

KLASIFIKASI TINGKAT PENJUALAN VIDEO GAME DENGAN MENGGUNAKAN METODE K – NEAREST NEIGHBORS

Nadhif Nurul Fajri Adzani¹, Wina Witanti², Fajri Rakhmat Umbara³
^{1,2,3}Informatika, Sains dan Informatika, Universitas Jenderal Achmad Yani
Email: [1nadhifna@gmail.com](mailto:nadhifna@gmail.com), [2wni@if.unjani.ac.id](mailto:wni@if.unjani.ac.id), [3fajriumbara@gmail.com](mailto:fajriumbara@gmail.com)

ABSTRACT

Classification of Video Game Sales Levels Using the K – Nearest Neighbors Method has a function to classify video games based on sales, and requires variables, such as genre, platform, publisher, best seller. Problems that occur on game sales platforms such as Steam, Epic games, etc. This is when gamers buy the game and it turns out that the game does not meet the expectations of the gamer who bought the game, as a result the game is no longer played. Therefore, the solution created here is to classify video games based on characteristics using the KNN method, where later video games will be divided based on their characteristics, and several games will be displayed according to their character classification, so that it is hoped that it can minimize the incidence of game buyers / gamers who regret not having according to their expectations

Keywords: Classification, Video games, Gamers, K-Nearest Neighbor, KNN.

ABSTRAK

Klasifikasi Tingkat Penjualan Video Game Dengan Menggunakan Metode K – Nearest Neighbors memiliki fungsi untuk mengklasifikasikan video game berdasarkan penjualannya, dan memerlukan variabel, seperti genre, platform, publisher, best seller. Permasalahan yang terjadi di Platform penjualan game seperti di Steam, Epic games, etc. Adalah dimana saat gamers membeli game tersebut dan ternyata game tersebut tidak sesuai dengan ekspektasi dari gamers yang membeli game tersebut alhasil game tidak lagi dimainkan. Oleh karena itu, solusi yang dibuat disini yaitu klasifikasi video game berdasarkan karakteristik yang menggunakan metode KNN, dimana nantinya video game akan dibagi berdasarkan karakteristiknya, dan akan ditampilkan beberapa game sesuai klasifikasi karakternya, sehingga diharapkan dapat meminimalisir kejadian pembeli game / gamers yang menyesal karena tidak sesuai dengan ekspektasi mereka

Kata Kunci: Klasifikasi, Video games, Gamers, K-Nearest Neighbor, KNN.

Riwayat Artikel :

Tanggal diterima : 13-11-2023

Tanggal revisi : 15-11-2023

Tanggal terbit : 16-11-2023

DOI :

<https://doi.org/10.31949/infotech.v9i2.7371>

INFOTECH journal by Informatika UNMA is licensed under CC BY-SA 4.0

Copyright © 2023 By Author



1. PENDAHULUAN

1.1. Latar Belakang

Video game adalah hal yang paling melekat pada setiap orang, khususnya anak muda / remaja (Novrialdy 2019).

Pada saat ini banyak game yang dibandrol dengan harga yang cukup tinggi, gamers tidak ingin lagi mengambil keputusan yang salah, gamers diharuskan untuk lebih teliti dalam memilih genre game sesuai dengan karakteristiknya dengan cara melihat review dan klasifikasi dimana jika sudah sesuai dengan yang diharapkan, barulah game tersebut di checkout dari berbagai platform penjualannya. (Goumagias et al. 2014)

Klasifikasi dilakukan untuk menampilkan beberapa game sesuai dengan penjualannya, yang nantinya akan menampilkan berdasarkan yang terdekat dengan spesifikasi tersebut, yang dimana diperlukan alat pengukur yang akan mengukur hubungan antara variabel acak yang memiliki keterkaitan terdekat, seperti kNN (K- Nearest Neighbor) (Gao, Oh, and Viswanath 2018). Kemudian, adapun penelitian KNN yang telah dilakukan sebelum penelitian ini dibuat yaitu yang memiliki judul “Klasifikasi Sepeda Motor Berdasarkan Karakter dari Konsumen dengan menggunakan metode K – Nearest Neighbour. (Putra et al. 2017)

Serupa dengan penelitian yang akan dilakukan oleh penulis, diperlukan pengelompokan berupa pengelompokan berdasarkan karakteristik dari pemainnya, dimana setiap pemain pasti memiliki ketertarikan game yang berbeda, dari segi genre saja memiliki banyak jenisnya, seperti strategi, olahraga, RPG, dan lain – lain., sehingga perlu ada klasifikasi untuk menyesuaikan hal tersebut (Caesar 2015) . KNN pada permasalahan ini sendiri merupakan sebuah metode pengelompokan, dimana lebih tepatnya KNN ini merupakan sebuah metode untuk melakukan klasifikasi berdasarkan kelompok data uji berdasarkan k data yang terdekatnya. (Putra et al. 2017) Dengan menggunakan metode KNN ini, nantinya setelah diklasifikasikan akan menghasilkan data berupa beberapa data valid yang masuk dalam klasifikasi yang telah ditentukan sebelumnya yang mendekati / bertetangga (Whidhiasih, Wahanani, and Supriyanto 2013)

Sehingga, sesuai dengan permasalahan yang ada, maka dibuatlah penelitian ini, yang dimana berfungsi untuk melakukan Klasifikasi Game berdasarkan penjualannya menggunakan metode KNN, yang nantinya akan menampilkan data sebanyak K, berdasarkan genre, platform, best seller. Sesuai dengan yang pengguna butuhkan.

1.2. Tinjauan Pustaka

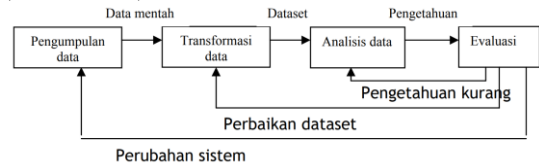
a. K- Nearest Neighbor (KNN)

K-Nearest-Neighbours (kNN) adalah metode klasifikasi non-parametrik, yaitu sederhana dan efektif dalam berbagai permasalahan. Untuk record dari data t yang akan diklasifikasikan, variabel k-nya diambil dari data tengga yang terdekat, dan ini akan

membentuk lingkungan t. Suara yang terbanyak di antara catatan di lingkungan data yang ada biasanya digunakan untuk memutuskan klasifikasi untuk t dengan atau tanpa pertimbangan dari nilai variabel berdasarkan jarak. (Isman, Andani Ahmad, and Abdul Latief 2021)

b. Data Mining

Data Mining memiliki beberapa definisi. Definisi data mining yaitu sebagai berikut : (1) menemukan pola yang berbeda diantara sejumlah besar data yang telah disimpan. Ini adalah metode yang sangat dibutuhkan dan merupakan evolusi logis dari teknologi database. (2) Menganalisis pemrosesan otomatis atau semi-otomatis dari sejumlah besar data untuk mencari pola dan hukum yang signifikan. (Beebe 2023)



Gambar 1 Proses Data Mining

c. Transformasi Data

Transformasi data adalah proses data mining dimana format / tipe data yang digunakan disesuaikan dengan format atau tipe data sehingga dapat digunakan untuk metode algoritma data mining yang akan digunakan, ada pula pembagian data menjadi dua bagian data diantaranya data uji dan data latih. (Dwi Handoko et al. n.d.)

d. Analisis Data

Analisis data dilakukan untuk memastikan bahwa data yang digunakan sesuai dengan penelitian yang akan dilakukan, menunjukkan keterhubungan data dari berbagai lapisan yang ada pada data tersebut, sehingga tidak ada lagi data yang salah, atau tidak relevan. (Ayub n.d.)

Pada analisis data terdapat beberapa proses yang disebut dengan pre-processing data, pre-processing data merupakan sebuah teknik untuk menghapus, memperbaiki, dan juga memanipulasi data. Berfungsi untuk meningkatkan kinerja dari sistem, dan juga memastikan agar sistem dapat berjalan dengan baik. (Firdaus 2017)

Pre-processing data memiliki beberapa proses tahapan yaitu Data cleaning, Data integration, Data transformation, dan Data reduction. (Nabila, Rahman Isnain, and Abidin 2021)

e. Proses Data Mining

Langkah pertama dalam proses data mining yaitu menentukan metode dari data mining tersebut, diantaranya yaitu regresi, klasifikasi, dan klasterisasi. Lalu setelah mendapatkan metode yang diinginkan, peneliti diharuskan memilih algoritma yang akan digunakan sesuai dengan metode yang dipilih dari data mining tersebut, lalu setelah itu dilakukan implementasi dalam proses data mining. (Ayub n.d.)

f. Evaluasi

Evaluasi dilakukan untuk melakukan penilaian pada proses *data mining* yang dilakukan, apa saja kekurangannya, yang harus dikembangkan, data yang digunakan. Agar selanjutnya dapat dilakukan perbagikan pada proses data mining sesuai dengan apa yang telah dievaluasi. (Divisive et al. n.d.)

g. Klasifikasi

Klasifikasi merupakan penyusunan atau pengelompokan data sesuai dengan sistematis atau golongan yang telah ditetapkan menurut pembagian kelas – kelas, dimana setiap data memiliki kesamaan dan perbedaan ciri – cirinya masing – masing(Dwi Handoko et al. n.d.)

h. K- Nearest Neighbor (KNN) Classification

Konsep dasar method KNN ini adalah algoritma untuk melakukan klasifikasi berdasarkan tetangga terdekat. (Delany 2021). Dalam kasus ini, menggunakan rumus Euclidean Distance. Rumus Euclidean Distance ini mirip dengan rumus pythagoras, bedanya yaitu data yang digunakan lebih dari 2, yaitu sebagai berikut

$$d_i = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

Keterangan:

d = Jarak

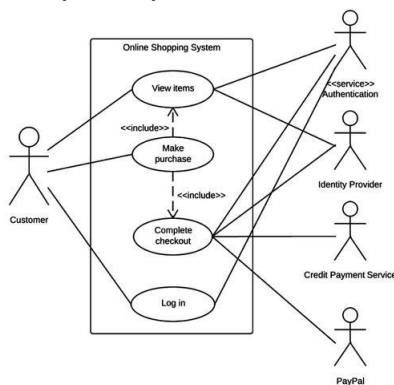
i = Variabel data

n = Jumlah data

qi = Data uji atau data testing

i. Use Case Diagram

Use case adalah deskripsi operasi sistem yang dilihat dari sudut pandang aktor yang menggunakannya. *Use case* menguraikan fungsi dan data yang akan diproses oleh system. *Use case* bekerja dengan menggunakan scenario yang merupakan deskripsi dari urutan atau pemrosesan yang menjelaskan apa yang dilakukan oleh user terhadap sistem maupun sebaliknya. (Setiyani 2021)



Gambar 2. Use Case

j. Black Box Testing

Metode *Black box* Testing merupakan salah satu metode yang mudah digunakan karena hanya memerlukan batas bawah dan batas atas dari data

yang di harapkan,Estimasi banyaknya data uji dapat dihitung melalui banyaknya field data entri yang akan diuji, aturan entri yang harus dipenuhi serta kasus batas atas dan batas bawah yang memenuhi. Dan dengan metode ini dapat diketahui jika fungsionalitas masih dapat menerima masukan data yang tidak diharapkan maka menyebabkan data yang disimpan kurang valid. Tujuan dari black box testing yaitu untuk menentukan apakah fungsi dari perangkat lunak tersebut memenuhi spesifikasi yang diperlukan, uji perangkat lunak terhadap fungsi-fungsinya saja tanpa harus menguji keseluruhan dari kode program. Metode pengujian kotak hitam mudah diterapkan karena hanya membutuhkan perangkat lunak yang sudah dibangun.(Nur Cholifah and Melati Sagita 2018)

k. Equivalence Partitioning

Equivalence partitioning merupakan Teknik atau metode dari *black box testing* yang akan dilakukan pada penelitian ini. Pada Teknik *black box* tidak akan menguji desain coding pada program, dan berfungsi untuk memastikan setiap fitur pada perangkat lunak tersebut telah memenuhi kualifikasi dan spesifikasi. (Ningrum et al. 2019)

Tabel 1 *Equivalence Partitioning*

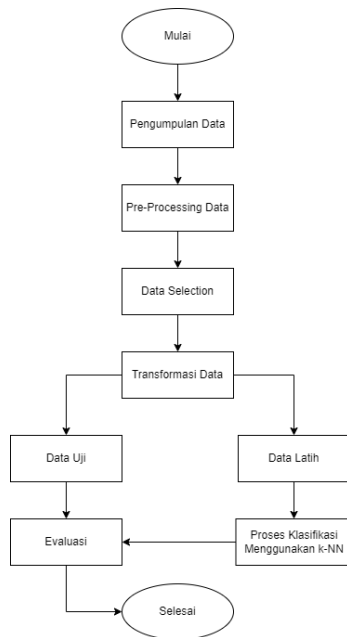
No	Id	Hasil Pengujian
1	T001	Berhasil
2	T002	Gagal
3	T003	Berhasil

l. Supervised Learning

Metode KNN adalah metode yang digunakan pada penelitian ini, metode ini termasuk ke dalam *supervised learning*, karena pada penelitian ini adalah penelitian data mining klasifikasi. Dimana terdapat proses melatih mesin dengan menggunakan data yang "diberi label". Artinya beberapa data sudah diberi label dengan jawaban yang benar. Metode ini bekerja dengan mencari k pola (diantara semua pola latih disemua kelas) yang berdekatan atau bertetangga dengan pola data uji kemudian menentukan kelas keputusan berdasarkan jumlah pola terbanyak. Proses pelatihan KNN menghasilkan k yang memberikan akurasi tertinggi dalam menggeneralisasi data yang akan datang. Jadi proses pelatihan proses pelatihan pada dasarnya adalah melakukan observasi terhadap sejumlah k sampai dihasilkan k yang paling tepat dengan masukan pelatihan data. (Pamungkas, Prasetya, and Kharisudin 2019)

1.3. Metodologi Penelitian

Metodologi Penelitian pada penelitian ini dapat dilihat pada gambar 3.



Gambar 3. Alur Metode Penelitian

a. Pengumpulan Data

Pada metode penelitian, tahapan awal yang dilakukan yaitu pengumpulan data yang dapat bertujuan untuk mendapatkan dan mengumpulkan informasi mengenai dataset penjualan game pada website kaggle.com. Data yang didapat yaitu berdasarkan ulasan dari pemain game pada platform penjualan game.

b. Data Mining

Metode *data mining* pada penelitian ini yaitu menggunakan metode algoritma klasifikasi k-Nearest Neighbor, yang dimana proses ini dimulai dengan pengumpulan data menggunakan dataset dari situs web kaggle.com, dengan dataset “Video Game Sales”. Data tersebut dibuat pada tahun 2016, dan terdiri dari 16.598 record data. Tahap selanjutnya yaitu pre-processing, pada tahap ini terdapat 2 (dua) proses pengolahan data yaitu berupa pembersihan data (data cleaning) dan seleksi data (data selection). Setelah proses itu selesai, tahap selanjutnya yaitu transformasi data berupa merubah data numerik menjadi nominal, dan juga menentukan data untuk menjadi sebuah data uji dan juga data latih. Yang dimana nantinya data latih digunakan untuk melatih model dari algoritma menggunakan metode k-Nearest Neighbor, dan data uji akan digunakan untuk mengetahui performa dari model algoritma yang telah dilatih.

c. Pre-Processing Data

Pre-processing adalah proses menempatkan data yang dapat dipahami dengan cara yang dapat merubah format yang dapat dipahami dari data mentah. Langkah ini sangat penting dalam proses data mining karena menggunakan data mentah akan membuat prosedur fungsi menjadi tidak efektif..(Yang n.d.) Ada pula beberapa proses yang

dilakukan pada tahap pre-processing ini, yaitu diantaranya:

a) Data Cleaning

Data cleaning merupakan Langkah pertama dalam tahapan *pre-processing* data yang dimana berfungsi untuk mencari *value* yang hilang, data yang berantakan, dan data yang tidak konsisten.(Tae et al. 2019) Pada penelitian ini, data cleaning menggunakan metode interpolasi pada Microsoft Excel, kolom Year diisi oleh tahun rilis dari Game tersebut.

b) Data Selection

Tahapan berikutnya yaitu Data selection ini merupakan sebuah tahapan untuk meminimalisir jumlah data yang akan digunakan pada proses data mining dengan data asli yang masih teridentifikasi, pada proses seleksi data ini akan dilakukan pemilihan atribut – atribut yang digunakan pada penelitian ini yaitu *Name, Platform, Year, Genre, Publisher, NA_Sales, EU_Sales, JP_Sales, Other_Sales, Global_Sales*.

d. Transformasi Data

Pada penelitian ini dilakukan transformasi data agar dapat dilakukan penelitian sesuai dengan metode yang digunakan, yang pertama dilakukan yaitu merubah data karakter dan numerik menjadi nominal. Karena berdasarkan metode yang digunakan pada penelitian hanya dapat dilakukan pada data yang bersifat nominal, beberapa data yang berupa karakter yaitu ada *Genre, publisher* dan *platform* yang diubah menjadi tipe nominal dari setiap karakternya, dan ada pula data numerik yang tersedia pada dataset ini yang digunakan dalam penelitian ini yaitu *Global_sales*. Dan pada data yang hilang, dilakukan penanganan menggunakan metode interpolasi pada Excel. Tahapan berikutnya dilakukan split data untuk membagi data uji dan data latih, pembagian yang dilakukan yaitu 80% untuk data uji dan 20% untuk data latih.

a) Split Data

Karena proses klasifikasi dalam penelitian ini dibagi menjadi dua bagian yaitu data uji dan data latih, maka *split data* merupakan tahapan yang harus dilakukan dalam penelitian ini. Data latih digunakan untuk menyempurnakan model data untuk metode k-Nearest Neighbor. Skenario yang memiliki 80% data latih dan 20% data uji dimasukkan dalam tahap ini.

e. Proses Klasifikasi Menggunakan Metode k-NN

Setelah tahap – tahap sebelumnya dikerjakan, selanjutnya yaitu menerapkan metode data mining yang akan digunakan dan telah dibahas sebelumnya, Teknik klasifikasi yang digunakan pada penelitian ini yaitu menggunakan metode k-NN terhadap data yang sebelumnya telah dilakukan pre-processing, seleksi, dan telah ditransformasi pada tahap sebelumnya.

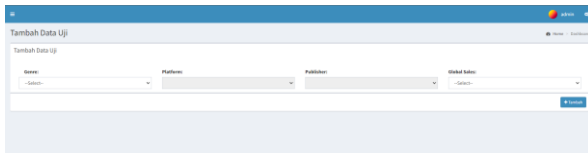
f. Evaluasi

Metode evaluasi yang digunakan pada penelitian ini yaitu menggunakan evaluasi Confusion Matrix. Yang dimana evaluasi ini bertujuan untuk mengetahui kemampuan model algoritma klasifikasi yang telah dibuat yaitu algoritma metode k-NN. Dimana klasifikasi k-nn merupakan sebuah supervised learning, sehingga sangat dianjurkan untuk menggunakan metode confusion matrix sebagai proses evaluasinya.

2. PEMBAHASAN

2.1. Input Data Uji

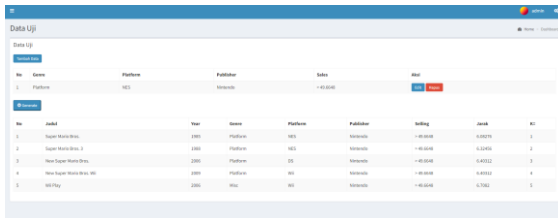
Pada proses input data uji, terdapat form yang harus di-input oleh user sesuai data latih yang tersedia, proses input data uji dapat dilihat pada gambar berikut.



Gambar 4. Input Data Uji

2.2. Generate Hasil Data Uji

Generate hasil data uji adalah hasil klasifikasi antara data uji yang di-input dengan data latih yang tersedia dalam sistem, Hasil Generate Klasifikasi dapat dilihat pada gambar berikut



Gambar 5. Generate Hasil Data Uji

2.3. Pengujian Perhitungan

Pada perhitungan ini, data uji yang akan dihitung ini berupa sebagai berikut:

Genre: Platform (1)

Plaform: NES (12)

Publisher: Nintendo (463)

Sales: 40.24 (5)



Gambar 6. Pengujian Perhitungan

Lalu, digunakan rumus seperti berikut:

$$d = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

pi: Sample Data

qi: Data Testing

i: Variabel Data

n: dimensi data

d: jarak

Implementasi perhitungan pada program klasifikasi penjualan video game dapat dilihat pada gambar berikut.

```
$hasil = v(((($a['u_genre'] - $b['data_genre_val'])^2 + ($a['u_platform'] - $b['data_platform_val'])^2 + ($a['u_publisher'] - $b['data_publisher_val'])^2 + ($a['u_selling'] - $b['data_selling_val'])^2 + ($a['u_selling'] - $b['data_selling_na_val'])^2 + ($a['u_selling'] - $b['data_selling_eu_val'])^2 + ($a['u_selling'] - $b['data_selling_jp_val'])^2 + ($a['u_selling'] - $b['data_selling_ot_val'])^2))
$a = data uji
$b = data kaset
```

Gambar 7. Implementasi Perhitungan

Lalu akan ditampilkan hasil klasifikasi yaitu dapat dilihat pada gambar berikut ini.

No	Judul	Tahun	Genre	Platform	Publisher	Selling	Jarak	K1
1	Super Mario Bros.	2000	Platform	NES	Nintendo	+40.0000	0.0000	1
2	Super Mario Bros. 2	2000	Platform	NES	Nintendo	+40.0000	0.2000	2
3	New Super Mario Bros.	2000	Platform	DS	Nintendo	+40.0000	0.4000	3
4	New Super Mario Bros. Wii	2000	Platform	Wii	Nintendo	+40.0000	0.4000	4
5	Wii Play	2000	Wii	Wii	Nintendo	+40.0000	0.7000	5

Gambar 8. Hasil Klasifikasi

Berikut ini pembuktian dari perhitungan data uji yang telah di-input dan hasil klasifikasinya K1:

$$\sqrt{(1 - 1)^2 + (12 - 12)^2 + (463 - 463)^2 + (5 - 5)^2} = 0$$

K2:

$$\sqrt{(1 - 1)^2 + (12 - 12)^2 + (463 - 463)^2 + (5 - 5)^2} = 0$$

K3:

$$\sqrt{(1 - 1)^2 + (12 - 11)^2 + (463 - 463)^2 + (5 - 5)^2} = 1$$

K4:

$$\sqrt{(1 - 1)^2 + (12 - 13)^2 + (463 - 463)^2 + (5 - 5)^2} = 1$$

K5:

$$\sqrt{(1 - 3)^2 + (12 - 13)^2 + (463 - 463)^2 + (5 - 5)^2} = 2,23$$

2.4. Uji Akurasi Perhitungan

Pada perhitungan yang telah dilakukan, selanjutnya ada uji akurasi dari perhitungannya menggunakan Confusion Matrix, yang nantinya akan mendapatkan hasil berupa nilai akurasi, presisi dan nilai recall. Berikut contoh hasil dari generate klasifikasi.

No	Genre	Platform	Publisher	Sales	K
1	Action	PS3	Take-Two Interactive	~140.000	5
2	Adventure	PS3	Ubisoft	~1.50.000	5

No	Judul	Year	Genre	Platform	Publisher	Selling	Jarak	K
1.00	Grand Theft Auto IV	2008	Action	PS3	Take-Two Interactive	10,07	0	1 (Target Release)
2.00	Grand Theft Auto V	2013	Action	PS3	Take-Two Interactive	21,4	1	2 (Relevansi)
3.00	Grand Theft Auto: San Andreas	2004	Action	PS2	Take-Two Interactive	20,81	1.41421	3 (Relevansi)
4.00	Grand Theft Auto V	2014	Action	PS4	Take-Two Interactive	11,98	2	4 (Meningkatkan Relevansi)
5.00	The Elder Scrolls IV: Oblivion	2006	Role-Playing	X360	Take-Two Interactive	8,39	2,23057	5 (Meningkatkan Relevansi)
6.00	Assassin's Creed	2007	Adventure	PS3	Ubisoft	8,03	0	1 (Target Release)
7.00	Assassin's Creed III	2012	Action	PS3	Ubisoft	6,5	1	2 (Relevansi)
8.00	Assassin's Creed III	2012	Action	X360	Ubisoft	9,3	1,41421	3 (Relevansi)
9.00	Watch Dogs	2014	Action	PS2	Ubisoft	8,14	2,23057	4 (Meningkatkan Relevansi)
10.00	Assassin's Creed Unity	2014	Action	X360	Ubisoft	9,46	1,60002	5 (Meningkatkan Relevansi)

Gambar 9. Contoh Data uji dan hasil generate klasifikasi

2.5. Confusion Matrix

Confusion Matrix berfungsi untuk mencari nilai akurasi, presisi dan recall. Berikut ini cara untuk mencari nilai akurasi.

	True Positif	True Negatif	True Negatif	True Negatif	Cost
Practical Target Release	2	0	0	0	
Practical Release	4	0	0	0	
Practical Missing Release	4	0	0	0	
Class Recall	20 %		NaN %	NaN %	

Gambar 10. Confusion Matrix

a) Nilai Recall (sensitifitas) merupakan rasio prediksi benar positif yang dibandingkan dengan seluruh data yang benar positif. Menghitung nilai Recall yaitu sebagai berikut.

$$Recall = \frac{496}{496 + 3} = 0,993 * 100 = 99,3\%$$

Dengan data yang ada didapatkan nilai recall yaitu 99,3%

b) Nilai Precision (presisi) adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Menghitung nilai precision yaitu sebagai berikut

$$Precision = \frac{496}{496 + 1} = 0,997 * 100 = 99,7\%$$

Dari data yang ada dalam sistem, didapatkan nilai precision yaitu 99,7%

c) Nilai akurasi adalah tingkat akurasi dari confusion matriks dari data yang ada, nilai akurasi dapat dilihat pada perhitungan berikut.

$$Akurasi = \frac{496 + 1}{496 + 1 + 3 + 0} = 0,994 * 100 = 99,4\%$$

3. KESIMPULAN

Dalam penelitian ini tahapan Klasifikasi Penjualan Video Game Menggunakan Metode k-NN yang dimana terdapat variabel Nama Game, Genre, Publisher, Platform, tahun rilis, Na_Sales, Eu_Sales, Jp_Sales, Other_Sales, dan Global_Sales menjadi attribute utama dalam proses Klasifikasi Penjualan Video Game. Penelitian ini juga akan menghasilkan Hasil klasifikasi berupa video game yang telah diklasifikasi berdasarkan input data uji berupa Genre, Publisher, Platform, dan Sales. Dataset yang digunakan didapatkan dari website kaggle.com, dimana data yang tersedia berjumlah 12600 data, setelah melewati data selection, cleaning data, dan pre-processing data, Data uji kemudian dihitung menggunakan metode k-NN, dan dicari nilai terkecil

atau terdekat dari input yang dilakukan, proses perhitungannya merupakan hasil klasifikasi yang ditampilkan sesuai dengan input yang dilakukan yaitu berjumlah k= 5, dimana k1 merupakan nilai yang paling sesuai, dan seterusnya.

PUSTAKA

Ayub, Mewati. *Proses Data Mining Dalam Sistem Pembelajaran Berbantuan Komputer*.

Beebe, Nelson H F. 2023. *Time Series Analysis Title Word Cross-Reference*.

Caesar, Rio. 2015. "Kajian Pustaka Perkembangan Genre Games Dari Masa Ke Masa." *Journal of Animation & Games Studies* 1(2): 113–1134.

Delany, Sarah Jane. 2021. "K-Nearest Neighbour Classifiers - A Tutorial." 54(6).

Divisive, Algoritma Hirarki et al. *Irawan-Penerapan Data Mining Untuk Evaluasi Data Penjualan Menggunakan Metode Clustering Dan PENERAPAN DATA MINING UNTUK EVALUASI DATA PENJUALAN MENGGUNAKAN METODE CLUSTERING DAN ALGORITMA HIRARKI DIVISIVE* Yuda Irawan.

Dwi Handoko, Fajar et al. "TRANSFORMASI DATA MENJADI INFORMASI PADA BISNIS INTELIJEN." 2(3): 2022. <https://doi.org/10.38035/jihhp.v2i3>.

Firdaus, Diky. 2017. 6 *Jurnal Penggunaan Data Mining Dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer*.

Gao, Weihao, Sewoong Oh, and Pramod Viswanath. 2018. "Demystifying Fixed k-Nearest Neighbor Information Estimators." *IEEE Transactions on Information Theory* 64(8): 5629–61.

Goumagias, Nikolaos et al. 2014. "A Phylogenetic Classification of the Video-Game Industry's Business Model Ecosystem." *IFIP Advances in Information and Communication Technology* 434: 285–94.

Isman, Andani Ahmad, and Abdul Latief. 2021. "Perbandingan Metode KNN Dan LBPH Pada Klasifikasi Daun Herbal." *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)* 5(3): 557–64.

Nabila, Zulfa, Auliya Rahman Isnain, and Zaenal Abidin. 2021. "ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS." *Jurnal Teknologi dan Sistem Informasi (JTISI)* 2(2): 100. <http://jim.teknokrat.ac.id/index.php/JTISI>.

Ningrum, Fadhila Cahya et al. 2019. "Pengujian Black Box Pada Aplikasi Sistem Seleksi

- Sales Terbaik Menggunakan Teknik Equivalence Partitions.” 4(4).
<http://openjournal.unpam.ac.id/index.php/informatika>.
- Novrialdy, Eryzal. 2019. “Kecanduan Game Online Pada Remaja: Dampak Dan Pencegahannya.” *Buletin Psikologi* 27(2): 148.
- Nur Cholifah, Wahyu, and Sri Melati Sagita. 2018. 3 Jurnal String *PENGUJIAN BLACK BOX TESTING PADA APLIKASI ACTION & STRATEGY BERBASIS ANDROID DENGAN TEKNOLOGI PHONEGAP*.
- Pamungkas, F S, B D Prasetya, and I Kharisudin. 2019. “Perbandingan Metode Klasifikasi Supervised Learning Pada Data Bank Customers Menggunakan Python.” *PRISMA, Prosiding Seminar Nasional Matematika* 3: 689–94.
<https://journal.unnes.ac.id/sju/index.php/prisma/>.
- Putra, Nanda Agung et al. 2017. “Klasifikasi Sepeda Motor Berdasarkan Karakteristik Konsumen Dengan Metode K-Nearest Neighbour Pada Big Data Menggunakan Hadoop Single Node Cluster.” *Jurnal Teknologi Informasi dan Ilmu Komputer* 4(2): 81.
- Setiyani, Lila. 2021. *Implementasi Cybersecurity Pada Operasional Organisasi*.
- Tae, Ki Hyun et al. 2019. “Data Cleaning for Accurate, Fair, and Robust Models: A Big Data - AI Integration Approach.”
<http://arxiv.org/abs/1904.10761>.
- Whidhiasih, R. N., N. A. Wahanani, and Supriyanto. 2013. “Klasifikasi Buah Belimbing Berdasarkan Citra RED-GREEN-BLUE.” *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic* 1(1): 29–35.
- Yang, Hui. *Data Preprocessing-Chapter 3*.
<http://cose-stor.sfsu.edu/~huiyang>.
- Ayub, Mewati. *Proses Data Mining Dalam Sistem Pembelajaran Berbantuan Komputer*.
- Beebe, Nelson H F. 2023. *Time Series Analysis Title Word Cross-Reference*.
- Caesar, Rio. 2015. “Kajian Pustaka Perkembangan Genre Games Dari Masa Ke Masa.” *Journal of Animation & Games Studies* 1(2): 113–1134.
- Delany, Sarah Jane. 2021. “K-Nearest Neighbour Classifiers - A Tutorial.” 54(6).
- Divisive, Algoritma Hirarki et al. *Irawan-Penerapan Data Mining Untuk Evaluasi Data Penjualan Menggunakan Metode Clustering Dan PENERAPAN DATA MINING UNTUK EVALUASI DATA PENJUALAN MENGGUNAKAN METODE CLUSTERING DAN ALGORITMA HIRARKI DIVISIVE Yuda Irawan*.
- Dwi Handoko, Fajar et al. “TRANSFORMASI DATA MENJADI INFORMASI PADA BISNIS INTELIJEN.” 2(3): 2022.
<https://doi.org/10.38035/jihhp.v2i3>.
- Firdaus, Diky. 2017. 6 Jurnal *Penggunaan Data Mining Dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer*.
- Gao, Weihao, Sewoong Oh, and Pramod Viswanath. 2018. “Demystifying Fixed κ -Nearest Neighbor Information Estimators.” *IEEE Transactions on Information Theory* 64(8): 5629–61.
- Goumagias, Nikolaos et al. 2014. “A Phylogenetic Classification of the Video-Game Industry’s Business Model Ecosystem.” *IFIP Advances in Information and Communication Technology* 434: 285–94.
- Isman, Andani Ahmad, and Abdul Latief. 2021. “Perbandingan Metode KNN Dan LBPH Pada Klasifikasi Daun Herbal.” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)* 5(3): 557–64.
- Nabila, Zulfa, Auliya Rahman Isnain, and Zaenal Abidin. 2021. “ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS.” *Jurnal Teknologi dan Sistem Informasi (JTSI)* 2(2): 100.
<http://jim.teknokrat.ac.id/index.php/JTSI>.
- Ningrum, Fadhila Cahya et al. 2019. “Pengujian Black Box Pada Aplikasi Sistem Seleksi Sales Terbaik Menggunakan Teknik Equivalence Partitions.” 4(4).
<http://openjournal.unpam.ac.id/index.php/informatika>.
- Novrialdy, Eryzal. 2019. “Kecanduan Game Online Pada Remaja: Dampak Dan Pencegahannya.” *Buletin Psikologi* 27(2): 148.
- Nur Cholifah, Wahyu, and Sri Melati Sagita. 2018. 3 Jurnal String *PENGUJIAN BLACK BOX TESTING PADA APLIKASI ACTION & STRATEGY BERBASIS ANDROID DENGAN TEKNOLOGI PHONEGAP*.
- Pamungkas, F S, B D Prasetya, and I Kharisudin. 2019. “Perbandingan Metode Klasifikasi Supervised Learning Pada Data Bank Customers Menggunakan Python.” *PRISMA, Prosiding Seminar Nasional Matematika* 3: 689–94.
<https://journal.unnes.ac.id/sju/index.php/prisma/>.
- Putra, Nanda Agung et al. 2017. “Klasifikasi Sepeda Motor Berdasarkan Karakteristik Konsumen Dengan Metode K-Nearest Neighbour Pada Big Data Menggunakan Hadoop Single Node Cluster.” *Jurnal*

Teknologi Informasi dan Ilmu Komputer 4(2):
81.

Setiyani, Lila. 2021. *Implementasi Cybersecurity Pada Operasional Organisasi*.

Tae, Ki Hyun et al. 2019. "Data Cleaning for Accurate, Fair, and Robust Models: A Big Data - AI Integration Approach." <http://arxiv.org/abs/1904.10761>.

Whidhiasih, R. N., N. A. Wahanani, and Supriyanto. 2013. "Klasifikasi Buah Belimbing Berdasarkan Citra RED-GREEN-BLUE." *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic* 1(1): 29–35.

Yang, Hui. *Data Preprocessing-Chapter 3*. <http://cose-stor.sfsu.edu/~huiyang>.