

BAB III HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

Pada penelitian ini, dilakukan analisis terhadap data banjir di Kota Samarinda, Data yang digunakan dalam penelitian ini mencakup berbagai variabel yang mempengaruhi kejadian banjir di Kota Samarinda, data tersebut terdiri atas 1095 Dataset dengan 19 Atribut dan 1 Atribut sebagai label didalamnya. Data ini memiliki karakteristik berdimensi tinggi yang memerlukan teknik seleksi fitur untuk meningkatkan akurasi model prediksi. Proses seleksi fitur dilakukan menggunakan algoritma *Relief* yang dioptimalkan dengan PSO. *Relief* berfungsi untuk menilai kepentingan setiap fitur berdasarkan kemampuan mereka dalam membedakan kelas target, dalam hal ini kejadian banjir dan tidak banjir. Hasil seleksi fitur menunjukkan bahwa dari sekian banyak variabel, terdapat 9 fitur utama yang paling signifikan dalam memprediksi kejadian banjir di Samarinda. Setelah seleksi fitur, model dioptimalkan menggunakan PSO untuk menentukan parameter terbaik yang memaksimalkan kinerja model.

3.1.1. Pengumpulan Data

Penelitian ini menggunakan total data banjir sebanyak 1095 *record*. yang didapatkan dari BPBD dan BMKG Kota Samarinda tahun 2021-2023 dalam dataset tersebut, terdapat 49 *record* yang menunjukkan adanya kejadian banjir, sementara 841 *record* menunjukkan tidak adanya kejadian banjir dengan label kelas 0 menunjukkan tidak terjadi banjir, sedangkan label kelas 1 menunjukkan terjadi banjir.

Data yang diperoleh dari BMKG yang memiliki 11 fitur sedangkan data yang diperoleh dari BPBD memiliki 9 fitur. Data yang didapatkan dari BMKG meliputi tanggal, temperatur maksimum (Tx), temperatur minimum (Tn), temperatur rata-rata (Tavg), kelembaban rata-rata (RH_avg), curah hujan (RR), lamanya penyinaran matahari (ss), kecepatan angin maksimum (ff_x), arah angin maksimum (ddd_x), kecepatan angin rata-rata (ff_avg), dan arah angin terbanyak (ddd_car). Sedangkan data yang didapatkan dari BPBD meliputi tanggal, jam kejadian, jenis bencana, lokasi wilayah, luas area m2, objek terkena bencana, korban, kerugian, dan keterangan. hasil penggabungan data dari pendekatan *excel* tersebut dapat dilihat dalam tabel berikut :

Tabel 3. 1 Data Yang Diperoleh Dari BMKG

Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	ddd_car
01-01-2021	23	33,2	26,5	88	1,8	3,3	4	280	2	W
02-01-2021	23,2	30,8	27,1	88	7	6,4	2	140	1	C
03-01-2021	24,8	32,7	27,3	84	2	1,2	5	290	2	NW
04-01-2021	24,6	31,8	28,1	84	2,7	5,4	3	300	2	NW
05-01-2021	24,6	31,4	27,4	83	10,5	1,8	4	300	2	W
...
27-12-2023	24,2	32	27,6	83	0,3	2,8	4	60	2	NE
28-12-2023	24	32	28	82	1	6,4	5	70	2	C
29-12-2023	23,7	32,7	28,7	77	3,5	5,9	5	60	2	NE
30-12-2023	24,2	32,6	28,3	82		10,4	4	60	1	NE
31-12-2023	24,6	32,4	28,3	84	0	6,9	4	90	2	E

Selain data dari BMKG, terdapat data dari bpbd yang ditunjukkan pada Tabel 3,2, sebagai berikut.

Tabel 3. 2 Data Yang Diperoleh Dari BPBD

NO	TANGGAL	JAM KEJADIAN	JENIS BENCANA	LOKASI/ WILAYAH KELURAHAN/ KECAMATAN	LUAS AREA M ²	JUMLAH OBYEK YANG TERKENA BENCANA	korban					JUMLAH JIWA	KERUGIAN (Rp)	KETERANGAN
							KL	KS	KH	KM	KK			
1	03 Januari 2021	-	Banjir	Jl. Irigasi RT. 50 Kel. Rawa Makmur Kec. Palaran (Dataran Rendah) Wilayah Handil Bakti RT. 1, RT. 2, RT 3 (Dataran Rendah)	±	-	Jalan mejadi Susah Untuk Di lalui Dan Mengganggu aktivitas warga	-	-	-	-	-	Rp. -	Genangan Air Penyebab Air Sungai Mahakam pasang Dan Lokasi Banjir adalah Dataran Rendah
2	Selasa, 05 Januari 2021	-	Pohon Tumbang	Jl. Kesehatan Dalam Kel. Temindung Permai	-	-	Jalan mejadi Susah Untuk Di lalui Dan Mengganggu aktivitas warga	-	-	-	-	-	-	Penyebab : Hujan Deras dan Angin Kencang
3	Kamis. 07 Januari 2021	-	TANAH LONGSOR	Jl Wiraswasta Gang Bukit Indah Rt.15 Kel.Sidodadi	-	-	1 Tiang listrik roboh	-	-	-	-	-	-	Akibat hujan intensitas deras dan angin kencang
19	Rabu, 13 Desember 2023	Pukul 21.33	Pohon Tumbang	Jl. Gunung Tabur Kel. Gunung Kelua Kec. Samarinda Ulu	Akses jalan tertutup	Penyebab Hujan deras disertai angin kencang Upaya: Melakukan Pemangkas Dampak: Angin Kencang Upaya: Melakukan Pemangkas
20	Jum'at, 15 Desember 2023	Pukul 15.00	Pohon Tumbang	Jl. Balai Kota Samarinda Kel. Bugis Kec. Samarinda Kota	Dampak:- Mengenal kanopi Parkiran Bus Pemkot Samarinda	Upaya: Melakukan Pemangkas

3.1.2. Data Preparation

a. Data Integration

Data yang didapatkan dari BPBD dan BMKG selanjutnya akan digabungkan menjadi satu untuk mempermudah pengolahan data sehingga data yang didapatkan lebih lengkap untuk mengetahui informasi mengenai penyebab terjadinya bencana banjir. Setelah kedua data digabungkan maka didapatkan 19 atribut dan 1 atribut yang dijadikan sebagai kelas.

Proses digabungkannya kedua data yang telah didapatkan menggunakan perangkat lunak *Microsoft Office Excel*. Adapun data dari BPBD yang digabungkan yaitu tanggal, jam kejadian, jenis bencana, lokasi wilayah, luas area, objek terkena bencana, korban, kerugian, dan Keterangan. Kemudian pada data dari BMKG yang memiliki fitur temperatur minimum, temperatur maksimum, temperatur rata-rata, kelembaban rata-rata, curah hujan, lamanya penyinaran matahari, kecepatan angin maksimum, arah angin saat kecepatan maksimum, kecepatan angin rata-rata, dan arah angin terbanyak.

Tabel 3. 3 Hasil Data *Integration*

No	Atribut	Tipe Data	Keterangan
1	Tanggal	<i>date</i>	Tanggal Kejadian
2	Jam Kejadian	<i>String</i>	Jam kejadian
3	Jenis Bencana	<i>string</i>	Bencana alam yang terjadi
4	Lokasi wilayah	<i>string</i>	Tempat terjadinya banjir
5	Luas Area M2	<i>numeric</i>	Luas area yang terdampak
6	Objek terkena bencana	<i>string</i>	Kerugian fasilitas yang terdampak bencana
7	Korban	<i>numeric</i>	Jumlah korban terdampak bencana
8	Kerugian	<i>numeric</i>	Nominal kerugian
9	Keterangan	<i>numeric</i>	Detail kejadian bencana
10	Tn	<i>numeric</i>	Temperatur minimum (°C)
11	Tx	<i>numeric</i>	Temperatur maksimum (°C)
12	Tavg	<i>numeric</i>	Temperatur rata-rata (°C)
13	RH_avg	<i>numeric</i>	Kelembaban rata-rata (%)
14	RR	<i>numeric</i>	Curah hujan (mm)
15	Ss	<i>numeric</i>	Lamanya penyinaran matahari (jam)
16	ff_x	<i>numeric</i>	Kecepatan angin maksimum (m/s)
17	ddd_x	<i>numeric</i>	Arah angin saat kecepatan maksimum (°)
18	ff_avg	<i>numeric</i>	Kecepatan angin rata-rata (m/s)
19	ddd_car	<i>string</i>	Arah angin terbanyak (°)

b. Data Selection

Tahap ini dilakukan proses pemilihan atribut-atribut yang telah dipilih secara manual pada Data Banjir dari BMKG dan BPBD yang diperoleh sebanyak 19 Atribut yang terpilih dan 1 atribut sebagai target atau kelas.

	Tanggal	Temperatur- minimum	Temperatur- maksimum	Temperature- rata-rata	Kelembaban	Curah- hujan	Lama- penyinaran- matahari	Kecepatan- angin	Arah- angin- maksimum	Kecepatan- angin- rata-rata	Arah- angin- terbanyak	terjadi- banjir
0	01-01-2021	23.0	33.2	26.5	88.0	1.8	3.3	4.0	280.0	2.0	W	tidak banjir
1	02-01-2021	23.2	30.8	27.1	88.0	7.0	6.4	2.0	140.0	1.0	C	tidak banjir
2	03-01-2021	24.8	32.7	27.3	84.0	2.0	1.2	5.0	290.0	2.0	NW	banjir
3	04-01-2021	24.6	31.8	28.1	84.0	2.7	5.4	3.0	300.0	2.0	NW	tidak banjir
4	05-01-2021	24.6	31.4	27.4	83.0	10.5	1.8	4.0	300.0	2.0	W	tidak banjir
...
1090	27-12-2023	24.2	32.0	27.6	83.0	0.3	2.8	4.0	60.0	2.0	NE	tidak banjir
1091	28-12-2023	24.0	32.0	28.0	82.0	1.0	6.4	5.0	70.0	2.0	C	tidak banjir
1092	28-12-2023	23.7	32.7	28.7	77.0	3.5	5.9	5.0	60.0	2.0	NE	tidak banjir
1093	30-12-2023	24.2	32.6	28.3	82.0	NaN	10.4	4.0	60.0	1.0	NE	tidak banjir
1094	31-12-2023	24.6	32.4	28.3	84.0	0.0	6.9	4.0	90.0	2.0	E	tidak banjir

1095 rows x 12 columns

Gambar 3. 1 Hasil *Data Selection*

3.1.1. Data Cleaning

Pada tahap ini akan menggunakan bahasa pemrograman *Python* serta *library Pandas*, dengan fungsi *dropna()* untuk menghapus baris yang memiliki nilai kosong pada dataset Banjir dengan jumlah awal 1095 *record*.

	Tanggal	Temperatur- minimum	Temperatur- maksimum	Temperature- rata-rata	Kelembaban	Curah- hujan	Lama- penyinaran- matahari	Kecepatan- angin	Arah- angin- maksimum	Kecepatan- angin- rata-rata	Arah- angin- terbanyak	terjadi- banjir
0	01-01-2021	23.0	33.2	26.5	88.0	1.8	3.3	4.0	280.0	2.0	W	tidak banjir
1	02-01-2021	23.2	30.8	27.1	88.0	7.0	6.4	2.0	140.0	1.0	C	tidak banjir
2	03-01-2021	24.8	32.7	27.3	84.0	2.0	1.2	5.0	290.0	2.0	NW	banjir
3	04-01-2021	24.6	31.8	28.1	84.0	2.7	5.4	3.0	300.0	2.0	NW	tidak banjir
4	05-01-2021	24.6	31.4	27.4	83.0	10.5	1.8	4.0	300.0	2.0	W	tidak banjir
...
1090	27-12-2023	24.2	32.0	27.6	83.0	0.3	2.8	4.0	60.0	2.0	NE	tidak banjir
1091	28-12-2023	24.0	32.0	28.0	82.0	1.0	6.4	5.0	70.0	2.0	C	tidak banjir
1092	28-12-2023	23.7	32.7	28.7	77.0	3.5	5.9	5.0	60.0	2.0	NE	tidak banjir
1093	30-12-2023	24.2	32.6	28.3	82.0	NaN	10.4	4.0	60.0	1.0	NE	tidak banjir
1094	31-12-2023	24.6	32.4	28.3	84.0	0.0	6.9	4.0	90.0	2.0	E	tidak banjir

1095 rows x 12 columns

Gambar 3. 2 *Dataset* sebelum dibersihkan

Pada tampilan data, Sebelum proses pembersihan data dilakukan, terdapat nilai-nilai kosong pada data. Oleh karena itu, diperlukan penghapusan data yang tidak lengkap. Jumlah total data sebelum pembersihan adalah 1095 baris. Setelah melalui proses *data cleaning*, jumlahnya berkurang menjadi 890 baris. Dengan demikian, terdapat 205 baris data kosong yang telah dihapus selama proses pembersihan data.

```

Tanggal                0
Temperatur-minimum    73
Temperatur-maksimum   10
Temperature-rata-rata 6
Kelembaban            7
Curah-hujan          133
Lama-penyinaran-matahari 8
Kecepatan-angin       2
Arah-angin-maksimum   2
Kecepatan-angin-rata-rata 2
Arah-angin-terbanyak 2
terjadi-banjir        0
dtype: int64

```

Gambar 3.3 Nilai kosong pada tiap atribut sebelum *data cleaning*

Pada gambar 3.3 merupakan nilai kosong pada tiap atribut sebelum *data cleaning*, dimana curah hujan yang terbesar nilai kosongnya, sebesar 133. Selanjutnya akan dilakukan proses pemeriksaan seperti yang ada pada gambar 3.4.

```

Jumlah data sebelum pembersihan data kosong: 1095
Jumlah data setelah pembersihan data kosong: 890

```

Gambar 3.4 Jumlah Sebelum dan sesudah *Data Cleaning*

Setelah dilakukan pemeriksaan pada data yang ada, berikut hasil dari *cleaning data*.

No	Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	ddd_car	terjadi_banjir
1	01-01-2021	23	33,2	26,5	88	1,8	3,3	4	280	2	8	0
2	02-01-2021	23,2	30,8	27,1	88	7	6,4	2	140	1	0	0
3	03-01-2021	24,8	32,7	27,3	84	2	1,2	5	290	2	4	1
4	04-01-2021	24,6	31,8	28,1	84	2,7	5,4	3	300	2	4	0
5	05-01-2021	24,6	31,4	27,4	83	10,5	1,8	4	300	2	8	0
....
876	26-12-2023	24,0	30	26,9	89	50	7,0	3,0	260	1	6	0
877	27-12-2023	24,2	32	27,6	83	0,3	2,8	4,0	60	2	3	0
878	28-12-2023	24,0	32,0	28,0	82	1,0	6,4	5	70	2	0	0
879	29-12-2023	23,7	32,7	28,7	77	3,5	5,9	5,0	60,0	2,0	3	0
890	31-12-2023	24,6	32,4	28,3	84	0	6,9	4	90	2	1	0

Gambar 3.5 Hasil Data *Cleaning*

Gambar 3.5 merupakan hasil proses data cleaning, selanjutnya setelah proses pembersihan data seperti yang ditunjukkan pada Gambar 3.5, jumlah data dengan nilai kosong telah dihilangkan, dan data yang tersisa setelah proses pembersihan berjumlah 890 *record*. Kemudian dilakukan pemeriksaan kembali, untuk memastikan bahwa tidak ada lagi data kosong pada data banjir, dengan hasil pemeriksaan sebagai yang ditunjukkan pada gambar 3.6.

```

Jumlah nilai yang hilang setelah pembersihan:
Tanggal                0
Temperatur-minimum    0
Temperatur-maksimum   0
Temperature-rata-rata 0
Kelembaban            0
Curah-hujan          0
Lama-penyinaran-matahari 0
Kecepatan-angin       0
Arah-angin-maksimum   0
Kecepatan-angin-rata-rata 0
Arah-angin-terbanyak 0
terjadi-banjir        0
dtype: int64

```

Gambar 3.6 Jumlah Nilai Kosong Tiap Kolom Setelah Pembersihan

3.1.2. Data Transformation

Pada tahap transformasi data, dilakukan perubahan terhadap data kategorikal menjadi numerik. Data yang diubah pada tahap ini meliputi 'Arah-angin-terbanyak' dan 'terjadi_banjir'. Dalam proses ini, perubahan data kategorikal menjadi numerik untuk variabel 'Arah-angin-terbanyak' menggunakan library `'LabelEncoder()'` dari `sklearn`, sedangkan perubahan untuk variabel 'terjadi_banjir' menggunakan fungsi `'replace()'` dari Python.

Tabel 3. 4 Dataset Sebelum ditransformasi

No	Arah-angin-terbanyak	terjadi-banjir
0	W	Tidak Banjir
1	C	Tidak Banjir
2	NW	Banjir
3	NW	Tidak Banjir
4	W	Tidak Banjir
...		
1089	SE	Tidak Banjir
1090	NE	Tidak Banjir
1091	C	Tidak Banjir
1092	NE	Tidak Banjir
1094	E	Tidak Banjir

Pada Tabel 3.4 merupakan tampilan data pada kolom atribut 'Arah-angin-terbanyak' dan 'terjadi_banjir' sebelum transformasi data menunjukkan bahwa data berbentuk kategorikal.

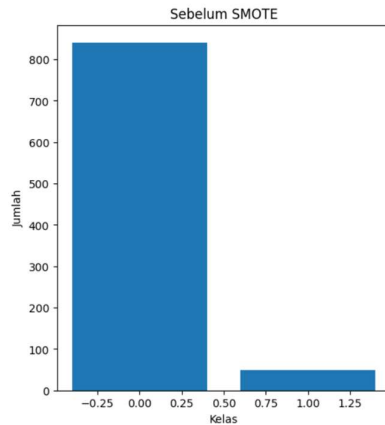
Tabel 3. 5 Dataset Setelah ditransformasi

No	Arah-angin-terbanyak	Terjadi-banjir
0	8	0
1	0	0
2	4	1
3	4	0
4	8	0
...
1090	6	0
1091	3	0
1092	0	0
1093	3	0
1094	1	0

Pada Tabel 3.5 merupakan tampilan data pada kolom atribut 'Arah-angin-terbanyak' dan 'terjadi_banjir' setelah *transformasi* data menunjukkan bahwa data yang sebelumnya berbentuk kategorikal telah diubah menjadi *numeric* untuk memudahkan proses klasifikasi.

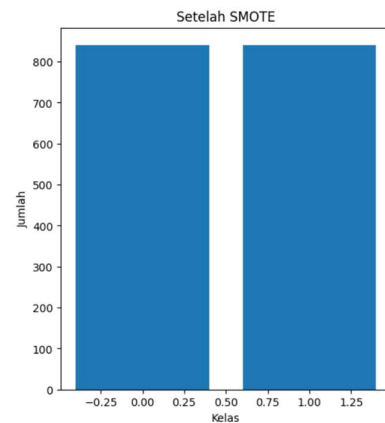
3.1.3. Data Balancing

Proses penyeimbangan data dilakukan menggunakan modul Python `'imblearn.over_sampling'` dengan mengimpor fungsi SMOTE (*Sampling Strategy*) untuk melakukan *oversampling*. Teknik *oversampling* ini diterapkan untuk menyamakan jumlah sampel antara kelas minoritas dan kelas mayoritas dalam dataset banjir yang memiliki ketidakseimbangan kelas. Dengan menerapkan SMOTE, ketidakseimbangan kelas pada data banjir dapat diatasi secara efektif.



Gambar 3. 7 Jumlah Kelas Sebelum Penerapan *SMOTE*

Pada Gambar 3.7 terdapat perbedaan jumlah kelas dimana kategori Tidak Banjir (0) berjumlah 841 data dan kategori terjadi-banjir (1) berjumlah 49 data.



Gambar 3. 8 Jumlah Kelas Sesudah Penerapan *SMOTE*

Pada Gambar 3.8 Setelah *oversampling* SMOTE diterapkan, jumlah sampel di kedua kelas menjadi seimbang, masing-masing dengan 841 sampel. Dengan demikian, setelah penerapan SMOTE, total jumlah data meningkat menjadi 1682 *record*.

3.1.4. Hasil Permodelan KNN Dengan *Pyhton*

Dalam penelitian ini, tahap pemodelan menggunakan algoritma KNN dengan nilai $k=3$, $k=5$, $k=7$, $k=11$, $k=13$, dan $k=15$, serta pembagian data menggunakan *10-fold cross-validation*. Evaluasi dilakukan menggunakan *Confusion Matrix*. Hasil evaluasi rata-rata *Confusion Matrix* untuk seluruh *fold* dari masing-masing nilai k dapat dilihat pada Tabel 3.6.

Tabel 3. 6 Hasil Evaluasi *Confusion Matrix*

Nilai K	TN	FP	FN	TP	Mean Accuracy
K=3	66	1.9	18	82	88,23%
K=5	62	2.6	22	82	85,38%
K=7	60	3.2	24	81	84,07%
K=9	58	3.6	26	80	82,52%
K=11	57	4.4	27	80	81,27%
K=13	56	4.9	28	79	80,62%
K=15	54	5.1	30	79	79,31%

Dapat dilihat pada tabel 3.6 bahwa hasil akurasi rata-rata dari masing-masing nilai K pada keseluruhan *fold* dengan total nilai TN = 66 adalah yang paling tinggi yang didapatkan dari k=3, nilai FP = 5.1 adalah yang paling tinggi didapatkan pada k=15, nilai FN = 28 memiliki nilai tinggi yang didapatkan pada K=15, dan nilai TP = 82 adalah nilai yang tinggi pada k=7 dan k=5. bisa dilihat juga bahwa akurasi rata-rata terbaik yang didapatkan dari keseluruhan *fold* adalah k=3 sebesar 88,23%, sedangkan untuk akurasi terendah adalah k=15 dengan nilai akurasi rata-rata 79,31%.

3.1.5. Hasil Permodelan KNN + PSO

PSO bertindak sebagai metode optimasi untuk mengoptimalkan parameter-parameter algoritma klasifikasi dengan tujuan meningkatkan kinerja keseluruhan model. Dalam konteks algoritma KNN, parameter-parameter yang dioptimalkan meliputi jumlah jumlah tetangga (*n_neighbors*), kedalaman maksimum pohon (*max_depth*), jumlah minimum sampel untuk memisah internal node (*min_samples_split*), dan jumlah minimum sampel di *leaf node* (*min_samples_leaf*). Pengoptimalan ini dilakukan untuk memastikan model mencapai kinerja yang optimal dalam hal akurasi, sehingga model dapat menghasilkan prediksi yang lebih akurat dan refektif.

Tabel 3. 7 Hasil Evaluasi k-nearest neighbors (KNN) + PSO

Nilai K	TN	FP	FN	TP	Mean Accuracy
K=3	84	15	0.6	69	90,85%
K=5	83	17	0.8	67	89,42%
K=7	84	20	0.5	64	87,70%
K=9	83	22	0.9	62	86,50%
K=11	83	22	0.8	62	86,39
K=13	83	23	1.3	61	85,55%
K=15	83	24	60	1.4	84,84%

Dapat dilihat pada tabel 3.7 bahwa hasil akurasi rata-rata dari masing-masing nilai K pada keseluruhan *fold* mengalami kenaikan akurasi.

3.1.6. Hasil Permodelan KNN + Relief

Pemodelan ini menggunakan algoritma KNN dan seleksi fitur menggunakan *Relief*. *Relief* sebagai metode seleksi fitur yang akan melakukan perangkingan fitur-fitur dalam dataset berdasarkan skor kepentingan (*importance score*), yang menunjukkan kemampuan fitur dalam membedakan antara kelas-kelas, di mana fitur dengan skor kepentingan yang lebih tinggi memiliki pengaruh yang besar terhadap hasil klasifikasi. Sebaliknya, fitur dengan skor kepentingan yang rendah memiliki pengaruh yang lebih kecil terhadap klasifikasi dan mungkin kurang relevan untuk model tersebut.

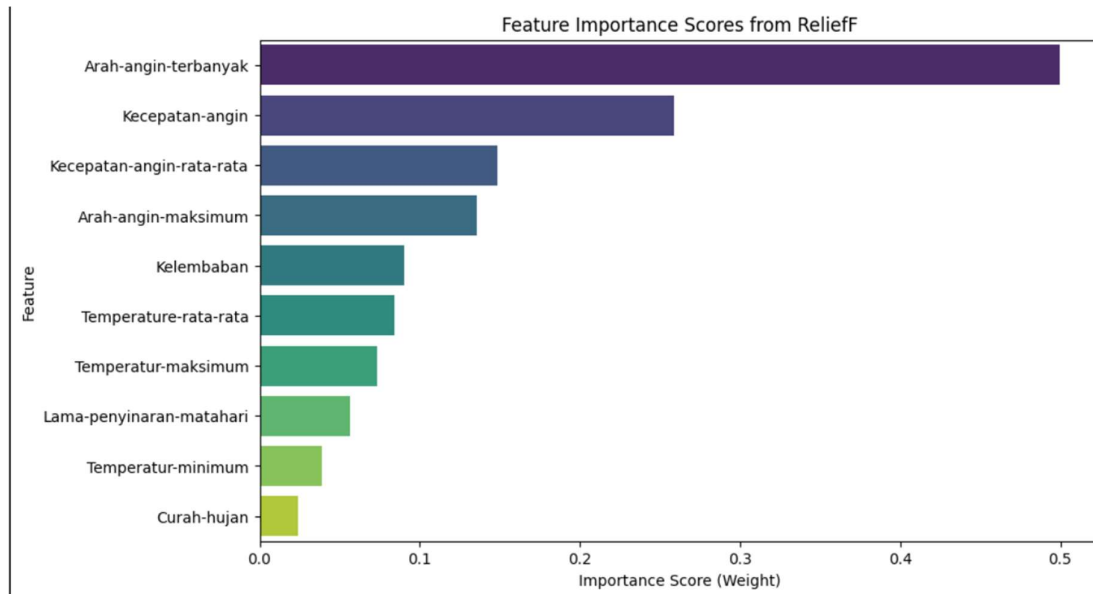
```

⇒ Fitur terbaik beserta bobotnya:
      Feature      Weight
0      Arah-angin-terbanyak  0.499401
1      Kecepatan-angin      0.259026
2      Kecepatan-angin-rata-rata  0.148464
3      Arah-angin-maksimum  0.135455
4      Kelembaban          0.090396
5      Temperature-rata-rata  0.084353
6      Temperatur-maksimum  0.073162
7      Lama-penyinaran-matahari  0.056167
8      Temperatur-minimum  0.039113
9      Curah-hujan         0.024114

```

Gambar 3. 9 Hasil Perangkingan relief berdasarkan (importance score)

Pada Gambar 3.10 menjelaskan Grafik setiap fitur berdasarkan skor kepentingan (*importance score*) dari Fitur yang memiliki Skor terbanyak hingga terendah.



Gambar 3. 10 Grafik *scores* dari *Relief*

Implementasi *Relief* dilakukan melalui pendekatan bahasa pemrograman *Python*. Berdasarkan Gambar 3.9 maka dilakukan penentuan atribut yang akan digunakan yang dapat dilihat pada tabel 3.8.

Tabel 3. 8 Penentuan Atribut yang Digunakan

No	Atribut	Skor	Hasil
1.	Arah-angin-terbanyak	0.499401	Digunakan
2.	Kecepatan-angin	0.259026	Digunakan
3.	Kecepatan-angin-rata-rata	0.148464	Digunakan
4.	Arah-angin-maksimum	0.135455	Digunakan
5.	Kelembaban	0.090396	Tidak Digunakan
6.	Temperature-rata-rata	0.084353	Tidak Digunakan
7.	Temperature-maksimum	0.073162	Tidak Digunakan
8.	Lama-penyiraman-matahari	0.056167	Tidak Digunakan
9.	Temperature-minimum	0.039113	Tidak Digunakan
10.	Curah-hujan	0.024114	Tidak Digunakan

Dapat dilihat pada tabel 3.8 dari 10 atribut, yang akan di ambil adalah 4 atribut dengan nilai Skor tertinggi. 4 atribut tersebut yaitu Arah-angin-terbanyak, kecepatan-angin, Kecepatan-angin-rata-rata dan Arah-angin-maksimum.

Setelah itu, dilakukan evaluasi ulang menggunakan *Confusion Matrix* dan melihat kembali hasil akurasi rata-rata pada tiap nilai K dengan pembagian data yang sama pada nilai K nya, yaitu 10-fold untuk melihat apakah ada perbedaan akurasi setelah melakukan seleksi fitur. Hasil evaluasi *Confusion Matrix* dapat dilihat pada tabel 3.9.

Tabel 3. 9 Evaluasi *Confusion Matrix* Setelah Seleksi Fitur

Nilai K	TN	FP	FN	TP	<i>Mean Accuracy</i>
K=3	74	5.6	9.8	78	89.59%
K=5	72	6.1	79	12	87.69%
K=7	70	6.4	78	14	85.67%
K=9	69	6.9	77	16	84.90%
K=11	68	8.1	76	16	82.64%
K=13	66	9.8	74	18	81.09%
K=15	65	11	73	19	80.20%

Pada Tabel 3.9, terdapat perubahan dalam hal peningkatan akurasi rata-rata pada masing-masing nilai K pada keseluruhan fold setelah menerapkan seleksi fitur *Relief*.

3.1.7. Hasil Permodelan KNN + *Relief* + PSO

Pada pemodelan ini, akan menerapkan optimasi PSO terhadap model klasifikasi untuk meningkatkan performa model KNN yang telah diintegrasikan dengan seleksi fitur *Relief*. Optimasi ini mencakup penyesuaian berbagai parameter, termasuk jumlah tetangga terdekat (*n_neighbors*), ukuran *leaf* (*leaf_size*), dan parameter jarak (*p*) dalam algoritma KNN.

Tabel 3. 10 Evaluasi *KNN+ Relief + PSO*

Nilai K	TN	FP	FN	TP	<i>Mean Accuracy</i>
K=3	66	1.9	18	82	90.84%
K=5	62	2.6	22	82	89.95%
K=7	60	3.2	24	81	87.75%
K=9	58	3.6	26	80	86.68%
K=11	57	4.4	27	80	85.55%
K=13	56	4.9	28	79	83.23%
K=15	54	5.1	30	79	82.11%

Pada Tabel 3.1 penerapan KNN dengan seleksi fitur *Relief* dan optimasi PSO mengalami peningkatan akurasi rata-rata pada masing-masing nilai K pada keseluruhan *fold*.

3.1.8. Perbandingan Hasil

Berikut adalah perbandingan hasil dari berbagai model klasifikasi yang telah diterapkan dalam penelitian ini. Penelitian ini mengevaluasi performa model *k-nearest neighbors* (KNN) dalam beberapa konfigurasi: KNN dasar tanpa optimasi, KNN yang dioptimalkan dengan *Particle Swarm Optimization* (PSO), KNN yang menggunakan seleksi fitur dengan algoritma *Relief*, dan kombinasi KNN yang menggunakan seleksi fitur *Relief* serta optimasi PSO. Pada tabel 3.11 akan menyajikan hasil akurasi dari masing-masing model untuk memberikan gambaran yang jelas tentang efektivitas setiap pendekatan dalam meningkatkan kinerja klasifikasi.

Tabel 3. 11 Perbandingan hasil akurasi dari setiap model KNN

Nilai K	<i>KNN</i>	<i>KNN+PSO</i>	Status	<i>KNN+Relief</i>	status	<i>KNN+Relief+PSO</i>	Status
K=3	88,23%	90,85%	+2.62%	89.59%	1,36%	90.84%	2,61%
K=5	85,38%	89,42%	+4.04%	87.69%	2,31%	89.95%	4,57%
K=7	84,07%	87,70%	+3,63%	85.67%	1,60%	87.75%	3,68%
K=9	82,52%	86,50%	+ 3.98%	84.90%	2,38%	86.68%	4,16%
K=11	81,27%	86,39	+ 5.12%	82.64%	1,37%	85.55%	4,28%
K=13	80,62%	85,55%	+ 4.93%	81.09%	0,47%	83.23%	2,61%
K=15	79,31%	84,84%	+ 5.53%	80.20%	0,89%	82.11%	2,80%

Dalam penelitian ini, berbagai metode optimasi diterapkan pada algoritma KNN untuk meningkatkan performa model. Hasil perbandingan akurasi pada table 3.11 menunjukkan bahwa pada K=3,k=5,k=7,k=9,k=11,k=13,k=15 Secara keseluruhan, model KNN dasar mengalami peningkatan akurasi setelah dioptimasi dengan PSO, dibandingkan dengan penggunaan seleksi fitur *Relief* atau kombinasi keduanya. Peningkatan terbesar terlihat pada KNN+PSO, terutama pada nilai K yang lebih tinggi.

3.2. Pembahasan

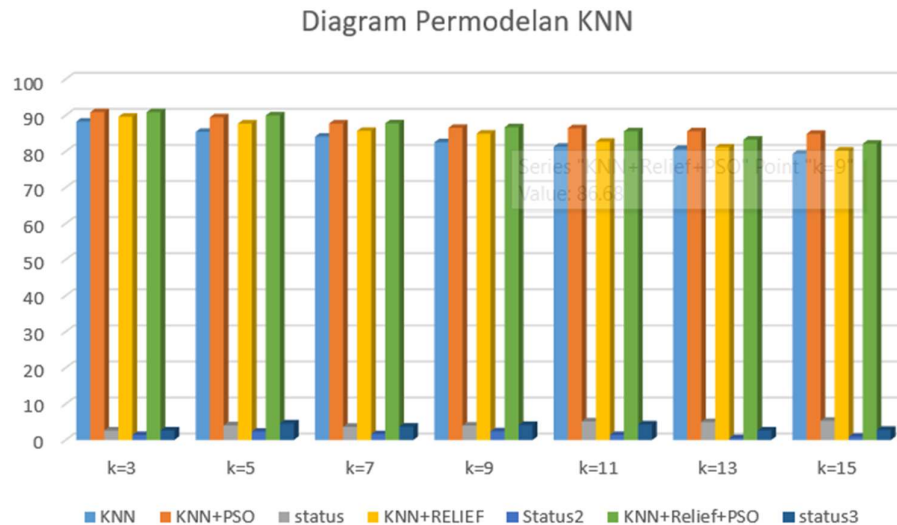
Penelitian ini menggunakan data banjir Kota Samarinda yang diperoleh dari BPBD dan BMKG untuk periode 2021-2023. Data yang terkumpul kemudian melalui beberapa tahapan pengolahan, termasuk pengumpulan data, persiapan data (*data preparation*), transformasi data (*data transformation*), pembersihan data (*data cleaning*), penyeimbangan data (*data balancing*), pembagian data, dan pemodelan yang melibatkan beberapa model klasifikasi untuk mengevaluasi hasilnya. Teknik *oversampling* SMOTE digunakan untuk penyeimbangan data, dan pembagian data training dan testing dilakukan menggunakan *K-fold Cross Validation* dengan K=10.

Setelah dilakukan Analisis, menunjukkan bahwa PSO secara signifikan meningkatkan akurasi algoritma KNN melalui optimasi parameter yang lebih tepat, sementara seleksi fitur dengan *Relief* juga memberikan peningkatan akurasi. Kombinasi *Relief* dan PSO memperlihatkan sinergi dalam pengoptimalan parameter dan pemilihan fitur yang relevan, menghasilkan peningkatan akurasi terbesar kedua setelah KNN dan PSO. Metode kombinasi ini menegaskan pentingnya penyesuaian parameter dan seleksi fitur untuk memperbaiki performa model secara signifikan, terutama pada nilai k yang lebih tinggi.

Peningkatan akurasi oleh PSO dapat dijelaskan oleh beberapa faktor utama. Pertama, optimasi parameter dengan PSO memungkinkan penyesuaian yang sangat tepat terhadap parameter model seperti jumlah tetangga terdekat (k), ukuran leaf, dan parameter jarak, dengan menggunakan pendekatan berbasis populasi untuk eksplorasi ruang parameter secara efektif. Kedua, PSO membantu mengatasi *overfitting* dengan mencari keseimbangan optimal antara bias dan varians, sehingga model dapat menangkap pola yang lebih relevan dalam data tanpa terlalu terpengaruh oleh noise. Ketiga, Meskipun seleksi fitur Relief meningkatkan akurasi dengan memilih fitur yang paling relevan dan mengurangi noise, sebagian peningkatannya tidak sebesar yang dicapai oleh PSO. Hal ini dikarenakan Relief hanya memilih fitur tanpa menyesuaikan parameter model, sementara PSO memberikan manfaat lebih efektif melalui optimasi parameter langsung. kombinasi antara seleksi fitur Relief dan optimasi PSO tetap memberikan manfaat, namun tidak selalu lebih baik daripada optimasi parameter langsung dengan PSO.

Dengan demikian, kemampuan PSO dalam menyesuaikan parameter model dan mengatasi *overfitting* menjelaskan mengapa model KNN yang dioptimalkan dengan PSO secara konsisten memberikan peningkatan akurasi tertinggi dibandingkan model KNN dasar, model KNN dengan seleksi fitur *Relief*, dan model KNN dengan kombinasi seleksi fitur *Relief* dan PSO.

Gambar 3. 11 Diagram Permodelan KNN



3.2.1. Fitur Terpilih

Dalam penelitian ini, seleksi fitur menggunakan algoritma *Relief* diterapkan pada data banjir Kota Samarinda untuk meningkatkan performa model klasifikasi *k-Nearest Neighbors* (KNN), dengan fitur-fitur terpilih meliputi arah-angin-terbanyak (0.499401), kecepatan-angin (0.259026), kecepatan-angin-rata-rata (0.148464), dan arah-angin-maksimum (0.135455). Pada model yang menggabungkan seleksi fitur dan optimasi parameter menunjukkan performa prediksi banjir yang lebih baik. Peningkatan terbesar pada model KNN dengan seleksi fitur *Relief* terlihat pada $K=9$, dengan akurasi meningkat dari 82.52% menjadi 84,90 %, mengindikasikan bahwa pemilihan fitur yang relevan berhasil meningkatkan performa model klasifikasi dalam memprediksi kejadian banjir di Kota Samarinda.

Penelitian sebelumnya juga mendukung hasil ini, seperti penelitian oleh (Intan & Sari, 2023) yang menggunakan seleksi fitur *gain ratio* mengidentifikasi ‘Kelembaban’, ‘Temperatur-minimum’, dan ‘Temperatur-maksimum’ sebagai fitur paling berpengaruh, menunjukkan peningkatan akurasi algoritma klasifikasi KNN dengan peningkatan sebesar 5.95%. Sementara itu, penelitian oleh (Dilla Evitasari et al., 2023) yang menerapkan seleksi fitur menggunakan *algoritma Genetik* (GA) pada data banjir Kota Samarinda menemukan bahwa ‘Kelembapan’, ‘Lama-penyinaran-matahari’, dan ‘Kecepatan-angin’ memiliki pengaruh paling signifikan terhadap prediksi banjir, meningkatkan akurasi algoritma klasifikasi SVM sebesar 13.45%.

3.2.2. Peningkatan Akurasi

Dalam penelitian ini, algoritma KNN diterapkan untuk klasifikasi data banjir dengan menggunakan metode *oversampling* SMOTE dalam proses pra-pemrosesan data. Hasil menunjukkan bahwa model dasar KNN mencapai berbagai tingkat akurasi pada nilai k yang berbeda, dengan peningkatan akurasi yang signifikan setelah menerapkan seleksi fitur *Relief* dan optimasi PSO. KNN PSO memberikan peningkatan sebesar 2-5%, dengan seleksi fitur *Relief* peningkatan sebesar 1-2% dan KNN dengan

kombinasi *Relief* dan PSO peningkatan sebesar 2-5%. Penggunaan algoritma KNN dalam klasifikasi data banjir yang menggunakan metode *oversampling* SMOTE tanpa seleksi fitur dan optimasi menunjukkan peningkatan akurasi yang signifikan setelah penerapan optimasi. Model KNN yang menggabungkan metode *oversampling* SMOTE dan seleksi fitur *Relief* juga memperlihatkan peningkatan akurasi yang berarti setelah optimasi. Namun, model KNN dengan seleksi fitur memiliki kecepatan pemrosesan yang lebih cepat, meskipun peningkatan akurasinya tidak sebesar yang dicapai oleh optimasi PSO. Hal ini mengindikasikan bahwa PSO memberikan pengaturan parameter yang lebih tepat dan peningkatan performa keseluruhan model dibandingkan dengan model yang hanya menggunakan seleksi fitur *Relief*. Hasil Penelitian ini ditujukan kepada Badan Penanggulangan Bencana Daerah (BPBD), Dinas Pekerjaan Umum dan Penataan Ruang (PUPR), Pemerintah Daerah Kota Samarinda, serta institusi penelitian dan akademik yang berkontribusi terhadap pengembangan ilmu pengetahuan dan teknologi untuk penanggulangan bencana. Diharapkan bahwa hasil dari penelitian ini dapat memberikan informasi yang berguna dan pengetahuan yang bermanfaat yang berguna untuk pengambilan keputusan memprediksi bencana banjir sehingga dapat membuat keputusan yang lebih tepat dalam meningkatkan kesiapsiagaan dan respons terhadap banjir, serta membantu dalam perencanaan dan mitigasi bencana banjir di Kota Samarinda.