

BAB V

HASIL DAN ANALISIS PENELITIAN

1.1 Pendahuluan

Pada bab sebelumnya pengembangan perangkat lunak telah dilakukan yang selanjutnya akan digunakan sebagai alat penelitian yang digunakan untuk mengetahui kinerja dari algoritma *k*-Nearest Neighbor (kNN) dan *Modified k-Nearest Neighbor* (MkNN). Pada bab ini penulis akan membahas hasil dari penelitian dengan menggunakan perangkat lunak yang telah dikembangkan, hasil pengujian dari algoritma *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor* untuk dapat mengetahui kinerja algoritma berupa nilai *precision*, *recall*, *accuracy*, dan mencatat waktu komputasi serta memori yang dibutuhkan.

1.2 Konfigurasi percobaan

Pengujian yang dilakukan pada perangkat lunak untuk mengetahui kinerja dari algoritma *k-Nearest Neighbor* (kNN) dan *Modified k-Nearest Neighbor* (MkNN) akan menggunakan dataset *website phishing* yang memiliki atribut sebanyak 30 atribut dengan data yang berjumlah 2456 data, kemudian dataset tersebut akan disimpan dalam format .csv. Pengujian pada algoritma *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor* akan diketahui nilai *k* yang terbaik pada penelitian ini, sehingga nilai *k* yang akan digunakan sebagai ruang solusi adalah $k=3$, $k=5$, $k=7$ dan $k=9$. Masing-masing nilai *k* akan dihitung nilai *precision*, *recall*, *accuracy* serta mengetahui waktu komputasi dan memori yang digunakan dari pengujian

yang dilakukan sebanyak 10-fold. Sehingga akan dilakukan perbandingan dan analisis hasil kinerja algoritma *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor* pada pengklasifikasian data menggunakan dataset *website phishing*.

1.3 Hasil Pengujian

Hasil pengujian masing-masing algoritma akan dijelaskan pada subbab berikutnya, sebelumnya pada konfigurasi percobaan telah dijelaskan bahwa algoritma *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor* akan diuji menggunakan dataset *website phishing*. Pada subbab 5.3.1 akan menampilkan hasil pengujian klasifikasi menggunakan algoritma *k-Nearest Neighbor*, 5.3.2 akan menampilkan hasil pengujian klasifikasi menggunakan algoritma *Modified k-Nearest Neighbor* dan 5.3.3 akan menampilkan hasil pengujian klasifikasi *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor*.

1.3.1 Hasil Pengujian Menggunakan Algoritma *k-Nearest Neighbor*

Pengujian menggunakan algoritma *k-Nearest Neighbor* dilakukan dengan menginputkan nilai k yang berbeda yaitu $k=3$, $k=5$, $k=7$ dan $k=9$, dan dilakukan sebanyak 10-fold dengan menggunakan teknik *cross validation*. Sehingga akan diketahui nilai terbaik dari *precision*, *recall*, *accuracy*, waktu komputasi dan memori yang dibutuhkan algoritma. Hasil pengujian algoritma *k-Nearest Neighbor* dapat dilihat pada table V-1 dan V-2, untuk hasil pengujian dari nilai *precision*, *recall*, *accuracy* dapat dilihat pada table V-1, sedangkan hasil pengujian terhadap waktu komputasi dan memori yang dibutuhkan dapat dilihat pada table V-2.

Table V-1. Hasil Pengujian Nilai *Precision*, *Recall* dan *Accuracy* Menggunakan Algoritma *k-Nearest Neighbor* dengan nilai parameter *k* yang berbeda-beda

<i>k-fold</i>	<i>k=3</i>			<i>k=5</i>			<i>k=7</i>			<i>k=9</i>		
	Accuracy	Recall	Precision									
1	0.9228	0.9115	0.9196	0.9472	0.9381	0.9464	0.9472	0.9381	0.9464	0.9431	0.9292	0.9459
2	0.9309	0.9355	0.8878	0.9553	0.9462	0.9362	0.9553	0.9462	0.9362	0.9431	0.9247	0.9247
3	0.9553	0.9649	0.9402	0.9634	0.9649	0.9565	0.9634	0.9649	0.9565	0.9553	0.9649	0.9402
4	0.9431	0.9600	0.9302	0.9634	0.9760	0.9531	0.9634	0.9760	0.9531	0.9553	0.9520	0.9597
5	0.9390	0.9266	0.9352	0.9472	0.9358	0.9444	0.9472	0.9358	0.9444	0.9309	0.9174	0.9259
6	0.9431	0.9770	0.8763	0.9512	0.9540	0.9121	0.9512	0.9540	0.9121	0.9472	0.9195	0.9302
7	0.9633	0.9569	0.9652	0.9714	0.9655	0.9739	0.9714	0.9655	0.9739	0.9592	0.9569	0.9569
8	0.9388	0.9328	0.9407	0.9551	0.9496	0.9576	0.9551	0.9496	0.9576	0.9347	0.9328	0.9328
9	0.9306	0.9135	0.9223	0.9306	0.9135	0.9223	0.9306	0.9135	0.9223	0.9102	0.8750	0.9100
10	0.9388	0.9474	0.9231	0.9510	0.9474	0.9474	0.9510	0.9474	0.9474	0.9429	0.9298	0.9464
Rata-rata	94.0557 %	0.9426	0.9240	95.3580 %	0.9491	0.9450	94.1354 %	0.9348	0.9348	94.2173 %	0.9302	0.9373

Table V-2. Hasil Pengujian Waktu Komputasi dan Memori yang dibutuhkan Algoritma *k-Nearest Neighbor* dengan nilai parameter *k* yang berbeda-beda.

<i>k=3</i>			<i>k=5</i>			<i>k=7</i>			<i>k=9</i>		
<i>k-fold</i>	Waktu komputasi (detik)	Memori (byte)	<i>k-fold</i>	Waktu komputasi (detik)	Memori (byte)	<i>k-fold</i>	Waktu komputasi (detik)	Memori (byte)	<i>k-fold</i>	Waktu komputasi (detik)	Memori (byte)
1	43.9214	148455424	1	44.0536	147386368	1	44.3585	147390464	1	46.1699	131702784
2	43.6169	148459520	2	43.3222	147542016	2	43.7391	147390464	2	43.5161	131702784
3	44.0313	148459520	3	43.2013	147542016	3	43.5052	147390464	3	42.7136	131702784
4	44.0724	148459520	4	43.1916	147542016	4	43.5385	147390464	4	42.7385	31702784
5	43.6742	148459520	5	43.3018	147382272	5	43.4365	147374080	5	42.8809	131702784
6	43.7548	148443136	6	43.4698	147382272	6	43.6063	147374080	6	43.1531	75595776
7	43.4904	148459520	7	43.0531	147382272	7	43.4396	147619840	7	42.7460	75563008
8	43.4663	148459520	8	43.1131	147382272	8	43.2909	147619840	8	42.5643	75563008
9	43.5198	148480000	9	43.1021	147382272	9	43.2420	147619840	9	42.6668	75563008
10	43.4538	148480000	10	43.1716	147382272	10	43.4612	131387392	10	42.6697	75563008
Rata-Rata	43.7001	148461568	Rata-Rata	43.2980	147430604	Rata-Rata	43.5618	145855692	Rata-Rata	43.1819	103636172

1.3.2 Hasil Pengujian menggunakan algoritma *Modified k-Nearest Neighbor*

Pengujian menggunakan algoritma *Modified k-Nearest Neighbor* (MkNN) pada dasarnya memiliki kesamaan dengan algoritma *k-Nearest Neighbor* namun pada algoritma MkNN memiliki tambahan perhitungan dari algoritma kNN tradisional. Penelitian dilakukan dengan menginputkan nilai k yang berbeda sebagai runag solusi penelitian yaitu $k=3$, $k=5$, $k=7$ dan $k=9$ yang akan dilakukan sebanyak 10 *fold* dengan menggunakan teknik *cross validation*. Sehingga dapat diketahui nilai terbaik dari nilai *precision*, *recall*, *accuracy*, waktu komputasi serta memori yang dibutuhkan algoritma. Hasil pengujian menggunakan algoritma *Modified k-Nearest Neighbor* dapat dilihat pada table V-3. dan table V-2, untuk hasil pengujian terhadap nilai *precision*, *recall*, *accuracy* dapat dilihat pada table V-3, sedangkan hasil pengujian terhadap waktu komputasi dan memori yang dibutuhkan dapat dilihat pada table V-3.

Table V-3. Hasil Pengujian Nilai *Precision*, *Recall* dan *Accuracy* Menggunakan algoritma *Modified k-Nearest Neighbor* (MkNN) dengan nilai parameter *k* yang berbeda-beda

<i>k-fold</i>	K=3			K=5			K=7			K=9		
	Accuracy	Recall	Precision									
1	0.9390	0.9292	0.9375	0.9593	0.9646	0.9478	0.9472	0.9469	0.9386	0.9472	0.9381	0.9464
2	0.9512	0.9355	0.9355	0.9553	0.9462	0.9362	0.9553	0.9355	0.9457	0.9553	0.9462	0.9362
3	0.9715	0.9825	0.9573	0.9675	0.9737	0.9569	0.9675	0.9737	0.9569	0.9634	0.9649	0.9565
4	0.9512	0.9680	0.9380	0.9675	0.9760	0.9606	0.9675	0.9600	0.9756	0.9634	0.9760	0.9531
5	0.9553	0.9450	0.9537	0.9593	0.9541	0.9541	0.9472	0.9358	0.9444	0.9472	0.9358	0.9444
6	0.9512	0.9770	0.8947	0.9553	0.9655	0.9130	0.9553	0.9655	0.9130	0.9512	0.9540	0.9121
7	0.9796	0.9828	0.9744	0.9714	0.9655	0.9739	0.9673	0.9569	0.9737	0.9714	0.9655	0.9739
8	0.9510	0.9496	0.9496	0.9510	0.9412	0.9573	0.9347	0.9412	0.9256	0.9551	0.9496	0.9576
9	0.9347	0.9135	0.9314	0.9265	0.9038	0.9216	0.9224	0.8942	0.9208	0.9306	0.9135	0.9223
10	0.9429	0.9474	0.9310	0.9469	0.9474	0.9391	0.9469	0.9386	0.9469	0.9510	0.9474	0.9474
Rata-Rata	95.2767 %	0.9530	0.9403	95.6015 %	0.9538	0.9460	95.1126 %	0.9448	0.9441	94.8691 %	0.9377	0.9439

Table V-4. Hasil Pengujian Waktu Komputasi Dan Memory Yang Dbutuhkan Algoritma *Modified k-Nearest Neighbor* (MkNN) dengan nilai parameter k yang berbeda-beda.

$k=3$			$k=5$			$k=7$			$k=9$		
<i>fold</i>	Waktu Komputasi (detik)	Memori (byte)	<i>k-fold</i>	Waktu Komputasi (detik)	Memori (byte)	<i>fold</i>	Waktu Komputasi (detik)	Memori (byte)	<i>fold</i>	Waktu Komputasi (detik)	Memori (byte)
1	454.2340	146096128	1	421.3488	146747392	1	473.4130	148291584	1	44.0536	147386368
2	448.4459	146477056	2	453.7401	147017728	2	486.8465	143699968	2	43.3222	147542016
3	435.3409	146321408	3	419.6396	147017728	3	465.1608	143486976	3	43.2013	147542016
4	432.7476	146325504	4	418.4779	147013632	4	467.0733	143400960	4	43.1916	147542016
5	461.5305	146259968	5	416.8014	147013632	5	472.5853	143446016	5	43.3018	147382272
6	457.7259	146264064	6	417.2410	147025920	6	473.7660	143450112	6	43.4698	147382272
7	421.5541	146522112	7	417.2634	147292160	7	472.1389	143630336	7	43.0531	147382272
8	421.3728	146522112	8	416.8753	147292160	8	483.3248	143618048	8	43.1131	147382272
9	426.3971	146518016	9	417.0268	147292160	9	530.4581	115281920	9	43.1021	147382272
10	428.0259	146518016	10	419.5370	147288064	10	525.5484	68468736	10	43.1716	147382272
Rata-Rata	438.7375	146382438	Rata-Rata	421.7951	147100057	Rata-Rata	485.0315	133677465	Rata-Rata	501.8383	133931827

1.3.3 Hasil Pengujian algoritma k-Nearest Neighbor dan Modified k-Nearest Neighbor.

Hasil Pengklasifikasi menggunakan algoritma *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor* menggunakan dataset *website phishing* berdasarkan penelitian yang telah dilakukan terjadi peningkatan terhadap nilai *precision*, *recall*, *accuracy*, waktu komputasi serta memori yang dibutuhkan pada algoritma *Modified k-Nearest Neighbor* untuk pengklasifikasian dataset *website phishing*. Perbandingan hasil pengklasifikasian kedua algoritma dilakukan terhadap nilai *k* yang berbeda sebagai ruang solusi yaitu *k*=3, 5, 7 dan 9 yang dapat dilihat pada table V.5.

Tabel V-5. Hasil Perbandingan Kinerja Algoritma k-Nearest Neighbor dan Modified k-Nearest Neighbor.

<i>k</i>	<i>k-Nearest Neighbor (kNN)</i>					<i>Modified k-Nearest Neighbor (MkNN)</i>				
	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>	<i>Waktu</i>	<i>Memori</i>	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>	<i>Waktu</i>	<i>Memori</i>
3	94.0557 %	0.9426	0.9240	43.7001 s	148461568 byte	95.2767 %	0.9530	0.9403	438.7375 s	146382438 byte
5	95.3580 %	0.9491	0.9450	43.2980 s	147430604 byte	95.6015 %	0.9538	0.9460	421.7951 s	147100057 byte
7	94.1354 %	0.9348	0.9348	43.5618 s	145855692 byte	95.1126 %	0.9448	0.9441	485.0315 s	133677465 byte
9	94.2173 %	0.9302	0.9373	43.1819 s	103636172 byte	94.8691 %	0.9377	0.9439	501.8383 s	133931827 byte

1.4 Analisis Hasil Pengujian

Hasil dari pengujian yang dilakukan didapatkan hasil berupa nilai *precision*, *recall*, *accuracy* serta waktu komputasinya dan memori yang dibutuhkan algoritma baik untuk algoritma *k-Nearest Neighbor* atau algoritma *Modified k-Nearest Neighbor* pada pengklasifikasian dataset *website phishing*. Pengujian yang dilakukan terhadap dua algoritma tersebut menggunakan nilai *k* yang berbeda-beda yaitu 3, 5, 7 dan 9 untuk mengetahui nilai *k* yang terbaik pada kedua algoritma tersebut.

Berdasarkan penelitian yang penulis lakukan dengan menggunakan dataset *website phishing* yang memiliki 30 atribut dan 2456 record data yang akan dievaluasi menggunakan *10-fold cross validation* dimana data akan dibagi menjadi 10 *fold* dengan rincian 9 *fold* untuk data latih dan 1 *fold* untuk data uji. Hasil dari pengujian dengan menggunakan algoritma *k-Nearest Neighbor* (kNN) didapatkan hasil pada table V-1 dan V-2, serta hasil pengujian menggunakan algoritma *Modified k – Nearest Neighbour* (MkNN) didapat hasil pada tabel V-3, dan V-4. Dari kedua algoritma tersebut *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor* akan dilakukan analisis dan perbandingan terhadap nilai akurasi, recall, precision, waktu komputasi dan besarnya memori yang dibutuhkan algoritma pada pengklasifikasian dataset *website phishing* yang didapatkan hasil pada table V-5.

Untuk analisis hasil pengujian klasifikasi *k-Nearest Neighbor* pada klasifikasi *website phishing* dapat dilihat pada subbab 5.4.1, dan analisis hasil pengujian klasifikasi *Modified k-Nearest Neighbor* pada klasifikasi *website phishing* dapat dilihat pada subbab 5.4.2.

1.4.1 Analisis hasil pengujian klasifikasi k-Nearest Neighbor pada klasifikasi *website phishing*

Dari hasil pengujian yang telah dilakukan dengan menggunakan berbagai nilai k yang berbeda yaitu $k=3$, $k=5$, $k=7$ dan $k=9$, dan dilakukan pengujian sebanyak 10-fold menggunakan teknik cross validation. Pada pengujian menggunakan algoritma k-Nearest Neighbor menghasilkan rata-rata nilai akurasi, recall dan precision terbaik yaitu pada nilai $k=5$, dengan nilai akurasi sebesar 95.35%, nilai recall 0.9491 dan nilai precision 0.9450. Untuk waktu komputasi terbaik yaitu rata-rata waktu 43.1819 detik dan memori terbaik yaitu rata-rata 103636172 byte.

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa penentuan nilai parameter k pada pengklasifikasian data menggunakan algoritma k-Nearest Neighbor (kNN) memiliki pengaruh yang cukup signifikan terhadap nilai *precision*, *recall* dan *accuracy* dari ruang solusi pada nilai parameter k yang diinputkan yaitu $k=3$, $k=5$, $k=7$ dan $k=9$ yang menghasilkan nilai *precision*, *recall* dan *accuracy* terbaik untuk digunakan pada pengklasifikasian dataset website phishing adalah $k=5$, hal tersebut berdasarkan hasil uji yang telah dilakukan pada algoritma k-Nearest Neighbor.

1.4.2 Analisis hasil pengujian klasifikasi Modified k-Nearest Neighbor pada klasifikasi *website phishing*.

Dari hasil pengujian yang telah dilakukan dengan menggunakan berbagai nilai k yang berbeda yaitu $k=3$, $k=5$, $k=7$ dan $k=9$, dan dilakukan pengujian sebanyak 10-fold menggunakan teknik *cross validation*. Pada pengujian menggunakan algoritma *Modified k-Nearest Neighbor* menghasilkan rata-rata nilai *precision*, *recall* dan *accuracy* terbaik yaitu pada nilai $k = 5$, dengan nilai *accuracy* sebesar 95.60%, nilai *recall* 0.9538 dan nilai *precision* 0.9460 dengan waktu komputasi terbaik rata-rata waktu 421.7951 detik dan memori terbaik dengan rata-rata 133677465 byte.

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa algoritma *Modified k-Nearest Neighbor* berdasarkan penelitian yang telah penulis lakukan terjadi peningkatan terhadap kualitas hasil klasifikasi dibandingkan menggunakan algoritma *k-Nearest Neighbor*, peningkatan tersebut terjadi pada nilai *precision*, *recall* dan *accuracy*. Namun pada algoritma *Modified k-Nearest Neighbor* waktu komputasi dan memori yang dibutuhkan juga terjadi penambahan yang cukup signifikan. Hal tersebut dikarenakan pada algoritma *Modified k-Nearest Neighbor* terjadi penghitungan terhadap nilai validitas data, sehingga setiap data latih akan dilakukan perhitungan yang menyebabkan penambahan waktu komputasi yang dibutuhkan oleh algoritma *Modified k-Nearest Neighbor* dalam melakukan klasifikasi data.

1.5 Kesimpulan

Dari hasil percobaan klasifikasi menggunakan algoritma *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor* pada klasifikasian data menggunakan dataset *website phishing* terjadi peningkatan presentase terhadap nilai *precision*, *recall* dan *accuracy* pada algoritma *Modified k-Nearest Neighbor*. Namun berdasarkan percobaan yang telah dilakukan penulis dapat disimpulkan bahwa dengan menggunakan dataset *website phishing* yang diimplementasikan dengan menggunakan algoritma *Modified k-Nearest Neighbor* tidak mengalami peningkatan yang cukup signifikan terhadap kinerja algoritma berdasarkan hasil uji yang telah penulis lakukan, dan terjadi peningkatan terhadap waktu komputasi serta memori yang dibutuhkan algoritma *Modified k-Nearest Neighbor*, sehingga dapat di nyatakan bahwa algoritma *Modified k-Nearest Neighbor* tidak lebih baik dibandingkan algoritma *k-Nearest Neighbor*, dikarenakan pada penelitian menggunakan dataset yang tidak sama dengan penelitian yang dilakukan oleh parvin, sehingga hasil pengklasifikasian yang penulis hasilkan tidak sama dengan penelitian yang parvin hasilkan