

**Klasifikasi Data FILM untuk Menentukan Rekomendasi FILM untuk Anak
Menggunakan Algoritma C5.0**

Tugas Akhir

Diajukan Untuk Memenuhi
Persyaratan Guna Meraih Gelar Sarjana Strata 1
Informatika Universitas Muhammadiyah Malang



**Ansyari Aqshal Raditya JS
201810370311055**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MALANG**

2024

LEMBAR PERSETUJUAN

**Klasifikasi data film untuk menentukan rekomendasi film untuk anak
menggunakan algoritma C5.0**

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1

Informatika Universitas Muhammadiyah Malang

Menyetujui,

Malang, 22 Maret 2025

Dosen Pembimbing 1

Dosen Pembimbing 2



Didih Rizki Chandranegara S.kom.,

M.Kom

NIP. 180302101992PNS.

Zamah Sari ST., MT.

NIP. 10814100555PNS.

LEMBAR PENGESAHAN

**Klasifikasi data film untuk menentukan rekomendasi film untuk
anak menggunakan algoritma C5.0**

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

Ansyari Aqshal Raditya Js

201810370311055

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji
pada tanggal 22 Maret 2025

Menyetujui,

Dosen Penguji 1



Ir. Syaifuddin S.Kom., M.Kom., IPM,

ASEAN Eng

NIP. 10816120590PNS.

Dosen Penguji 2



Ir. Wildan Suharso S.Kom., M.Kom

NIP. 10817030596PNS.

Mengetahui,
Dekan Jurusan Informatika



Dr. Galih Wasis Wicaksono S.kom. M.Cs.

NIP. 10814100541PNS.

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NAMA : Ansyari Aqshal Raditya Js

NIM : 201810370311055

FAK./JUR. : Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul “**Klasifikasi data film untuk menentukan rekomendasi film untuk anak menggunakan algoritma C5.0**” beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,
Dosen Pembimbing



Didih Rizki Chandranegara S.kom.,
M.Kom

Malang, 22 Maret 2025
Yang Membuat Pernyataan



Ansyari Aqshal Raditya Js

ABSTRAK

Film merupakan media yang berpengaruh besar dalam kehidupan sehari-hari, baik secara positif maupun negatif. Salah satu permasalahan yang muncul adalah ketidaksesuaian film dengan Batasan usia, terutama bagi anak-anak. Penelitian ini bertujuan untuk menerapkan metode data mining menggunakan algoritma C5.0 dalam mengklasifikasikan film berdasarkan Batasan usia agar memudahkan orang tua dalam memilih film yang sesuai untuk anak-anak. Penelitian ini menggunakan data dari situs Kaggle dan melalui beberapa tahapan proses, mulai dari preprocessing data, penerapan *feature selection*, splitting dataset, sampai dengan pembangunan model dan analisis. Dalam penelitian ini, dilakukan dua skenario percobaan. Skenario pertama menggunakan atribut yang dipilih melalui *feature selection* chi-square, menghasilkan tingkat akurasi sebesar 89,7%. Skenario kedua menggunakan atribut yang dipilih oleh *feature selection* pearson correlation, menghasilkan tingkat akurasi sebesar 91,7%. Meskipun ada peningkatan akurasi pada skenario kedua dibandingkan dengan skenario pertama, tingkat akurasi ini masih lebih rendah dibandingkan dengan tingkat akurasi tanpa menggunakan *feature selection*, yang mencapai 91,8%. Algoritma C5.0 tanpa *feature selection* memberikan hasil akurasi tertinggi, dengan tingkat akurasi sebesar 91,8%. Hal ini menegaskan bahwa penggunaan *feature selection* tidak selalu meningkatkan akurasi klasifikasi dan bahwa jumlah data yang lebih besar serta variasi data yang lebih kompleks dapat berkontribusi pada performa model yang lebih baik.

Kata Kunci: Film, Algoritma C5.0, Data Mining, Klasifikasi, Pruning, *Feature selection*, Preprocessing

ABSTRACT

Films are a significant medium that greatly influences everyday life, both positively and negatively. One of the challenges that arise is the mismatch between films and age restrictions, especially for children. This study aims to apply data mining techniques using the *C5.0 Algorithm* to classify films based on age restrictions to assist parents in selecting appropriate films for their children. This research uses data from the Kaggle website and follows several stages, including data preprocessing, *feature selection*, dataset splitting, model building, and analysis. Two experimental scenarios were conducted in this study. The first scenario used attributes selected through chi-square *feature selection*, resulting in an accuracy of 89.7%. The second scenario used attributes selected by the Pearson correlation *feature selection* method, achieving an accuracy of 91.7%. Although there was an improvement in accuracy in the second scenario compared to the first, this accuracy was still lower than the accuracy obtained without using *feature selection*, which reached 91.8%. The *C5.0 Algorithm* without *feature selection* provided the highest accuracy, with an accuracy rate of 91.8%. This emphasizes that using *feature selection* does not always improve classification accuracy and that a larger dataset and more complex data variations can contribute to better model performance.

Keywords: Film, *C5.0 Algorithm*, Data Mining, Classification, Pruning, *Feature selection*, Preprocessing

KATA PENGANTAR

Puji syukur alhamdulillah senantiasa Penulis ucapkan atas kehadiran Allah SWT Yang telah melimpahkan Rahmat, Taufik serta Hidayah-Nya sehingga pada akhirnya mampu menyelesaikan tugas akhir dengan judul “**Klasifikasi Data FILM untuk Menentukan Rekomendasi FILM untuk Anak Menggunakan Algoritma C5.0**”, meskipun masih terdapat banyak kekurangan. Shalawat serta salam semoga selalu tercurah kepada junjungan kita Nabi Besar SAW.

Penyusunan Tugas Akhir ini diajukan untuk memenuhi syarat akademis dalam rangka menyelesaikan Studi S1 Program Studi Teknik Informatika di Fakultas Teknik Universitas Muhammadiyah Malang.

Penulis menyadari bahwa dalam penyusunan tugas akhir ini tidak lepas dari bimbingan dan bantuan berbagai pihak. Oleh karena itu dalam kesempatan ini ucapkan terima kasih penulis sampaikan kepada:

1. Orang tua tercinta Ayah Syarifudin dan Ibu Siti Aminah yang selalu memberikan semangat, doa, kebahagiaan, nasehat, motivasi, materi, dan semua hal baik yang tidak bisa disebutkan secara rinci.
2. Bapak Didih Rizki Chandranegara, S.Kom., M.kom., selaku Dosen Pembimbing 1 dan Bapak Zamah Sari, S.T., M.T., selaku Dosen Pembimbing 2 yang selalu bersedia meluangkan waktu dan tenaga untuk memberikan arahan, petunjuk, serta saran dengan sabar karena saya sadari saya masih lambat dalam memahami sesuatu.
3. Seluruh sahabat Kelas B angkatan 2018 yang menjadi sahabat seperjuangan dalam menuntut ilmu, yang tak pernah mengenal budaya, kasta, dan materi.
4. Seluruh teman – teman yang tidak bisa disebutkan satu persatu yang memberikan motivasi selama penyusunan tugas akhir ini.
5. Penulis menyadari masih banyak kekurangan dalam penulisan tugas akhir ini. Untuk itu, penulis sangat mengharapkan saran yang membangun agar tulisan ini dapat berguna untuk perkembangan ilmu pengetahuan kedepannya.

Malang, 27 Februari 2025

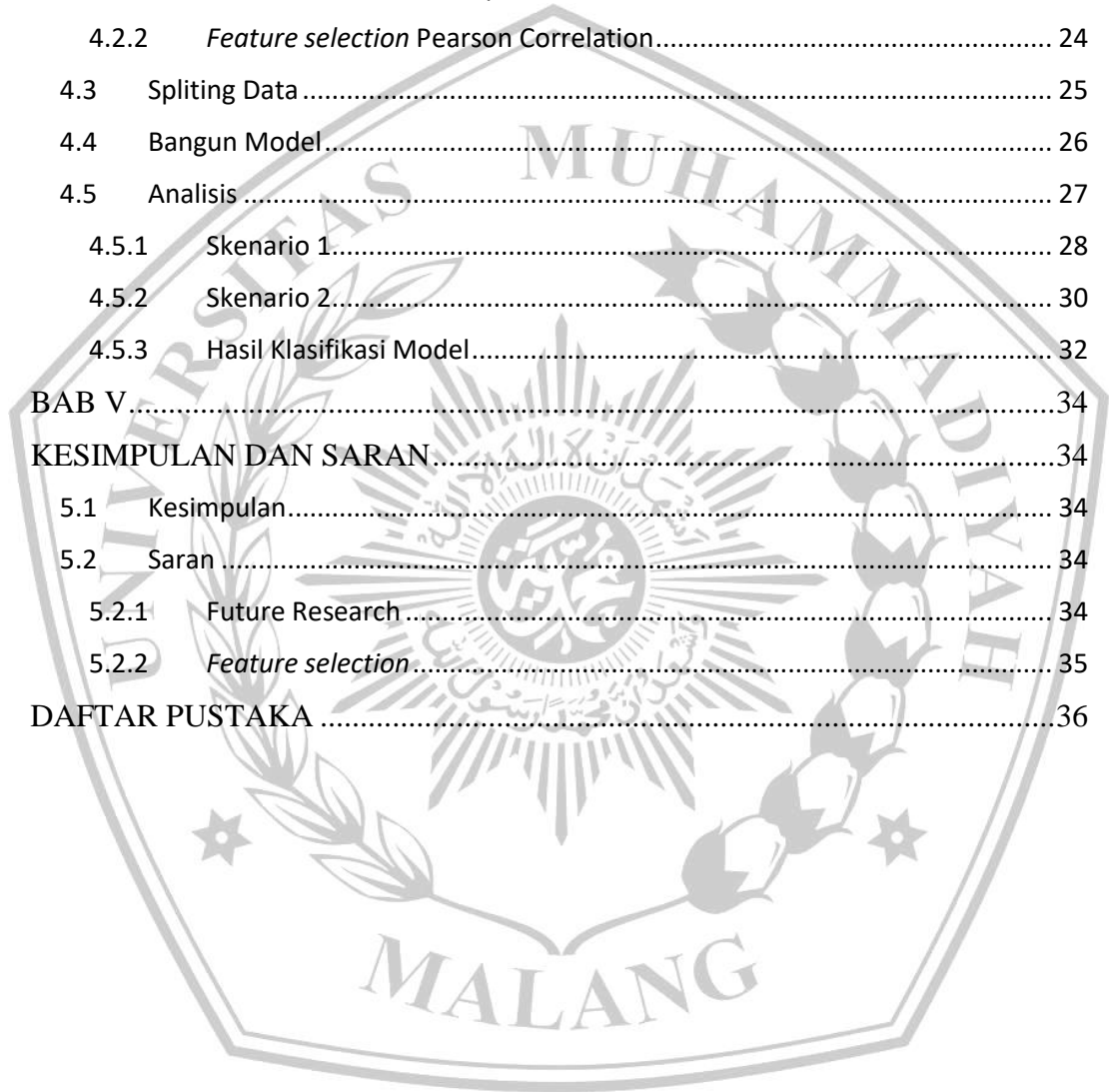


Ansyari Aqshal Raditya JS

DAFTAR ISI

LEMBAR PERSETUJUAN.....	i
LEMBAR PENGESAHAN.....	ii
LEMBAR PERNYATAAN.....	iii
ABSTRAK.....	iv
ABSTRACT.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	1
DAFTAR TABEL.....	2
BAB I.....	3
PENDAHULUAN.....	3
1.1 Latar Belakang.....	3
1.2 Rumusan Masalah.....	5
1.3 Tujuan Penelitian.....	5
1.4 Batasan Masalah.....	6
BAB II.....	7
TINJAUAN PUSTAKA.....	7
2.1 Studi Literatur.....	7
2.2 Data Mining.....	7
2.3 Klasterisasi.....	8
2.4 <i>Decision Tree Algorithm</i>	8
2.5 <i>Feature selection Algorithm</i>	11
BAB III.....	12
METODOLOGI PENELITIAN.....	12
3.1 Metode Penelitian.....	12
3.2 Alur Penelitian.....	13
3.2.1 Preprocessing Data.....	13
3.2.2 <i>Feature selection</i>	14
3.2.3 Splitting data.....	14
3.2.4 Bangun Model.....	15
3.2.5 Analisis.....	15

BAB IV	17
HASIL DAN PEMBAHASAN.....	17
4.1 Preprocessing Data	17
4.1.1 Pengecekan Dataset.....	17
4.1.2 Pengolahan Dataset	17
4.2 <i>Feature selection</i>	19
4.2.1 <i>Feature selection</i> Chi-Square	19
4.2.2 <i>Feature selection</i> Pearson Correlation.....	24
4.3 Spliting Data	25
4.4 Bangun Model.....	26
4.5 Analisis	27
4.5.1 Skenario 1.....	28
4.5.2 Skenario 2.....	30
4.5.3 Hasil Klasifikasi Model.....	32
BAB V.....	34
KESIMPULAN DAN SARAN.....	34
5.1 Kesimpulan.....	34
5.2 Saran	34
5.2.1 Future Research	34
5.2.2 <i>Feature selection</i>	35
DAFTAR PUSTAKA	36



DAFTAR GAMBAR

Gambar 4.1 Potongan Code Feature selection	23
Gambar 4.2 Potongan Code Feature selection Pearson Correlation	24
Gambar 4. 3 Hasil dari code Algoritma Feature selection Pearson Correlation	25
Gambar 4.4 Potongan Code proses Splitting Dataset	25
Gambar 4.5 Potongan Code Model C5.0	26
Gambar 4.6 Potongan Code untuk menentukan nilai alpha.....	26
Gambar 4.7 Potongan Code untuk menentukan nilai alpha terbaik.....	27
Gambar 4.8 Potongan Code K-Fold Cross Validation.....	28
Gambar 4.9 Loss and Accuracy Algoritma C5.0 Menggunakan Feature selection Chi-Square Test Method	29
Gambar 4.10 Grafik Tingkat Accuracy Algoritma C5.0 Menggunakan Feature selection Chi-Square Test Method pada setiap iterasi	30
Gambar 4.11 Loss and Accuracy <i>Algorithm</i> C5.0 using Feature selection Pearson Correlation Method.....	31
Gambar 4.12 Grafik Tingkat Accuracy Algoritma C5.0 Menggunakan Feature selection Pearson Correlation Method pada setiap iterasi	32

DAFTAR TABEL

Tabel 4.1 Mapping Rating Film Dari Dataset dengan berbagai lembaga sertifikasi ..	18
Tabel 4.2 Perbandingan data pada variabel date sebelum dan sesudah diubah	20
Tabel 4.3 Perbandingan data pada variabel rate sebelum dan sesudah diubah.....	20
Tabel 4.4 Perbandingan data pada variabel Votes sebelum dan sesudah diubah.....	21
Tabel 4.5 Perbandingan Data pada variabel Certificate sebelum dan sesudah diubah21	
Tabel 4.6 Hasil perhitungan Metode Feature selection Chi-Square Test.....	23
Tabel 4.7 Hasil Splitting dataset	25
Tabel 4.8 Perbandingan Hasil Performa Skenario 1	29
Tabel 4.9 Perbandingan Hasil Performa Skenario 2	31
Tabel 4.10 Perbandingan hasil performa klasifikasi algoritma C5.0.....	33



DAFTAR PUSTAKA

- [1] A. C. Wijaya, N. A. Hasibuan, and P. Ramadhani, "Implementasi Algoritma C5.0 Dalam Klasifikasi Pendapatan Masyarakat (Studi Kasus : Kelurahan Mesjid Kecamatan Medan Kota)," *Inf. dan Teknol. Ilm.*, vol. 13, pp. 192–198, 2018.
- [2] D. Wintana, Hikmatulloh, N. Ichsan, J. J. Purnama, and A. Rahmawati, "Klasifikasi Penentuan Penerima Manfaat Program Keluarga Harapan (Pkh)," *Kumpul. J. Ilmu Komput.*, vol. 06, no. 03, pp. 254–264, 2019.
- [3] Y. Yusuf, "Perbandingan Performansi Algoritma Decision Tree C5.0, Cart ," *Seminar*, vol. 2007, no. Snati, pp. 0–3, 2007.
- [4] P. W. Kastawan, D. M. Wiharta, and M. Sudarma, "Implementasi Algoritma C5.0 pada Penilaian Kinerja Pegawai Negeri Sipil," *Maj. Ilm. Teknol. Elektro*, vol. 17, no. 3, p. 371, 2018, doi: 10.24843/mite.2018.v17i03.p11.
- [5] N. Benediktus and R. S. Oetama, "The Decision Tree C5.0 Classification Algorithm for Predicting Student Academic Performance," *Ultim. J. Tek. Inform.*, vol. 12, no. 1, pp. 14–19, 2020, doi: 10.31937/ti.v12i1.1506.
- [6] I. Kurniawan and R. A. Saputra, "Penerapan Algoritma C5.0 Pada Sistem Pendukung Keputusan KelayakanPenerimaan BerasMasyarakat Miskin," *J. Inform.*, vol. 4, no. 2, pp. 236–240, 2017.
- [7] M. Pardede, E. Buulolo, and E. Ndruru, "Implementasi Algoritma C5.0 Pada Kelulusan Peserta Ujian Kemahiran Berbahasa Indonesia (Ukbi) Pada Balai Bahasa Sumatera Utara," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 64–72, 2019, doi: 10.30865/komik.v3i1.1569.
- [8] N. Sagala and H. Tampubolon, "Komparasi Kinerja Algoritma Data Mining pada Dataset Konsumsi Alkohol Siswa," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 4, no. 2, p. 98, 2018, doi: 10.23917/khif.v4i2.7061.
- [9] P. A. Melakukan, T. Kriminal, J. Y. Sudarso, and J. Timur, "671-Article Text-1912-1-10-20210112," no. 02, 2019.
- [10] S. Desti, "Dampak Tayangan Film Di Televisi Terhadap Perilaku Anak," *Dampak Tayangan Film di Telev. Terhadap Perilaku Anak J. Komunikologi*, vol. 2, no. 1, pp. 1–7, 2005.
- [11] D. Transaksi Bongkar Muat di Provinsi Riau, I. Kamila, U. Khairunnisa, P. Studi Sistem Informasi, and F. Sains dan Teknologi UIN Sultan Syarif Kasim Riau, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, pp. 119–125, 2019.
- [12] T. Informatika, T. Industri, M. Fahmi, Z. Naimah, and N. Setiani, "Klasterisasi Judul Buku dengan Menggunakan Metode K-Means Deka Dwinavinta Candra Nugraha."
- [13] N. Rofiqo, A. P. Windarto, and D. Hartama, "KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer) PENERAPAN CLUSTERING PADA PENDUDUK YANG MEMPUNYAI KELUHAN KESEHATAN DENGAN DATAMINING K-MEANS," [Online]. Available: <http://ejurnal.stmik->

- budidarma.ac.id/index.php/komik.
- [14] Y. Darmi, A. Setiawan, J. Bali, K. Kampung Bali, K. Teluk Segara, and K. Bengkulu, "PENERAPAN METODE CLUSTERING K-MEANS DALAM PENGELOMPOKAN PENJUALAN PRODUK," 2016.
- [15] H. H. Patel and P. Prajapati, "Study and Analysis of Decision Tree Based Classification Algorithms," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 74–78, 2018, doi: 10.26438/ijcse/v6i10.7478.
- [16] N. Nahar and F. Ara, "Liver Disease Prediction by Using Different Decision Tree Techniques," *Int. J. Data Min. Knowl. Manag. Process*, vol. 8, no. 2, pp. 01–09, 2018, doi: 10.5121/ijdkp.2018.8201.
- [17] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, and E. R. Kaburuan, "Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes," *2018 Int. Conf. Orange Technol. ICOT 2018*, pp. 1–6, 2018, doi: 10.1109/ICOT.2018.8705796.
- [18] N. M. Abdulkareem, A. Mohsin Abdulazeez, D. Qader Zeebaree, and D. A. Hasan, "COVID-19 World Vaccination Progress Using Machine Learning Classification Algorithms," *Qubahan Acad. J.*, vol. 1, no. 2, pp. 100–105, 2021, doi: 10.48161/qaj.v1n2a53.
- [19] Ş. Çelik, "The comparison of the model performances of Naïve Bayes , C4 . 5 and C5 . 0 Algorithms : Implementation on fish consumption habits," no. February, 2021.
- [20] V. Rahmayanti, Y. Azhar, and A. E. Pramudita, "Penerapan algoritma C5.0 pada analisis faktor-faktor pengaruh kelulusan tepat waktu mahasiswa Teknik Informatika UMM," *J. Repos.*, vol. 1, no. 2, p. 131, 2020, doi: 10.22219/repositor.v1i2.545.
- [21] A. Z. Amrullah, A. Sofyan Anas, and M. A. J. Hidayat, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *Jurnal*, vol. 2, no. 1, pp. 40–44, 2020, doi: 10.30812/bite.v2i1.804.
- [22] J. LING, I. P. E. N. KENCANA, and T. B. OKA, "Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square," *E-Jurnal Mat.*, vol. 3, no. 3, p. 92, 2014, doi: 10.24843/mtk.2014.v03.i03.p070.
- [23] N. T. Romadloni and Hilman F Pardede, "Seleksi Fitur Berbasis Pearson Correlation Untuk Optimasi Opinion Mining Review Pelanggan," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 505–510, 2019, doi: 10.29207/resti.v3i3.1189.
- [24] S. Anisah, A. S. Honggowibowo, and A. Pujiastuti, "Klasifikasi Teks Menggunakan Chi Square Feature selection Untuk Menentukan Komik Berdasarkan Periode, Materi Dan Fisikdengan Algoritma Naivebayes," *Compiler*, vol. 5, no. 2, pp. 59–66, 2016, doi: 10.28989/compiler.v5i2.171.
- [25] A. Sikri, N. P. Singh, and S. Dalal, "INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Chi-Square Method of Feature selection : Impact of Pre- Processing of Data," *Intell. Syst. Appl. Issn2147-6799*, vol. 11, pp. 241–248, 2023.



UNIVERSITAS
MUHAMMADIYAH
MALANG



FAKULTAS TEKNIK

INFORMATIKA

informatika.umm.ac.id | informatika@umm.ac.id

FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Ansyari Aqshal Raditya JS
 NIM : 201810370311055
 Judul TA : Klasifikasi Data Film untuk Menentukan Rekomendasi Film untuk Anak Menggunakan Algoritma C5.0

Hasil Cek Plagiarisme dengan Turnitin

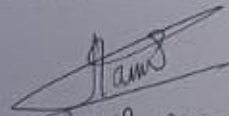
No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	10%
2.	Bab 2 – Daftar Pustaka	25 %	23%
3.	Bab 3 – Analisis dan Perancangan	25 %	15%
4.	Bab 4 – Implementasi dan Pengujian	15 %	0%
5.	Bab 5 – Kesimpulan dan Saran	5 %	4%
6.	Makalah Tugas Akhir	20%	2%

*) Hasil cek plagiarism diisi oleh pemeriksa (staf TU)

*) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)

Mengetahui,

Pemeriksa (Staff TU)


 (.....)



Kampus 1
 Jl. Semarang 1 Malang, Jawa Timur
 T: +62 341 501 512 (Pusat)
 F: +62 341 542 514

Kampus 2
 Jl. Sarunggi Industri No. 108 Malang, Jawa Timur
 T: +62 341 501 144 (Pusat)
 F: +62 341 542 514

Kampus 3
 Jl. Raya Tlogomas No. 749 Malang, Jawa Timur
 T: +62 341 461 210 (Pusat)
 F: +62 341 461 410
 E: webmaster@umm.ac.id