

THE APPLICATION OF K-NEAREST NEIGHBOURS ALGORITHMS FOR THE CLASSIFICATION OF FASHION TREND

Nugroho Dwi Saputro¹, Setyoningsih Wibowo², Mochamad Fadjar Darmaputra³

¹Fakultas Teknik dan Informatika, universitas PGRI Semarang
email: nugputra1@gmail.com

²Fakultas Teknik dan Informatika, universitas PGRI Semarang
email: ninink.1623@gmail.com,

³Fakultas Pendidikan Ilmu Sosial Keolahragaan, universitas PGRI Semarang
email: fadjar.darmaputra@gmail.com

ABSTRACT

The creative economy in various countries is currently driving the nation's economy. Fashion is an activity related to the design, production, consultation and distribution of fashion products. Fashion manufacturers consist of the fashion industry, clothing and accessories, the fashion industry, manufacturers of bags, shoes and accessories. In apparel creation, trend becomes an important aspect. Indonesian designers today still follow the European trend which is influenced by the type of season. Decision making on the basis of accurate data and information will result in a decision to classify fashion trends in the creative industry can be done by adopting a data mining approach. In this study the accuracy of the k-Nearest Neighbor algorithm has never been used to classify clothing models. The purpose of this study is to find out how much accuracy is obtained for clothing trend classification. From the classification of clothing models obtained a pretty good value. From testing, showing the accuracy of the AUC goes into the good category.

Keyword: Fesyen, Clasification, K-Nearest Neighbour (K-NN) .

1. PENDAHULUAN

Industri fesyen merupakan salah satu pengembangan ekonomi kreatif. Ekonomi kreatif di berbagai negara saat ini menjadi pendorong perekonomian bangsa. Fesyen atau mode sendiri merupakan kegiatan yang terkait dengan desain, produksi, konsultasi serta distribusi produk fesyen. Produsen fesyen terdiri dari industri fesyen pakaian dan aksesoris, industri fesyen produsen tas, sepatu dan aksesoris. Dalam kreasi pakaian jadi, *trend* menjadi aspek penting. Perancang indonesia sekarang ini masih mengikuti *trend* Eropa yang dipengaruhi oleh jenis musim. Empat musim dalam satu tahun, menyebabkan muncul 4 trend. Pelaku di model bisnis *mass production*, yang berdasar pada volume biasanya memilih atau memodifikasi beberapa desain yang sedang *trend*, kemudian diproduksi dalam jumlah besar [1].

Mode merupakan bentuk nomina yang bermakna ragam cara atau bentuk terbaru pada suatu waktu tertentu. Mode yang dikenakan oleh seseorang mampu mencerminkan siapa pemakainya. Mode bersifat temporer atau hanya waktu tertentu. Suatu mode akan hilang sendirinya seiring munculnya mode baru. Tren merupakan suatu pandangan, gagasan, ide maupun mode yang sedang digandrungi oleh masyarakat, dapat digemari oleh remaja, anak-anak, dewasa maupun para orang tua. Tren adalah sesuatu yang sedang "menjamur" atau sedang disukai dan digandrungi oleh orang banyak. Cirinya mudah saja, apabila kita sedang berjalan di tempat - tempat umum atau tempat berbelanja tertentu dan melihat ada sebuah kesamaan pada 5 - 25 orang, maka itulah tren yang sedang berlaku [2].

Dalam dunia bisnis yang selalu dinamis dan penuh persaingan, para pelakunya harus senantiasa

memikirkan cara untuk terus bertahan dan jika mungkin mengembangkan bisnis mereka. Untuk mencapai hal itu, dapat diringkaskan tiga kebutuhan bisnis, yaitu:

- a. Penambahan jenis maupun peningkatan kapasitas produk.
- b. Pengurangan biaya operasi perusahaan.
- c. Peningkatan efektifitas pemasaran dan keuntungan.

Memenuhi kebutuhan-kebutuhan di atas, banyak cara yang dapat ditempuh. Salah satunya adalah dengan memikirkan teknik-teknik pemasaran yang efektif dengan biaya yang minimal. Memprediksi pola kebiasaan konsumen di masa yang akan datang merupakan salah satu solusi dari masalah ini yang akan berkaitan dengan penemuan dan pemanfaatan pola yang tersembunyi dari data yang kemungkinan sudah dimiliki oleh perusahaan [3].

Pengambilan keputusan atas dasar data dan informasi yang akurat akan menghasilkan keputusan untuk klasifikasi trend fesyen pada industri kreatif dapat dilakukan dengan mengadopsi dari pendekatan data mining. Menurut Pang-Ning Tan data mining merupakan suatu proses yang dilakukan secara otomatis untuk menemukan informasi yang berguna dalam repositori data yang besar [4]. Bagian integral dari penemuan pengetahuan dalam database (*Knowledge Discovery in Database*), yang memerlukan ekstraksi informasi berharga dari data mentah. Berdasarkan data historis, model dapat dilatih untuk diklasifikasikan [5].

Klasifikasi merupakan 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. Model itu sendiri bisa berupa aturan “jika-maka”, berupa *decision tree*, *formula matematis* atau *neural network*. Metode-metode klasifikasi antara lain C4.5, *RainForest*, *Naïve Bayesian*, *neural network*, *genetic algorithm*, *fuzzy*, *case-based reasoning*, dan *k-Nearest Neighbor* [6].

Penelitian Krisandi mengkaji tentang Algoritma k-NN dan kemudian mengaplikasikan Algoritma k-NN dalam klasifikasi data. Data yang digunakan adalah data hasil produksi kelapa sawit (Tonase) dari 50

kelompok tan. Berdasarkan hasil penelitian data terklasifikasi dalam 6 cluster berdasarkan kemiripan hasil produksi dari 50 kelompok tani yang ada di KUD. HIMADO. Hasil produksi yang dominan adalah produksi dari kelompok tani kelapa sawit yang terletak pada C1 dengan 17 anggota dengan persentase 34% yaitu kelompok 1, 2, 33, 34, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49,50 untuk nilai $k=7$ [7.].

Jumadi dan Edi Winarko melakukan penelitian mengklasifikasikan dengan hasil teks berita yang tak-terkelompokkan pada saat pengklasteran oleh stc (*suffix treeclustering*), tetapi dengan menambahkan algoritma *K-Nearest Neighbor* mampu mengelompokkan dokumen pada kelompok yang sudah ada.[8.]

Berdasarkan latar belakang diatas algoritma *k-Nearest Neighbor* banyak digunakan pada penelitian diberbagai bidang, seperti penelitian yang di lakukan oleh Krisnadi yang mengklasifikasikan kelapa sawit dan oleh Jumadi menklasifikasikan teks berita. Banyak komparasi algoritma yang dilakukan peneliti-peneliti sebelumnya, algoritma *k-Nearest Neighbor* dalam mengklasifikasi dengan berbagai data memiliki nilai terbaik diantara algoritma lainnya, akan tetapi untuk klasifikasi data model pakaian belum pernah dilakukan.

Tujuan penelitian ini adalah untuk mengetahui pola trend fesyen yang diinginkan pasar dan mencari tahu tingkat akurasi yang didapat dengan menggunakan algoritma *K-Nearest Neighbor*.

Manfaat penelitian ini adalah Hasil penelitian ini diharapkan dapat digunakan untuk pengambilan keputusan dalam membantu menentukan kebijakan yang strategi dan efektif untuk penentuan *trend* fesyen

2. TINJAUAN PUSTAKA

A. Model Pakain

Berdasarkan dataset model pakaian yang didapat toko pakaian Nyetrik kode barang, jenis pakaian, ukuran, garis leher, lengan baju, bahan, motif dan sebagai label adalah variabel kategori yang terdiri dari Kasual dan Pesta. Semua atribut tersebut selain hasil merupakan detail pakaian dalam dataset yaitu antara lain Kategori, Ukuran, Garis leher, Lengan baju, Bahan, dan Motif.

B. Algoritma *K-Nearest-Neighbours*

Algoritma KNN merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran di proyeksikan k ruang berdimensi banyak, dimana masing-masing dimensi mempresentasikan fitur dari data. Ruang ini di bagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Sebuah titik pada ruangan ini ditandai dengan kelas c , jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat titik tersebut. *K-Nearest Neighbour* merupakan metode yang bersifat supervised, dimana hasil dari *query instance* yang baru di klasifikasikan berdasarkan mayoritas kategori pada KNN [9]. Ketepatan algoritma KNN berdasarkan pada ada atau tidak adanya fitur – fitur yang tidak relevan, atau jika bobot fitur tersebut setara dengan relevansinya terhadap klasifikasi. Algoritma *K Nearest Neighbour* memiliki kelebihan yaitu dapat menghasilkan data yang kuat atau jelas dan efektif jika digunakan pada data yang besar. Dari beberapa kelebihan tersebut, algoritma *K Nearest Neighbour* juga memiliki kekurangan yaitu membutuhkan nilai K sebagai parameter, jarak dari data percobaan tidak dapat jelas dengan tipe jarak yang digunakan dan dengan atribut yang digunakan untuk memperoleh hasil yang terbaik, maka harus menggunakan semua atribut atau satu atribut yang telah pasti, dan perhitungan harga sangat tinggi karena percobaan ini membutuhkan perhitungan jarak dari beberapa *query* untuk semua data percobaan. Rumus perhitungan jarak yang paling banyak digunakan yaitu euclidean distance, dua contoh data sederhananya $(x_1, x_2, x_3, \dots, x_p)$ dan $(u_1, u_2, u_3, \dots, u_p)$, maka didapatkan rumus,

$$\sqrt{(x_1 - u_1)^2 + (x_2 - u_2)^2 + \dots + (x_p - u_p)^2} \quad (9)$$

Dimana, $x_1, x_2, x_3, \dots, x_p$ adalah prediksi dari contoh data pertama dan $u_1, u_2, u_3, \dots, u_p$ adalah prediksi dari contoh data kedua.

Dalam implementasinya terdapat dua jenis metode *k nearest neighbour* yang nantinya akan digunakan pada penelitian ini yaitu:

1. Simple Unweighted Voting

Simple unweighted voting biasanya digunakan untuk mengklasifikasi data yang baru masuk berdasarkan kedekatan jaraknya dengan histori data yang telah ada.

$$Y_{new} = \frac{\sum_i^n y_i}{n} \quad (10)$$

- Tentukan nilai k (jumlah tetangga terdekat)
- Hitung jarak antara data yang di uji dengan data training
- Urutkan berdasarkan jarak terdekat dan tentukan apakah termasuk dalam k (jumlah tetangga terdekat)
- Hitung rata rata dari data yang termasuk dalam K .

Weighted Voting

Weighted voting ini biasanya digunakan untuk estimasi dan prediksi yang merupakan salah satu kegunaan algoritma *K Nearest Neighbour* adalah mengestimasi dan memprediksi sesuatu.

$$Y_{new} = \frac{\sum_i w_i y_i}{\sum_i w_i} \quad W_i = \frac{1}{d(x,y)^2} \quad (11)$$

- Tentukan jumlah k (jumlah tetangga terdekat)
- Hitung jarak antara data yang diuji dengan data training
- Urutkan berdasarkan jarak terdekat dan tentukan apakah termasuk dalam k (jumlah tetangga terdekat)
- Menghitung bobot dari data yang termasuk dalam K . [9]

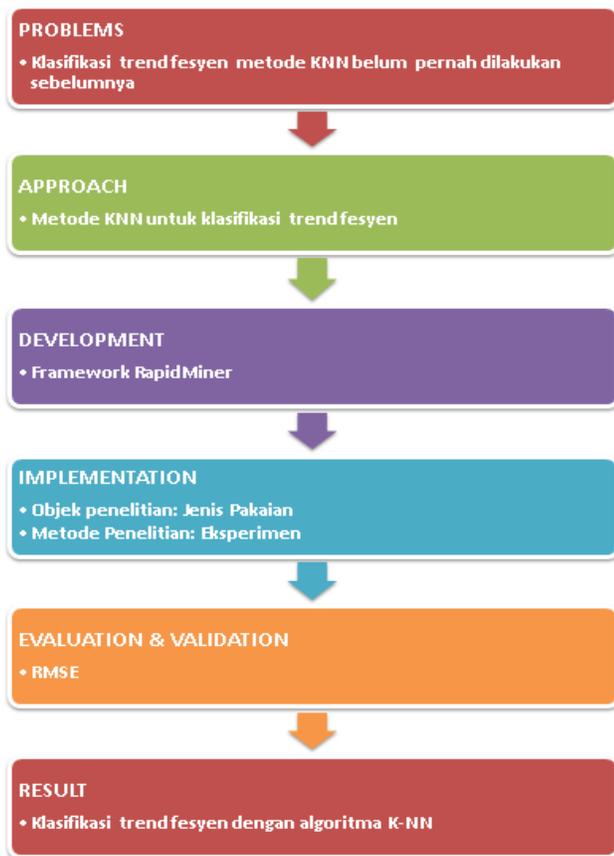
Rumus menghitung *Euclidean*

$$d(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2} \quad (12)$$

Rumus menghitung *Manhattan city*

$$d(x, y) = \sum_i^n |x_i - y_i| \quad (13)$$

A. Kerangka Penelitian



Gambar 1 Kerangka berpikir

3. METODE PENELITIAN

A. Pengumpulan data

Pengumpulan data merupakan langkah awal dari suatu penelitian. Berdasarkan dataset model pakaian yang didapat toko pakaian Nyetrik, data yang di dapat adalah data model pakaian dengan hasil yang didapat sebanyak 173 model pakaian dengan dengan kategori sebagai labelnya, yang terdiri dari Kasual dan Pesta. Dengan atribut dari setiap model pakaian yang dipakai adalah Kode barang, kategori, jenis pakaian, ukuran, garis leher, lengan baju, Bahan, Motif.

B. Preprocessing

Data yang diperoleh akan di *preprocessing* terlebih dahulu supaya data berkualitas dengan cara manual.

Teknik dalam *preprocessing* [10] yaitu :

- 1) *Data cleaning* bekerja membersihkan nilai kosong, tidak konsisten atau tupel kosong (*missing value* dan *noisy*).
- 2) *Data integration* menyatukan tempat penyimpanan (arsip) yang berbeda dalam satu arsip.
- 3) *Data reduction* jumlah atribut yang digunakan untuk data training terlalu besar sehingga ada beberapa atribut yang tidak diperlukan dihapus. Dalam penelitian ini akan menggunakan teknik Data cleaning, data yang nilai atributnya kosong akan dihapus supaya data menjadi lebih valid dan berkualitas. Dalam penelitian jumlah record/data model pakaian adalah 173, setelah melakukan data cleaning maka dapat ditemukan data dengan atribut yang kosong berjumlah 12. Jadi data yang diolah dan diteliti sebanyak Metode yang diusulkan adalah strategi yang digunakan untuk menyelesaikan permasalahan yang dihadapi. Dalam menyelesaikan permasalahan akan digunakan penerapan metode KNN

C. Eksperimen dan Pengujian model

Eksperimen dan pengujian metode adalah proses penghitungan dan langkah-langkah untuk menguji metode yang digunakan sehingga didapatkan hasil yang dapat membuktikan bahwa metode yang digunakan merupakan metode yang paling akurat. Pada tahap ini dilakukan eksperimen dan pengujian model yaitu menghitung dan mendapatkan *rule-rule* pada model algoritma yang diusulkan dengan menggunakan contoh kasus. Pada tahap ini digunakan parameter algoritma K-NN yang dapat menambah akuarasi hasil.

Pada tahap ini dilakukan eksperimen dan pengujian model yaitu menghitung dan mendapatkan *rule-rule* pada model algoritma yang diusulkan dengan menggunakan contoh kasus. Setelah itu, di uji rule sehingga mendapatkan hasil yang lebih baik.

Tabel 1 Data kasus *training*

No	x11	x12	x13	x14	x15	x16	Y
1	1	2	2	1	1	10	Pesta

2	5	3	5	3	1	5	Kasual
3	5	3	5	3	1	10	Kasual
4	5	3	5	4	4	7	Kasual
5	1	1	5	3	5	10	Pesta

Tabel 2 Data kasus *testing*

No	x21	x22	x23	x24	x25	x26	Y
1	1	3	5	1	1	7	?

Menghitung kuadrat jarak *Euclid* (*query instance*) masing-masing objek terhadap sampel data atau *training sample* dengan $k=2$, yang diberikan dengan menggunakan rumus:

$$\left(\sum_{k=1}^m (x_{ik} - x_{ij})^2 \right)$$

Euclid

$$= ((1 - 1)^2 + (2 - 3)^2 + (2 - 5)^2 + (1 - 1)^2 + (1 - 1)^2 + (10 - 7)^2)^{1/2}$$

$$= ((0)^2 + (-1)^2 + (-3)^2 + (0)^2 + (0)^2 + (3)^2)^{1/2}$$

$$= (19)^{1/2}$$

$$= 4,35$$

Sehingga didapat nilai euclid sebesar 4,35 untuk data pakaian *training record* ke 1. Tabel Berikut hasil dari perhitungan euclid semua data pakaian *training*:

Tabel 3 *Square Instance to Query Distance*

x11	x12	x13	x14	x15	x16	Euclid
1	2	2	1	1	10	4,36

5	3	5	3	1	5	4,90
5	3	5	3	1	10	5,39
5	3	5	4	4	7	5,83
1	1	5	3	5	10	5,74

Untuk langkah selanjutnya mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *Euclid* terkecil. Terlihat pada tabel 10 sudah di beri label urutan nilai tertinggi ke nilai terkecil.

Tabel 3 Mengurutkan Objek ke dalam Kelompok ke Jarak Euclid Terkecil

x11	x12	x13	x14	x15	x16	Euclid	Jarak terkecil
1	2	2	1	1	10	4,36	1
5	3	5	3	1	5	4,90	2
5	3	5	3	1	10	5,39	3
5	3	5	4	4	7	5,83	5
1	1	5	3	5	10	5,74	4

Pada Tabel ini merupakan tahap mengumpulkan label class Y dan (klasifikasi nearest neighbor).

D. Evaluasi dan validasi hasil

Dalam penelitian ini dilakukan pengujian obyektifnya menggunakan 10 *fold cross validation*. *Cross validation* adalah metode pengamatan dalam kumpulan data yang akan digunakan untuk pengujian dan pelatihan, tetapi tidak pada waktu yang sama. Dalam *cross validation* akan mengulang pengujian sebanyak 10 kali dan hasil pengukuran adalah nilai rata-rata dari 10 kali pengujian.

Dalam pengujian ini dilakukan pengujian obyektifnya menggunakan 10 *fold cross validation*. Untuk pengukurannya menggunakan metode confusion matrix dan kurva ROC (*Receiver Operating Characteristic*) untuk mengukur hasil proses prediksi. Sedangkan untuk

merepresentasikan grafis yang menentukan klasifikasi mana yang lebih baik, digunakan metode yang menghitung luas daerah dibawah kurva ROC yang disebut AUC (*Area Under the ROC Curve*)[11]. AUC memiliki nilai antara 0.0 dan 1.0 dengan tingkat keakuratan klasifikasi sebagai berikut:

- 0.90 – 1.00 sangat baik,
- 0.80 – 0.90 baik,
- 0.70 – 0.80 sama,
- 0.60 – 0.70 rendah dan
- 0.50 – 0.60 gagal.

4. HASIL DAN PEMBAHASAN

Awal pengolahan data dibagi menjadi dua untuk pengujian yaitu setiap variabel ini di bagi lagi untuk data *testing* dan data *training*. Pengolahan data pada penelitian ini menggunakan Algoritma K-NN, dengan menggunakan penghitungan data sebagai berikut: Penghitungan data menggunakan tools rapidminer, yang terlebih dahulu menentukan data inputan sebagai data training dan data testing dalam memproses data tersebut. Penelitian ini menggunakan 10% dari dataset yaitu sebanyak 16 data untuk pengujian atau *testing*. Untuk data sisanya sebanyak 90% atau 145 data akan digunakan untuk membangun model.

Percobaan pertama ini menggunakan Algoritma *K-Nearest Neighbor*. Pada algoritma ini parameter untuk memaksimalkan hasil eksperimen menggunakan parameter nilai K. Eksperimen ini nilai K= 1 sampai 3 dengan besarnya k-fold (*number of validation*) k = 10. Berikut ini hasil dari percobaan yang telah dilakukan dengan memasukkan nilai K serta nilai range (k-fold) yang telah ditentukan untuk menguji.

Tabel 4 Hasil eksperimen parameter K pada K-NN

Nilai K	Accuracy	AUC
1	72,01%	0,843
2	72,01%	0,844
3	73,26%	0,847

Untuk evaluasi menggunakan *ROC curve* sehingga menghasilkan nilai AUC (*Area Under Curve*) untuk model algoritma K-NN menghasilkan nilai 0.847 dengan nilai diagnosa *Good Classificatio*.



Gambar 2 Nilai AUC dalam grafik *ROC curve* (Algoritma K-NN)

5. KESIMPULAN

Setelah Penelitian yang dilakukan dengan jumlah input data model menunjukkan bahwa penggunaan algoritma *K-NN* untuk klasifikasi model pakaian dapat menghasilkan nilai yang cukup baik. Dari pengujian, menunjukkan penggunaan algoritma *K-NN* menghasilkan klasifikasi *performance* keakurasian AUC masuk ke kategori yang baik. Sedangkan Algoritma *K-Nearest Neighbor* juga menghasilkan klasifikasi *performance* keakurasian AUC masuk ke kategori baik.

6. SARAN

Dalam Penelitian ini dilakukan menggunakan algoritma *K-Nearest Neighbor*. Mencoba algoritma lain yang menghasilkan tingkat akurasi tinggi. Hasil penelitian ini diharapkan bisa digunakan sebagai acuan untuk klasifikasi data dan meningkatkan akurasi.

7. REFRENSI

- [1] Industri, P., Menuju, K., Ekonomi, V., & Indonesia, K. (2015). EKONOMI KREATIF INDONESIA 2025
- [2] Arum Ratnisty Diany S, Tren Mode Remaja Putri (Studi Analisis Isi Produksi Pesan Rubrik Mode Majalah Gadis Dalam Pembentukan Tren Mode Remaja Putri Tahun 2011), Universitas Sebelas Maret, Surakarta. 2014.
- [3] Solusi, S. (2002). DATA MINING SEBAGAI SOLUSI BISNIS, 7(1), 44-56.

- [4] Tan, P., & Steinbach, M. (2006). Introduction to Data Mining Instructor ' s Solution Manual. *Names*.
- [5] Dhanasaputra, N. & Santosa, B. Pengembangan Algoritma CAT Swarm Optimization (CSO) Untuk Klasifikasi (2006). Pengembangan algoritma.
- [6] Larose D. *Discovering Knowledge in Data*. USA:John Wiley's and Son ;2005.
- [7] Krisandi, DKK. (2013).Algoritma Knearest Neighbor Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada PT. Minamas Kecamatan Parindu.
- [8] Jumadi, Edi Winarko. (2015). Penggunaan KNN (k-Nearst Neighbor) untuk Klasifikasi Teks Berita yang tak- Terkelompokkan Pada saat Pengklasteran oleh STC (*Suffix Tree Clustering*)
- [9] M.V.Subha, "Classification of Stock Index movement using k-Nearest Neighbours (k-NN) algorithm," *Wseas Transactions on Information Science and Applications*, p. 10, 2012.
- [10] Davies, and Paul Beynon, 2004, "*Database Systems Third Edition*", Palgrave Macmillan, New York.
- [11] Florin Gorunescu, *Data mining concepts models and technique*. berlin: Springer, 2011.