

**MODEL OPTIMASI *KNN*-PSORF DALAM MENANGANI
HIGH DIMENSIONAL DATA BANJIR KOTA SAMARINDA**

SKRIPSI

**Diajukan Oleh:
Anggiq Karisma Aji Restu
2011102441089**



**PROGRAM STUDI S1TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
JULI 2024**

**MODEL OPTIMASI KNN-PSORF DALAM MENANGANI
HIGH DIMENSIONAL DATA BANJIR KOTA SAMARINDA**

SKRIPSI

Diajukan Sebagai Salah Satu Persyaratan Untuk Memperoleh Gelar
Sarjana Komputer Fakultas Sains dan Teknologi
Universitas Muhammadiyah Kalimantan Timur

Diajukan Oleh:
Anggiq Karisma Aji Restu
2011102441089



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UIMVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
JULI 2024**

LEMBAR PERSETUJUAN

**Model Optimasi KNN-PSORF Dalam Menangani *High Dimensional Data*
Banjir Kota Samarinda**

SKRIPSI

Diajukan oleh:

**Anggiq Karisma Aji Restu
2011102441089**

**Disetujui untuk diujikan
Pada tanggal 22 Juli 2024**

Pembimbing



**Taqfirul Azhima/Yoga Siswa, S.Kom, M.Kom
NIDN. 1118038805**

29/6/2024

**Mengetahui,
Koordinator Skripsi**



**Abdul Rahim, S.Kom, M.Cs.
NIDN. 1115039601**

LEMBAR PENGESAHAN

Model Optimasi KNN-PSORF Dalam Menangani *High Dimensional* Data
Banjir Kota Samarinda

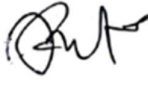

SKRIPSI

Diajukan oleh:

Anggiq Karisma Aji Restu

2011102441089


Diseminarkan dan Diujikan
Pada Tanggal ..3.Juli 2024

Penguji I	Penguji II
 <u>Wawan Joko Pranoto, S.Kom. M.Ti</u> NIDN. 1102057701	 <u>Taghfirul Azhima Yoga Siswa, S.Kom. M.Kom</u> NIDN. 1118038805

Mengetahui,
Ketua

Program Studi Teknik Informatika




Wansyah, S.Kom. M. TI
NIDN.1118019203

PERNYATAAN KEASLIAN PENELITIAN

Saya yang bertanda tangan di bawah ini:

Nama : Anggiq Karisma Aji Restu
NIM : 2011102441089
Program Studi : Teknik Informatika
Judul Penelitian : Model Optimasi KNN-PSORF Dalam Menangani High Dimensional Data Banjir Kota Samarinda

Menyatakan bahwa Skripsi yang saya tulis ini benar-benar hasil karya saya sendiri, dan bukan merupakan hasil plagiasi/falsifikasi/fabrikasi baik sebagian atau seluruhnya.

Atas pernyataan ini, saya siap menanggung resiko atau sanksi yang dijatuhkan kepada saya apa bila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam skripsi saya ini, atau klaim dari pihak lain terhadap keaslian karya saya ini.

Samarinda, 14 juli 2024
Yang membuat pernyataan



Anggiq Karisma Aji Restu
NIM: 2011102441127

ABSTRAK

Banjir adalah fenomena alam yang sering terjadi di Indonesia, termasuk di Kota Samarinda yang mengalami masalah banjir dalam tiga tahun terakhir dengan dampak ribuan rumah sebanyak 27.000 jiwa terkena banjir. Untuk memprediksi bencana banjir dibutuhkan teknologi machine learning menggunakan metode klasifikasi data mining. Namun, pada proses klasifikasi seringkali terjadi permasalahan yang berkaitan dengan data berdimensi tinggi ini dapat menyebabkan overfitting dan ketidakseimbangan kelas yang menyebabkan bias pada kelas yang dominan dengan mengabaikan kelas minoritas. Penelitian ini bertujuan untuk meningkatkan nilai akurasi klasifikasi pada data banjir Kota Samarinda menggunakan algoritma K-Nearest Neighbor (KNN) yang dikombinasikan seleksi fitur relief dan optimasi Particle Swarm Optimization (PSO). Metode validasi yang digunakan adalah 10-fold cross-validation, sementara evaluasi kinerja model dilakukan menggunakan confusion matrix. Data yang digunakan diperoleh dari BPBD dan BMKG Kota Samarinda pada rentang tahun 2021-2023, dengan 19 fitur dan total 1095 record. Hasil seleksi fitur Relief didapatkan empat fitur penting, yaitu arah angin maksimum, kecepatan angin, kecepatan angin rata-rata, dan arah angin maksimum. Evaluasi rata-rata dengan nilai $k=3$, $k=5$, $k=7$, $k=11$, $k=13$, dan $k=15$ menunjukkan penerapan seleksi fitur Relief dan optimasi PSO, efektif dalam meningkatkan akurasi pada algoritma k-Nearest Neighbor pada data banjir dengan hasil akurasi KNN dan PSO memberikan peningkatan sebesar 2-5%, KNN dengan seleksi fitur Relief memberikan peningkatan sebesar 1-2% dan KNN dengan kombinasi Relief dan PSO memberikan peningkatan sebesar 2-5%. Kombinasi model KNN, Relief, PSO diharapkan dapat memberikan performa yang optimal dalam klasifikasi data banjir Kota Samarinda.

Kata kunci: K-Nearest Neighbor, Relief, Banjir, 10-Fold Cross-Validation, klasifikasi

ABSTRACT

Floods are a natural phenomenon that frequently occurs in Indonesia, including in Samarinda City which has faced flood issues over the past three years, affecting thousands of homes and around 27,000 residents. Predicting flood disasters requires machine learning technology using data mining classification methods. However, classification processes often encounter issues related to high-dimensional data, which can lead to overfitting and class imbalance, thereby biasing dominant classes while neglecting minority classes. This research aims to enhance classification accuracy in Samarinda City's flood data using the K-Nearest Neighbor (KNN) algorithm combined with Relief feature selection and Particle Swarm Optimization (PSO) optimization. The validation method employed is 10-fold cross-validation, with performance evaluation using a confusion matrix. Data sourced from Samarinda City's Disaster Management Agency (BPBD) and Meteorology, Climatology, and Geophysics Agency (BMKG) spans from 2021 to 2023, comprising 19 features and a total of 1095 records. Relief feature selection identified four crucial features: maximum wind direction, wind speed, average wind speed, and maximum wind speed direction. Average evaluations with k values of 3, 5, 7, 11, 13, and 15 demonstrate that Relief feature selection and PSO optimization effectively enhance accuracy in the K-Nearest Neighbor algorithm for flood data, with KNN and PSO yielding improvements of 2-5%. Relief feature selection alone improves accuracy by 1-2%, while combining Relief with PSO provides a 2-5% enhancement. The combined KNN, Relief, PSO model is expected to deliver optimal performance in classifying Samarinda City's flood data.

Keywords: K-Nearest Neighbor, Relief, Flood, 10-Fold Cross-Validation, classification

PRAKATA

Dengan nama Allah Yang Maha Pengasih dan Penyayang, segala puji hanya bagi-Nya. Shalawat serta salam semoga tercurahkan kepada Rasulullah Muhammad SAW, yang telah membawa petunjuk serta rahmat bagi seluruh alam. Prakata ini dibuat sebagai ungkapan terima kasih dan penghargaan yang tulus dari penulis kepada semua pihak dalam penyelesaian skripsi ini, sehingga Peneliti dapat menyelesaikan skripsi ini dengan judul "Model Optimasi KNN-PSORF Dalam Menangani High Dimensional Data Banjir Kota Samarinda". Kami ingin mengucapkan terima kasih kepada banyak pihak yang telah membantu dan mendukung penyelesaian penelitian ini. Terutama kepada:

1. Bapak Agus Triyono, Ibu Natlisa Dwi Kristiani, dan Adik tercinta yang selalu memberikan doa serta dukungan kepada penulis.
2. Bapak Taghfirul Azhima Yoga Siswa, S.Kom, M.Kom selaku Dosen Pembimbing, yang telah memberikan bimbingan, arahan, dan pengarahan yang diberikan selama proses studi akhir ini.
3. Bapak Wawan Joko Pranoto, S.Kom, M.Ti selaku Dosen Penguji seminar proposal penelitian dan sidang skripsi penulis.
4. Bapak Arbansyah, S.Kom., M.TI selaku ketua Program Studi S1 Teknik Informatika.
5. Prof. Ir. Sarjito, M.T., Ph.D., selaku Dekan Fakultas Sains & Teknologi Universitas Muhammadiyah Kalimantan Timur
6. Dr. Muhammad Musiyam, M.T selaku Rektor Universitas Muhammadiyah Kalimantan Timur.
7. Seluruh Bapak dan Ibu Dosen Program Studi Teknik Informatika Universitas Muhammadiyah Kalimantan Timur dan staff yang penulis banggakan dan hormati.
8. Winda Wahyuni yang selalu memberikan semangat, dukungan dan masukan dalam penyusunan laporan dari awal hingga penyelesaian studi akhir.
9. Teman-teman yang selalu mengganti nama grub yang beranggotakan 12 orang yang menjadi seperti keluarga sendiri dan tidak lupa selalu memberikan semangat dan dukungan selama menjalani proses perkuliahan, semoga kalian sukses semua saudara.

Kami sadar bahwa penelitian ini masih jauh dari kesempurnaan, oleh karena itu segala saran dan kritik membangun sangat kami harapkan guna perbaikan di masa yang akan datang.

Samarinda, 29 Juni 2024
Yang membuat pernyataan



Anggiq Karisma Aji Restu
NIM: 2011102441089

DAFTAR ISI

HALAMAN JUDUL.....	iii
LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN.....	iv
PERNYATAAN KEASLIAN PENELITIAN.....	v
ABSTRAK	vi
ABSTRACT	vii
PRAKATA.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	14
1.1. Latar Belakang	14
1.2. Rumusan Masalah.....	3
1.3. Tujuan Penelitian	3
1.4. Manfaat Penelitian	3
1.5. Batasan Masalah	4
BAB II METODOLOGI PENELITIAN	5
2.1. Objek Penelitian.....	5
2.2. Prosedur Penelitian	5
2.2.1. Identifikasi Masalah	5
2.2.2. Pengumpulan Data	6
2.2.3. Data Pre-Processing.....	6
2.2.4. Pembagian Data.....	10
2.2.5. <i>Modelling</i>	11
2.2.5. Evaluasi	15
BAB III HASIL DAN PEMBAHASAN.....	16
3.1. Hasil Penelitian.....	16
3.1.1. Pengumpulan Data	16
3.1.2. Data Preparation.....	18
3.2. Pembahasan.....	26

BAB IV KESIMPULAN DAN SARAN.....	29
4.1. Kesimpulan	29
4.2. Saran	29
DAFTAR RUJUKAN	30
DAFTAR RIWAYAT HIDUP	33
LAMPIRAN	34

DAFTAR TABEL

Tabel	Halaman
2. 1 Fitur Dataset Banjir BMKG dan BPBD Kota Samarinda	6
2. 2 Data Selection	7
2. 3 Parameter Data Cleaning	8
2. 4 Pemeriksaan nilai yang hilang setelah penghapusan	8
2. 5 Parameter Data Transformation	9
2. 6 Parameter Data Balancing	10
2. 7 Parameter Pembagian Data	11
2. 8 Parameter Persiapan Insialisasi Model KNN	12
2. 9 Parameter Mengimpor dan menerapkan PSO Pada Model	13
2. 10 Parameter Relief	14
3. 1 Data Yang Diperoleh Dari BMKG	16
3. 2 Data Yang Diperoleh Dari BPBD	17
3. 3 Hasil Data Integration	18
3. 4 Dataset Sebelum ditranformasi	21
3. 5 Dataset Setelah ditranformasi	21
3. 6 Hasil Evaluasi Confusion Matrix	23
3. 7 Hasil Evaluasi k-nearest neighbors (KNN) + PSO	23
3. 8 Penentuan Atribut yang Digunakan	25
3. 9 Evaluasi Confusion Matrix Setelah Seleksi Fitur	25
3. 10 Evaluasi KNN+ Relief + PSO	26
3. 11 Perbandingan hasil akurasi dari setiap model KNN	26

DAFTAR GAMBAR

Gambar	Halaman
2. 1 Diagram Alur Penelitian.....	5
2. 2 Proses Data Cleaning.....	7
2. 3 Pemeriksaan nilai yang hilang setelah penghapusan.....	8
2. 4 Sebelum Transformasi Data	8
2. 5 Proses Encoding	9
2. 6 Data Balancing	9
2. 7 Pembagian Data.....	10
2. 8 Persiapan Inisialisasi Model KNN	11
2. 9 Menginstall PSO.....	12
2. 10 Mengimpor dan menerapkan PSO Pada Model	13
2. 11 Menginstall Relief	13
3. 1 Hasil Data Selection	19
3. 2 Dataset sebelum dibersihkan	19
3. 3 Nilai kosong pada tiap atribut sebelum data cleaning	20
3. 4 Jumlah Sebelum dan sesudah Data Cleaning	20
3. 5 Hasil Data Cleaning.....	20
3. 6 Jumlah Nilai Kosong Tiap Kolom Setelah Pembersihan	21
3. 7 Jumlah Kelas Sebelum Penerapan SMOTE	22
3. 8 Jumlah Kelas Sesudah Penerapan SMOTE.....	22
3. 9 Hasil Perangkingan relief berdasarkan (importance score)	24
3. 10 Grafik scores dari Relief.....	24
3. 11 Diagram Permodelan KNN	27

DAFTAR LAMPIRAN

Lampiran	Halaman
Lampiran 1 Codingan.....	34
Lampiran 2 Surat Pengantar Pengambilan Data BMKG.....	40
Lampiran 3 Surat Pengantar Pengambilan Data BPBD	41
Lampiran 4 Surat Letter of Acceptance Jurnal.....	42
Lampiran 5 Lampiran Kartu bimbingan.....	43