	0.50	0.45	1005
Weighted Avg	0.61	0.71	0.45	1005

Tabel 12. RMSE

Tahap berikutnya pembuatan model sebelum dilakukannya prediksi dengan membagi *dataframe* menjadi 2 *dataset* baru. Visualisasi dari model terdapat pada gambar 13.



Gambar 13. Data Train/Test Split

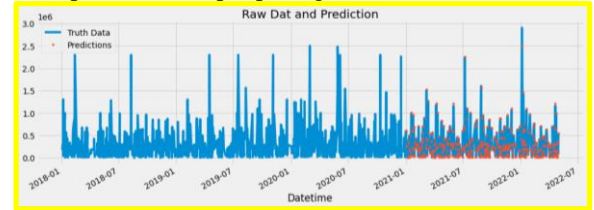
Tahap berikutnya *regression* yaitu untuk memprediksi hubungan antara dua variabel, setelah dibuatnya model kemudian model tersebut dijalankan dengan menggunakan *machine learning regression*. Proses *regression* dapat dilihat pada gambar 14.

```

XGBRegressor
early_stopping_rounds=50, enable_categorical=False,
eval_metric=None, gamma=0, gpu_id=-1, grow_policy='depthwise',
importance_type=None, interaction_constraints='',
learning_rate=0.01, max_bin=256, max_cat_to_onehot=4,
max_delta_step=0, max_depth=3, max_leaves=0, min_child_weight
=1,
missing=nan, monotone_constraints=('',), n_estimators=1000,
n_jobs=0, num_parallel_tree=1, objective='reg:linear',
predictor='auto', random_state=0, reg_alpha=0, ...
    
```

Gambar 14. Regression

Tahap berikutnya adalah visualisasi perbandingan antara hasil prediksi dengan *dataframe*. Visualisasi hasil prediksi terdapat pada gambar 15.



Gambar 15. Raw Data Prediction

#### 4. DISKUSI

Berdasarkan perbandingan dengan penelitian-penelitian terdahulu diatas, keunggulan dari penelitian kami terletak pada teknik pengambilan data, dimana dataset yang digunakan pada penelitian kami belum digunakan oleh penelitian-penelitian sebelumnya, dan dengan *classification* dan *forecasting* dengan algoritma *gaussian Naïve bayes* sebagai perbandingan dengan dataset Tokopedia yang didapatkan melalui *data mining* pada situs Tokopedia. Tujuannya adalah untuk analisis *forecasting* dan *TSA (Time Series Analysis)* stok barang, harga, *profit* TB. Sinar Jaya.

#### 5. KESIMPULAN

Penggunaan *decision support* dan *intelligence system* menggunakan data mining dengan metode *classification*, *forecasting* dengan algoritma *gaussian naïve bayes* pada TB. Sinar Jaya dapat memberikan solusi bagi TB. Sinar Jaya dalam menentukan strategi marketing dan pemanfaatan teknologi informasi. Dengan sistem ini pun dapat dijadikan media dalam memberikan informasi kepada TB. Sinar Jaya mengenai penerapan algoritma *gaussian naïve bayes* dalam menentukan *classification* dan *forecasting* dengan perbandingan terhadap toko online yaitu Tokopedia dengan *insight* berupa *taker* dan *non taker*. Dataset yang telah dibagi, dilatih untuk digunakan untuk memprediksi *taker* dan *non taker rate* dari hasil prediksi dataset pada tahun berikutnya dengan kategori stok, harga, barang terjual dan *profit* TB. Sinar Jaya.

Namun penelitian ini memiliki kekurangan. Secara khusus ada beberapa aspek yang dapat di kembangkan salah satunya adalah dengan penambahan algoritma *chi square* atau *K-means* untuk dapat meningkatkan akurasi dari hasil prediksi.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada perusahaan TB. Sinar Jaya dan Fakultas Ilmu Komputer Universitas Mercu Buana yang telah mendukung penelitian ini.

## DAFTAR PUSTAKA

- [1] P. N. Ardiansa, M. N. Tentua, and M. Fairuzabadi, "PERAMALAN PENJUALAN MENGGUNAKAN DATA MINING BERBASIS WEB," *Jurnal Dinamika Informatika*, vol. 7, pp. 57–64, 2018.
- [2] E. Sutoyo and A. Almaarif, "Educational Data Mining for Predicting Student Graduation Using the Naïve Bayes Classifier Algorithm," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 1, pp. 95–101, Feb. 2020, doi: 10.29207/resti.v4i1.1502.
- [3] A. Ambarwari, Q. Jafar Adrian, and Y. Herdiyeni, "Analysis of the Effect of Data Scaling on the Performance of the Machine Learning Algorithm for Plant Identification," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 1, pp. 117–122, Feb. 2020, doi: 10.29207/resti.v4i1.1517.
- [4] K. Wabang, Oky Dwi Nurhayati, and Farikhin, "Application of The Naïve Bayes Classifier Algorithm to Classify Community Complaints," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 5, pp. 872–876, Nov. 2022, doi: 10.29207/resti.v6i5.4498.
- [5] A. P. Wibawa *et al.*, "Naïve Bayes Classifier for Journal Quartile Classification," *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*, vol. 7, no. 2, p. 91, Jun. 2019, doi: 10.3991/ijes.v7i2.10659.
- [6] B. S. Prakoso, D. Rosiyadi, H. S. Utama, and D. Aridarma, "Klasifikasi Berita Menggunakan Algoritma Naive Bayes Classifier Dengan Seleksi Fitur Dan Boosting," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 3, no. 2, pp. 227–232, Aug. 2019, doi: 10.29207/resti.v3i2.1042.
- [7] S. Saputra, A. Yudhana, and R. Umar, "Implementation of Naïve Bayes for Fish Freshness Identification Based on Image Processing," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 3, pp. 412–420, Jun. 2022, doi: 10.29207/resti.v6i3.4062.
- [8] SUHENDRA and I. RANGGADARA, "Naive Bayes Algorithm with Chi Square and NGram Feature for Reviewing Laptop Product on Amazon Site," *International Research Journal of Computer Science*, vol. 4, pp. 28–33, 2017.
- [9] A. R. Lubis, M. Lubis, Al-Khowarizmi, and D. Listriani, "Big Data Forecasting Applied Nearest Neighbor Method," in *2019 International Conference on Sustainable Engineering and Creative Computing (ICSECC)*, Aug. 2019, pp. 116–120. doi: 10.1109/ICSECC.2019.8907010.
- [10] Z. Bi, Y. Han, C. Huang, and M. Wang, "Gaussian Naive Bayesian Data Classification Model Based on Clustering Algorithm," in *Proceedings of the 2019 International Conference on Modeling, Analysis, Simulation Technologies and Applications (MASTA 2019)*, 2019. doi: 10.2991/masta-19.2019.67.
- [11] A. Nugroho and Y. Religia, "Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 3, pp. 504–510, Jun. 2021, doi: 10.29207/resti.v5i3.3067.
- [12] Mila Desi Anasanti, Khairunisa Hilyati, and Annisa Novtariyani, "The Exploring feature selection techniques on Classification Algorithms for Predicting Type 2 Diabetes at Early Stage," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 5, pp. 832–839, Nov. 2022, doi: 10.29207/resti.v6i5.4419.
- [13] S. Sendari, I. A. E. Zaeni, D. C. Lestari, and H. P. Hariyadi, "Opinion Analysis for Emotional Classification on Emoji Tweets using the Naïve Bayes Algorithm," *Knowledge Engineering and Data Science*, vol. 3, no. 1, pp. 50–59, Aug. 2020, doi: 10.17977/um018v3i12020p50-59.
- [14] I. Santoso, Windu Gata, and Atik Budi Paryanti, "Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 3, no. 3, pp. 364–370, Dec. 2019, doi: 10.29207/resti.v3i3.1084.
- [15] Evi Purnamasari, D. P. Rini, and Sukemi, "Feature Selection using Particle Swarm Optimization Algorithm in Student Graduation Classification with Naive Bayes Method," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 3, pp. 469–475, Jun. 2020, doi: 10.29207/resti.v4i3.1833.
- [16] D. M. Tarigan, Dian Palupi Rini, and Samsuryadi, "Feature Selection in Classification of Blood Sugar Disease Using Particle Swarm Optimization (PSO) on C4.5 Algorithm," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 3, pp. 569–575, Jun. 2020, doi: 10.29207/resti.v4i3.1881.

