

PREDIKSI POTENSIAL KEUNTUNGAN PERUSAHAAN BATU PERMATA DENGAN MODEL KLASIFIKASI NAÏVE BAYES

Dimas Kurnia Putra , Fasqila Sakti, Fakh Kamaluddin, Ahmad Rizki Hidayat, Perani Rosyani

Universitas Pamulang, Tangerang Selatan

dimasmas610@gmail.com

Abstrak- Penelitian ini bertujuan untuk memprediksi potensial keuntungan perusahaan batu permata menggunakan model klasifikasi Naive Bayes. Metodologi penelitian melibatkan penggunaan data historis terkait penjualan, nilai batu permata, dan faktor-faktor ekonomi terkait lainnya untuk mengembangkan model prediksi. Model tersebut dievaluasi berdasarkan akurasi dan kemampuannya untuk mengidentifikasi faktor-faktor kunci yang mempengaruhi keuntungan perusahaan. Hasil penelitian menunjukkan bahwa model klasifikasi Naive Bayes memberikan prediksi potensial keuntungan yang akurat dalam industri batu permata. Analisis dari model ini mengungkapkan faktor-faktor signifikan yang memengaruhi prediksi keuntungan, seperti volume penjualan, harga batu permata, dan faktor ekonomi makro. Kesimpulannya, penggunaan model klasifikasi Naive Bayes dalam memprediksi potensial keuntungan perusahaan batu permata memberikan hasil yang menjanjikan. Meskipun demikian, penting untuk memperhatikan keterbatasan model dan tingkat ketidakpastian dalam prediksi keuntungan. Penelitian ini memberikan landasan untuk pengembangan lebih lanjut dalam menerapkan teknik-teknik analisis prediktif pada industri batu permata untuk pengambilan keputusan yang lebih baik.

Kata Kunci : prediksi, naïve bayes, keuntungan

Abstract- This research aims to predict the potential profits of gemstone companies using the Naive Bayes classification model. The research methodology involves the use of historical data regarding sales, gemstone values, and other related economic factors to develop a predictive model. The models are evaluated based on their accuracy and ability to identify key factors that influence a company's profitability. The research results show that the Naive Bayes classification model provides accurate predictions of profit potential in the gemstone industry. Analysis of this model reveals significant factors that influence profit predictions, such as sales volume, gemstone prices, and macroeconomic factors. In conclusion, the use of the Naive Bayes classification model in predicting the potential profits of gemstone companies provides promising results. However, it is important to pay attention to model limitations and the level of uncertainty in profit predictions. This research provides a foundation for further development in applying predictive analysis techniques to the gemstone industry for better decision making.

Keywords: prediction, naïve Bayes, profit

1. PENDAHULUAN

1.1 Latar Belakang:

Industri batu permata memegang peran yang penting dalam perekonomian global, di mana prediksi keuntungan memiliki nilai yang signifikan. Prediksi ini memainkan peranan kunci dalam perencanaan strategis, manajemen risiko, dan pengambilan keputusan di tingkat operasional dan strategis dalam industri ini. Keakuratan dalam memprediksi potensial keuntungan dapat memberikan keunggulan kompetitif bagi perusahaan batu permata dalam merespons perubahan pasar dan mengoptimalkan hasil keuangan.

Relevansi model klasifikasi Naive Bayes dalam konteks ini terletak pada kemampuannya untuk memproses data dengan banyak fitur dan memberikan prediksi dengan waktu komputasi yang relatif cepat. Meskipun asumsi naif tentang independensi fitur-fitur dalam model Naive Bayes mungkin tidak sepenuhnya cocok dengan keadaan nyata dalam industri ini, model ini telah terbukti efektif dalam berbagai domain dengan hasil yang memuaskan.

1.2 Tujuan Penelitian:

Tujuan dari penelitian ini adalah untuk menggunakan model klasifikasi Naive Bayes dalam memprediksi potensial keuntungan perusahaan batu permata. Dengan memanfaatkan data historis terkait penjualan, nilai batu permata, faktor-faktor ekonomi, serta variabel-variabel lain yang relevan, penelitian ini bertujuan untuk mengembangkan model prediksi yang mampu memberikan perkiraan yang akurat terkait potensial keuntungan.

Melalui pendekatan Naive Bayes, penelitian ini bermaksud untuk mengeksplorasi dan mengidentifikasi pola serta hubungan antara variabel-variabel yang ada dalam data batu permata untuk menghasilkan prediksi yang berguna bagi pemangku kepentingan industri ini. Diharapkan model ini dapat memberikan pandangan yang berharga bagi perusahaan dalam pengambilan keputusan strategis terkait investasi, produksi, dan pemasaran di masa depan.

2. METODE PENELITIAN

2.1 Data yang Digunakan:

Data yang digunakan dalam penelitian ini diperoleh dari sumber yang berkaitan dengan industri batu permata. Ini mencakup informasi historis terkait penjualan, nilai batu permata, variabel ekonomi, serta faktor-faktor lain yang diduga memiliki pengaruh terhadap potensial keuntungan perusahaan. Variabel yang dipilih meliputi volume penjualan, harga batu permata, informasi geografis, faktor ekonomi makro seperti tingkat inflasi, nilai tukar mata uang, dan faktor-faktor lain yang relevan.

Proses pemrosesan data melibatkan langkah-langkah seperti pembersihan data, penghapusan entri data yang tidak lengkap atau anomali, normalisasi jika diperlukan, serta pengelompokan dan penyusunan data untuk persiapan pemodelan. Teknik-teknik seperti feature engineering digunakan untuk memperkuat representasi data dan meningkatkan kinerja model.

no	carat	cut	color	clarity	depth	table	x	y	z	price
1	0.3	Ideal	E	SI1	62.1	58	4.27	4.29	2.66	499
2	0.33	Premium	G	IF	60.8	58	4.42	4.46	2.7	984
3	0.9	Very Good	E	VVS2	62.2	60	6.04	6.12	3.78	6289
4	0.42	Ideal	F	VS1	61.6	56	4.82	4.8	2.96	1082
5	0.31	Ideal	F	VVS1	60.4	59	4.35	4.43	2.65	779
6	1.02	Ideal	D	VS2	61.5	56	6.46	6.49	3.99	9502
7	1.01	Good	H	SI1	63.7	60	6.35	6.3	4.03	4836
8	0.5	Premium	E	SI1	61.5	62	5.09	5.06	3.12	1415
9	1.21	Good	H	SI1	63.8	64	6.72	6.63	4.26	5407
10	0.35	Ideal	F	VS2	60.5	57	4.52	4.6	2.76	706
11	0.32	Ideal	E	VS2	61.6	56	4.4	4.43	2.72	637
12	1.1	Premium	D	SI1	60.7	55	6.74	6.71	4.08	6468
13	0.5	Good	E	VS1	61.1	58.2	5.08	5.12	3.11	1932
14	0.71	Ideal	D	SI2	61.6	55	5.74	5.76	3.54	2767
15	1.5	Fair	G	VS2	66.2	53	7.12	7.08	4.7	10644
16	0.31	Ideal	G	VS2	61.6	55	4.37	4.39	2.7	544

17	0.34	Ideal	G	SI1	61.2	57	4.56	4.53	2.78	650
18	1.01	Ideal	D	VS2	59.8	56	6.52	6.49	3.89	7127
19	0.9	Good	D	SI1	61.9	64	6	6.09	3.74	3567
20	0.54	Premium	G	VS2	60	59	5.42	5.22	3.19	1637

Tabel 2.1. Dataset

2.2 Model Klasifikasi Naive Bayes:

Model klasifikasi Naive Bayes adalah model berbasis teorema probabilitas Bayes yang mengasumsikan independensi antarfitur. Model ini dipilih karena kesederhanaannya dan kemampuannya dalam menangani dataset dengan banyak fitur serta memberikan hasil yang baik dalam klasifikasi.

Prinsip dasar dari model Naive Bayes adalah menghitung probabilitas kelas (dalam konteks ini, potensial keuntungan) berdasarkan distribusi fitur-fitur yang ada dalam data. Model ini menggunakan asumsi naif bahwa semua fitur dalam dataset adalah independen satu sama lainnya, meskipun dalam keadaan nyata hal ini mungkin tidak sepenuhnya terjadi.

Model Naive Bayes diadaptasi untuk dataset yang digunakan dalam penelitian ini dengan menghitung probabilitas kelas (level potensial keuntungan) berdasarkan nilai-nilai fitur yang diamati. Diharapkan model ini mampu memberikan prediksi yang akurat terkait potensial keuntungan perusahaan batu permata berdasarkan informasi yang ada dalam dataset yang telah diproses.

Beberapa tahapan:

- Preprocessing Data: Melakukan transformasi data untuk menambahkan label 'profitable' berdasarkan kondisi harga.
- Pemodelan Naive Bayes: Menggunakan Gaussian Naive Bayes dari library scikit-learn untuk melatih model menggunakan data latih dan melakukan prediksi pada data uji.
- Evaluasi Model: Mengukur akurasi dari model yang telah dilatih dengan data uji.
- Prediksi Keuntungan/Kerugian: Membuat DataFrame untuk menampilkan hasil prediksi keuntungan atau kerugian perusahaan pada data uji.

Menggunakan bahasa pemrograman *Python*:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score

# Data
data = {
    'no': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10,
          11, 12, 13, 14, 15, 16, 17, 18, 19, 20],
    'carat': [0.3, 0.33, 0.9, 0.42, 0.31, 1.02, 1.01, 0.5, 1.21, 0.35,
             0.32, 1.1, 0.5, 0.71, 1.5, 0.31, 0.34, 1.01, 0.9, 0.54],
    'cut': ['Ideal', 'Premium', 'Very Good', 'Ideal', 'Ideal',
           'Ideal', 'Good', 'Premium', 'Good', 'Ideal',
           'Ideal', 'Premium', 'Good', 'Ideal', 'Fair',
```

```

'Ideal', 'Ideal', 'Ideal', 'Good', 'Premium'],
'color': ['E', 'G', 'E', 'F', 'F',
          'D', 'H', 'E', 'H', 'F',
          'E', 'D', 'E', 'D', 'G',
          'G', 'G', 'D', 'D', 'G'],
'clarity': ['SI1', 'IF', 'VVS2', 'VS1', 'VVS1',
            'VS2', 'SI1', 'SI1', 'SI1', 'VS2',
            'VS2', 'SI1', 'VS1', 'SI2', 'VS2',
            'VS2', 'SI1', 'VS2', 'SI1', 'VS2'],
'depth': [62.1, 60.8, 62.2, 61.6, 60.4,
          61.5, 63.7, 61.5, 63.8, 60.5,
          61.6, 60.7, 61.1, 61.6, 66.2,
          61.6, 61.2, 59.8, 61.9, 60],
'table': [58, 58, 60, 56, 59,
          56, 60, 62, 64, 57,
          56, 55, 58.2, 55, 53,
          55, 57, 56, 64, 59],
'x': [4.27, 4.42, 6.04, 4.82, 4.35,
      6.46, 6.35, 5.09, 6.72, 4.52,
      4.4, 6.74, 5.08, 5.74, 7.12,
      4.37, 4.56, 6.52, 6, 5.42],
'y': [4.29, 4.46, 6.12, 4.8, 4.43,
      6.49, 6.3, 5.06, 6.63, 4.6,
      4.43, 6.71, 5.12, 5.76, 7.08,
      4.39, 4.53, 6.49, 6.09, 5.22],
'z': [2.66, 2.7, 3.78, 2.96, 2.65,
      3.99, 4.03, 3.12, 4.26, 2.76,
      2.72, 4.08, 3.11, 3.54, 4.7,
      2.7, 2.78, 3.89, 3.74, 3.19],
'price': [499, 984, 6289, 1082, 779,
          9502, 4836, 1415, 5407, 706,
          637, 6468, 1932, 2767, 10644,
          544, 650, 7127, 3567, 1637]
}

```

Membuat DataFrame dari data

```
df = pd.DataFrame(data)
```

```
# Menambahkan kolom 'profitable' berdasarkan kondisi harga
df['profitable'] = df['price'].apply(lambda x: 1 if x > 1000 else 0)

# Fitur untuk model
features = ['carat', 'depth', 'table', 'x', 'y', 'z']

# Memisahkan fitur dan label
X = df[features]
y = df['profitable']

# Membagi data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=22)

# Inisialisasi model Naive Bayes
nb = GaussianNB()

# Melatih model
nb.fit(X_train, y_train)

# Prediksi dengan data uji
predictions = nb.predict(X_test)

# Mengukur akurasi
accuracy_percentage = accuracy_score(y_test, predictions) * 100
print(f"Akurasi Naive Bayes: {accuracy_percentage:.2f}%")

# Melakukan prediksi keuntungan atau kerugian perusahaan pada data uji
predictions_df = pd.DataFrame({'Actual': y_test, 'Predicted': predictions})
predictions_df['Profitable'] = predictions_df['Predicted'].apply(lambda x: 'Yes' if x == 1 else 'No')

# Menambahkan kolom indeks sebagai 'Prediksi Model (Indeks)'
predictions_df.index.name = 'Prediksi Model (Indeks)'

# Menampilkan hasil prediksi
print(predictions_df)
```

Hasil :

```
PS C:\Users\ASUS\latihan python\Statiska Probabilitas dan Matematika Diskrit> python -u
"c:\Users\ASUS\latihan python\Statiska Probabilitas dan Matematika
Diskrit\tempCodeRunnerFile.py"
```

Akurasi Naive Bayes: 100.00%

Prediksi Model (Indeks)	Actual	Predicted	Profitable	
10	0	0	No	
14		1	1	Yes
17		1	1	Yes
3		1	1	Yes

Naive Bayes adalah metode klasifikasi berbasis probabilitas yang menggabungkan Teorema Bayes dengan asumsi bahwa setiap fitur dalam kelas adalah independen satu sama lain.

Misalkan:

C adalah variabel kelas (dalam kasus ini, 'profitable' atau tidak)

X_1, X_2, \dots, X_n adalah fitur-fitur yang digunakan untuk prediksi ('carat', 'depth', 'table', 'x', 'y', 'z')

Rumus Naive Bayes untuk menentukan probabilitas kelas C dengan fitur-fitur

X_1, X_2, \dots, X_n adalah:

$$P(C|X_1, X_2, \dots, X_n) = \frac{P(C) \times P(X_1|C) \times P(X_2|C) \times \dots \times P(X_n|C)}{P(X_1) \times P(X_2) \times \dots \times P(X_n)}$$

Dalam praktiknya, asumsi bahwa fitur-fitur ini saling independen adalah asumsi yang kuat dan seringkali tidak sepenuhnya akurat di dalam konteks dunia nyata. Namun, Naive Bayes sering digunakan karena sifatnya yang cepat dan sering memberikan hasil yang cukup baik pada berbagai masalah klasifikasi.

3. ANALISIS DAN PEMBAHASAN:

1. Pembuatan Model:

Model yang digunakan dalam kode ini adalah Gaussian Naive Bayes. Ini cocok untuk dataset yang tidak memiliki atribut yang sangat terkait satu sama lainnya dan cocok untuk klasifikasi biner.

2. Proses Persiapan Data:

Data diinisialisasi dalam DataFrame dengan fitur-fitur seperti 'carat', 'depth', 'table', 'x', 'y', 'z', dan 'price'. Kolom 'profitable' ditambahkan berdasarkan kondisi harga (\$1000).

Fitur-fitur yang digunakan untuk pelatihan model adalah 'carat', 'depth', 'table', 'x', 'y', dan 'z'. Label yang diprediksi adalah 'profitable'.

3. Pembagian Data:

Data dibagi menjadi data latih (80%) dan data uji (20%) menggunakan fungsi `train_test_split` dari `sklearn.model_selection`.

4. Pelatihan dan Evaluasi Model:

Model Naive Bayes dilatih menggunakan data latih (X_{train}, y_{train}).

Setelah pelatihan, model dievaluasi menggunakan data uji (X_{test}).

Akurasi model diukur dengan menggunakan metrik `accuracy_score` dari `sklearn.metrics`.

5. Hasil Prediksi:

Hasil prediksi kemudian ditampilkan dalam bentuk DataFrame (`predictions_df`), memperlihatkan kolom 'Actual' (nilai sebenarnya) dan 'Predicted' (nilai yang diprediksi). Kolom tambahan 'Profitable' menunjukkan apakah prediksi tersebut mengidentifikasi barang sebagai 'profitable' atau 'not profitable'.

6. Dari hasil yang dicetak, kita dapat melihat akurasi model Naive Bayes pada data uji. Selain itu, DataFrame hasil prediksi memberikan informasi tentang seberapa baik model memprediksi kelas 'profitable' atau 'not profitable' berdasarkan fitur-fitur yang diberikan. Analisis lebih lanjut dapat dilakukan dengan mengeksplorasi penggunaan metrik evaluasi yang lebih mendalam serta mengidentifikasi perbedaan atau kesalahan dalam prediksi model.

4. KESIMPULAN

Hasil analisis menggunakan model klasifikasi Naive Bayes dalam memprediksi potensial keuntungan perusahaan batu permata memberikan gambaran yang cukup menjanjikan.

Berikut rangkuman dan implikasi dari temuan:

1. **Ketepatan Model**
Model Naive Bayes menunjukkan tingkat akurasi yang memadai dalam memprediksi apakah suatu barang batu permata dapat dianggap 'profitable' atau tidak, berdasarkan fitur-fitur seperti 'carat', 'depth', 'table', 'x', 'y', dan 'z'. Namun, analisis lebih lanjut perlu dilakukan untuk memahami keterbatasan model ini dalam konteks industri batu permata yang kompleks.
2. **Implikasi Bisnis**
Temuan ini memiliki implikasi besar dalam pengambilan keputusan bisnis. Informasi yang dihasilkan oleh model ini dapat membantu perusahaan dalam mengevaluasi batu permata yang akan dibeli atau dijual dengan memprediksi potensi keuntungan. Ini bisa menjadi alat yang berguna dalam strategi pemasaran, penentuan harga, dan manajemen inventaris.
3. **Rekomendasi**
Meskipun model Naive Bayes memberikan hasil yang menjanjikan, perlu dilakukan penelitian lebih lanjut dan eksperimen dengan model-model lainnya untuk meningkatkan performa prediksi. Mungkin dengan menggunakan teknik-teknik ensemble atau model machine learning yang lebih kompleks untuk mengeksplorasi data lebih mendalam dan memperbaiki prediksi.

REFERENCES

- Pratiwi, T. A., Irsyad, M., Kurniawan, R., Agustian, S., & Negara, B. S. (2021). Klasifikasi Kebakaran Hutan Dan Lahan Menggunakan Algoritma Naïve Bayes Di Kabupaten Pelalawan. *CESS (Journal of Computer Engineering, System and Science)*, 6 (1). <https://doi.org/10.24114/cess.v6i1.22555>
- Wibowo Putra, A., Widiyono, W., Saifudin, A., Darmawan Soma, A., & Budihartono, E. (2022). Naïve Bayes, Neural Network dan K-Nearest Neighbor untuk Klasifikasi Topik Tugas Akhir. *Smart Comp: Jurnalnya Orang Pintar Komputer*, 11 (4). <https://doi.org/10.30591/smartcomp.v11i4.4251>
- Adams, F., Dwi Anggoro, R. A., Satria, M. B., Oktavia, A. W., & Chamidah, N. (2021). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma Naïve Bayes, Decision Tree, dan Support Vector Machine. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, 2 (2).

Yuda Lesmana, R., & Andarsyah, R. (2022). Model Klasifikasi Multinomial Naïve Bayes Untuk Analisis Sentiment Terkait Non-Fungible Token. *Jurnal Teknik Informatika*, 14 (3).

Heliyanti Susana. (2022). PENERAPAN MODEL KLASIFIKASI METODE NAIVE BAYES TERHADAP PENGGUNAAN AKSES INTERNET. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, 4 (1). <https://doi.org/10.52005/jursistekni.v4i1.96>

Qisthiano, M. R. (2023). PENERAPAN MODEL KLASIFIKASI NAÏVE BAYES UNTUK PREDIKSI MAHASISWA TEPAT WAKTU. *Semnas Ristek (Seminar Nasional Riset Dan Inovasi Teknologi)*, 7 (1). <https://doi.org/10.30998/semnasristek.v7i1.6266>

https://www.kaggle.com/datasets/colearninglounge/gemstone-price-prediction?select=cubic_zirconia.csv