

Model Prediksi Penjualan Jenis Produk Tekstil Menggunakan Algoritma *K-Nearest Neighbor* (K-NN)

Wiyana Yusuf¹, Rahma Witri², Christina Juliane³

Sistem Informasi Bisnis, STMIK LIKMI Bandung
Bandung, Indonesia

e-mail: wiyana.ys@gmail.com¹, rahmawitri2@gmail.com², christina.juliane@likmi.ac.id³

ABSTRAK

Indonesia merupakan negara yang memiliki industri tekstil yang besar, mulai dari industri pembuatan serat sintetis (*fiber making*), industri pemintalan benang (*spinning*), industri pertenunan (*weaving*), industri perajutan (*knitting*), industri pencelupan, pencetakan dan penyempurnaan sampai dengan industri pakaian jadi (*garment*). Dalam dunia perstekstilan kita perlu memperhatikan pasar terhadap jenis kain yang akan diproduksi secara luas kepada masyarakat. Akan tetapi, tidak ada proses survey pasar pada perusahaan ini yang diakibatkan oleh kurangnya karyawan sehingga perusahaan hanya memproduksi barang tergantung dengan pesanan dari *customer*. Oleh karena itu diperlukan pemantauan terhadap permintaan oleh *customer* menggunakan klasifikasi *data mining* sehingga perusahaan dapat menentukan jenis produk mana yang akan tinggi permintaan dengan cara mengolah data penjualan perusahaan tersebut. Penelitian ini bertujuan untuk membantu perusahaan dalam menentukan arah untuk produksi serta penjualan di masa yang akan datang. Dalam penelitian ini algoritma yang dipilih adalah klasifikasi *data mining* dengan *K-Nearest Neighbor* dengan mengklaster data dari nilai $k=1$ hingga $k=5$. Hasil yang diperoleh dengan nilai $k=1$ mendapatkan akurasi 86.9% dengan RMSE 0.362. Serta nilai *class recall & class precision* pada *confusion matrix* yang didapat cukup baik sehingga model algoritma yang digunakan pada penelitian ini bekerja dengan baik.

Kata kunci: *data mining*, tekstil, *k-nearest neighbor*, akurasi, *confusion matrix*

ABSTRACTS

Indonesia is a country that has a large textile industry, ranging from the synthetic fiber manufacturing industry (*fiber making*), the yarn spinning industry (*spinning*), the weaving industry (*weaving*), knitting industry (*knitting*), dyeing, printing and refinement industries to the clothing industry. so (*garments*). In the world of textiles, we need to pay attention to the market for the types of fabrics that will be widely produced to the public. However, there is no market survey process at this company which is caused by a lack of employees so that the company only produces goods depending on orders from customers. Therefore, it is necessary to monitor demand by customers using data mining classification so that companies can determine which types of products will be in high demand by processing the company's sales data. This study aims to assist the company in determining the direction for production and sales in the future. In this study, the algorithm chosen is data mining classification with *K-Nearest Neighbor* by clustering data from values $k=1$ to $k=5$. The results obtained with a value of $k=1$ get an accuracy of 86.9% with an RMSE of 0.362. And the value of *class recall & class precision* in the *confusion matrix* is quite good so that the algorithm model used in this study works well.

Keywords: *data mining*, textile, *K-Nearest Neighbor*, accuracy, *confusion matrix*

1. PENDAHULUAN

Kegiatan pertekstilan secara sederhana telah dikenal sejak lama oleh masyarakat



Indonesia melalui kegiatan tenun dan batik, terutama untuk lingkungan yang terbatas. Dalam perkembangannya saat ini, kegiatan tekstil terus meluas perannya, tidak hanya untuk keperluan seni-budaya dan kebutuhan pakaian di lingkungan terbatas tapi produk sandang ini sudah dijadikan sumber penghasilan. Pertumbuhan industri perstekstilan juga mengalami perkembangan yang cukup luas untuk memenuhi kebutuhan dalam dan luar negeri sehingga mampu memberikan kontribusi penting bagi perekonomian nasional.

Dengan demikian, penting bagi suatu perusahaan untuk memiliki *outlook* terhadap pasar global sebagai bentuk perencanaan kedepannya. Selain itu, keterbatasan jumlah karyawan juga berpengaruh untuk mencegah pola monoton dari hasil yang diproduksi. Oleh karena itu, dilakukan pemodelan dengan tujuan melakukan prediksi penjualan jenis produk tekstil menggunakan sebuah algoritma sebagai salah satu bentuk proses survey pasar untuk perusahaan ini di waktu mendatang. Hal ini didukung dengan belum adanya penerapan proses pemodelan tersebut pada perusahaan ini karena jumlah SDM yang terbatas.

Dalam penelitian ini penulis melakukan analisa terhadap model prediksi penjualan jenis produk tekstil dimana jika ditinjau dari sebelumnya ada beberapa algoritma yang mampu mendukung dengan nilai akurasi yang tinggi seperti *Neural Network* (Yunita, 2017), *Naïve Bayes* (Annur, 2018), dan juga *K-Nearest Neighbour* (Cholil, Handayani, & Tria, 2021). Pada prosesnya, penulis juga melakukan *trial* terhadap 3 algoritma tersebut dan didapatkan hasil akurasi untuk *Neural Network* 68.66%, *Naïve Bayes* 75.16%, dan *K-Nearest Neighbour* 82.44% dengan tambahan algoritma *Deep Learning* dengan nilai akurasi sebesar 59.52%. Dilihat dari hasil *tracing* prediksi dengan beberapa algoritma tersebut, akurasi tertinggi ada pada algoritma *K-Nearest Neighbour*. Algoritma *K-Nearest Neighbor* biasa digunakan untuk melakukan prediksi (*forecasting*) pada harga barang (Bode, 2017), peramalan cuaca (Harun, Pelangi, & Yuliyanti, 2020), kelulusan mahasiswa (Hakim, Rizal, & Ratnasari, 2019), dan kelayakan kredit (Wahyuningsih & Utari, 2018).

Sebelum penulis memilih algoritma yang akan digunakan pada penelitian ini, penulis mencoba beberapa algoritma yaitu *K-Nearest Neighbor*, *Deep Learning*, *Naive Bayes*, dan *Neural Network* untuk dilihat nilai akurasinya dimana algoritma dengan nilai akurasi tertinggi akan digunakan pada penelitian ini. Setelah proses percobaan yang dilakukan, didapatkan nilai akurasi dari setiap algoritma yang digunakan. Dapat dilihat pada Tabel 1 nilai akurasi tertinggi didapatkan saat menggunakan algoritma *K-Nearest Neighbor*.

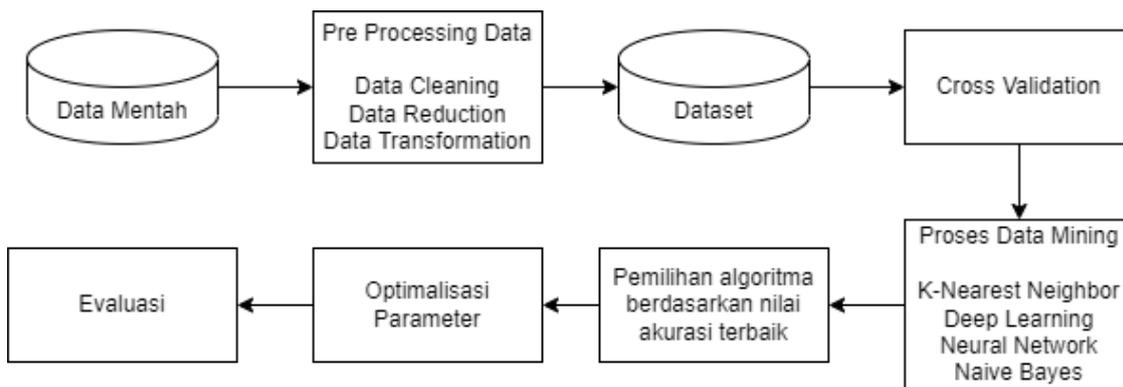
Tabel 1. Nilai akurasi dari setiap algoritma yang digunakan

Algoritma	Nilai Akurasi
<i>K-Nearest Neighbor</i>	82.44%
<i>Deep Learning</i>	59.52%
<i>Naïve Bayes</i>	75.16%
<i>Neural Network</i>	68.66%

Oleh karena itu, dalam penelitian ini digunakan algoritma *K-Nearest Neighbor* untuk memprediksi jenis produk dalam penjualan serta proses analisa yang belum diterapkan sebelumnya terhadap nilai *class recall* dan *class precision* untuk melihat performa model *data mining* yang digunakan. Hal ini didukung dengan metode klasifikasi yang juga digunakan pada penelitian ini dari penelitian sebelumnya oleh (Cholil et al., 2021).

2. METODE PENELITIAN

Dalam penelitian ini akan dilakukan eksperimen dengan metode klasifikasi data mining yaitu *K-Nearest Neighbor* terhadap data penjualan sebuah perusahaan tekstil dan berfokus pada dua jenis produk yaitu tenun dan rajut. Data akan diujicobakan pada model yang dibuat dengan menggunakan algoritma *K-Nearest Neighbor* lalu dilakukan pengujian menggunakan *k-fold cross validation*, serta evaluasi dan validasi hasil dengan *confusion matrix*. Langkah-langkah penelitian ini digambarkan pada Gambar 1.



Gambar 1. Langkah-langkah penelitian

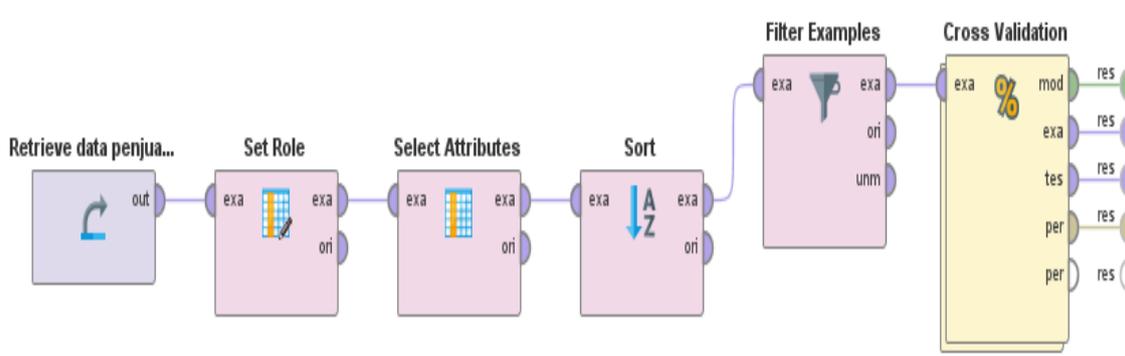
Data yang digunakan merupakan data sekunder yang didapat dari *database* salah satu perusahaan tekstil yang terletak di Padalarang, Bandung Barat. Data tersebut menghimpun transaksi penjualan yang terjadi pada perusahaan tersebut sebanyak 14.478 data dengan *range* penjualan tahun 2018-2021. Dengan atribut tanggal penjualan, kategori produk, dan jumlah penjualan. Dapat dilihat pada Tabel 2 untuk contoh 14 sampel dari keseluruhan 14.478 data penjualan pada tanggal 5 Januari 2019, dengan 2 kategori kain yang terjual yaitu kain rajut dan kain tenun dengan jumlahnya masing-masing.

Tabel 2. Data Penjualan (*transformed*)

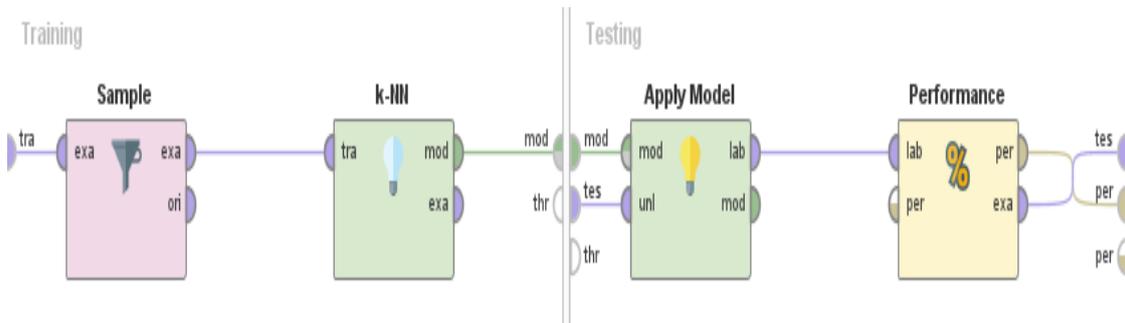
Row No.	Kategori	Jumlah	Tanggal
1	Rajut	1029.560	Jan 5, 2019
2	Rajut	3073.400	Jan 5, 2019
3	Rajut	1522.330	Jan 5, 2019
4	Rajut	3046.810	Jan 5, 2019
5	Rajut	1529.810	Jan 5, 2019
6	Tenun	1984.700	Jan 5, 2019
7	Tenun	179	Jan 5, 2019
8	Tenun	3256	Jan 5, 2019
9	Tenun	26289	Jan 5, 2019
10	Tenun	103	Jan 5, 2019
11	Tenun	363	Jan 5, 2019
12	Tenun	4063	Jan 5, 2019
13	Tenun	216	Jan 5, 2019
14	Tenun	10752.3	Jan 5, 2019
..

Model yang didesain pada aplikasi Rapidminer meliputi tahap *pre-processing data*, *data transformation*, dan *data mining*. Tahap *pre-processing* pada model, dilakukan proses pemilihan atribut yang berfungsi untuk mengurangi atribut-atribut yang tidak digunakan pada proses *data mining*. Lalu, data masuk ke tahap *data transformation* agar bentuk data disesuaikan sebelum dilakukannya proses data mining. Tahap akhir, data masuk ke kedalam *operator cross validation* agar data divalidasi serta akan dibagi secara acak kedalam *k* bagian dan masing-masing bagian akan dilakukan proses klasifikasi. Nilai *k* pada *operator cross validation* berfungsi untuk banyaknya percobaan yang akan dilakukan. Pada penelitian ini, nilai *k* yang digunakan adalah 10 atau *10-fold cross validation*.

Dapat dilihat pada Gambar 2. Sebelum data diolah menggunakan algoritma, data diolah kembali menggunakan operator *sample* agar jumlah data dengan kategori tenun dan rajut lebih seimbang. Hal ini dilakukan karena jumlah data dengan kedua kategori tersebut tidak seimbang, data dengan kategori tenun lebih banyak muncul dibandingkan rajut. Nilai *sample ratio per-class* yang digunakan pada operator *sample* 1.0 untuk tenun dan 0.46 untuk rajut.



Gambar 2. Model pada aplikasi Rapidminer



Gambar 3. Algoritma ditempatkan di dalam operator cross validation

3. HASIL DAN PEMBAHASAN

Algoritma *K-Nearest Neighbor* dilakukan 5 kali percobaan dengan nilai *k* yang berbeda. Nilai *k* yang digunakan yaitu *k*=1, *k*=2, *k*=3, *k*=4, dan *k*=5 untuk mengetahui nilai *k* yang paling optimal untuk data dan model yang digunakan. Hasil dari kelima percobaan tersebut dapat dilihat pada Tabel 3.

Pada Tabel 3 dapat dilihat bahwa nilai akurasi tertinggi didapatkan dengan nilai *k*=1, serta *confusion matrix* yang didapatkan dapat dilihat pada Gambar 4. Berdasarkan kelima percobaan tersebut, tingkat akurasi yang diperoleh dengan nilai *k* pada Tabel 3 tergolong baik karena nilai akurasi berada di atas 80%.

Tabel 3. Hasil percobaan dengan nilai *k*=1 sampai *k*=5

Nilai <i>k</i>	Nilai Akurasi	Standard Deviation	Root Mean Square Error (RMSE)
1	86.9%	0.84%	0.362
2	86.89%	0.82%	0.333
3	84.94%	0.74%	0.33
4	85.25%	0.67%	0.338
5	82.44%	0.89%	0.345

accuracy: 86.90% +/- 0.84% (micro average: 86.90%)

	true Rajut	true Tenun	class precision
pred. Rajut	8795	697	92.66%
pred. Tenun	1199	3787	75.95%
class recall	88.00%	84.46%	

Gambar 4. Confussion matrix yang didapatkan dari hasil percobaan dengan nilai *k*=1.

Jumlah *True Positive* (TP) yang didapatkan sebanyak 8795 *records* yang diklasifikasikan sebagai Rajut dan jumlah *False Negative* (FN) sebanyak 697 *records* yang diklasifikasikan sebagai Rajut tetapi Tenun. Sedangkan untuk *class* Tenun, jumlah *False Positive* (FP) didapatkan sebanyak 3787 *records* yang diklasifikasikan sebagai Tenun, dan *True Negative* (TN) berjumlah 1199 *records* yang diklasifikasikan sebagai Tenun ternyata Rajut.

	<u>Rajut</u>	<u>Tenun</u>	
<u>Rajut</u>	TP = 8795	FP = 697	Predicted Class
<u>Tenun</u>	FN = 1199	TN = 3787	
	Actual Class		

Gambar 5. Visualisasi *confussion matrix*.

Berdasarkan Gambar 5 sebagai visualisasi dari *confussion matrix*, dapat dilihat bahwa nilai *class precision* untuk *class* Rajut pada Gambar 4 menandakan bahwa model yang telah didesain pada penelitian ini dapat memprediksi 92.66% dari 9492 (TP + FP) data dengan benar, sedangkan nilai *class recall* menandakan bahwa ada 88% dari 9994 (TP + FN) data yang dengan benar diklasifikasikan. *Class* Tenun memiliki nilai *class precision* sebesar 75.95% dari 4986 data dan *class recall* sebesar 84.46% dari 4484 data.

Nilai akurasi yang didapat pada Tabel 3 menandakan bahwa metode ini dapat digunakan untuk melakukan prediksi. Pada penelitian lainnya yang memiliki masalah serupa, untuk melakukan prediksi terhadap suatu penjualan metode K-NN banyak digunakan seperti pada penelitian untuk memprediksi penjualan produk elektronik terlaris oleh (Amalia Rizki, 2018) dimana nilai akurasi yang didapatkan yaitu 92,51%, implementasi KNN pada data penjualan oleh (Hermawan & Agung, 2017) dengan nilai akurasi sebesar 58.33% dengan keakuratan prediksi sebesar 88,54% dan prediksi untuk penjualan produk unilever menggunakan metode knn oleh (Alfani, Rozi, & Sukmana, 2021) dimana nilai akurasi tertinggi yang didapatkan sebesar 86,66%.

4. KESIMPULAN

Berdasarkan hasil percobaan dapat, nilai akurasi yang didapatkan pada penelitian ini yaitu 86.9% dengan jumlah prediksi benar 12582 dari 14478 *records*, tetapi nilai akurasi tidak dapat dijadikan acuan bahwa model yang dibuat dapat memprediksi dengan baik terutama saat kondisi *class* yang tidak *balance*. Hal ini dapat menyebabkan nilai akurasi sangat tinggi tetapi nilai *class recall* dan *class precision* rendah. Oleh karena itu, perlu diamati bahwa nilai *class recall* dan *class precision* memegang peranan terhadap performa model yang dibuat untuk melakukan klasifikasi terhadap data yang digunakan.

Saran atau usul yang dapat diberikan untuk penelitian selanjutnya antara lain, dapat digunakannya algoritma klasifikasi lain seperti *Random Forest*, penting juga untuk penyesuaian nilai-nilai parameter pada model yang dibuat seperti nilai *k* pada algoritma K-NN, serta melakukan analisa terhadap nilai *class precision*, *class recall*, AUC, ROC, serta *F1-Score* agar model yang dibuat bisa lebih terlihat performanya.

5. REFERENSI

Alfani, A., Rozi, F., & Sukmana, F. (2021). Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*.

Amalia Rizki, Y. (2018). Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode K-Nearest Neighbor (Studi Kasus : PT. Bintang Multi Sarana Palembang). Retrieved from <http://eprints.radenfatah.ac.id/3302/>

Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10, 6. Retrieved from <http://103.226.139.203/index.php/ILKOM/article/view/303>

Bode, A. (2017). K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika. *ILKOM Jurnal Ilmiah*, 9(2), 188–195. <https://doi.org/10.33096/ilkom.v9i2.139.188-195>

Cholil, S., Handayani, T., & Tria, A. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi

- Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 6(2), 118–127.
- Hakim, L. A. R., Rizal, A. A., & Ratnasari, D. (2019). Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-NN). *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, 1(1), 30–36. <https://doi.org/10.35746/jtim.v1i1.11>
- Harun, R., Pelangi, K., & Yuliyanti, L. (2020). Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma K Nearest Neighbor (KNN). *MISI (Jurnal Manajemen Informatika & Sistem Informasi)*. Retrieved from <http://e-journal.stmiklombok.ac.id/index.php/misi/article/view/125/84>
- Hermawan, F., & Agung, H. (2017). Implementasi Metode K-Nearest Neighbor Pada Aplikasi Data Penjualan PT. Multitek Mitra Sejati.
- Wahyuningsih, S., & Utari, D. R. (2018). Perbandingan Metode K-Nearest Neighbor, Naive Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit. *Konferensi Nasional Sistem Informasi 2018 STMIK Atma Luhur Pangkalpinang, 8 – 9 Maret 2018*, 619–623.
- Yunita. (2017). Seleksi Fitur Menggunakan Backward Elimination Pada Prediksi Cuaca. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 2(1), 26–37.