

**ANALISIS PENGARUH *GAIN RATIO* UNTUK ALGORITMA
K-NEAREST NEIGHBOR PADA KLASIFIKASI
DATA BANJIR DI KOTA SAMARINDA**

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan mencapai gelar
Sarjana Komputer

DISUSUN OLEH :

SEPTIA INTAN PERMATA SARI

1911102441128



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
SAMARINDA
2023**

Analisis Pengaruh *Gain Ratio* untuk Algoritma *K-Nearest Neighbor* pada Klasifikasi Data Banjir di Kota Samarinda

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan mencapai gelar
Sarjana Komputer

Disusun Oleh:

Septia Intan Permata Sari

1911102441128



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
SAMARINDA
2023**

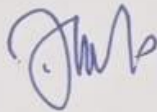
HALAMAN PENGESAHAN

**ANALISIS PENGARUH *GAIN RATIO* UNTUK ALGORITMA
K-NEAREST NEIGHBOR PADA KLASIFIKASI
DATA BANJIR DI KOTA SAMARINDA**

**DISUSUN OLEH:
SEPTIA INTAN PERMATA SARI
1911102441128**

Telah melaksanakan ujian skripsi dan dinyatakan lulus,
Pada tanggal 05 Juli 2023

Dosen Pembimbing



Wawan Joko Pranoto, S.Kom, M.TI

NIDN. 1102057701

Penguji



Naufal Azmi Verdikha, S.Kom., M.Eng

NIDN. 1114048801

Dekan



Prof. Drs. Sarlito, MT., Ph.D.
NIDN. 0610116204

Ketua Program Studi



Prof. Drs. Ratipah, S.Kom., M.Cs
NIDN. 1124098902

PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Septia Intan Permata Sari

NIM : 1911102441128

Konsentrasi : Sistem Informasi

Menyatakan bahwa skripsi dengan judul berikut:

Analisis Pengaruh *Gain Ratio* Untuk Algoritma *K-Nearest Neighbor* Pada Klasifikasi Banjir di Kota Samarinda.

Dosen Pembimbing : Wawan Joko Pranoto, S.Kom, M.Ti

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar Sarjana Komputer, baik di Universitas Muhammadiyah Kalimantan Timur (UMKT) maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab UMKT.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi Muhammadiyah

Samarinda, 25 Juni 2023



yang Menyatakan,

Septia Intan Permata Sari

1911102441128

ABSTRAK

Berdasarkan data yang diperoleh dari BMKG dan BPBD Kota Samarinda, diketahui bahwa telah terjadi bencana banjir pada periode tahun 2019 - 2020 di Kota Samarinda. Penelitian ini bertujuan untuk melakukan klasifikasi data banjir di Kota Samarinda menggunakan algoritma *K-Nearest Neighbor* dan pembagian data menerapkan teknik *5-Fold Cross-Validation* serta perhitungan rumus jarak *Euclidean Distance*. Kemudian, dilakukan seleksi fitur pada algoritma KNN menggunakan metode *Gain Ratio* guna mengetahui pengaruhnya terhadap akurasi dari KNN. Hasil penelitian menunjukkan bahwa peningkatan akurasi tertinggi setelah menerapkan *Gain Ratio* didapatkan oleh K=7 dengan persentase kenaikan akurasi sebesar 5,95%, diikuti oleh K=5 dengan persentase kenaikan akurasi 5,81%, K=3 dengan persentase kenaikan akurasi 5,68%, K=9 sebesar 3,61%, K=11 sebesar 2,44%, dan K=13 sebesar 1,23%. Hanya ada satu akurasi yang tidak mengalami peningkatan atau penurunan akurasi, yaitu K=15.

Kata kunci: Banjir, *K-Nearest Neighbor*, *5-Fold Cross-Validation*, *Euclidean Distance*, *Gain Ratio*

ABSTRACT

Based on data obtained from the BMKG and BPBD Samarinda City, it is known that there has been a flood disaster in the 2019 - 2020 period in Samarinda City. This study aims to classify flood data in Samarinda City using the K-Nearest Neighbor algorithm and dividing data using the 5-Fold Cross-Validation technique and calculating the Euclidean Distance formula. Then, feature selection is carried out in the KNN algorithm using the Gain Ratio method to determine its effect on the accuracy of the KNN. The results showed that the highest increase in accuracy after applying the Gain Ratio was obtained by K=7 with a percentage increase in accuracy of 5.95%, followed by K=5 with a percentage increase in accuracy of 5.81%, K=3 with a percentage increase in accuracy of 5.68 %, K=9 of 3.61%, K=11 of 2.44%, and K=13 of 1.23%. There is only one accuracy that does not experience an increase or decrease in accuracy, namely K=15.

Keywords: Flood, K-Nearest Neighbor, 5-Fold Cross-Validation, Euclidean Distance, Gain Ratio

KATA PENGANTAR



Puji dan syukur dipanjatkan kepada Tuhan Yang Maha Esa, karena atas rahmatnya Skripsi yang berjudul “Implementasi Optimasi Gain Ratio Untuk Algoritma K-Nearest Neighbor Pada Klasifikasi Data Banjir Di Kota Samarinda” ini diajukan sebagai salah satu syarat untuk memenuhi persyaratan mencapai gelar Sarjana Komputer di Universitas Muhammadiyah Kalimantan Timur.

Penulis menyadari bahwa dalam penyelesaian Proposal ini mendapat bantuan dari berbagai pihak, baik langsung maupun tidak langsung. Oleh karena itu pada kesempatan ini penulis menyampaikan terima kasih kepada:

1. Bapak Prof. Dr. H. Bambang Setiaji, selaku Rektor Universitas Muhammadiyah Kalimantan Timur.
2. Bapak Prof. Ir. Sarjito, M.T.,Ph.D., selaku Dekan Fakultas Sains & Teknologi.
3. Ibu Asslia Johar Latipah, M.Cs., selaku ketua Program Studi S1 Teknik Informatika
4. Bapak Naufal Azmi Verdikha, S.Kom., M.Eng., selaku dosen penguji satu yang telah menyediakan waktu, tenaga dan pemikirannya untuk membimbing dan mengarahkan saya dalam proses penyelesaian tugas akhir ini.
5. Bapak Wawan Joko Parnoto, S.Kom, M.Kom, selaku dosen penguji dua serta pembimbing yang telah menyediakan waktu, tenaga dan pikiran untuk mengarahkan saya selama penyusunan tugas akhir ini.
6. Seluruh dosen S1 Teknik Informatika Universitas Muhammadiyah Kalimantan Timur yang telah memberikan ilmunya kepada penulis.
7. Ibu Samsiah dan Bpk. Asmuransyah selaku kedua orang tua, saya ucapkan terima kasih tiada terkira atas dukungan dan doa sehingga selama penyusunan tugas akhir ini peneliti saya tetap semangat menyelesaikan tugas akhir ini.

8. Saudara/i saya, Alm.Malaika Hesty Mandasari,Linda Mayang Sari S.P,Nabilla Auliyah Haqim,dan serta keponakan saya Alubna Husna Fadmasari ,saya ucapkan terima kasih atas dukungan moral maupun materil dan doa sehingga selama penyusunan tugas akhir ini peneliti saya tetap semangat menyelesaikan tugas akhir ini.
9. Sinta Nuriah, S.Keb yang telah menjadi teman yang selalu mendukung dan memberikan semangat selama penyusunan tugas akhir ini
10. Kepada rekan-rekan bimbingan seperjuangan saya yaitu Yuliana Dilla Evitasari , Suci Mawaddah, Kamirul Hakim, Putra Saymen dan Muhammad Fath Thoriq, Rahmat Ramadhani yang telah banyak membantu dan berjuang Bersama-sama serta memberikan saran dan semangat sehingga dapat menyelesaikan skripsi ini.
11. Kepada teman-teman S1 Teknik Informatika Angkatan 2019
12. Semua pihak yang tidak dapat penulis sebutkan satu persatu, terimakasih atas bantuan semangat serta nasehat dan dorongan dalam proses penyusunan skripsi ini.

Penulis menyadari skripsi ini masih jauh dari sempurna, oleh karena itu dengan segala hormat dan kerendahan hati, penulis berharap saran dan kritik yang membangun untuk hasil yang dituangkankan dalam skripsi ini bisa bermanfaat bagi pihak-pihak yang memerlukan.

Samarinda, 25 Juni 2023



Septia Intan Permata Sari

19111022441128

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	iii
PERNYATAAN KEASLIAN SKRIPSI	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah	3
1.5 Manfaat Penelitian	4
BAB 2 TINJAUAN PUSTAKA.....	6
2.1 Banjir.....	6
2.2 Data Mining	6
2.2.1 Klasifikasi	7
2.2.2 <i>K-Nearest Neighbor</i>	7
2.2.3 Data Preprocessing	8
2.2.4 <i>K-Fold Cross-validation</i>	10

2.2.5	<i>Confusion Matrix</i>	10
2.2.6	CRISP-DM	11
2.2.7.	Gain Ratio	12
2.3	Penelitian Terdahulu	13
BAB 3 METODOLOGI PENELITIAN		17
3.2	Tahap Pelaksanaan Penelitian.....	17
3.2.1	Pengumpulan data	18
3.2.2	Data Pre-Processing	19
3.2.3	Pembagian data	20
3.2.4	Modelling	20
3.2.5	Evaluasi.....	21
BAB 4 HASIL DAN PEMBAHASAN		22
4.1	Pengumpulan Data	22
4.2	Data Preparation	23
4.2.1	Data Integration.....	23
4.4.2	Data Selection	24
4.4.3	Data Transformation	25
4.4.4	Data Cleaning	27
4.4.5	Data Balancing.....	28
4.3	PEMBAHASAN.....	33
BAB 5 PENUTUP.....		35
5.1	Kesimpulan	35
5.2	Saran.....	36
DAFTAR PUSTAKA.....		37
LAMPIRAN		41

DAFTAR TABEL

Tabel 2. 1 confusion matrix.....	10
Tabel 2. 2 Penelitian Terdahulu	14
Tabel 3. 1 Atribut Data BMKG Kota Samarinda	18
Tabel 4. 1. Sebagian Data BMKG Kota Samarinda	22
Tabel 4. 2 hasil data integration.....	24
Tabel 4. 3 Sebagian hasil data selection	25
Tabel 4. 4 Hasil Transformasi Data Sebagian	25
Tabel 4. 5 Hasil Data Sebagian Setelah Ditransformasi	26
Tabel 4. 6 Sebagian data nilai NaN.....	27
Tabel 4. 7 Nilai NaN diganti nilai rata-rata.....	28
Tabel 4. 8 Sebagian Hasil Data Balancing.....	29
Tabel 4. 9 Hasil Evaluasi Confusion Matrix	30
Tabel 4. 10 Penentuan Atribut yang Digunakan	31
Tabel 4. 11 Evaluasi Confusion Matrix Setelah Seleksi Fitur.....	32
Tabel 4. 12 Perbandingan Hasil Akurasi Rata-rata.....	32

DAFTAR GAMBAR

Gambar 2. 1 Alur CRISP-DM	11
Gambar 3. 1 alur penelitian	17
Gambar 4. 1 Sebagian Data BPBD Kota Samarinda Bulan November 2021	23
Gambar 4. 2 Hasil Gain Ratio	31
Gambar 4. 3 Diagram Akurasi KNN+GR	34

DAFTAR LAMPIRAN

Lampiran 1 Riwayat Hidup	42
Lampiran 2 Data BMKG dan BPBD Banjir Di Kota Samarinda	43
Lampiran 3 Data Setelah Hasil Seleksi, Integrasi, Transformasi	45
Lampiran 4 Seluruh Data Setelah Melalui Tahap Data Balacing	46
Lampiran 5 Codingan.....	47
Lampiran 6 Surat Ijin Penelitian	54