

Analisis Perbandingan Algoritma SVM Dan KNN Untuk Klasifikasi Anime Bergenre Drama

Vika Vitaloka Pramansah¹, Dadang Iskandar Mulyana², Titi Silfia³

¹Program Studi Sistem Informasi, STIKOM CKI

²Program Studi Teknik Informatika, STIKOM CKI

³Program Studi Teknik Informatika, STIKOM CKI

¹vitalokavika@gmail.com, ²mahvin2012@gmail.com, ³fiviasilviatiti1101@gmail.com

Abstract

There are many genres of anime such as drama, action, romance, comedy, and so on. However, because there are so many anime genres, it is quite difficult for viewers to find anime whose genre they like, such as the drama genre which tells about everyday human life which is quite light in nature. From these problems, a classification method is needed to classify anime that belongs to the drama genre. Classification is a common method in data mining, an object whose class/label is unknown can go through the classification method so that its class can be estimated [7]. Classification has several algorithms including Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). SVM and KNN algorithms have been widely used and have a good level of accuracy. In this study, a comparative analysis will be carried out between the two algorithms, the dataset used is 12,294 data and 2 genre classes, namely drama and non-drama, the attribute of the anime dataset is 7. The results obtained in this study indicate that the K-Nearest Neighbors Algorithm (KNN) get a training accuracy value of 100% and a test accuracy value of 84%. And also the Support Vector Machine (SVM) algorithm gets a training accuracy value of 83% and a test accuracy value of 82%. The results of the accuracy values of the two algorithms indicate that the K-Nearest Neighbors (KNN) algorithm has a better testing accuracy than the Support Vector Machine (SVM) with a fairly thin difference between the two algorithms.

Keywords: Support Vector Machine, K-Nearest Neighbors, anime, data mining, classification.

Abstrak

Terdapat banyak genre anime seperti drama, aksi, romansa, komedi, dan lain sebagainya. Namun, dikarenakan genre anime itu banyak, penonton cukup kesulitan untuk mencari anime yang genrenya mereka sukai seperti genre drama yang menceritakan kehidupan manusia sehari-hari yang sifatnya cukup ringan. Dari permasalahan tersebut, maka dibutuhkan suatu metode klasifikasi untuk mengklasifikasikan anime yang tergolong ke dalam genre drama. Klasifikasi dalam data mining merupakan metode yang umum, suatu objek yang sebelumnya belum diketahui kelas/labelnya dapat melalui metode klasifikasi agar kelasnya dapat diperkirakan [7]. Klasifikasi memiliki beberapa algoritma diantaranya Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN). Penggunaan algoritma SVM dan KNN telah banyak digunakan dan tingkat akurasi yang baik. Dalam penelitian ini akan menganalisa perbandingan diantara kedua algoritma tersebut pada dataset yang digunakan berjumlah 12.294 data dan 2 kelas genre yaitu drama dan non drama, atribut dataset anime berjumlah 7. Hasil penelitian ini, menunjukkan bahwa algoritma dengan K-Nearest Neighbors (KNN) yang menghasilkan nilai akurasi training sebesar 100% dan nilai akurasi testing sebesar 84%. Dan juga hasil dari algoritma Support Vector Machine (SVM) menghasilkan nilai akurasi training sebesar 83% dan nilai akurasi testing sebesar 82%. Hasil nilai akurasi kedua algoritma tersebut menunjukkan bahwa algoritma K-Nearest Neighbors (KNN) memiliki akurasi testing yang lebih baik dari Support Vector Machine (SVM) dengan selisih keduanya cukup tipis.

Kata kunci: Support Vector Machine, K-Nearest Neighbors, anime, data mining, klasifikasi.

1. Pendahuluan

Klasifikasi dalam data mining merupakan metode yang umum, suatu objek yang sebelumnya belum diketahui kelas/labelnya dapat melalui metode klasifikasi agar kelasnya dapat diperkirakan [7]. Klasifikasi memiliki beberapa algoritma diantaranya Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN). Penggunaan algoritma SVM dan KNN telah banyak digunakan dan tingkat akurasi yang dihasilkannya baik.

Dalam penelitian ini, akan menganalisa perbandingan diantara kedua algoritma tersebut yaitu algoritma SVM dan KNN yang akan dilatih dan diuji menggunakan dataset anime yang jumlahnya 12.294 data dan ada 2 kelas genre yaitu drama dan non drama.

Algoritma yang pertama yaitu KNN memiliki beberapa keunggulan, seperti mudah dimengerti, mudah dipelajari, pelatihannya efisien, sederhana, cepat, serta efektif bahkan jika ukuran data pelatihan cukup besar [8]. Namun, K-Nearest Neighbors juga memiliki kelemahan yaitu nilai kfold nya. Agar setiap data dapat dijadikan sebagai data training dan data testing maka kita membutuhkan Cross Validation karena Cross Validation merupakan salah satu metode yang digunakan agar kita dapat mensimulasikan semua data yang ada [9]. Pada penelitian ini, Cross Validation yang digunakan adalah 10. Algoritma kedua yaitu Support Vector Machine, suatu teknik baru untuk melakukan prediksi, baik dalam kasus regresi maupun klasifikasi yang sangat populer pada saat ini [10].

Penggunaan algoritma Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) dalam pengklasifikasian data menunjukkan hasil yang baik. Penggunaan data anime dipilih karena anime memiliki beragam genre atau konsep ceritanya yang membuat anime menarik untuk ditonton. Terdapat banyak genre anime seperti drama, aksi, romansa, komedi, dan lain sebagainya. Namun, dikarenakan genre anime itu banyak, penonton cukup kesulitan untuk mencari anime yang genrenya mereka sukai seperti genre drama yang menceritakan kehidupan manusia sehari-hari yang sifatnya cukup ringan. Dari permasalahan tersebut, maka dibutuhkan suatu metode klasifikasi untuk mengklasifikasikan anime yang tergolong ke dalam genre drama.

Sebelumnya, penelitian pengklasifikasian dengan metode Support Vector Machine (SVM) maupun K-Nearest Neighbors (KNN) telah dilakukan. Yaitu dengan judul penelitian “Implementasi Algoritma Ant Tree Miner Untuk Klasifikasi Jenis Fauna” oleh Wilda Imama Sabilla, Yunita Ardilla, dan Nurissaidah Ulinuha, penelitian ini menggunakan dataset sebanyak 101 data dengan 16 atribut dan label kelas sebanyak 7. Menggunakan Ant Tree Miner dan memiliki hasil nilai akurasi sebesar 95,09% [1]. Penelitian selanjutnya berjudul “Optimasi Prediksi Bencana Banjir menggunakan Algoritma SVM untuk penentuan Daerah Rawan Bencana Banjir” oleh Saruni Dwiasnati dan

Yudo Devianto, dalam penelitian ini dataset yang digunakan adalah dataset yang berasal dari website resmi BNPB, BPS Kabupaten Bandung dan BMKG yang berkisar pada tahun 2016 - 2017 di wilayah Kabupaten Bandung. Penelitian ini menggunakan algoritma Support Vector Machine (SVM) dengan nilai akurasi sebesar 85,71%, lalu tingkat akurasi yang dihasilkan dengan menggunakan algoritma SVM berbasis PSO adalah 97,62 [2]. Penelitian selanjutnya berjudul “Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes” oleh Andi Maulida Argina A. M, dalam penelitian ini menggunakan K-Nearest Neighbors (KNN) yang mendapatkan hasil akurasi tertinggi yaitu 39% [3]. Penelitian selanjutnya berjudul “Memprediksi Pembayaran Biaya Kursus Siswa Enter (English Center) Dengan Algoritma K-Nn” oleh Hadiani¹, Sri Rahmawati, Eka Kurniawan, Ilham Gata dan Windu, dalam penelitian ini menggunakan K-Nearest Neighbors (KNN) dimana evaluasi kinerja algoritma tersebut menggunakan metode Cross Validation sebesar 84.69%, Confusion Matrix dan Kurva ROC k5, k10, k15, k20 berturut-turut 84.69%, 83.93%, 83.36%, 83.36% serta menghasilkan akurasi dan nilai AUC sebesar 0.962, 0.937, 0.922, 0.916 [5]. Penelitian selanjutnya berjudul “Analisis Algorithms Support Vector Machine dengan Naive Bayes Kernel pada Klasifikasi Data” oleh Dhany, Hanna Willa Izhari, dan Fahmi, dalam penelitian ini menggunakan Naive Bayes dimana dataset pelatihan yang digunakan sebanyak 170 dataset serta nilai presisi yang didapatkan Support Vector Machine yaitu 89.66%, Naive bayes menghasilkan presisi sebesar 89.29%. Sedangkan Class Recall yang didapatkan yaitu Naive Bayes 89.29% dan Support Vector Machine 92.86% dan perbedaan nilai akurasinya mencapai 3.22% [6].

Tujuan penelitian ini untuk menganalisa perbandingan pengklasifikasian anime bergenre drama menggunakan algoritma Support Vector Machine (SVM) dan algoritma K-Nearest Neighbors (KNN). Dataset yang digunakan berjumlah 12.294 data dan 2 kelas genre yaitu drama dan non drama, atribut dataset anime berjumlah 7.

2. Metode Penelitian

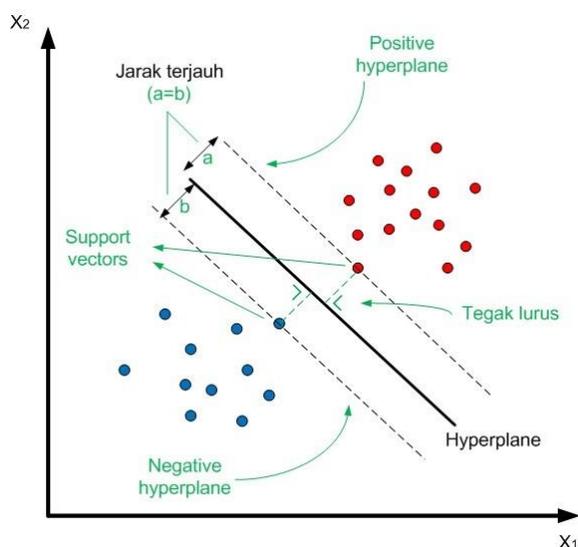
2.1 Data Mining

Data mining merupakan pemisahan pola yang menarik dari suatu data yang besar menggunakan matematika, statistik, machine learning maupun kecerdasan buatan. Berdasarkan fungsinya, data mining dikelompokkan menjadi klasifikasi, clustering, deskripsi, estimasi, asosiasi, dan prediksi. Persiapan data dalam data mining meliputi pemilihan data, dibersihkan, dikompresi, normalisasi dan integrasi dengan file untuk menggali data yang terintegrasi yang memudahkan identifikasi informasi bernilai. Namun, semakin besar data yang diolah maka akan semakin besar pula waktu prosesnya [4].

2.2 Support Vector Machine (SVM)

Support Vector Machine merupakan salah satu metode unggulan dibidang klasifikasi, yang dirilis pada tahun 1992, dengan berprinsipkan Structural Risk Minimization (SRM), SVM berkembang pesat karena dengan memisahkan 2 buah kelas pada suatu ruang sehingga algoritma ini dapat menemukan hyperplane terbaik untuk dapat diimplementasikan sebagai klasifikator. Support Vector Machine memiliki prinsip dasar yaitu linear classifier, yang selanjutnya dikembangkan agar masalah non-linear dapat dikerjakan oleh SVM [11]. Dimana konsep kernel trick dimasukkan pada ruang kerja yang berdimensi tinggi.

Pada input space, SVM berusaha untuk menemukan hyperplane yang dinilai paling baik [12]. Berbeda dengan strategi jaringan saraf tiruan, dimana jaringan saraf tiruan berusaha mencari hyperplane pemisah antar kelas yang ada.

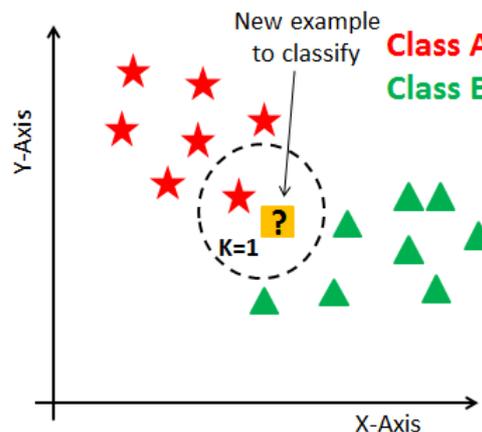


Gambar 1. Visualisasi SVM.

Support Vectors yaitu datapoint yang paling dekat dengan hyperplane, hyperplane yaitu bidang keputusan atau ruang yang dibagi antara satu set objek yang memiliki kelas yang berbeda, dan margin yaitu celah antara dua garis pada titik data dari kelas yang berbeda yang dihitung sebagai jarak tegak lurus dari garis ke vektor pendukung. Semakin besar margin maka akan dianggap baik, begitupun sebaliknya semakin kecil margin maka akan dianggap buruk.

2.3 K-Nearest Neighbors (KNN)

K-Nearest Neighbors (K-NN) termasuk kelompok instance-based learning yang merupakan salah satu teknik *lazy learning*. KNN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing.



Gambar 2. Visualisasi KNN.

Algoritma yang pertama yaitu KNN memiliki beberapa keunggulan, seperti mudah dimengerti, mudah dipelajari, pelatihannya yang cukup sederhana, cepat, serta efektif apabila ukuran data pelatihan cukup besar. Namun, KNN juga memiliki kelemahan, yaitu nilai kfold. Agar setiap data dapat berkesempatan menjadi data training dan data testing maka dibutuhkan Cross Validation karena Cross Validation adalah salah satu metode yang digunakan untuk mengevaluasi kinerja model. Pada penelitian ini, Cross Validation yang digunakan adalah 10.

Dalam algoritma KNN dibutuhkan rumus untuk menghitung jarak terdekat, nilai jarak pada metode K-NN dapat di hitung dengan menggunakan rumus jarak *Euclidean* [13], berikut rumus jarak *Euclidean* :

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2}$$

Rumus 1. Jarak *Euclidean*

Diketahui:

- x = data 1
- y = data 2
- d = jarak *Euclidean*
- i = iterasi

Dimana $x = x^1, x^2, \dots, x^n$, dan $y = y^1, y^2, \dots, y^n$. n mewakili nilai atribut dari dua record.

2.4 Sklearn Learn

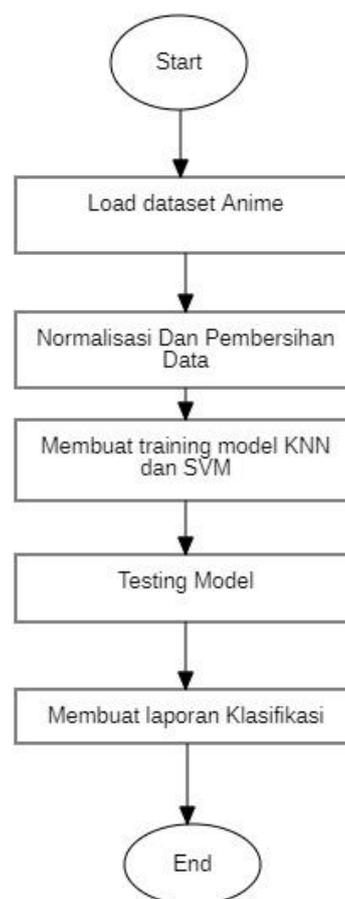
Scikit-Learn adalah salah satu pustaka *Machine Learning* yang bersifat *open source* berbasis *Python* yang dapat digunakan dalam *Data Science*. Kelebihan *Scikit Learn* adalah penggunaannya yang cukup mudah serta memiliki kecepatan saat melakukan banyak pekerjaan berbeda dalam dataset. Pustaka ini dapat dijalankan dengan *SciPy* dan *NumPy*.

2.5 Anime

Anime merupakan animasi yang dibuat sedemikian rupa yang ditayangkan di televisi, film, maupun platform sejenis lainnya yang berasal dari negara Jepang. Anime merupakan ciri khas seni dan hiburan dari negara tersebut dan merupakan hiburan yang sangat unik dimana banyak juga warga negara lain menonton Anime ini untuk sekadar hiburan maupun membuat karya seni lain tentang anime serta mempelajarinya untuk kebutuhan pembelajaran seperti Machine Learning yang terdapat dalam penelitian ini.



Gambar 3. Contoh Anime (Naruto).



Gambar 4. Diagram alir klasifikasi anime.

2.6 Diagram Alir

Pada penelitian ini, alur pengklasifikasian anime bergenre drama adalah sebagai berikut :

Dataset yang digunakan berjumlah 12.294 data dan 2 kelas genre yaitu drama dan non drama, atribut dataset anime berjumlah 7. Dimana dataset akan dinormalisasikan nilainya menggunakan Min-Max Scalar agar mempercepat proses training dan testing.

2.7 Metode Pengumpulan Data

Teknik pengumpulan data dan informasi yang digunakan dalam penelitian ini yaitu dengan melakukan studi literatur yang akan menghasilkan data sekunder. Dataset merupakan data publik yang dapat diambil dari website Kaggle. Atribut yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1 berikut :

Nama Atribut	Tipe Data
anime_id	Nominal
name	Nominal
genre	Biner
type	Nominal
episodes	Nominal

rating	Nominal
members	Nominal

Tabel 1. Atribut data.

2.8 Metode Pengujian

Confusion matrix digunakan sebagai metode pengujian dalam penelitian klasifikasi anime menggunakan algoritma SVM dan KNN ini. Metode ini merepresentasikan hasil klasifikasi yang menggunakan matriks yang dapat dilihat pada Tabel 2.

Klasifikasi benar	Diklasifikasikan sebagai	
	+	-
+	True Positive (TP).	False Negative (FN).
-	False Positive (FP).	True Negative (TN).

Tabel 2. Confusion Matrix

Penjelasan tabel diatas antara lain True Positive yaitu jumlah record positif yang berhasil diklasifikasikan sebagai positif. False positive yaitu record positif yang salah diklasifikasikan menjadi negatif. False negative yaitu record negatif yang salah diklasifikasikan sebagai positif. True negative yaitu record negatif yang berhasil diklasifikasikan sebagai record negatif.

Metode pengujian yang digunakan dalam penelitian ini menghasilkan perhitungan sebagai berikut :

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

Rumus 2. Precision

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

Rumus 3. Recall

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \times 100\%$$

Rumus 4. Accuracy

$$F1\ Score = \frac{2TP}{2TP+FP+FN} \times 100\%$$

Rumus 5. F1 Score

Rumus Precision diatas merupakan perbandingan antara True Positive (TP) dengan banyaknya data yang diprediksi positif.

Rumus Recall diatas merupakan perbandingan antara True Positive (TP) dengan banyaknya data yang sebenarnya positif.

Rumus Accuracy atau akurasi diatas yaitu perbandingan antara jumlah data yang benar dengan banyaknya seluruh data.

Rumus F1 Score diatas yaitu harmoni rata-rata dari precision dan recall.

3. Hasil dan Pembahasan

3.1 Dataset

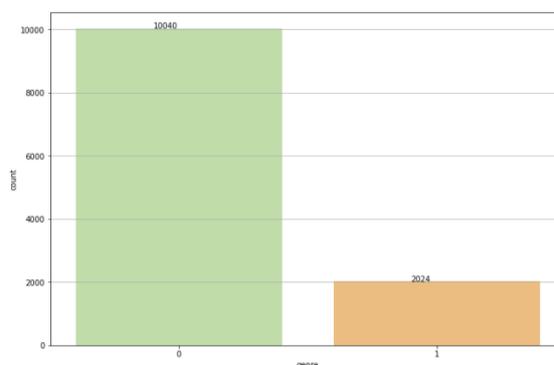
Dataset yang digunakan berjumlah 12.294 data dan 2 kelas genre yaitu drama dan non drama, atribut dataset anime berjumlah 7 yang diambil dari Kaggle sebagai data publik. Dataset dibagi menjadi dua bagian yaitu untuk data training sebanyak 80% dan untuk data testing sebanyak 20% dari dataset yang ada.

3.2 Klasifikasi Anime

Dalam mengklasifikasi anime yang harus kita lakukan adalah load dataset yang kita butuhkan yaitu data anime. Kita dapat normalisasi dataset yang kita miliki dengan *Min-Max Scalar*, *Min-Max Scalar* akan menormalisasikan nilai jika nilai variabel yang kita miliki rentangnya sama. Hal ini dilakukan agar tidak ada nilai yang terlalu besar maupun nilainya terlalu kecil. Hal ini dapat membuat analisis statistik menjadi efisien dan lebih mudah.

Kita juga harus melakukan pembersihan data, dikarenakan tidak semua baris data memiliki nilai yang lengkap. Terkadang terdapat nilai null, nilai NaN atau nilai yang tidak ditentukan, nilai negatif dan sebagainya. Nilai-nilai tersebut akan menghambat pengklasifikasian dan membuat *error*.

Kita dapat memvisualisasikan dataset yang kita miliki seperti gambarsebagai berikut :



Gambar 5. Visualisasi jumlah anime disetiap tipe kelas.

Dari grafik diatas berikut detail jumlah anime dari setiap kelas yang ada :

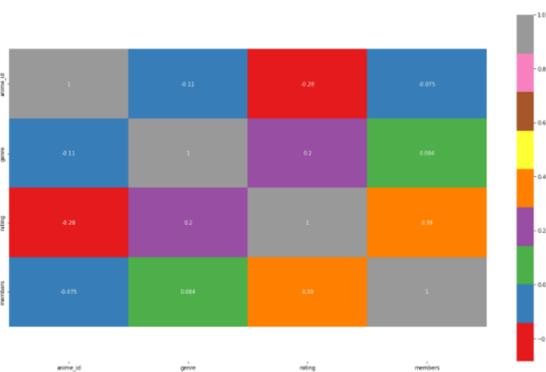
Tipe kelas (angka)	Tipe kelas (huruf)	Jumlah anime
1	Drama	2.024

2	Non Drama	10.040
---	-----------	--------

Tabel 3. Pesebaran anime berdasarkan kelas sebelum ditraining.

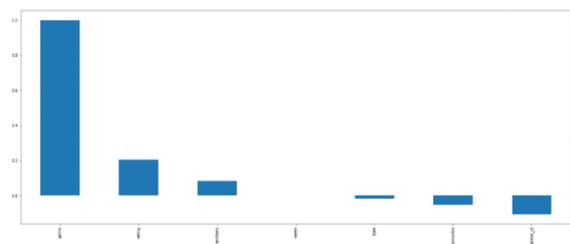
Kita juga bisa memvisualisasikan data kita sesuai dengan korelasinya, dapat menggunakan Heat Map. Matriks korelasi merupakan suatu matriks yang di dalamnya terdapat korelasi – korelasi dimana korelasi yaitu pembelajaran yang membahas tentang hubungan antara dua variabel atau lebih. Korelasi merupakan salah satu teknik statistika yang banyak digunakan oleh peneliti umumnya tertarik terhadap peristiwa-peristiwa yang terjadi dan mencoba menghubungkannya. Dalam penelitian ini menggunakan

Heatmap, yaitu Sebuah peta yang merepresentasikan grafis dari data di mana nilai-nilai individu yang terkandung dalam matriks yang direpresentasikan sebagai warna .



Gambar 6. Heatmap dataset.

Berikut adalah grafik bar dari dataset anime berdasarkan atribut :



Gambar 7. Grafik Bar.

Setelah itu, dapat dilakukan pemodelan dengan Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN). Pemodelan dengan Support Vector Machine (SVM) menghasilkan nilai akurasi training sebesar 83% dan akurasi testing sebesar 82%.

Detail laporan klasifikasi anime menggunakan SVM adalah sebagai berikut :

<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F1 Score</i>
41%	50%	82%	45%

Tabel 4. Laporan klasifikasi SVM

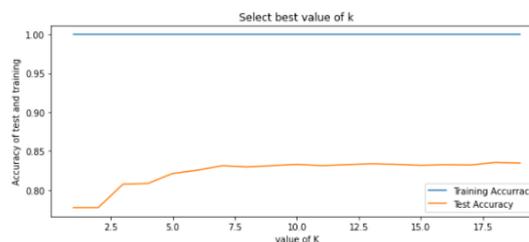
Algoritma K-Nearest Neighbors (KNN) menghasilkan nilai akurasi training sebesar 100% dan nilai akurasi testing sebesar 84%.

Detail laporan klasifikasi anime menggunakan KNN adalah sebagai berikut :

<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>F1 Score</i>
74%	57%	84%	58%

Tabel 5. Laporan klasifikasi KNN

Berikut grafik akurasi testing dan training KNN dengan kfold = 10. Dalam melakukan training dan testing KNN, digunakan *GridSearchCV*. Dimana *GridSearchCV* adalah salah satu bagian dari modul scikit-learn yang berguna untuk melakukan validasi terhadap lebih dari hyperparameter dan satu model masing-masing secara sistematis dan otomatis. Berikut grafik akurasi testing dan training algoritma KNN :



Gambar 8. Grafik akurasi testing dan training Algoritma KKN

Grafik diatas didapatkan dari iterasi k dimana nilai k yaitu 1-19. Berikut detail akurasi testing dan training yang didapatkan :

<i>k</i>	<i>Akurasi testing</i>	<i>Akurasi training</i>
1	78%	100%
2	78%	100%
3	81%	100%
4	81%	100%
5	82%	100%
6	83%	100%
7	83%	100%
8	83%	100%
9	83%	100%

10	83%	100%
11	83%	100%
12	83%	100%
13	83%	100%
14	83%	100%
15	83%	100%
16	83%	100%
17	83%	100%
18	84%	100%
19	84%	100%

Tabel 6. Detail akurasi testing dan training KNN menggunakan k (1-19)

Dari data tabel diatas, didapatkan akurasi training terbaik KNN sebesar 100% dan akurasi testing terbaik sebesar 84%.

4. Kesimpulan

Klasifikasi dalam data mining merupakan salah satu penggunaan metode yang objek paling umum, suatu objek yang belum diketahui kelas/labelnya dapat melalui metode klasifikasi agar kelasnya dapat diperkirakan. Klasifikasi memiliki beberapa algoritma diantaranya Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN). Penggunaan algoritma SVM dan KNN telah banyak digunakan dan dan tingkat akurasinya yang baik.

Dalam penelitian ini, akan menganalisa perbandingan diantara kedua algoritma tersebut yaitu algoritma SVM dan KNN yang akan dilatih dan diuji menggunakan dataset anime yang jumlahnya 12.294 data dan ada 2 kelas genre yaitu drama dan non drama.

Didapatkan hasil dari algoritma dengan K-Nearest Neighbors (KNN) yang menghasilkan nilai akurasi training sebesar 100% dan nilai akurasi testing sebesar 84%. Dan juga hasil dari algoritma Support Vector Machine (SVM) menghasilkan nilai akurasi training sebesar 83% dan nilai akurasi testing sebesar 82%.

Pada penelitian klasifikasi anime ini, algoritma K-Nearest Neighbors (KNN) memiliki akurasi testing yang lebih baik dari Support Vector Machine (SVM) dengan selisih keduanya cukup tipis, diharapkan kedepannya dilakukan optimasi model agar hasil yang didapatkan lebih baik lagi.

Daftar Rujukan

- [1] Y. Ardilla, W. I. Sabilla, and N. Ulinuha, "Implementasi Algoritma Ant Tree Miner Untuk Klasifikasi Jenis Fauna," *Infotekmesin*, vol. 12, no. 2, pp. 150–154, 2021, doi: 10.35970/infotekmesin.v12i2.616.
- [2] S. Dwiasnati and Y. Devianto, "Optimasi Prediksi Bencana Banjir menggunakan Algoritma SVM untuk penentuan Daerah Rawan Bencana Banjir," *Pros. SISFOTEK*, pp. 202–207, 2021, [Online]. Available: <http://seminar.iaii.or.id/index.php/SISFOTEK/article/view/283>.
- [3] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neighbors pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.
- [4] A. A. Karim, H. Azis, and Y. Salim, "Kinerja Metode C4.5 dalam Penyaluran Bantuan Dana Bencana," *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 2, pp. 84–87, 2018.
- [5] S. Hadiani¹, E. Rahmawati, I. Kurniawan, and W. Gata, "Memprediksi Pembayaran Biaya Kursus Siswa Enter (English Center) Dengan Algoritma K-Nn," *Semin. Nas. Teknol. Inf. Univ. Ibn Khaldun Bogor*, no. 18, p. 31908575, 2018, [Online]. Available: www.antaranews.com, 2016.
- [6] H. W. Dhany and F. Izhari, "Analisis Algorithms Support Vector Machine Dengan Naive Bayes Kernel Pada Klasifikasi Data," *J. Tek. Dan Inform.*, vol. 6, no. 2, pp. 30–35, 2019, [Online]. Available: <http://jurnal.pancabudi.ac.id/index.php/Juti/article/view/675>.
- [7] A. A. P. Henny Leidiyana, "Pemodelan Klasifikasi Dalam Meningkatkan Proses," *J. Tek. Inform. Univ. Muhammadiyah Tangerang*, vol. ISSN: 254, no. Januari 2020, pp. 7–14, 2020.
- [8] B. Hardiyanto and F. Rozi, "Prediksi penjualan sepatu menggunakan metode k- nearest neighbor," vol. 04, pp. 13–18, 2020.
- [9] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 269–274, 2019, doi: 10.33096/ilkom.v11i3.489.269-274.
- [10] N. I. Widiastuti, E. Rainarli, and K. E. Dewi, [1] R. Sistem, P. N. Andono, and E. H. Rachmawanto, "JURNAL RESTI Evaluasi Ekstraksi Fitur GLCM dan LBP Menggunakan Multikernel SVM," vol. 1, no. 10, pp. 1–9, 2021.
- [11] T. E. Puspitawati, S. Basuki, V. Rahmayanti, and S. Nastiti, "Algoritma Support Vector Machine (SVM) Untuk Identifikasi Komponen Abstrak Pada Jurnal Ilmiah Berbasis Teknik Klasifikasi," vol. 3, no. 5, pp. 441–448, 2021.
- [12] H. Nalatissifa, W. Gata, S. Diantika, and K. Nisa, "Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 578, 2021, doi: 10.32493/informatika.v5i4.7575.
- [13] S. R. Raysyah, Veri Arinal, and Dadang Iskandar Mulyana, "Klasifikasi Tingkat Kematangan Buah Kopi Berdasarkan Deteksi Warna Menggunakan Metode Knn Dan Pca," *JSiI (Jurnal Sist. Informasi)*, vol. 8, no. 2, pp. 88–95, 2021, doi: 10.30656/jsi.v8i2.3638.