

ANALISA DATA MINING UNTUK MENGETAHUI *OUTPUT* PRODUKSI YANG *ACCEPT* PADA PT. MULIA KERAMIK INDAH RAYA MENGGUNAKAN KLASIFIKASI *NAÏVE BAYES*

Wahyu Hadikristanto, Suhardian

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa

wahyu.hadikristanto@pelitabangsa.ac.id

Disetujui, 26 September 2018

Abstraksi

PT.Mulia Keramik Indah Raya adalah perusahaan yang bergerak dibidang produksi ubin dalam berbagai jenis desain dan ukuran. PT.Mulia Keramik Indah Raya mampu memproduksi ubin sesuai dengan gaya hidup modern bagi seluruh pelanggan di dunia dan mampu memenuhi permintaan pasar dan berhasil menempatkan diri sebagai pemain terkemuka di industri ubin. Karena banyaknya permintaan pasar dapat menghasilkan informasi yang sangat banyak berupa keluaran produksi per harinya. Perusahaan mengalami kesulitan dalam menentukan hasil produksi yang dihasilkan. Kemudian akan dilakukan uji hasil produksi yang masuk dengan menggunakan salah satu metode klasifikasi yaitu metode *naive bayesian*. maka perlu dilakukannya penelitian untuk dapat memprediksi hasil keluar produksi yang masuk. sebagai tolak ukur manajemen perusahaan untuk mengetahui tingkat produksi seperti apa yang harus dilakukan kedepannya untuk mencapai target produksi.

Kata kunci: *Data Mining, Klasifikasi, Naïve Bayes dan RapidMiner*

Abstract

PT.Mulia Keramik Indah Raya is a company engaged in the production of tiles in various types of designs and sizes of PT. Mulia Keramik Indah Raya is able to produce tiles in accordance with modern lifestyles for all customers in the world and able to meet market demand and succeed in placing themselves as leading players in the tile industry. Because a lot of market demand can produce information that is very much in the form of production output per day. The company has difficulty in determining the outcome of the incoming production. Then the results of the incoming production will be tested using one of the classification methods, namely the Naive Bayesian method. hence it is necessary to conduct research to be able to predict the outcome of the incoming production. as a benchmark for company management to find out what level of production should be done in the future to achieve production targets.

Keyword: Data Mining, Classification, Naive Bayes and RapidMiner

1. Pendahuluan

Kemampuan berkompetisi antara perusahaan industri kini semakin tinggi, sehingga setiap perusahaan dituntut untuk selalu memperbaiki kinerja sistem industri yang berjalan. Kompetensi global yang terjadi dapat diantisipasi dengan peningkatan output yang sering digunakan oleh produksi dalam menyatakan jumlah unit yang berhasil diproduksi. Dalam produksi dapat diketahui output yang berhasil diproduksi per hari sangat lah penting karena semakin cepatnya *line leader* maupun manajemen mengetahui output produksi semakin cepat pula *line leader* dapat mengambil tindakan agar output produksi tercapai.

PT.Mulia Keramik Indah Raya adalah perusahaan yang bergerak dibidang produksi ubin dalam berbagai jenis desain dan ukuran. PT.Mulia Keramik Indah Raya mampu memproduksi ubin sesuai dengan gaya hidup modern bagi seluruh pelanggan di dunia dan mampu memenuhi permintaan pasar dan berhasil menempatkan diri

sebagai pemain terkemuka di industri ubin. Banyaknya data permintaan pasar dapat menghasilkan informasi yang sangat banyak berupa output produksi per harinya. Data-data tersebut dengan menggunakan algoritma *machine learning* dapat dimanfaatkan untuk mengetahui informasi penting berupa klasifikasi mengenai output produksi yang *accept* dan menentukan peningkatan mutu produksi per harinya. Metode *machine learning* yang digunakan pada penelitian ini yaitu metode Naive Bayes, untuk mengklasifikasi output produksi yang *accept*.

2. Tinjauan Pustaka

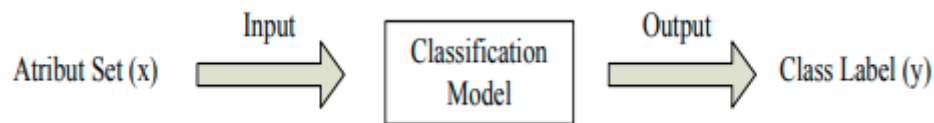
2.1. Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisa dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. *Knowledge Discovery in databases* (KDD) adalah penerapan metode saintifik pada data mining. Dalam konteks ini data mining merupakan satu langkah dari proses KDD (Hermawati Fajar Astuti, 2013, p. 3).

2.2. Metode Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data ke dalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu.

Model itu sendiri bisa berupa aturan “jika-maka”, berupa pohon keputusan, atau formula matematis.



Gambar 1. Blok Diagram Model Klasifikasi

2.3. Metode Naive Bayes

Naive Bayes merupakan salah satu algoritma dalam teknik data mining yang menerapkan teori Bayes dalam klasifikasi. Teorema keputusan *Bayes* adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*). *Naive Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu (Ridwan et al., 2013).

Naive Bayes Classifier merupakan salah satu algoritma klasifikasi dalam data mining yang memiliki kecepatan proses yang baik dan tingkat akurasi yang cukup tinggi. Dalam proses klasifikasi, *Naive Bayes Classifier* mengadopsi teorema Bayesian untuk memetakan suatu data terhadap class dengan memperhitungkan probability dari attribute data tersebut (Muktamar et al., 2015).

Naive Bayes Classifier merupakan salah satu algoritma yang digunakan dalam klasifikasi data mining. Klasifikasi atau yang disebut supervised learning adalah menentukan sebuah record data baru ke salah satu dari beberapa kategori (atau kelas) yang telah didefinisikan sebelumnya. *Naive Bayes Classifier* berbasis pada probabilitas sederhana yang berdasar pada penerapan teorema bayes (atau aturan bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam *Naive Bayes*, model yang digunakan adalah “model fitur independen” (Agustianto, Permanasari, & Hidayah, 2015).

Klasifikasi *Naive Bayes* adalah klasifikasi statistik yang bisa memprediksi probabilitas sebuah kelas. Klasifikasi *Naive Bayes* ini dihitung berdasarkan teorema Bayes sebagai berikut ini:

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

Keterangan

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

$P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X (posteriori probability)

$P(H)$: Probabilitas hipotesis H (prior probability)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

3. Metode/ Model Penelitian

A. Penentuan Objek Penelitian

Penelitian ini bertujuan untuk memprediksi hasil produksi yang *accept*. Penelitian ini dipilih karena semakin cepatnya *line leader* maupun manajemen mengetahui output produksi semakin cepat pula *line leader* dapat mengambil tindakan agar output produksi tercapai.

B. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data primer yang diambil langsung dari pihak perusahaan yang bersangkutan karena data ini bersifat *private* jadi tidak sembarang orang bisa mendapatkan data tersebut. Jumlah data hasil output produksi yang akan dijadikan dataset berjumlah 500 data yang diambil dari hasil data produksi 2017/2018. Data tersebut digunakan sebagai data training dan juga data testing.

1) Data *Training*

Data training adalah data yang digunakan untuk perhitungan probabilitas dari data berdasarkan data pembelajaran yang dilakukan. Data training yang digunakan adalah data sampel yang di dapat dari primer

2) Data *Testing*

Data testing merupakan data yang akan atau sedang terjadi dan dipergunakan sebagai bahan uji yang sebelumnya sudah didapatkan pada data training. Data testing tersebut juga menggunakan data primer.

C. Penentuan *Attribute*

Atribut-atribut yang digunakan untuk proses data mining data ini mengacu pada tujuan penelitian. Ada dua jenis variabel yang ditentukan, yaitu :

1) *Variable dependen* (Y)

Variabel dependen (Y) merupakan variabel yang nilainya terikat, bisa disebut variabel terikat. Variabel Y yang digunakan yaitu kolom Keterangan.

2) *Variabel independen* (X)

Variabel independen (X) merupakan variabel yang nilainya tidak tergantung pada nilai dari variabel lainnya atau bisa disebut sebagai variabel bebas. Variabel X yang digunakan adalah Nama Operator, model df1, model df2, model df3, dan totalPerShift

D. Data *Cleaning*

Pembersihan data perlu dilakukan supaya data yang digunakan valid sesuai kebutuhan. Sehingga dari nilai class data dalam atribut tidak terjadi ketidakkonsistenan data dalam pengujian.

E. Penggunaan Metode Naïve Bayes

Dalam perhitungan naive bayes, data set dibagi menjadi 500 data training dan 100 data testing. Berikut ini adalah perhitungan pengujian menggunakan naive bayes adalah sebagai berikut

1) Pahami setiap isi record class pada data testing.

2) Usahakan buat tabel perhitungan pada masing – masing atribut agar mudah dibaca dan dipahami.

3) Menentukan jumlah kasus yang dibagi menjadi 2 class label yaitu *target* dan *belum*.

4) Setelah itu menghitung P(H) dengan cara menghitung peluang target dan belum dibagi dengan jumlah semua data yang ada di field “keterangan”. Misalkan P(keterangan = target) = 265/500.

5) Kemudian menentukan P(X | H) dengan menghitung class yang ditentukan dengan class label dibagi dengan class label yang ditentukan. Misalkan P(nama operator = maulana | keterangan = target) = 96/265.

6) Ulangi langkah ke – 4 sampai semua atribut terhitung semuanya.

7) Hitunglah P(X) yang sudah ditentukan pada data testing dengan cara menghitung dengan cara menghitung seluruh class atribut yang sudah ditentukan lalu dikalikan antar masing – masing atribut. Contoh : P(X) = 0,0101616

8) Jika sudah diketahui semua masing – masing, maka masukkan ke peluang class label masing – masing dengan rumus

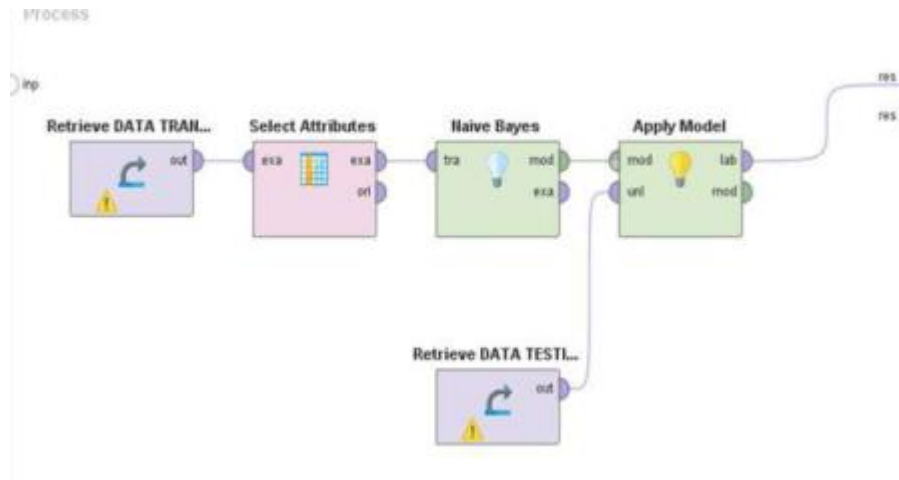
$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

- 9) Proses penentuan prediksi akan dipastikan dengan cara memilih mana yang lebih besar. Jika nilai $P(\text{keterangan} = \text{target} | X)$ lebih besar daripada nilai $P(\text{keterangan} = \text{belum} | X)$, maka prediksinya adalah target. Jika nilai $P(\text{keterangan} = \text{belum} | X)$ lebih besar daripada nilai $P(\text{keterangan} = \text{target} | X)$, maka prediksinya adalah belum.

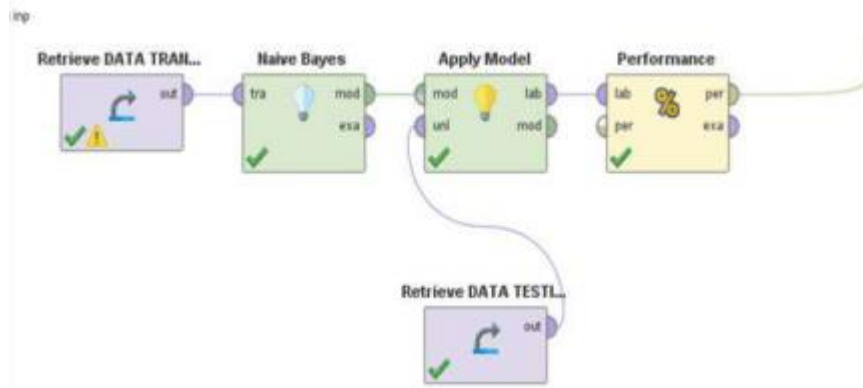
4. Hasil dan Pembahasan

A. Implementasi Software Rapid Miner

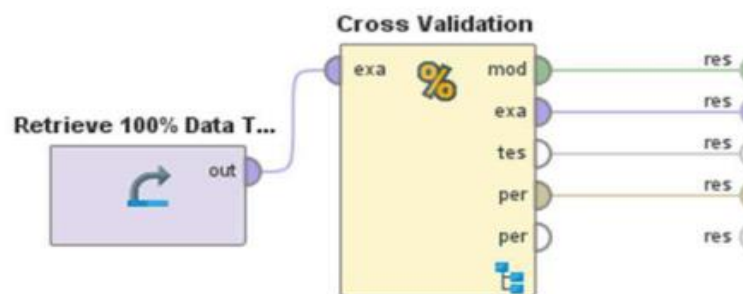
Penggunaan software ini bisa mengimpor sebuah informasi yang terdapat dari berbagai macam sumber database untuk diperiksa dan dianalisis di dalam sebuah aplikasi. Rapidminer sebagai solusi untuk memprediksi dan menganalisis komputasi statistik. Gambar 2 hingga Gambar 5 merupakan hasil penelitian yang dilakukan menggunakan aplikasi Rapid Miner.



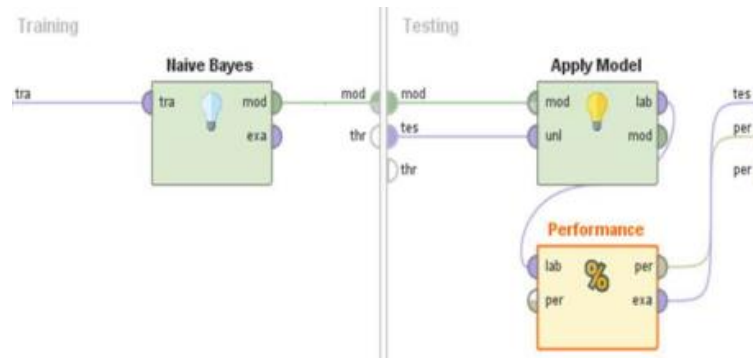
Gambar 2. Pemodelan prediksi Naïve Bayes menggunakan RapidMiner



Gambar 3. Proses performa



Gambar 4 Proses *cross validation*



Gambar 5 Proses susunan dalam box cross validation

B. Hasil Penelitian

Setelah perbandingan aktual dan prediksi dihitung sesuai golongannya, maka langkah selanjutnya adalah proses penyusunan data pada tabel confusion matrix. Pada proses penyusunan data pada tabel confusion matrix pada penelitian ini menampilkan nilai accuracy, recall dan precision. Maka hasil tabel confusion matrix pada penelitian ini adalah sebagai berikut :

	Aktual Target	Aktual Belum	Total
Pred Target	49	8	57
Pred Belum	0	43	43
Total	49	51	100

$$Accuracy = \frac{\text{Positif Target} + \text{Positif Belum}}{\text{Semua Prediksi}} = \frac{(49 + 43)}{(49 + 8 + 0 + 43)} = \frac{92}{100}$$

$$= 0,92 = 92,00 \%$$

5. Kesimpulan

Kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut :

- 1) Analisa data mining dengan metode klasifikasi dan algoritma naive bayes dapat mengetahui pola hasil output produksi yang *accept* pada perusahaan.
- 2) Evaluasi hasil klasifikasi naive bayes dalam prediksi output menggunakan 20% data testing dan 100% data testing dengan penyusunan tabel confusion matrix memiliki tingkat akurasi sebesar 92,00 % dan data testing tingkat akurasi sebesar 85,33 %. Jika dibandingkan kedua perbandingan tersebut, maka yang paling baik adalah memakai prediksi dengan 20% data testing.

Daftar Pustaka

Agustianto, K., Permasari, A. E., & Hidayah, I. (2015). TEACHER MODELING UNTUK MENDUKUNG ADAPTIVE LEARNING DALAM PROSES PEMBELAJARAN FACE TO FACE LEARNING ENVIRONMENTS (KASUS : KELAS X BIDANG STUDI KEAHLIAN TIK SMK NEGERI DI KABUPATEN PONOROGO), 6–8.

Asmono, R. T., Wahono, R. S., & Syukur, A. (2015). Absolute Correlation Weighted Naïve Bayes for Software Defect Prediction. *Journal of Software Engineering*, 1(1), 38–45.

- Asriningtyas, Y. & Mardhiyah, R., 2014. Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa. *Jurnal Informatika*, Volume 8.1.
- Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4. 5 (Studi Kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama*, 11(2), 130–138.
- Hermawati, Fajar Astuti. (2013). *Datamining*. Yogyakarta: CV ANDI OFFSET Iskandar, D., & Suprpto, Y. K. (2013). Perbandingan akurasi klasifikasi tingkat kemiskinan antara algoritma C4 . 5 dan Naïve Bayes Classifier, 11(1), 14–17.
- Jiawei, H., Kamber, M., Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. San Francisco, CA, itd: Morgan Kaufmann. [https://doi.org/ 10.1016/ B978-0-12-381479-1.00001-0](https://doi.org/10.1016/B978-0-12-381479-1.00001-0)
- Khadafy, A. R., & Wahono, R. S. (2015). THESIS - Penerapan Naïve Bayes Untuk Mengurangi Data Noise Pada Klasifikasi Multi Kelas Dengan Decision Tree. *Journal of Intelligent Systems*, 1(2), 11–17.
- Larose, D. T. (2014). *Discovering knowledge in data*. Amerika : Public Studio. Mukhtar, B. A., Setiawan, N. A., Adji, T. B., No, J. G., Ugm, K., & Yogyakarta, D.I. (2015).
- Nuraeni, N. (2017). Penentuan Kelayakan Kredit Dengan Algoritma Naïve Bayes Classifier : Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC, III(1), 9–15.
- Paramitha, F. A. (2015). Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ),1, 260–270.
- Putri, A. N. (2017). PENERAPAN NAIVE BAYESIAN UNTUK PERANGKINGAN KEGIATAN DI FAKULTAS TIK UNIVERSITAS SEMARANG, 8(2), 603–610.
- Riana Aprilia, Apriatni E P, dan S. (2016). Influence Leadership and Health and Safety (K3) on Employee Performance PT. PLN (Persero) UPJ-Central Semarang Engineering, 4(3), 201–208.
- Riang, S. C. (2016). Teknik Data Mining Menggunakan Classification Dalam Sistem Penunjang Keputusan Peminatan SMA Negeri 1 Polewali. *IJNS –Indonesian Journal on Networking and Security - Volume 5 No 1 – 2016 –Ijns.Org Menggunakan*, 5(1), 48–54.
- Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier, 7(1), 59–64.
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga, 2(3), 207–217. *ISH*, (June).
- Wijayatun, R., & Sulisty, Y. (2016). Prediksi Rating Film Menggunakan Metode Naïve Bayes, 8(2), 60–63.
- Winarko & Sumarni Adi, E. (2015). Klasifikasi Data NAP (Nota Analisis Pembiayaan) untuk Prediksi Tingkat Keamanan Pemberian Kredit (Studi Kasus : Bank Syariah Mandiri Cabang Luwuk Sulawesi Tengah), 9(1), 1–12.