

**PERBANDINGAN KINERJA ALGORITMA *k*-NEAREST  
NEIGHBOR DAN MODIFIED *k*-NEAREST NEIGHBOR PADA  
KLASIFIKASI WEBSITE PHISHING**

*Diajukan Sebagai Syarat  
untuk Menyelesaikan Pendidikan Program Strata-1  
pada Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI*



**Oleh :**

**MOH TRISNO  
NIM : 09021181520037**

**Jurusan Teknik Informatika  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2020**

**LEMBAR PENGESAHAN TUGAS AKHIR**


**PERBANDINGAN KINERJA ALGORITMA *k*-NEAREST NEIGHBOR DAN  
MODIFIED *k*-NEAREST NEIGHBOR PADA KLASIFIKASI WEBSITE  
*PHISHING***

Oleh :

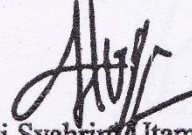
**Moh Trisno**  
NIM. 09021181520037

Indralaya, Juni 2020

Pembimbing I,

  
Rifkie Primartha, M.T  
NIP. 197802232006042002

Pembimbing II,

  
Alvi Syahrini Utami, M.Kom.  
NIP. 19781222 200604 2 003

Mengetahui,  
Ketua Jurusan Teknik Informatika,

  
  
Rifkie Primartha, M.T  
NIP. 197802232006042002

## TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Selasa tanggal 16 Juni 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Moh Trisno  
NIM : 09021181520037  
Judul Skripsi : Perbandingan Kinerja Algoritma *k- Nearest Neighbor* dan *Modified k- Nearest Neighbor* pada Klasifikasi *Website Phishing*.

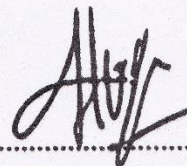
### 1. Pembimbing I

Rifkie Primartha, M.T  
NIP.19770601 200912 1 004



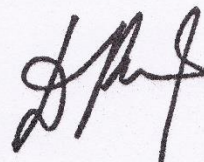
### 2. Pembimbing II

Alvi Syahrini Utami, M.Kom.  
NIP. 19781222 200604 2 003



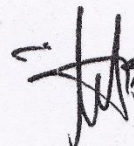
### 3. Penguji I

Dian Palupi Rini, M.Kom.,Ph.D  
NIP. 19780223 200604 2 002



### 4. Penguji II

Muhammad Ali Buchari, M.T  
NIP. 19880330 2019031 007



Mengetahui,  
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T  
NIP.197706012009121004

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Moh Trisno  
NIM : 09021181520037  
Program Studi : Teknik Informatika  
Judul Skripsi : Perbandingan kinerja algoritma *k- Nearest Neighbour* dan *Modofied k- Nearest Neighbour* pada Klasifikasi *Website Phishing*.

Hasil Pengecekan Software *iThenticate/Turnitin* : 14 %

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, Juni 2020



Moh Trisno  
NIM. 09021181520037

## KATA PENGANTAR

Puji syukur kepada Allah atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini yang berjudul “**Perbandingan Kinerja Algoritma *k*-Nearest Neighbor dan Modified *k*-Nearest Neighbor Pada Klasifikasi Website Phishing**”. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Orang tuaku, Kirommaidi, S. Pd dan Holilah, S. Pd, dan saudara-saudaraku yang selalu mendokan serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Rifkie Primartha, M.T selaku selaku Ketua Jurusan Teknik Informatika dan sebagai dosen pembimbing I, yang telah mengarahkan dan membimbing penulis dalam proses pengerjaan Tugas Akhir.
4. Ibu Alvi Syahrini Utami, M.Kom selaku pembimbing II, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses pengerjaan Tugas Akhir.
5. Bapak Danny Matthew Saputra, S.T., M.Sc selaku dosen pembimbing II pada Proposal skripsi saya, yang telah membimbing dan mengarahkan penulis dalam pengerjaan Proposal Tugas Akhir.
6. Bapak Muhammad Ali Buchari, S.Kom., M.T selaku dosen pembimbing akademik, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan serta sebagai dosen penguji II

yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.

7. Ibu Dian Palupi Rini, M.Kom., Ph.D selaku dosen penguji I yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
8. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
9. Kak Ricy serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
10. Teman – teman seperjuangan jurusan Teknik Informatika yang telah saling berbagi selama masa perkuliahan ini.
11. Sahabat-sahabatku satu almamater yang telah saling berbagi dan saling membantu selama ini.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Inderalaya, Juni 2020

Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PENGESAHAN .....	ii
HALAMAN TANDA LULUS UJIAN.....	iii
HALAMAN MOTO DAN PERSEMBAHAN .....	iv
HALAMAN PERNYATAAN .....	v
ABSTRAK.....	vi
ABSTRCT.....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR .....	xiv
DAFTAR TABEL.....	xv
DAFTAR LAMPIRAN.....	xvi
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan .....	I-1
1.2 Latar Belakang .....	I-1
1.3 Rumusan Masalah.....	I-5
1.4 Tujuan Penelitian .....	I-5
1.5 Manfaat Penelitian .....	I-5
1.6 Batasan Masalah .....	I-6
1.7 Sistematika Penulisan .....	I-6
1.8 Kesimpulan .....	I-7
BAB II KAJIAN LITERATUR .....	II-1
2.1 Pendahuluan .....	II-1
2.2 Landasan Teori.....	II-1
2.2.1 Website Phishing .....	II-1
2.2.2 Data Mining .....	II-2
2.2.3 Klasifikasi .....	II-2
2.2.4 Algoritma k-Nearest Neighbor .....	II-3
2.2.5 Perhitungan k Terbaik.....	II-4
2.2.6 Perhitungan Jarak Euclidean .....	II-5
2.2.7 Algoritma Modified k-Nearest Neighbor .....	II-6
2.2.8 k-Fold Cross Validation.....	II-8
2.2.9 Pengujian Algoritma .....	II-10
2.3 Penelitian Lain yang Relevan .....	II-12
2.4 Kesimpulan .....	II-14
BAB III METODOLOGI PENELITIAN .....	III-1
3.1 Pendahuluan .....	III-1
3.2 Unit Penelitian .....	III-1
3.3 Data .....	III-1
3.3.1 Jenis dan Sumber Data.....	III-1
3.3.2 Metode Pengumpulan Data.....	III-5

3.4 Tahapan Penelitian.....	III-5
3.4.1 Menetapkan Kerangka Kerja / Framework.....	III-6
3.4.2 Menetapkan Kriteria Pengujian .....	III-9
3.4.3 Menetapkan Format Data Pengujian.....	III-9
3.4.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian..	III-13
3.4.5 Melakukan Pengujian Penelitian.....	III-14
3.4.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan ....	III-15
3.5 Metode Pengembangan Perangkat Lunak.....	III-16
3.5.1 Konsep Rational Unified Proses (RUP) .....	III-16
3.5.2 Fase – fase RUP .....	III-17
3.6 Manajemen Proyek Penelitian .....	III-20
3.6 Kesimpulan .....	III-32
<b>BAB IV PENGEMBANGAN PERANGKAT LUNAK .....</b>	<b>IV-1</b>
4.1 Pendahuluan.....	IV-1
4.2 Rational Unified Proses (RUP).....	IV-1
4.2.1 Fase Inception .....	IV-1
4.2.1.1 Pemodelan Bisnis.....	IV-2
4.2.1.2 Kebutuhan Sistem .....	IV-5
4.2.1.3 Analisis dan Desain .....	IV-13
4.2.2 Fase Elaboration .....	IV-15
4.2.2.1 Pemodelan Bisnis.....	IV-15
4.2.2.2 Kebutuhan Sistem .....	IV-19
4.2.2.3 Analisis dan Desain .....	IV-23
4.2.3 Fase Construction .....	IV-26
4.2.3.1 Pemodelan Bisnis.....	IV-26
4.2.3.2 Kebutuhan Sistem .....	IV-28
4.2.4 Fase Transition .....	IV-29
4.2.4.1 Pemodelan Bisnis.....	IV-29
4.2.4.2 Kebutuhan Sistem .....	IV-29
4.2.4.3 Pengujian .....	IV-30
4.3 Kesimpulan .....	IV-36
<b>BAB V HASIL DAN ANALISIS PENELITIAN .....</b>	<b>V-1</b>
5.1 Pendahuluan.....	V-1
5.2 Konfigurasi Percobaan.....	V-1
5.3 Hasil Pengujian .....	V-2
5.3.1 Hasil Pengujian Menggunakan Algoritma kNN.....	V-2
5.3.2 Hasil Pengujian Menggunakan Algoritma MkNN .....	V-5
5.3.3 Hasil Pengujian Algoritma k-NN dan MkNN .....	V-8
5.4 Analisis Hasil Pengujian .....	V-9
5.4.1 Analisis Hasil Pengujian Klasifikasi kNN pada Klasifikasi Website Phishing .....	V-10
5.4.2 Analisis Hasil Pengujian Klasifikasi MkNN pada Klasifikasi website phishing.....	V-11
5.5 Kesimpulan .....	V-12



BAB VI KESIMPULAN DAN SARAN .....	VI-1
6.1 Pendahuluan .....	VI-1
6.2 Kesimpulan .....	VI-1
6.3 Saran .....	VI-2

DAFTAR PUSTAKA  
LAMPIRAN

## DAFTAR GAMBAR

	Halaman
Gambar II-1 Skema 10-fold cross validation.....	II-10
Gambar III-1 Diagram Tahapan Penelitian.....	III-6
Gambar III-2 Tahapan Pengujian Penelitian.....	III-14
Gambar III-3 Model Rational Unified Proses (RUP).....	III-16
Gambar III-4 Gantt Chart pada tahap menetapkan ruang lingkup serta unit penelitian.....	III-26
Gambar III-5 Gantt Chart menentukan dasar teori yang berkaitan dengan penelitian.....	III-26
Gambar III-6 Gantt Chart melakukan kriteria pengujian.....	III-27
Gambar III-7 Gantt Chart menentukan alat yang digunakan pada penelitian untuk fase inception.....	III-27
Gambar III-8 Gantt Chart menentukan alat yang digunakan pada penelitian untuk fase elaboration.....	III-28
Gambar III-9 Gantt Chart menentukan alat yang digunakan pada penelitian untuk fase construction.....	III-29
Gambar III-10 Gantt Chart menentukan alat yang digunakan pada penelitian untuk fase transition.....	III-30
Gambar III-11 Gantt Chart melakukan pengujian penelitian.....	III-31
Gambar III-12 Gantt Chart Pada tahap analisis hasil pengujian dan membuat kesimpulan.....	III-31
Gambar IV-1 Diagram Use Case.....	IV-3
Gambar IV-2 Diagram Sequence Melakukan Klasifikasi data menggunakan algoritma kNN.....	IV-17
Gambar IV-3 Diagram Sequence Melakukan Klasifikasi data menggunakan algoritma MkNN.....	IV-18
Gambar IV-4 Rancangan Antar Muka Perangkat Lunak.....	IV-19
Gambar IV-5 Rancangan memuat data.....	IV-20
Gambar IV-6 Rancangan pilih algoritma.....	IV-20
Gambar IV-7 Rancangan input nilai parameter k.....	IV-21
Gambar IV-8 Rancangan button Mulai.....	IV-21
Gambar IV-9 Rancangan menampilkan kinerja algoritma.....	IV-22
Gambar IV-10 Rancangan menampilkan hasil setiao k-Fold data.....	IV-22
Gambar IV-11 Diagram Aktivitas klasifikasi data menggunakan algoritma k-Nearest Neighbor.....	IV-24
Gambar IV-12 Diagram Aktivitas klasifikasi data menggunakan algoritma Modified k-Nearest Neighbor.....	IV-25
Gambar IV-13 Diagram Kelas.....	IV-27

## DAFTAR TABEL

		Halaman
Tabel II-1	Confusion Matrix.....	II-11
Tabel III-1	Deskripsi Atribut Dataset Website Phishing.....	III-2
Tabel III-2	Rancangan Hasil Pengujian <i>Confusion Matriks</i> Knn.....	III-10
Tabel III-3	Rancangan Hasil Pengujian <i>Confusion Matriks</i> MkNN.....	III-11
Tabel III-4	Rancangan Hasil Pengujian Waktu Komputasi dan Memori Algoritma k-Nearest Neighbor.....	III-11
Tabel III-5	Rancangan Hasil Pengujian Waktu Komputasi dan Memori Algoritma Modified k-Nearest Neighbor.....	III-12
Tabel III-6	Rancangan hasil Pengujian Kinerja Algoritma <i>k-Nearest Neighbor</i> dan <i>Modified k-Nearest Neighbor</i> .....	III-15
Tabel III-7	Penjadwalan Penelitian dalam Work Breakdown Structure.....	III-21
Tabel IV-1	Definisi Aktor Use Case.....	IV-4
Tabel IV-2	Definisi Use Case.....	IV-4
Tabel IV-3	Kebutuhan Fungsional.....	IV-5
Tabel IV-4	Kebutuhan Non-Fungsional.....	IV-6
Tabel IV-5	Skenario klasifikasi menggunakan algoritma kNN.....	IV-7
Tabel IV-6	Skenario klasifikasi menggunakan algoritma MkNN.....	IV-10
Tabel IV-7	Skenario evaluasi hasil klasifikasi.....	IV-13
Tabel IV-8	Implementasi Kelas-kelas.....	IV-28
Tabel IV-9	Rencana Pengujian Use Case klasifikasi dataset website phishing menggunakan algoritma kNN.....	IV-30
Tabel IV-10	Rencana Pengujian Use Case klasifikasi dataset website phishing menggunakan algoritma MkNN.....	IV-31
Tabel IV-11	Pengujian Use Case klasifikasi dataset website phishing menggunakan algoritma kNN.....	IV-32
Tabel IV-12	Pengujian Use Case klasifikasi dataset website phishing menggunakan algoritma MkNN.....	IV-35
Tabel V-1	Hasil Pengujian Nilai Akurasi, Recall dan Precision Menggunakan Algoritma kNN dengan nilai parameter k yang berbeda.....	V-3
Tabel V-2	Hasil Pengujian Waktu Komputasi dan Memori yang dibutuhkan Algoritma kNN dengan nilai parameter k yang berbeda.....	V-4
Tabel V-3	Hasil Pengujian Nilai Akurasi, Recall Dan Precision Menggunakan algoritma MkNN dengan nilai parameter k yang berbeda.....	V-6
Tabel V-4	Hasil Pengujian Waktu Komputasi Dan Memori yang Dibutuhkan Algoritma MkNN dengan nilai parameter k Yang berbeda.....	V-7
Tabel V-5	Hasil Perbandingan Kinerja Algoritma k-Nearest Neighbor dan Modified k-Nearest Neighbor.....	V-8

## DAFTAR LAMPIRAN

	Halaman
Diagram Kelas.....	L1-1
Kode Program.....	L2-1

**PERBANDINGAN KINERJA ALGORITMA *k*-NEAREST  
NEIGHBOR DAN MODIFIED *k*-NEAREST NEIGHBOR PADA  
KLASIFIKASI WEBSITE PHISHING**

**Oleh:**

**Moh. Trisno**

**09021181520037**

**ABSTRAK**

Algoritma *k*-Nearest Neighbor (kNN) merupakan algoritma klasifikasi yang dapat digunakan untuk mengklasifikasikan data website *phishing* namun pada algoritma terdapat banyak kelemahan seperti nilai *k* yang bias, komputasi kompleks, keterbatasan memori, dan mudah tertipu dengan atribut yang tidak relevan sehingga menyebabkan tingkat akurasi yang rendah. Kelemahan tersebut dapat diatasi oleh algoritma pengembangan dari kNN yaitu *Modified k-Nearest Neighbor* (MkNN) dimana MkNN dapat mengatasi masalah *outlier* pada kNN biasa. Sehingga dengan menggunakan algoritma kNN dan MkNN dapat diketahui kinerja klasifikasi website *phishing* berupa nilai akurasi, *recall*, *precision*, dan waktu komputasi serta memori yang digunakan. Untuk mengetahui kinerjanya, algoritma akan diujikan menggunakan *confusion matrix*. Dari hasil penelitian menunjukkan bahwa MkNN mempunyai pengaruh terhadap peningkatan nilai akurasi, *recall* dan *precision* dibandingkan algoritma kNN. Hasil evaluasi kinerja algoritma dengan menggunakan nilai *k* yang terbaik untuk kedua algoritma yaitu  $k = 5$ , sehingga MkNN mampu menghasilkan akurasi sebesar 95.60% dan hasil akurasi menggunakan kNN sebesar 95.35% pada klasifikasi website *phishing*. Namun peningkatan juga terjadi terhadap waktu komputasi yang diperlukan algoritma MkNN yang lebih besar dibandingkan algoritma kNN, *sedangkan* memori yang digunakan pada algoritma MkNN lebih kecil.

Kata Kunci: *k*-Nearest Neighbor, *Modified k*-Nearest Neighbor, Website *Phishing*, Klasifikasi, Akurasi, *Recall*, *Precision*

**COMPARISON OF K-NEAREST NEIGHBOR ALGORITHM  
PERFORMANCE AND MODIFIED K-NEAREST NEIGHBOR  
ALGORITHM FOR *PHISHING* WEBSITES CLASSIFICATION**

**By :**

**Moh. Trisno**

**09021181520037**

**ABSTRACT**

The k-Nearest Neighbor (kNN) algorithm is a classification algorithm that can be used to classify phishing websites, but the algorithm has many weaknesses such as the value of k bias, complex computing, memory limitations dan easily fooled by irrelevant attributes causes a low level off accuracy. This problem can be overcome by the development of kNN algorithm, namely Modified k- Nearest Neighbor (MkNN) where MkNN can overcome the problem of outliers in traditional kNN. So by using the kNN and MkNN Algorithms, it can be find the performance in the form of accuracy, recall, precision, the computing time and memory used. To find out its performance, the algorithm will be using confusion matrix. The result shows that MkNN had an effect on increasing the value of accuracy, recall and precision compared to the kNN algorithm. The results of the evaluation of the performance of the algorithm using the best k value for both algorithms is (k = 5) so that MkNN is able to produce an accuracy of 95.60 % and the accuracy result using KNN of 95.35 % on the classification of phishing websites. But the inscrease also occurs in the computational time required by MkNN algorithm, which is greater than the KNN algorithm, whereas the memory used in the MkNN algorithm is less.

Keywords : k-Nearest Neighbor, Modified k-Nearest Neighbor, Phishing Websites, Classification, Accuracy, Recall, Precision.

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Pendahuuan**

Pada bab ini membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian serta batasan masalah. Bab ini akan memberikan penjelasan umum mengenai keseluruhan penelitian. Pendahuluan dimulai dengan penjelasan mengenai tantangan dan tujuan proses implementasi data mining untuk melakukan klaifikasi data. Serta penelitian yang berkaitan dengan data mining pada klasifikasi data yang menjadi latar belakang dari penelitian ini.

#### **1.2 Latar Belakang Masalah**

Internet telah menjadi bagain penting dalam kehidupan masyarakat karena dapat mempermudah kegiatan ekonomi maupun sosial. Setiap tahunnya jumlah pengguna internet terus meningkat diiringi dengan jumlah populasi manusia. Berdasarkan data yang di *publish* oleh [internetworldstats.com](http://internetworldstats.com) menyatakan bahawa jumlah pengguna internet hingga akhir tahun 2019 mencapai 4.574 miliar pengguna atau 58.7% dari jumlah populasi penduduk.

*Phishing* merupakan kejahatan Internet yang relatif umum ditemukan. Masalah *phishing* adalah masalah yang sulit, karena faktanya bahwa sangat mudah bagi penyerang untuk membuat replika *website* yang sangat menyerupai seperti website perbankan yang terlihat bagus dan sangat meyakinkan bagi nasabah. APWG (*Anti-Phishing Working Group*) menyatakan bahwa kesadaran masyarakat terhadap

*website phishing* meningkat pertahunnya, akan tetapi jumlah *website phishing* tumbuh lebih cepat dan mengakibatkan kerugian yang ditimbulkan dari *website phishing*. Pada laporan APWG kuartal empat tahun 2019, *phishing activity trend* pada bulan desember terdapat 45.771 *website* yang terdeteksi sebagai *website phishing*, sedangkan Jumlah laporan e-mail *phishing* diterima oleh APWG dari konsumen sendiri dari bulan oktober sampai desember mengalami peningkatan hingga pada bulan desember berjumlah 45.072 laporan. APWG menyatakan bahwa bahwa situs SaaS dan webmail tetap ada target utama. (James, 2006) *Website phishing* diciptakan untuk memperoleh data pelanggan, informasi, atau hak akses. Oleh karena itu perlukan dilakukan pengelompokan dengan menggunakan teknik klasifikasi untuk mengelompokkan sebuah *website* apakah termasuk dalam kelompok *website* yang di kategorikan sebagai *website phishing* atau bukan.

Klasifikasi dalam Data Mining bersifat *supervised learning* dan sering diimplementasikan pada bidang statistik, pengenalan pola, *machine learning* dan lain-lain. Salah satu metode klasifikasi adalah *k-Nearest Neighbor* yang diperkenalkan oleh Fix dan Hog yang digunakan untuk mengklasifikasikan suatu objek baru berdasarkan kedekatan jarak suatu data dengan data yang lain. Prinsip kerja dari algoritma *k-Nearest Neighbor* adalah mencari jarak antara dua titik yaitu titik *training* dan titik *testing* (Yusara, 2016).

(Purwiantono & Tjahyanto, 2017) menerapkan model klasifikasi untuk mendeteksi situs *phishing* yang ada di Indonesia secara akurat dan presisi menggunakan 4 algoritma klasifikasi yaitu *Sequential Minimal Optimization*, *Naive Bayes*, *Bagging* dan *Multilayer Perceptron*. Disimpulkan bahwa *Multilayer*



*Perceptron* lebih unggul karena nilai akurasi, *Precision*, *Recall*, *F-Measure* dan *Time*. Algoritma *k-Nearest Neighbor* juga dapat diimplementasikan untuk mengklasifikasikan email spam, (Pratiwi & Ulama, 2016) membandingkan kinerja algoritma *Support Vector Machine* dan *k-Nearest Neighbor* untuk mengklasifikasi email spam dan disimpulkan bahwa algoritma *Support Vector Machine* lebih baik dibanding metode *k-Nearest Neighbor*.

Penelitian (Yusa, Utami, & Luthfi, 2016) melakukan evaluasi terhadap model algoritma klasifikasi yaitu *Decision Tree*, *k-Nearest Neighbor* (k-NN), dan *Naïve Bayes* yang didapatkan hasil bahwa *k-Nearest Neighbor* berada dibawah algoritma *Naïve Bayes* yang memiliki nilai akurasi, *error rate* dan kappabilitas statistic terbaik diantara model algoritma klasifikasi lainnya. Perdasarkan hasil penelitian tersebut mendukung pernyataan (Bhatia, 2010) yang menyatakan bahwa Algoritma *k-Nearest Neighbor* memiliki kelemahan yaitu Nilai *k* bias, komputasi kompleks, keterbatasan memori, dan mudah tertipu dengan atribut yang tidak relevan yang menyebabkan tingkat akurasinya rendah.

Dari kelemahan algoritma *k-Nearest Neighbor* banyak penelitian yang dilakukan bertujuan untuk memperbaiki dari kelemahan algoritma *k-Nearest Neighbor* dalam pengklasifikasian data. Salah satu modifikasi algoritma *k-Nearest Neighbor* yang telah dikembangkan yaitu algoritma *Modified k-Nearest Neighbor*. Algoritma *Modified k-Nearest Neighbor* dikembangkan oleh (Parvin, Alizadeh, & Minaei-Bidgoli, 2010) dengan tujuan mengatasi masalah pada tingkat akurasi algoritma *k-Nearest Neighbor* yang rendah dan meningkatkan performa algoritma *k-Nearest Neighbor*. Pada algoritma *Modified k-Nearest Neighbor* ditambahkan

proses perhitungan nilai validitas yang bertujuan untuk mengatasi permasalahan outlier dan perhitungan *weighted voting* pada semua data uji menggunakan validitas data. Dengan penambahan 2 proses pada algoritma *Modified k-Nearest Neighbor* diklaim dapat memperbaiki kelemahan dari algoritma *k-Nearest Neighbor*.

Penelitian menggunakan algoritma *Modified k- Nearest Neighbor* pernah dilakukan (Wafiyah, Hidayat, & Perdana, 2017) untuk mengklasifikasi penyakit demam. Untuk mengetahui kinerja dari algoritma (Gazalba & Reza, 2017) melakukan analisis perbandingan dari kinerja algoritma *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor* yang menghasilkan bahwa algoritma *Modified k-Nearest Neighbor* memiliki keunggulan serta nilai akurasi yang lebih baik dari pada *k-Nearest Neighbor* tradisional.

Berdasarkan uraian diatas diketahui bahwa algoritma klasifikasi *k- Nearest Neighbor* dan *Modified k- Nearest Neighbor* baik digunakan untuk proses pengklasifikasian data. Berdasarkan dari pernyataan tersebut, pada penelitian ini akan dilakukan penelitian terhadap perbandingan kinerja dari algoritma *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor* pada pengklasifikasian data menggunakan dataset *website phishing*.

### **1.3 Rumusan Masalah**

Rumusan masalah pada penelitian ini adalah apakah algoritma *Modified k-Nearest Neighbor* kinerjanya lebih baik dibandingkan algoritma *k-Nearest Neighbor* pada klasifikasi *website phishing* ? Untuk menjawab rumusan masalah tersebut, diuraikan beberapa pertanyaan penelitian sebagai berikut:

1. Bagaimana cara mengimplementasikan algoritma *k- Nearest Neighbor* dan *Modified k- Nearest Neighbor* untuk klasifikasi *website phishing*?
2. Bagaimana hasil pengujian dari kinerja algoritma *k- Nearest Neighbor* dan *Modified k- Nearest Neighbor* untuk klasifikasi *website phishing*?

#### **1.4 Tujuan Penelitian**

Tujuan dari penelitian yang dilakukan :

1. Mengetahui apakah algoritma *k- Nearest Neighbor* dan *Modified k- Nearest Neighbor* dapat diimplementasikan untuk klasifikasi *website phishing*.
2. Mengetahui bagaiman kinerja algoritma *k- Nearest Neighbor* dan *Modified k- Nearest Neighbor* untuk klasifikasi *website phishing*.

#### **1.5 Manfaat penelitian**

Manfaat yang di dapatkan dari penelitian yang dilakukan:

1. Memahami algoritma *k-Nearest Neighbor* dan *Modified k- Nearest Neighbor* untuk diimplementasikan pada klasifikasi *website phishing*
2. Dapat Mengetahui kinerja algoritma *Modified k- Nearest Neighbor* apakah lebih baik dibandingkan algoritma *k- Nearest Neighbor* untuk klasifikasi *website phishing*
3. Mengetahui nilai *accuracy*, *recall* dan *precision* serta mengetahui waktu komputasi dan besarnya memori yang dibutuhkan algoritma *k- Nearest Neighbor* dan *Modified k- Nearest Neighbor*.

## 1.6 Batasan Masalah

Batasan masalah dari penelitian yang dilakukan :

1. Menggunakan *euclidean distance* sebagai metode perhitungan jarak.
2. Dataset yang digunakan yaitu dataset *website phishing* yang diperoleh dari repositori publik dengan 30 atribut, dan klasifikasi hanya dikelompokkan dalam 2 kelompok yaitu *phishing*, atau tidak *phishing*.
3. Evaluasi kinerja dari algoritma akan dilakukan dengan menggunakan *confusion matrix*, serta mengetahui besarnya waktu komputasi yang dibutuhkan dan memori yang dibutuhkan. Kemudian hasil klasifikasi akan dibandingkan antara algoritma *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor*.

## 1.7 Sistematika Penulisan

Dalam penulisan tugas akhir ini, sistematika penulisan adalah sebagai berikut:

### **BAB I. PENDAHULUAN**

Bab ini akan dibahas latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, sistematika penulisan dan kesimpulan.

### **BAB II. KAJIAN LITERATUR**

Bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian, seperti hasil penelitian terdahulu yang relevan dengan penelitian ini, algoritma *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor*.

### **BAB III. METODOLOGI PENELITIAN**

Bab ini akan dibahas tahapan-tahapan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada

suatu kerangka kerja. Pada akhir bab akan berisi perancangan manajemen proyek pada pelaksanaan penelitian

#### **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Bab ini akan dibahas perancangan perangkat lunak yang akan dibangun pada penelitian ini.

#### **BAB V. HASIL DAN ANALISIS PENELITIAN**

Bab ini akan dibahas hasil pengujian berdasarkan tahapan yang telah direncanakan. Serta analisis diberikan sebagai basis dari kesimpulan yang diambil dalam penelitian ini.

#### **BAB VI. KESIMPULAN DAN SARAN**

Bab ini akan dibahas kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam pengembangan perangkat lunak ini selanjutnya.

### **1.8 Kesimpulan**

Beberapa penelitian mengatakan bahwa algoritma *k-Nearest Neighbor* memiliki tingkat akurasi yang rendah. Salah satu solusinya yaitu algoritma *Modified k-Nearest Neighbor* yang pada beberapa penelitian dinyatakan memiliki tingkat akurasi yang lebih baik dibandingkan *k-Nearest Neighbor*, Oleh karena itu, pada penelitian ini akan membahas apakah algoritma *Modified k-Nearest Neighbor* diterapkan untuk mengklasifikasikan *website phishing* terbukti kinerjanya lebih baik dari algoritma *k-Nearest Neighbor* berdasarkan nilai *accuracy*, *recall*, *precision*, waktu komputasi dan memori yang dibutuhkan.

## DAFTAR PUSTAKA

- Bhatia, N. (2010). Survey of nearest neighbor techniques. *arXiv preprint arXiv:1007.0085*.
- Eyupoglu, C. (2018). Breast cancer classification using k-nearest neighbors algorithm. *The Online Journal of Science and Technology-July*, 8(3).
- Frank, E., & Hall, M. A. (2011). *Data mining: practical machine learning tools and techniques*: Morgan Kaufmann.
- Gazalba, I., & Reza, N. G. I. (2017). *Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification*. Paper presented at the Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2017 2nd International conferences on.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*: Elsevier.
- Hassanat, A. B., Abbadi, M. A., Altarawneh, G. A., & Alhasanat, A. A. (2014). Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach. *International Journal of Computer Science and Information Security*, 12.
- Islam, M. J., Wu, Q. J., Ahmadi, M., & Sid-Ahmed, M. A. (2007). *Investigating the performance of naive-bayes classifiers and k-nearest neighbor classifiers*. Paper presented at the 2007 International Conference on Convergence Information Technology (ICCIT 2007).
- James, L. (2006). Phishing exposed. Tech target article sponsored by: Sunbelt software. *available at: searchexchange.com*.
- Karegowda, A. G., Jayaram, M., & Manjunath, A. (2012). Cascading k-means clustering and k-nearest neighbor classifier for categorization of diabetic patients.
- Mohammad, R. M., Thabtah, F., & McCluskey, L. (2015). Phishing websites features *School of Computing and Engineering*: University of Huddersfield.
- Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques*: Springer Science & Business Media.
- Parvin, H., Alizadeh, H., & Minaei-Bidgoli, B. (2010). *MKNN: Modified k-nearest neighbor*. *Global Journal of Computer Science and Technology*.

- Pratiwi, S. N. D., & Ulama, B. S. S. (2016). Klasifikasi Email Spam dengan Menggunakan Metode Support Vector Machine dan k-Nearest Neighbor. *Jurnal Sains dan Seni ITS*, 5(2).
- Purwiantono, F. E., & Tjahyanto, A. (2017). Model Klasifikasi Untuk Deteksi Situs Phising Di Indonesia: no.
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. *Encyclopedia of database systems*, 532-538.
- Wafiyah, F., Hidayat, N., & Perdana, R. S. (2017). Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Demam. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- Yusa, M., Utami, E., & Luthfi, E. T. (2016). Analisis Komparatif Evaluasi Performa Algoritma Klasifikasi pada Readmisi Pasien Diabetes. *Jurnal Buana Informatika*, 7(4).
- Yusra, Y. (2016). Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor. *Jurnal Sains dan Teknologi Industri*, 14(1), 79-85.
- Zaki, M. J., Meira Jr, W., & Meira, W. (2014). *Data mining and analysis: fundamental concepts and algorithms*: Cambridge University Press.