

## ANALISIS METODE K-NEAREST NEIGHBOR (KNN) DALAM KLASIFIKASI SISWA TERBAIK KELAS III-IPA PADA SMP NEGERI 3SATAP SELESAI

Suhendra<sup>1)</sup>, Juliana Naftali Sitompul,<sup>2)</sup> Magdalena Simanjuntak<sup>3)</sup>

<sup>123)</sup>STMIK KAPUTAMA

Jl.Veteran No.4A-9A, Binjai, Sumatra Utara, Telp:(061)8828840, Fax:  
(061)8828845

Email: [hendrasoe66@gmail.com](mailto:hendrasoe66@gmail.com)

### ABSTRACT

*In conducting the data classification process, the best students refer to the artificial intelligence method. as a classification process including K-Nearest Neighbor (K-NN). In responding to this, the research is compiling and trying to help SMP Negeri 3 Satap in class III-IPA to implement a system that is able to work quickly, precisely and efficiently and so as to take a decision for the classification of students in class 3 science. valid and it can really be said that the student belongs to the classification of achievement students. Classification is one of the substance of Data Mining which is often used to group data based on certain categories or variables. Among the classification methods that are often used to classify data in the form of documents are Naïve Bayes and K-Nearest Neighbor (KNN). This study discusses in detail the comparison of the two methods by measuring the accuracy and length of processing time with 4 trial scenarios. From the results of the study, Naïve Bayes outperformed KNN with an average accuracy of 76.81% Naïve Bayes and KNN K 3, 6, 9 respectively 72.47%, 71.12%, 71.56%. But if it only takes the highest accuracy, KNN is superior with the Naïve Bayes ratio of 72.22% with KNN of 77.78%. With notes, the difference can still increase directly proportional to the amount of treasury training documents. The average processing time required by Naïve Bayes is 0.31 seconds compared to KNN as much as 1.09 seconds.*

**Keywords:** *Classification, training\_data, test\_data.*

### ABSTRAK

Dalam melakukan proses klasifikasi data siswa terbaik mengacu pada metode kecerdasan buatan. sebagai proses klasifikasi diantaranya K-Nearest Neighbor (K-NN). Dalam menyikapi hal tersebut pada penelitian menyusun dan berusaha untuk membantu SMP negeri 3 Satap pada kelas III-IPA untuk menerapkan sistem yang mampu bekerja secara cepat, tepat dan efisien dan sehingga untuk mengambil sebuah keputusan untuk klasifikasi siswa pada kelas 3 IPA hasil yang di keluarkan tersebut valid dan benar – benar bisa dikatakan bahwa siswa tersebut termasuk klasifikasi siswa perprestasi. Klasifikasi merupakan salah satu substansi Data Mining yang sering digunakan untuk mengelompokkan data berdasarkan kategori atau variabel tertentu. Diantara metode klasifikasi yang sering digunakan untuk mengklasifikasikan data berupa dokumen adalah Naïve Bayes dan K-Nearest Neighbor (KNN). Penelitian ini membahas secara rinci perbandingan kedua metode tersebut dengan mengukur akurasi dan lama waktu proses dengan 4 skenario uji coba. Dari hasil penelitian, Naïve Bayes mengungguli KNN dengan rata-rata hasil akurasi Naïve Bayes sebesar 76.81% dan KNN bernilai K 3, 6, 9 berturut-turut sebesar 72,47%, 71,12%, 71,56%. Namun jika hanya mengambil akurasi tertinggi, KNN unggul dengan perbandingan Naïve Bayes sebesar 72.22% dengan KNN sebesar 77.78%. Dengan catatan, selisih masih bisa bertambah berbanding lurus dengan jumlah perbendaharaan dokumen latih. Rata-rata waktu proses yang dibutuhkan Naïve Bayes sebanyak 0.31 detik berbanding KNN sebanyak 1.09 detik.

**Kata Kunci:** Klasifikasi,data\_latih,data\_uji.

## 1. PENDAHULUAN

Dalam penelitian ini akan dilakukan proses penentuan siswa terbaik kelas 3 IPA pada SMP negeri 3 Satap Selesai, Pendidikan mempunyai suatu peranan yang sangat penting untuk kemajuan bangsa Indonesia, Sekolah sebagai institusi pendidikan, yang mengembangkan berbagai sistem pembinaan yang sifatnya mengembangkan motivasi siswa. Namun pada proses menentukan siswa berprestasi pada SMP negeri 3 Selesai siswa terbaik kelas 3 IPA masih bersifat manual sehingga untuk penentuan siswa terbaik pada SMP negeri 3 Satap Selesai masih kurang efisien.

Dalam melakukan proses klasifikasi data siswa terbaik mengacu pada metode kecerdasan buatan. sebagai proses klasifikasi diantaranya K-Nearest Neighbor (K-NN). Dalam menyikapi hal tersebut pada penelitian menyusun dan berusaha untuk membantu SMP negeri 3 Satap pada kelas 3 IPA untuk menerapkan sistem yang mampu bekerja secara cepat, tepat dan efisien dan sehingga untuk mengambil sebuah keputusan untuk klasifikasi siswa pada kelas 3 IPA hasil yang di keluarkan tersebut valid dan benar – benar bisa dikatakan bahwa siswa tersebut termasuk klasifikasi siswa berprestasi.

Klasifikasi adalah cara yang dilakukan sebagai teknik untuk membentuk model klasifikasi dari contoh data pelatihan. Klasifikasi akan menganalisis input data dan membentuk model dengan menggambarkan kelas data. Istilah ini menunjuk kepada sebuah metode untuk menyusun data secara sistematis atau menurut beberapa aturan atau kaidah yang telah ditetapkan (Mulak & Talhar, 2015).

Salah satu algoritma yang bisa digunakan dalam klasifikasi adalah algoritma *K-Nearest Neighbour*. Tujuan algoritma K-NN adalah mengklasifikasikan obyek baru berdasarkan atribut dan training *sample* (Maulana, 2017).

K-Nearest Neighbor (k-NN) adalah suatu metode klasifikasi klasik yang tidak memerlukan pengetahuan sebelumnya, label

sampel baru hanya ditentukan oleh tetangga terdekatnya

(Pan *et al.*, 2017).

K-Nearest Neighbor (KNN) metode ini merupakan sebuah metode yang menggunakan algoritma supervised dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN yang tujuannya untuk mengklasifikasikan data uji berdasarkan data latih. Sedangkan untuk mengatasi masalah perangkingan, di dalam penelitian ini digunakan metode Wighted Product (WP), karena Wighted Product merupakan salah satu metode penyelesaian multi kriteria [Diah, 2013],

dimana dalam penentuan siswa berprestasi memiliki banyak kriteria yang harus dipertimbangkan. Dalam pelaksanaannya, penentuan siswa berprestasi tersebut menggunakan beberapa kriteria yang akan digunakan dalam penilaian siswa berprestasi, setiap kriteria memiliki bobot atau nilai yang berbeda.

Berdasarkan penelitian sebelumnya yang dilakukan oleh Ari Sulistyio menggunakan metode K-Nearest Neighbor untuk penentuan jurusan di Sekolah Menengah Atas menggunakan metode K-Nearest Neighbor mempunyai keakuratan sebesar 79,68 % dengan menggunakan 7 kriteria. [Sulistyio, 2014] Sehingga pada penelitian ini akan dibuat sebuah penelitian untuk membantu guru dalam menentukan siswa berprestasi dengan menggunakan metode K-Nearest Neighbor dan Wighted Product untuk di implementasikan di SMP NEGERI 3 SATAP Selesai.

Penerapan metode *K-Nearest Neighbor* (KNN) telah banyak dilakukan oleh para peneliti sebelumnya tentang klasifikasi untuk tingkat kematangan buah tomat berdasarkan warna menggunakan *K-Nearest Neighbor* (KNN) oleh Trisnaningtyas, dkk. Dalam penelitiannya menyebutkan bahwa metode *K-Nearest Neighbor* (KNN) merupakan metode klasifikasi data bekerja berdasarkan pada jarak terdekat dari data training ke data testing dengan pemeriksaan jarak *Cityblock Distance* dan *Euclidean Distance*.

Dengan adanya masalah tersebut serta ada solusi untuk mengatasi keadaan seperti itu, maka penulis

tertarik untuk membuat si penulis melakukan sebuah klasifikasi data yang berjudul : “ Analisis Metode K-Nearest Neighbor (KNN) Dalam Klasifikasi siswa terbaik kelas III-IPA Pada SMP Negeri 3 Satap Selesai.”

#### Rumusan Masalah

Berdasarkan latar belakang tersebut, maka permasalahan yang muncul diantaranya :

Dengan inputan data siswa kelas III-IPA sebagai data latih dan target, bagaimana cara klasifikasi siswa berprestasi dengan menggunakan metode K-Nearest Neighbor (KNN)?

Dengan menggunakan metode KNN, apakah nilai K akan mempengaruhi penentuan Klasifikasi mahasiswa berprestasi?

#### Batasan Masalah

Berdasarkan perumusan masalah, maka dapat ditentukan Batasan masalah diantaranya sebagai berikut:

Pengujian metode dilakukan pada klasifikasi data siswa untuk menentukan data training dan data uji terhadap pengaruh jumlah K dalam metode K-Nearest Neighbor.

Klasifikasi data siswa digunakan dalam penelitian untuk penentuan siswa berprestasi.

#### Tujuan Penelitian

Adapun tujuan penelitian adalah sebagai berikut:

Membangun sebuah aplikasi klasifikasi siswa berprestasi di sekolah SMP negeri 3 Satap Selesai pada kelas 3 IPA dengan menggunakan metode *K-Nearest Neighbor* (KNN).

Untuk mengetahui proses dan menguji tingkat akurasi dari metode *K-Nearest Neighbor* dalam mengambil sebuah keputusan yang mampu mengidentifikasi siswa berprestasi.

#### Manfaat Penelitian

Manfaat yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

Membantu mengidentifikasi klasifikasi siswa berprestasi pada kelas 3 IPA di SMP negeri 3 Satap Selesai menggunakan system.

Manfaat dibidang ilmu diharapkan dapat menambah literatur mengenai bidang ilmu data mining terkhususnya metode KNN.

## 2. KLASIFIKASI DATA

Klasifikasi data adalah proses pengurutan dan pengelompokan data ke dalam berbagai jenis, bentuk atau kelas berbeda lainnya. Klasifikasi data memungkinkan pemisahan dan klasifikasi data sesuai dengan persyaratan kumpulan data untuk berbagai tujuan bisnis atau pribadi. Ini terutama merupakan proses manajemen data.

Analisis kelompok sebagai suatu metode untuk melakukan klasifikasi data menjadi beberapa kelompok dengan menggunakan metode pengukuran ukuran asosiasi, sehingga data yang sama berada dalam satu kelompok dan data yang memiliki perbedaan yang besar diletakkan dalam kelompok data lainnya. Masukan untuk sistem analisis kelompok adalah sebuah data set dan kesamaan ukuran antara kedua data tersebut. Sedangkan hasil dari analisis kelompok adalah sejumlah kelompok yang membentuk sebuah partisi atau struktur partisi dari kumpulan data dan deskripsi umum dari setiap kelompok, dimana hal ini sangat penting untuk analisis yang lebih dalam pada karakteristik yang terdapat pada data tersebut.

Pengelompokan data harus menggunakan sebuah pendekatan untuk mencari kesamaan dalam data sehingga mampu menempatkan data ke dalam kelompok-kelompok yang tepat. Pengelompokan data akan membagi kumpulan data ke dalam beberapa kelompok dimana kesamaan dalam sebuah kelompok lebih besar jika dibandingkan dengan kelompok-kelompok lainnya.

Terdapat dua metode pembelajaran yang tersedia pada model klasifikasi yaitu :

- a. Eager learning adalah proses pembelajaran pada data latih secara intensif sehingga model dapat melakukan prediksi ke label kelas output dengan benar. Beberapa metode bersifat eager learning antara lain, Neural Network, Bayesian,

decision tree, Support Vector Machine.

- b. Lazy learning adalah proses pembelajaran tanpa melakukan pelatihan dan hanya menyimpan nilai data latih untuk digunakan pada proses prediksi. Beberapa metode bersifat lazy learning antara lain: K-Nearest Neighbor, Regresi Linear, Fuzzy K-Nearest Neighbor.

Proses klasifikasi pada machine learning memiliki empat komponen (Saputra, 2018) yaitu :

1. Class  
Variabel dependen tetamu harus dalam bentuk yang mewakili label yang dimiliki oleh objek.
2. Energi  
Variabel independen yang diwakili oleh atribut karakteristik data. Misalnya, gaji, kehadiran, merokok, tekanan darah.
3. Dataset Pelatihan  
Satu set data yang memiliki kedua nilai komponen diatas yang digunakan untuk menentukan kelas yang sesuai berdasarkan energi.
4. Menguji dataset  
Satu set data baru yang akan diklasifikasi dengan model yang telah dibuat dan akan dievaluasi pada proses akurasi klasifikasi.

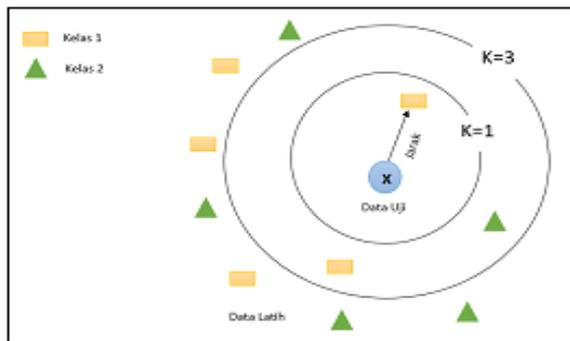
Pada proses klasifikasi, sebelum melakukan prediksi, perlu dilakukan proses pembelajaran terlebih dahulu. Proses pembelajaran tersebut memerlukan data. Data yang diperlukan pada saat proses klasifikasi terdiri atas dua jenis, yaitu :

- a. Data latih atau data training adalah data yang digunakan pada proses pembelajar dalam proses klasifikasi.
- b. Data uji atau data testing adalah data yang digunakan pada proses prediksi dalam proses klasifikasi.

### 3. K-NEAREST NEIGHBOUR

Klasifikasi KNN adalah metode non-parametrik sederhana untuk klasifikasi. Terlepas dari kesederhanaan algoritme, kinerjanya sangat baik, dan merupakan metode tolok ukur yang penting. Klasifikasi KNN membutuhkan metrik dan integer positif  $K$  [42]. Aturan KNN memegang posisi sampel pelatihan dan kelas mereka. Saat memutuskan tentang data masuk baru. Tujuan algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan nilai atribut dan data latih (Okfalisa et al, 2017).

Cover dan Hart memperkenalkan K-Nearest Neighbor pada tahun 1968. K-Nearest Neighbor adalah metode klasifikasi pemalas karena algoritma ini menyimpan semua nilai data pelatihan dan menunda proses pembentukan model klasifikasi hingga data uji diberikan untuk prediksi. (Mulak & Talhar, 2015). Gagasan dalam metode k-Nearest Neighbor adalah untuk mengidentifikasi sampel  $k$  dalam set pelatihan yang variabel independen  $x$  mirip dengan  $u$ , dan menggunakan sampel  $k$  ini untuk mengklasifikasikan sampel baru ini ke dalam kelas,  $v$ .  $F$  adalah fungsi yang halus, sebuah ide yang masuk akal adalah mencari sampel dalam data pelatihan kami yang berada di dekatnya (dalam hal variabel independen) dan kemudian menghitung  $v$  dari nilai  $y$  untuk sampel ini. Jarak atau ukuran ketidaksamaan dapat dihitung antara sampel dengan mengukur jarak menggunakan jarak Euclidean. Jarak Euclidean antara titik-titik adalah (Hu et al, 2016).



Pada gambar dapat dilihat bahwa terlebih dahulu dilakukan perhitungan jarak data uji dengan data latih untuk selanjutnya dilakukan diidentifikasi tetangga terdekat. Selanjutnya KNN akan memberikan titik ke kelas yang memiliki jarak terdekat dengan tetangganya. Pada gambar diatas data uji memiliki matrik jarak terdekat dengan kelas 1 sehingga data uji akan diklasifikasikan sebagai data kelas 1. Adapun langkah-langkah klasifikasi algoritma KNN adalah sebagai berikut:

1. Tentukan parameter nilai  $k$  = banyaknya jumlah tetangga terdekat
2. Hitung jarak antara data baru dengan semua data training
3. Urutkan jarak dan tetapkan tetangga terdekat berdasarkan jarak minimum ke- $k$
4. Periksa kelas dari tetangga terdekat
5. Gunakan mayoritas sederhana dari kelas tetangga terdekat sebagai nilai prediksi data baru

#### 4. METODOLOGI PENELITIAN

Perlu dilakukan pengujian dalam penentuan pembagian data training dan data uji terhadap pengaruh jumlah  $K$  dalam K-Nearest Neighbor pada suatu dataset berukuran besar. Dataset yang ada menggunakan UCI Machine Learning Repository. UCI Machine Learning Repository adalah sebuah koleksi database, domain teori, dan data generator yang digunakan oleh komunitas yang mempelajari mesin pembelajaran (machine learning), untuk keperluan analisis empiris

dari algoritma machine learning. Dataset yang tersedia pada UCI Machine Learning Repository digunakan oleh pelajar, pendidik, dan peneliti diseluruh dunia sebagai sumber utama dari data set pada machine learning. Jumlah data set yang tersedia pada UCI Machine Learning Repository pada saat ini sudah berjumlah 320 data set yang dapat digunakan sesuai dengan kebutuhan pada pembelajaran machine learning.

Machine learning memungkinkan manusia untuk memprogram komputer sehingga mesin dapat mengenali pola atau belajar dari apa yang dimasukkan ke dalamnya. Konsep pembelajaran adalah proses mengubah informasi menjadi pengetahuan. Algoritma machine learning adalah data pelatihan, yang mewakili pengalaman, dan hasilnya adalah beberapa keahlian, yang biasanya berupa program komputer lain yang dapat melakukan beberapa tugas (Ion C. Freeman, Ashley J. et al, 2018).

Machine learning memungkinkan dalam klasifikasi data, aplikasi ini mengenali pola dalam data baik dengan pelatihan atau tanpa pelatihan. Di klasifikasi data disebut clustering dalam machine learning. Beberapa contoh algoritma pengelompokan termasuk K-Means, Farthest-First Maksimalisasi-Ekspektasi (EM), dan lainnya (Ion C. Freeman, Ashley J. et al, 2018).

Machine Learning merupakan cabang ilmu artificial intelligence yang memiliki konsep bahwa komputer sebagai mesin memiliki kemampuan untuk melakukan adaptasi terhadap lingkungan yang baru dan mampu mendeteksi pola dari fakta yang ada. Definisi pembelajaran yang dilakukan mesin adalah ketika mesin dari pengalaman  $E$  terhadap tugas  $T$  dan mengukur peningkatan kinerja  $P$ , jika kinerja tugas  $T$  diukur oleh kinerja  $P$ , meningkatkan pengalaman  $E$ .

Proses penelitian ini terdapat beberapa

kegiatan, yaitu kegiatan-kegiatan yang terdapat pada penelitian. yaitu observasi lapangan pengumpulan data dan analisis data.

1. Observasi yang dilakukan pada penelitian ini adalah hal yang paling penting. Karena penulis dapat mengetahui tingkat visibilitas yang digunakan. Data-data yang telah dikumpulkan menjadi titik pantauan dalam observasi ini sehingga mendapat hasil yang digunakan.
2. Pengumpulan data literature, jurnal, paper, dan bacaan-bacaan lainnya yang berhubungan dengan algoritma klasifikasi data mining. Peneliti melakukan pengumpulan data dengan mengambil dataset UCI Machine Learning Repository. Peneliti menggunakan dataset iris.
3. Persiapan dan pemilihan data yang didapat seperti melihat struktur tabel yang ada pada database. Pemilihan data dilakukan karena tidak semua tabel serta data yang ada dalam database berhubungan dengan penelitian yang dilakukan, sehingga hanya data yang berkaitan dengan penelitian yang digunakan.
4. Pembersihan data dilakukan untuk memastikan bahwa tidak ada data yang terduplikasi, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Data yang telah bersih dari kesalahan dapat mempermudah penelitian dan mencegah adanya kesalahan pada penelitian.
5. Pada tahap ini menganalisis terhadap pembagian data training dan data uji dengan pengaruh jumlah K algoritma K-Nearest Neighbor.

Pada penelitian ini, untuk mengetahui kinerja dari metode yang digunakan maka akan digunakan sejumlah dataset. Adapun data yang digunakan terdiri dari 2 dataset yang berasal dari UCI *Machine Learning*

*Repository* diantaranya adalah data *ionosphere*, iris dan wine.

Iris merupakan salah satu dataset yang populer, dataset ini mungkin merupakan dataset yang paling terkenal dan dapat ditemukan dalam literatur pengenalan pola yang telah disusun oleh Fisher (1936). Dataset ini terdiri dari 150 *records* dengan 4 *attributes* dan 3 kelas, dimana setiap kelas data terdiri dari 50 *records* (Marshall. 1988).

Kumpulan data minat Iris atau kumpulan data Iris Fisher adalah kumpulan data multivarian yang diperkenalkan oleh ahli statistik dan biolog Inggris Ronald Fisher pada tahun 1936. Penggunaan beberapa pengukuran dalam masalah taksonomi misalnya analisis diskriminan linier. Disebut juga data Iris Anderson diatur karena Edgar Anderson mengumpulkan data untuk mengukur variasi morfologi bunga Iris dari tiga spesies terkait. Dua dari tiga spesies dikumpulkan semua dari padang rumput yang sama, dan mengambil pada hari yang sama dan diukur pada saat yang sama oleh orang yang sama dengan peralatan yang sama, dan dipetik pada hari yang sama dan diukur pada waktu yang sama oleh orang yang sama dengan peralatan yang sama.

### 3. RANCANGAN PENELITIAN

Penelitian ini menganalisis terhadap hasil studi literatur untuk mengetahui dan mendapatkan pemahaman mengenai metode yang penulis analisis yaitu pembagian data training dan data uji dengan pengaruh jumlah K dalam K-nearest Neighbor. Untuk lebih jelas dalam mendeskripsikan proses dalam penelitian ini maka akan dijelaskan tahapan demi tahapan dalam sub bab ini. Adapun tahapan-tahapannya secara garis besar dapat dilihat pada Gambar.

#### Gambar.1 Flowchart Rancangan Penelitian



dalam mengetahui klasifikasi siswa terbaik disekolah.

**Keterangan :**

Adapun dari gambar dapat dilihat flowchart rancangan penelitian dalam penyelesaian algoritma K-NN dijelaskan sebagai berikut :

Langkah pertama dalam melakukan analisis penelitian dilakukan penentuan dataset, adapun disini penulis menggunakan dataset Uci Machine Learning. Selanjutnya menentukan nilai K dan hitung jarak euclidean dari setiap data, setelah didapat hasil jarak setiap data maka diurutkan berdasarkan ascending. Setelah terurut maka dilakukan penentuan jarak terdekat sehingga di kelompokkan dan berpasangan. Maka selanjutnya cari kelas tetangga terbanyak sehingga mendapatkan hasil klasifikasi.

**4. HASIL DAN PEMBAHASAN**

Dalam perancangan aplikasi klasifikasi siswa terbaik kelas 9 pada SMP negeri 3 Satap Selesai ini, penulis menggunakan sistem aplikasi yang berbasis aplikasi desktop dengan visual basic sebagai bahasa pemrogramannya. Program yang penulis buat cukup sederhana dan mudah untuk digunakan karena admin hanya perlu mengklik tombol-tombol yang sudah tersedia sesuai dengan struktur. Adapun alasan diatas dapat menjadi tujuan untuk mempermudah pihak sekolah SMP negeri 3 Satap Selesai ini dalam

**Tabel 1. Data Latih**

N o	Nama	U H 1	U H 2	U H 3	U H 4	U TS	U AS	Klasifikasi
1	ADEL AMELIA	88,00	88,00	88,00	88,00	90,00	88,00	Tidak berprestasi
2	ALFIN KHAIRA AFWA PUTRA	85,00	85,00	85,00	85,00	88,00	86,00	Tidak berprestasi
3	ALZA HAFIZA	92,00	92,00	92,00	92,00	95,00	93,00	berprestasi
4	CAHYO RUDDI ANSYAH	90,00	90,00	90,00	90,00	90,00	90,00	Tidak berprestasi
5	CINTA JUWITA BR MANURUNG	90,00	90,00	90,00	90,00	90,00	90,00	Tidak berprestasi
6	DANDI PRAYUDI	80,00	80,00	80,00	80,00	80,00	80,00	Tidak berprestasi

								si
7	DENI KURNI AWAN	85,00	85,00	85,00	85,00	88,00	86,00	Tidak berprestasi
8	DIMAS SEYTO	83,00	83,00	83,00	83,00	85,00	83,00	Tidak berprestasi
9	DWI RAHA YU	95,00	95,00	95,00	95,00	95,00	95,00	berprestasi
10	ELSA ZAHIRA	95,00	95,00	95,00	95,00	98,00	96,00	berprestasi

**Tabel 2. Data Uji**

n	Nama	UH1	UH2	UH3	UH4	UTS	UAS	klasifikasi
1	X	95,00	70,00	80,00	90,00	70,00	95,00	?

**Penyelesaian**

Berikut merupakan langkah-langkah proses perhitungan metode K-Nearest Neighbor dalam klasifikasi data.

**Keterangan :**

- UH = Ulangan harian
- UTS = Ujian Tengah Semester
- UAS = Ujian Akhir Semester

1. Ditentukan nilai K = 4

2. Hitung jarak Euclidean data uji.

No	Nama	UH1	UH2	UH3	UH4	UTS	UAS	Euclidean distance (95,00, 70,00, 80,00, 90,00, 70,00, 95,00)
1	ADEL AMELIA	88,00	88,00	88,00	88,00	90,00	88,00	$\sqrt{(88,00 - 95,00)^2 + (88,00 - 70,00)^2 + (88,00 - 80,00)^2 + (88,00 - 90,00)^2 + (90,00 - 70,00)^2 + (88,00 - 95,00)^2} = 29,83$
2	ALFIN KHAIIRA AFWA PUTRA	85,00	85,00	85,00	85,00	88,00	86,00	$\sqrt{(85,00 - 95,00)^2 + (85,00 - 70,00)^2 + (85,00 - 80,00)^2 + (85,00 - 90,00)^2 + (88,00 - 90,00)^2 + (86,00 - 95,00)^2} = 21,44$
3	ALZA HAFIZA	92,00	92,00	92,00	92,00	95,00	93,00	$\sqrt{(92,00 - 95,00)^2 + (92,00 - 70,00)^2 + (92,00 - 80,00)^2 + (92,00 - 90,00)^2 + (95,00 - 70,00)^2 + (93,00 - 95,00)^2} = 35,63$
4	CAHYO RUDDIANSYAH	90,00	90,00	90,00	90,00	90,00	90,00	$\sqrt{(90,00 - 95,00)^2 + (90,00 - 70,00)^2 + (90,00 - 80,00)^2 + (90,00 - 90,00)^2 + (90,00 - 70,00)^2 + (90,00 - 95,00)^2} = 30,82$
5	CINTA JUWITA BR MANURUNG	90,00	90,00	90,00	90,00	90,00	90,00	$\sqrt{(90,00 - 95,00)^2 + (90,00 - 70,00)^2 + (90,00 - 80,00)^2 + (90,00 - 90,00)^2 + (90,00 - 70,00)^2 + (90,00 - 95,00)^2} = 30,82$
6	DANDI PRAYUDI	80,00	80,00	80,00	80,00	80,00	80,00	$\sqrt{(80,00 - 95,00)^2 + (80,00 - 70,00)^2 + (80,00 - 80,00)^2 + (80,00 - 90,00)^2 + (80,00 - 70,00)^2 + (80,00 - 95,00)^2} = 27,38$
7	DENI KURNIAWAN	85,00	85,00	85,00	85,00	88,00	86,00	$\sqrt{(85,00 - 95,00)^2 + (85,00 - 70,00)^2 + (85,00 - 80,00)^2 + (85,00 - 90,00)^2 + (88,00 - 70,00)^2 + (86,00 - 95,00)^2} = 22,09$
8	DIMAS SEYTO	83,00	83,00	83,00	83,00	85,00	83,00	$\sqrt{(83,00 - 95,00)^2 + (83,00 - 70,00)^2 + (83,00 - 80,00)^2 + (83,00 - 90,00)^2 + (85,00 - 70,00)^2 + (83,00 - 95,00)^2} = 27,20$
9	DWI RAHA YU	95,00	95,00	95,00	95,00	95,00	95,00	$\sqrt{(95,00 - 95,00)^2 + (95,00 - 70,00)^2 + (95,00 - 80,00)^2 + (95,00 - 90,00)^2 + (95,00 - 70,00)^2 + (95,00 - 95,00)^2} = 38,72$
10	ELSA ZAHIRA	95,00	95,00	95,00	95,00	98,00	96,00	$\sqrt{(95,00 - 95,00)^2 + (95,00 - 70,00)^2 + (95,00 - 80,00)^2 + (95,00 - 90,00)^2 + (98,00 - 70,00)^2 + (96,00 - 95,00)^2} = 40,29$

3. urutkan jarak secara ascending

No	Nama	UH1	UH2	UH3	UH4	UTS	UAS	Eucledian distance (95,00, 70,00, 80,00, 90,00, 70,00, 95,00)	Jarak Ascending
1	ADEL AMELIA	88,00	88,00	88,00	88,00	90,00	88,00	$\sqrt{(88,00-95,00)^2 + (88,00-70,00)^2 + (88,00-80,00)^2 + (88,00-90,00)^2 + (90,00-70,00)^2 + (88,00-95,00)^2} = 29,83$	5
2	ALFIN KHAIRA AFWA PUTRA	85,00	85,00	85,00	85,00	88,00	86,00	$\sqrt{(85,00-95,00)^2 + (85,00-70,00)^2 + (85,00-80,00)^2 + (85,00-90,00)^2 + (88,00-70,00)^2 + (86,00-95,00)^2} = 21,44$	1
3	ALZA HAFIZA	92,00	92,00	92,00	92,00	95,00	93,00	$\sqrt{(92,00-95,00)^2 + (92,00-70,00)^2 + (92,00-80,00)^2 + (92,00-90,00)^2 + (95,00-70,00)^2 + (93,00-95,00)^2} = 35,63$	8
4	CAHYO RUDDIANSYAH	90,00	90,00	90,00	90,00	90,00	90,00	$\sqrt{(90,00-95,00)^2 + (90,00-70,00)^2 + (90,00-80,00)^2 + (90,00-90,00)^2 + (90,00-70,00)^2 + (90,00-95,00)^2} = 30,82$	6
5	CINTA JUWITA BR MANURUNG	90,00	90,00	90,00	90,00	90,00	90,00	$\sqrt{(90,00-95,00)^2 + (90,00-70,00)^2 + (90,00-80,00)^2 + (90,00-90,00)^2 + (90,00-70,00)^2 + (90,00-95,00)^2} = 30,82$	7
6	DANDI PRAYUDI	80,00	80,00	80,00	80,00	80,00	80,00	$\sqrt{(80,00-95,00)^2 + (80,00-70,00)^2 + (80,00-80,00)^2 + (80,00-90,00)^2 + (80,00-70,00)^2 + (80,00-95,00)^2} = 27,38$	4
7	DENI KURNIAWAN	85,00	85,00	85,00	85,00	88,00	86,00	$\sqrt{(85,00-95,00)^2 + (85,00-70,00)^2 + (85,00-80,00)^2 + (85,00-90,00)^2 + (88,00-70,00)^2 + (86,00-95,00)^2} = 22,09$	2
8	DIMAS SEYTO	83,00	83,00	83,00	83,00	85,00	83,00	$\sqrt{(83,00-95,00)^2 + (83,00-70,00)^2 + (83,00-80,00)^2 + (83,00-90,00)^2 + (85,00-70,00)^2 + (83,00-95,00)^2} = 22,09$	3

8	DIMAS SEYTO	83,00	83,00	83,00	83,00	85,00	83,00	Tidak Berprestasi
9	DWI RAHAYU	95,00	95,00	95,00	95,00	95,00	95,00	Berprestasi
10	ELSA ZAHIRA	95,00	95,00	95,00	95,00	98,00	96,00	Berprestasi

### 3.1 Pembahasan

Data Siswa

**Data Siswa**

Nama:

UH1:  UH2:  UH3:  UH4:  UTS:  UAS:

ID	NAMA	UH1	UH2	UH3	UH4	UTS	UAS	KLASIFIKASI
11	ADEL AM...	88,00	88,00	88,00	88,00	90,00	88,00	Tidak ber...
12	ALFIN KH...	85,00	85,00	85,00	85,00	88,00	86,00	Tidak ber...
13	ALZA HAF...	92,00	92,00	92,00	92,00	95,00	93,00	berprestasi
14	CAHYO R...	90,00	90,00	90,00	90,00	90,00	90,00	Tidak ber...
15	CINTA JU...	90,00	90,00	90,00	90,00	90,00	90,00	Tidak ber...
16	DANDI PR...	80,00	80,00	80,00	80,00	80,00	80,00	Tidak ber...
17	DENI KUR...	85,00	85,00	85,00	85,00	88,00	86,00	Tidak ber...
18	DIMAS SE...	83,00	83,00	83,00	83,00	85,00	83,00	Tidak ber...
19	DWI RAH...	95,00	95,00	95,00	95,00	95,00	95,00	berprestasi
20	ELSA ZAH...	95,00	95,00	95,00	95,00	98,00	96,00	berprestasi

Tabel Hasil Variasi Nilai K Metode K-NN dan Cross Validation

Adapun analisis dari Tabel menyajikan informasi tingkat akurasi specificity algoritma K-Nearest Neighbor. Analisis dilakukan dengan melakukan

#### 4. cari kelas tetangga terbanyak

No	Nama	UH1	UH2	UH3	UH4	UTS	UAS	Klasifikasi
1	ADEL AMELIA	88,00	88,00	88,00	88,00	90,00	88,00	Tidak Berprestasi
2	ALFIN KHAIRA AFWA PUTRA	85,00	85,00	85,00	85,00	88,00	86,00	Tidak Berprestasi
3	ALZA HAFIZA	92,00	92,00	92,00	92,00	95,00	93,00	Berprestasi
4	CAHYO RUDDIANSYAH	90,00	90,00	90,00	90,00	90,00	90,00	Tidak Berprestasi
5	CINTA JUWITA BR MANURUNG	90,00	90,00	90,00	90,00	90,00	90,00	Tidak Berprestasi
6	DANDI PRAYUDI	80,00	80,00	80,00	80,00	80,00	80,00	Tidak Berprestasi
7	DENI KURNIAWAN	85,00	85,00	85,00	85,00	88,00	86,00	Tidak Berprestasi

N	Na	UH 1	UH 2	UH 3	UH 4	UT S	UA S	Klasifi kasi
1	x	95,00	70,00	80,00	90,00	70,00	95,00	Tidak Berprestasi

perhitungan Jumlah benar / jumlah data \* 100%.

## 5. KESIMPULAN

kesimpulan yang di dapat sebagai berikut:

1. Sistem yang dibangun dapat memudahkan dalam proses K-Nearest Neighbor untuk mengetahui kinerja dan peningkatan akurasi.
2. Nilai K 1 hingga 10 yang di uji persentase dari hasil analisis metode K-NN lebih tinggi dari hasil analisis metode K-NN. Dan dari nilai K yang telah di ujikan nilai K 2 dan nilai K 9 memiliki persentasi terbesar sehingga keakuratannya juga semakin tepat.
3. Adapun dari hasil pengujian metode K-Nearest Neighbor dalam klasifikasi data. Adapun pengujian penulis menggunakan variasi nilai K K-Nearest Neighbor 3,4,5,6,7,8,9. Memiliki persentase akurasi yang sangat baik di dibandingkan hanya K-NN. Hasil pengujian menunjukan metode K-Nearest Neighbor dalam klasifikasi data memiliki akurasi persentase yang baik ketika menggunakan data random. Persentase variasi nilai K K-Nearest Neighbor 3,4,5,6,7,8,9 memiliki persentase 100 % .

## 6. SARAN

adapun penulis memberikan saran untuk dapat dikembangkan untuk masa yang akan datang sebagai berikut:

1. Diharapkan kedepannya perlu dilakukan perbandingan terhadap algoritma lain, untuk menguji sejauh mana algoritma K-Nearest Neighbor masih dapat diandalkan untuk memproses dan menemukan pola hubungan asosiasi antar item pada database skala besar.
2. Diharapkan kedepannya adanya pengembangan baik berupa penambahan *fitur-fitur* aplikasi yang sesuai dibutuhkan dan *anti virus* yang berguna mengamankan sistem untuk melindungi data-data yang berada di dalam *database*.
3. Diharapkan sebaiknya ditambahkan fasilitas untuk *backup* data. Jadi, jika terjadi kerusakan pada *server* data tidak akan terhapus.

## DAFTAR PUSTAKA

- [1]. Boiculese, V.L., Dimitru, G., & Moscalu, M. 2013. Improving Recall of K-Nearest Neighbor Algorithm for Classes of Uneven Size. The 4th IEEE International Conference on E-Health and Bioengineering-EHB : 1-4.
- [2]. Boiculese, V.L., Dimitru, G., & Moscalu, M. 2013. Improving Recall of K-Nearest Neighbor Algorithm for Classes of Uneven Size. The 4th IEEE International Conference on E- Health and Bioengineering-EHB : 1-4.  
Priyanto Hidayatullah, 2012, "Visual Basic .NET Membuat Aplikasi Database dan Program Kreatif". Bandung : Informatika.  
Wardani, Sistrina H. (2016). Aplikasi Klasifikasi Jenis Tumbuhan Mangrove Berdasarkan Karakteristik Morfologi Menggunakan Metode K-Nearest Neighbor (KNN) Berbasis Web. Pontianak : Jurusan Sistem Komputer Fakultas MIPA Universitas Tanjungpura. ISSN : 2338-493X.  
Penentuan Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor dan Weighted Product (Studi Kasus : SMP Negeri 3 Mejayan) Jodi Irjaya Kartika1, Edy Santoso2, Sutrisno3 Program Studi Teknik Informatika, Fakultas Ilmu

Komputer, Universitas Brawijaya 2017.

Implementasi metode k-nearest neighbor (knn) dan simple additive weighting (saw) dalam pengambilan keputusan seleksi penerimaan anggota paskibraka (Studi Kasus : Dinas Pemuda dan Olahraga Provinsi Bengkulu) Asahar Johar T1 , Delfi Yanosma2 , Kurnia Anggriani3 Jurnal Pseudocode, Volume III Nomor 2, September 2016, ISSN 2355-5920.

<http://www.ketrare.com/2018/11/algorithm-k-nearest-neighbor-dan-contoh-soal.html> Rusmawan, Uus. 2011. Koleksi Program VB.NET. Jakarta : Elex Media.

- [3]. Priyanto Hidayatullah, 2012, “Visual Basic .NET Membuat Aplikasi Database dan Program Kreatif”. Bandung : Informatika.
- [4]. Wardani, Sistria H. (2016). Aplikasi Klasifikasi Jenis Tumbuhan Mangrove Berdasarkan Karakteristik Morfologi Menggunakan Metode K-Nearest Neighbor (KNN) Berbasis Web. Pontianak : Jurusan Sistem Komputer Fakultas MIPA Universitas Tanjungpura. ISSN : 2338-493X