

RANCANG BANGUN APLIKASI DATA MINING PREDIKSI KELULUSAN UJIAN NASIONAL MENGGUNAKAN ALGORITMA (KNN) K-NEAREST NEIGHBOR DENGAN METODE EUCLIDEAN DISTANCE PADA SMPN 2 PAGEDANGAN

Sri Mulyati ¹, Syepri Maulana Husein ², Ramdhan ³

Program Studi Informatika, Fakultas Teknik Universitas Muhammadiyah Tangerang
Jl. Perintis Kemerdekaan 1/33 Cikokol Kota Tangerang
Email : lilysrimulyati@gmail.com, syepriymaulana@ft-umt.ac.id, acengmon68@gmail.com

Abstrak

Dengan meningkatnya kualitas pendidikan di Indonesia sangat mempengaruhi nilai standar kelulusan Ujian Nasional (UN) untuk tingkat Sekolah Menengah Pertama (SMP) yang semakin tinggi. SMPN 2 Pagedangan merupakan salah satu SMP yang akan melaksanakan Ujian Nasional yang diselenggarakan oleh pemerintah. Dengan ini pihak sekolah melakukan persiapan untuk menghadapi Ujian Nasional dengan mengadakan ujian *Tryout*. Ujian *Tryout* merupakan salah satu cara guru untuk membantu para siswa siswi dalam mempersiapkan diri untuk menghadapi Ujian Nasional. Maka dari itu ujian *Tryout* bisa menjadi tolak ukur para siswa siswi dalam menghadapi Ujian Nasional dengan baik dan mendapatkan nilai yang bagus. Jika siswa siswi mendapatkan nilai *Tryout* nya yang tidak melampaui standar yang ditentukan, maka pihak sekolah harus cepat mengambil tindakan dengan membuat solusi lain yang tepat bagi siswa siswi nya. Dari hasil *Tryout* tersebut, para guru masih kesulitan untuk memprediksi kelulusan Ujian Nasional setiap tahunnya. Dikarenakan belum ada aplikasi yang menggunakan teknik untuk memprediksi yang ada untuk meningkatkan kelulusan ujian nasional. Dengan adanya masalah tersebut, maka ada solusi terbaik yang dapat membantu pihak sekolah dalam memprediksi kelulusan Ujian Nasional. Algoritma *K-Nearest Neighbor* dengan metode *Euclidean Distance* (pengukuran jarak) akan sangat membantu dalam masalah efisiensi waktu dan efektif dalam memprediksi kelulusan Ujian Nasional. Atribut yang digunakan dalam prediksi dan klasifikasi ini adalah nilai *Tryout* sedangkan atribut pada siswa yang digunakan adalah NIS, Nama, Ipa1, dan Ipa2. Dengan menggunakan nilai *Tryout* siswa siswi 3 tahun terakhir yang berjumlah 744 data. Setelah data tersebut diproses akan menghasilkan rata-rata akurasi sebesar 88.42%. Menggunakan nilai k yaitu 7 rata-rata akurasi 96.26%, presisi 96.17%, dan recall sebesar 97.32%. Dengan ini dapat disimpulkan bahwa dari pengujian yang telah dilakukan dengan menggunakan dataset siswa siswi SMPN 2 Pagedangan angkatan, 2016/2017, 2017/2018, 2018/2019 sistem dapat memprediksi dan mengklasifikasikan dengan baik dan cepat.

Kata Kunci : *K-Nearest Neighbor, Euclidean Distance, Klasifikasi, Tryout, Ujian Nasional.*

Abstract

With the improvement of the quality of education in Indonesia, it greatly influences the value of the National Exam (UN) graduation standard for the junior high school (SMP) level which is getting higher. SMPN 2 Pagedangan is one of the junior high schools that will carry out the National Examination organized by the government. With this the school makes preparations to face the National Examination by holding a Tryout exam. The Tryout exam is one way for teachers to help students prepare for the National Examination. Therefore the Tryout exam can be a benchmark for students in facing the National Examination well and getting good grades. If a student gets a Tryout value that does not exceed the specified standard, the school must quickly take action by making other solutions that are right for their students. From the results of the Tryout, teachers still find it difficult to predict the passing of the National Examination every year. Because there is no application that uses techniques to predict that is there to increase the passing of the national exam. With these problems, there is the best solution that can help the school in predicting the passing of the National Examination. K-Nearest Neighbor algorithm with Euclidean Distance method (distance measurement) will be very helpful in the problem of time efficiency and effective in predicting the passing of the National Examination. The attributes used in this prediction and classification are Tryout values while the attributes of the students used are NIS, Name, Ipa1, and

IPA2. By using the Tryout value of the last 3 years students totaling 744 data. After the data is processed it will produce an average accuracy of 88.42%. Using a k value of 7 is an average accuracy of 96.26%, a precision of 96.17%, and a recall of 97.32%. With this it can be concluded that from the testing that has been done using a dataset of students of SMPN 2 Pagedangan class, 2016/2017, 2017/2018, 2018/2019 the system can predict and classify well and quickly.

Keywords: *K-Nearest Neighbor, Euclidean Distance, Classification, Tryout, National Examination.*

PENDAHULUAN

Perkembangan teknologi saat ini mengalami perkembangan yang sangat pesat. Teknologi telah menjadi bagian dari kehidupan manusia dan telah banyak memberikan manfaat dan kemajuan diberbagai bidang. Teknologi sudah menjadi kebutuhan yang sangat diperlukan bagi manusia dalam menunjang segala aktifitas agar lebih efisien. Dalam era teknologi yang berkembang ini pun, segala sesuatu yang belum pernah terjadi dapat diprediksi menggunakan teknologi yang ada walaupun tingkat akurasinya tidak mencapai 100%.

SMPN 2 Pagedangan merupakan salah satu instansi pendidikan yang akan melaksanakan *Tryout* sebelum Ujian Nasional. *Tryout* tersebut merupakan salah satu cara guru untuk membantu para siswa dalam mempersiapkan Ujian Nasional yang akan diselenggarakan oleh pemerintah. Nantinya bagi siswa siswi yang tidak melampaui nilai BSNP (Badan Standar Nasional Pendidikan) dapat berkonsultasi dengan guru SMPN 2 Pagedangan.

Permasalahan yang ada pada SMPN 2 Pagedangan dalam memprediksi kelulusan siswanya belum ada aplikasi yang menggunakan teknik prediksi yang dapat memberikan tingkat akurasi yang cukup tinggi. Dengan ini data yang akan dikelola kurang maksimal dan mengurangi tingkat akurasi dari proses prediksi yang dilakukan.

Untuk mengatasi hal tersebut maka dibutuhkan sebuah sistem yang dapat digunakan untuk memprediksi kelulusan ujian nasional dengan mengimplementasikan ilmu data mining. menggunakan pendekatan klasifikasi. Perlunya data mining karena adanya sejumlah data besar yang dapat digunakan untuk menghasilkan informasi dan *knowledge* yang berguna. Informasi dan *knowledge* yang didapat tersebut dapat digunakan untuk mengetahui suatu pola dalam suatu data yang banyak, terlebih lagi besarnya kebutuhan untuk

mengubah data tersebut menjadi informasi yang berguna.

TINJAUAN PUSTAKA

A. Data Mining

Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Patut diingat bahwa kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu *Data Mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan database (Siburian, 2014)

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Menurut (Mustakim, Giantika O, 2016) Menyebutkan bahwa ada tiga karakteristik *Data Mining* yaitu :

- *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- *Data mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
- *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi

B. Algoritma KNN (*K-Nearest Neighbor*)

Algoritma K-Nearest Neighbor (K-NN) merupakan algoritma yang melakukan proses klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain. Nilai K pada KNN merupakan jumlah tetangga terdekat, jika K bernilai 1, maka kelas dari satu data latih yang terdekat akan menjadi kelas bagi data uji yang baru, jika K bernilai 3 akan diambil tiga data latih yang terdekat menjadi kelas untuk data uji yang baru. Salah satu masalah yang dihadapi K-NN adalah dalam pemilihan nilai K yang tepat. Pemilihan nilai K yang besar dapat mengakibatkan distorsi data yang besar pula. Hal ini dikarenakan setiap tetangga mempunyai bobot yang sama terhadap data uji, sedangkan K yang terlalu kecil bisa menyebabkan algoritma terlalu sensitive terhadap noise. K-NN merupakan teknik klasifikasi yang sederhana, tetapi mempunyai hasil kerja yang cukup bagus. (Lukman & Labellapansa, 2015) Algoritma ini sederhana, bekerja berdasarkan jarak terpendek dari query instance ke training sample untuk menentukan ketetanggaannya (Rizal, 2013). Langkah-langkah untuk menghitung metode k-Nearest Neighbor antara lain:

- Menentukan parameter k
- Menghitung jarak antara data yang akan dievaluasi dengan semua pelatihan
- Mengurutkan jarak terbentuk
- Menentukan jarak terdekat sampai urutan K
- Memasangkan kelas yang bersesuaian

C. Euclidean Distance

Euclidean Distance mengukur kedekatan antara dua buah objek yang digambarkan sebagai pengukuran garis lurus atau secara langsung (J.Han, J.Pei, 2012). Metode pengukuran ini cocok apabila diimplementasikan terhadap 14 data yang memiliki nilai atribut bersifat numerikal, khususnya dengan atribut kontinyu (Gorunescu, 2011). *Euclidean Distance* dihitung dengan persamaan dibawah ini:

$$d_{Euclidean}(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2}$$

Sebagai contoh, terdapat 2 buah objek yang akan diukur tingkat kedekatannya dengan menggunakan *Euclidean Distance*. Dimana 2 buah objek tersebut adalah sebagai berikut :

X = [8, 2, 4, 3]

Y = [1, 6, 7, 5]

Maka, jarak antara objek X dan Y jika dihitung dengan persamaan adalah:

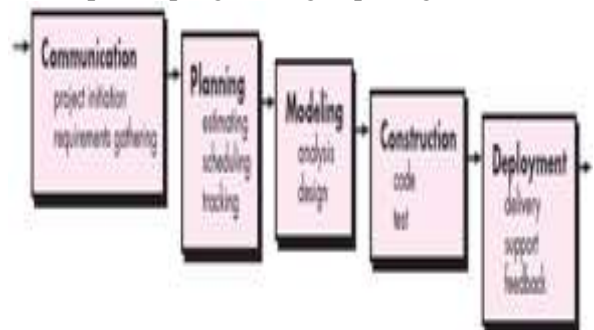
$$\begin{aligned} d_{Euclidean}(x, y) &= \sqrt{(8-1)^2 + (2-6)^2 + (4-7)^2 + (3-5)^2} \\ &= \sqrt{49 + 16 + 9 + 4} \\ &= \sqrt{84} \\ &= 9,17 \end{aligned}$$

Maka dari itu berarti tingkat kedekatan X dan Y adalah 9,17

D. SDLC (*Software Development Lyfe Cycle*)

Menurut Roger S. Pressman (2010: 20) *Software Development Life Cycle* adalah proses mengembangkan atau mengubah suatu sistem perangkat lunak dengan menggunakan model-model dan metodologi yang digunakan orang untuk mengembangkan sistem perangkat lunak.

Menurut Roger S. Pressman (2010:39) *Waterfall Model* sebuah proses perancangan yang secara berurutan dan sering digunakan dalam proses pengembangan perangkat lunak.



Gambar 1 Model *Waterfall* (Roger S. Pressman 2010)

METODOLOGI PENELITIAN

A. Jenis Penelitian

Model penelitian pada dasarnya merupakan cara ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu. Ada empat kata kunci yang perlu diperhatikan, yaitu:

- Cara ilmiah, kegiatan penelitian didasarkan pada ciri-ciri keilmuan, antara lain: rasional, empiris dan sistematis.
- Data, data yang diperoleh dalam penelitian itu adalah data empiris (teramati) yang mempunyai kriteria tertentu yaitu valid, reliabel, dan objektif.
- Tujuan, setiap penelitian mempunyai tujuan dan kegunaan tertentu. Secara umum tujuan penelitian ada tiga macam yaitu: penemuan, pembuktian, pengembangan.
- Kegunaan, melalui penelitian manusia dapat menggunakan hasilnya. Secara umum data yang telah diperoleh dari penelitian dapat digunakan untuk memahami, memecahkan dan mengantisipasi masalah.

Penelitian yang peneliti lakukan menggunakan metode penelitian campuran, dimana pengumpulan data akan dilakukan dengan cara metode survei menggunakan wawancara langsung ke perusahaan terkait.

Teori pengembangan sistem menggunakan metode pengembangan sistem menggunakan SLDC model *waterfall*. Tahapan analisis dan perancangan sistem menggunakan pendekatan berorientasi objek, implementasi hasil analisis dan perancangan berbasis PHP dan database MySQL. Kualitas sistem pendukung keputusan yang diimplementasikan diuji berdasarkan pengujian *black-box*. Hasil penelitian berupa aplikasi sistem pendukung keputusan pemilihan *supplier* bahan baku menggunakan metode SMART yang berguna untuk membuat perusahaan dalam pemilihan *supplier* menjadi lebih efisien.

B. Euclidean Distance

- Sumber data merupakan salah satu komponen utama dalam kegiatan Prediksi dan Klasifikasi, karena data dapat dikatakan baik apabila dapat mewakili keadaan suatu objek yang sedang diteliti dan sesuai

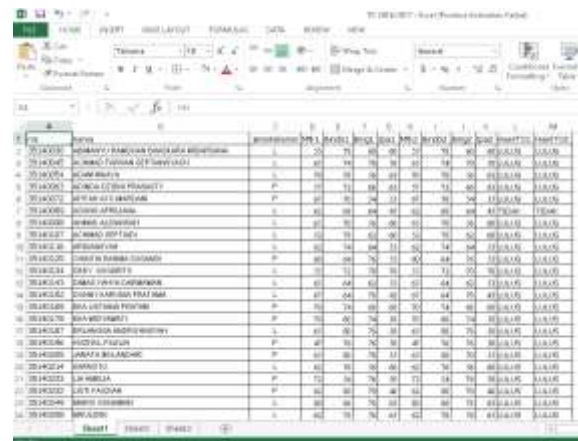
dengan keadaan yang ada. Pengumpulan data ini dilakukan dengan memperoleh data primer dan sekunder. Data primer diperoleh dari data hasil ujian *Tryout* dan hasil Ujian Nasional SMPN 2 Pagedangan. Dataset yang dipergunakan adalah data hasil Ujian *Tryout* dan hasil Ujian Nasional selama 3 tahun terakhir.

Tabel 1. Tahun Ajaran

No	Tahun Ajaran	Jumlah Data
1	2015/2016	231
2	2016/2017	278
3	2017/2018	295

- Persiapan data Format dataset yang digunakan untuk uji coba sistem yang dibuat. Dataset yang didapatkan dari SMPN 2 Pagedangan, format data harus disesuaikan dengan yang ada pada sistem yang dibuat agar dapat melakukan uji sistem

Tabel 2. Data Set



The screenshot shows a spreadsheet with columns labeled with codes like NIS, NAMA, MTK1, BINDO1, BING1, IPA1, MTK2, BINDO2, BING2, IPA2, HASILTO1, and HASILTO2. The rows contain numerical data for each of these attributes across multiple records.

Tahap-tahap yang dilakukan untuk mengubah dataset yang didapat bisa digunakan untuk pengujian adalah:

- File harus berekstensi .xls atau .xlsx.
- Memformat ulang urutan dan nama kolom pada dataset menjadi NIS, NAMA, MTK1, BINDO1, BING1, IPA1, MTK2, BINDO2, BING2, IPA2, HASILTO1, HASILTO2.
- Menempatkan nama kolom yang dimulai dari sel A1.

- Pemilihan data memilih atribut yang tepat untuk proses analisa data. Pemilihan atribut ini merupakan hasil diskusi dengan Guru Bidang Kurikulum SMPN 2 Pagedangan yang mendapatkan keputusan terkait atribut yang dipakai untuk kelulusan Ujian Nasional SMPN 2 Pagedangan.
- Transformasi data proses ini dilakukan perubahan data ke bentuk yang sesuai dengan algoritma klasifikasi yang digunakan. Yang sebelumnya diketahui bahwa algoritma tidak cocok dilakukan untuk atribut yang bertipe kategorikal, jadi perlu dilakukan transformasi data pada atribut hasil *Tryout* menjadi atribut numerikal, Selain attribute tersebut. Nilai hasil ujian nasional *tryout* pertama dan kedua juga di transformasikan menjadi nilai sesuai kebijakan yang ada, yaitu dengan cara menentukan nilai rata – rata dari nilai hasil ujian *tryout* pertama dan kedua Rata-rata Nilai Ujian Tryout pertama dan kedua:

$$(MTK1 + BINDO1 + BING1 + IPA1)/4$$

$$(MTK2 + BINDO2 + BING2 + IPA2)/4$$

Nama	meanTO1	meanTO2	HasilTO
Abimanyu	63,25	63,25	1
Achmad	66,75	66,75	1
Agung	59,75	59,75	0
Adinda	64,5	64,5	1

$$E = \sqrt{(j0 - i1)^2 + (j1 - i1)^2 + (j2 - i2)^2}$$

$$E = \text{Euclidean Distance}$$

$$J = \text{Dataset Model}$$

$$I = \text{Dataset Uji}$$

$$\text{Abimanyu}$$

$$E = \sqrt{(63,25 - 72,5)^2 + (63,25 - 75,50)^2 + (1 - 1)^2}$$

$$E = \sqrt{85,56 + 150,06 + (0)}$$

$$E = \sqrt{223,62}$$

$$E = 15,34$$

Achmad

$$E = \sqrt{(66,75 - 72,5)^2 + (66,75 - 75,50)^2 + (1 - 1)^2}$$

$$E = \sqrt{33,06 + 76,56 + (0)}$$

$$E = \sqrt{109,62}$$

$$E = 10,46$$

Agung

$$E = \sqrt{(59,75 - 72,5)^2 + (59,75 - 75,50)^2 + (1 - 0)^2}$$

$$E = \sqrt{162,56 + 248,06 + (1)}$$

$$E = \sqrt{2,893}$$

$$E = 20,28$$

Adinda

$$E = \sqrt{(64,5 - 72,5)^2 + (64,5 - 75,50)^2 + (1 - 1)^2}$$

$$E = \sqrt{(64)^2 + (121)^2 + (0)^2}$$

$$E = \sqrt{185}$$

$$E = 13,60$$

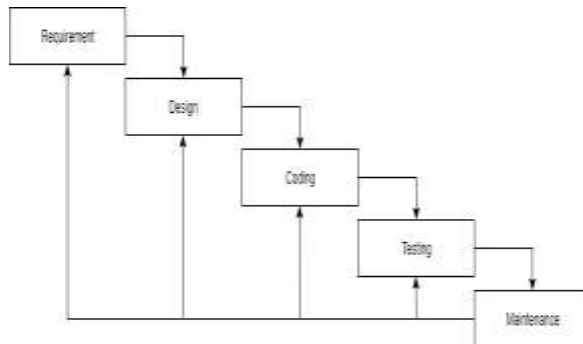
C. Metode Pengembang Sistem

Dalam pengembangan aplikasi ini penulis memilih untuk menggunakan metode pengembangan *System Development Life Cycle* (SDLC) model *Waterfall*. Berikut alasan penulis menggunakan model *waterfall*:

1. Aplikasi yang penulis coba kembangkan ini bukanlah sistem dengan skala teramat besar.
2. Aplikasi ini lebih cocok dikembangkan dengan proses yang terstruktur dan sekuensial.

Siklus SDLC ini dijalankan secara berurutan, mulai dari tahap awal sampai akhir. Setiap langkah yang telah selesai harus dikaji ulang (*review*), terutama dalam langkah analisis dan desain untuk memastikan bahwa tahapan tersebut telah dikerjakan dengan benar dan sesuai dengan harapan. Jika tidak maka tahap

tersebut perlu diulangi lagi atau kembali ke tahapan sebelumnya. Tetapi kadang ada beberapa langkah yang dapat dilakukan secara bersamaan, hal ini dilakukan untuk mempercepat mendapatkan hasil yang diinginkan.



Gambar 2 Tahap metode *waterfall* (Roger S. Pressman 2010: 20)

Secara garis besar mengenai tahapan-tahapan siklus SDLC model *waterfall* adalah sebagai berikut:

1. **Analisa Kebutuhan Sistem (*Requirement*)**
Merupakan proses pengumpulan kebutuhan perangkat lunak. Untuk memahami dasar dari program yang akan dibuat, seorang analisis harus mengetahui batasan masalah informasi, fungsi-fungsi yang dibutuhkan, kemampuan kinerja yang ingin dihasilkan dan perancangan antarmuka pemakai perangkat lunak tersebut.
2. **Perancangan (*Design*)**
Perancangan perangkat lunak merupakan proses bertahap yang memfokuskan pada empat bagian penting, yaitu: struktur data, arsitektur perangkat lunak, detail prosedur, dan karakteristik antarmuka pemakai.
3. **Pengkodean (*Coding*)**
Pengkodean perangkat lunak merupakan proses penulisan bahasa program agar piranti lunak tersebut dapat dijalankan oleh mesin.
4. **Pengujian (*Testing*)**
Proses ini akan menguji kode program yang telah dibuat dengan memfokuskan pada bagian dalam perangkat lunak. Tujuannya untuk memastikan bahwa semua pernyataan telah diuji dan memastikan juga bahwa input

yang digunakan akan menghasilkan output yang sesuai. Pada tahap ini pengujian dibagi menjadi dua bagian, pengujian internal dan pengujian eksternal. Pengujian internal bertujuan menggambarkan bahwa semua statement sudah dilakukan pengujian, sedangkan pengujian eksternal bertujuan untuk menemukan kesalahan serta memastikan output yang dihasilkan sesuai dengan yang diharapkan.

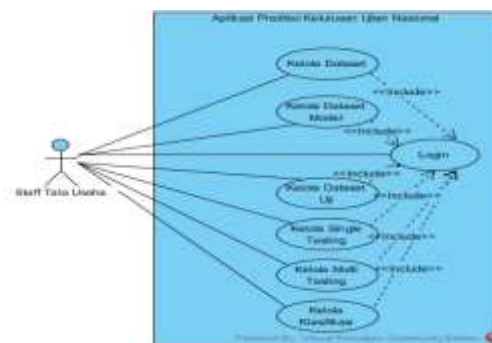
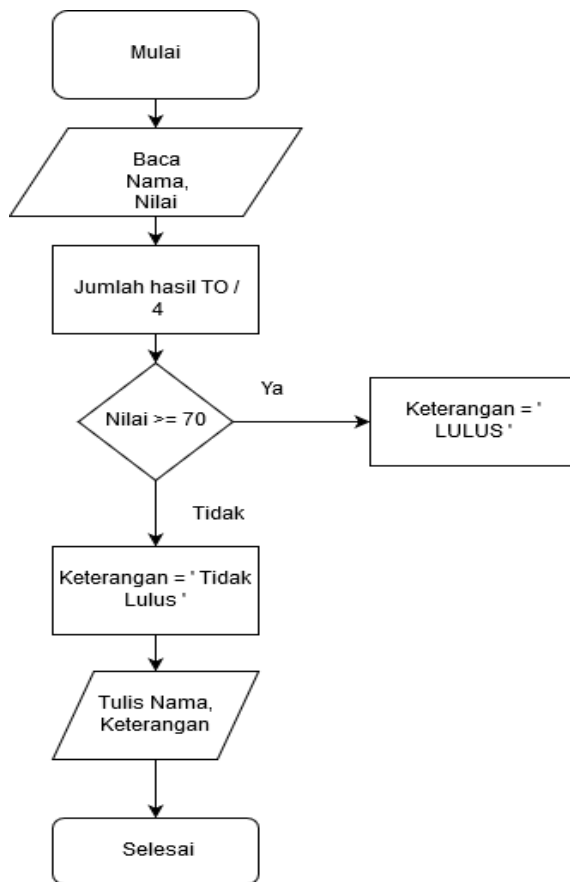
5. **Pemeliharaan (*Maintenance*)**

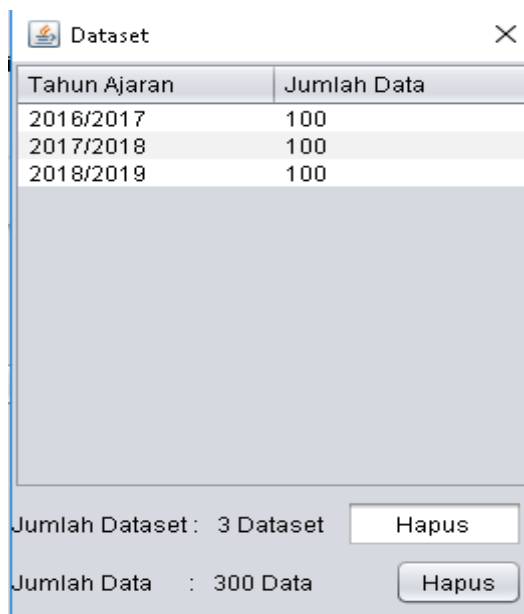
Proses ini dilakukan setelah perangkat lunak telah digunakan oleh pengguna. Perubahan akan dilakukan jika terdapat kesalahan, oleh karena itu perangkat lunak harus disesuaikan lagi untuk menampung perubahan kebutuhan yang diinginkan pengguna.

HASIL DAN PEMBAHASAN

A. Analisa Sistem Berjalan

Untuk mengetahui sistem yang sedang berjalan dan untuk mempelajari sistem yang ada, diperlukan suatu penggambaran aliran-aliran informasi dari bagian-bagian yang terkait baik dari dalam maupun dari luar sistem. Hal ini memudahkan kita untuk memahami informasi-informasi yang didapat dan dikeluarkan oleh sistem itu sendiri. Berikut adalah gambaran atau simulasi sistem berjalan :



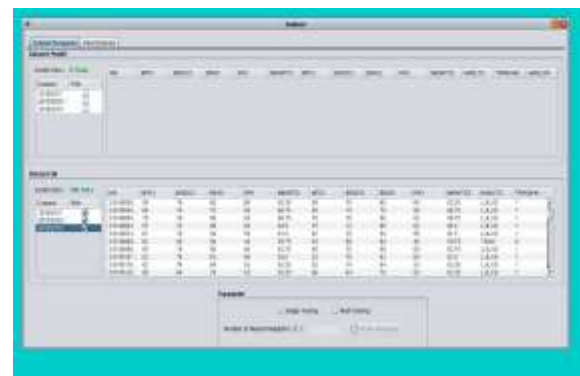


Tahun Ajaran	Jumlah Data
2016/2017	100
2017/2018	100
2018/2019	100

Jumlah Dataset : 3 Dataset Hapus

Jumlah Data : 300 Data Hapus

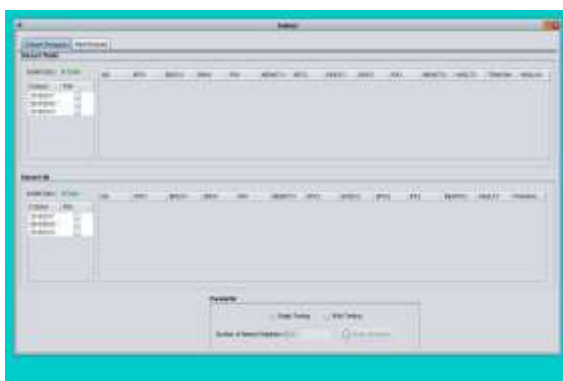
Gambar 8 Tampilan lihat dataset



Gambar 11. Tampilan kelola dataset uji



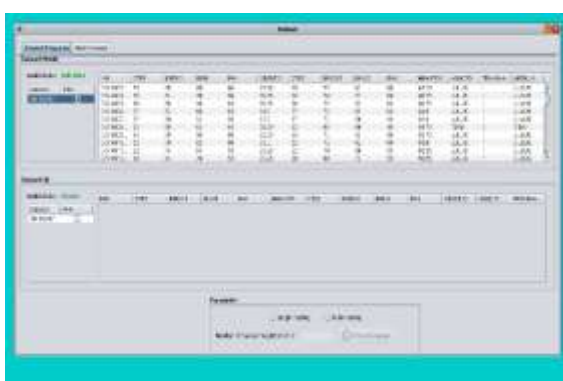
Gambar 12 Tampilan kelola *single testing*



Gambar 9 Tampilan menu evaluasi



Gambar 13 Tampilan hasil *single testing*



Tahun Ajaran	Jumlah Data
2016/2017	100
2017/2018	100
2018/2019	100

Jumlah Dataset : 3 Dataset Hapus

Jumlah Data : 300 Data Hapus

Gambar 10 tampilan kelola dataset model

Hasil Pengujian

Setelah dilakukan pengujian, pada tahap ini akan dijelaskan beberapa kelebihan dan kekurangan sistem yang telah dibangun, sebagai berikut:

Kelebihan

- Sistem dapat meng-*Import file* yang berekstensi *.xls atau *.xlsx.
- Sistem dapat melakukan pengujian dengan menggunakan banyak parameter nilai(k) yang terdapat pada *Multi Testing*, sehingga dapat mengetahui perbandingan akurasi.
- Sistem dapat memprediksi kelulusan Ujian Nasional data siswa dengan jumlah yang banyak.
- Sistem dapat digunakan di beberapa *platform* yang berbasis desktop.

- Waktu yang diperlukan dalam proses prediksi dan klasifikasi terbilang cukup cepat dibandingkan dengan sistem manual sebelumnya.
- 2.Kelemahan
- *User* harus memasukan *file spreadsheet* sesuai dengan format yang telah ditentukan agar terdeteksi oleh sistem.
- Proses pengujian masih harus memasukan nilai(k) untuk menentukan ketetanggan terdekat

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan terhadap permasalahan yang ada, maka penulis dapat menarik kesimpulan dan saran yang mungkin diperlukan, untuk mengembangkan system ini ketahap yang lebih baik kedepanya Melalui proses pengerjaan dan pengujian dalam penelitian ini, maka dari itu dapat disimpulkan beberapa hal, yaitu:

- Aplikasi prediksi kelulusan ujian nasional ini dibangun menggunakan algoritma *K-Nearest Neighbor* dengan metode *Euclidean Distance* menggunakan nilai *tryout* sehingga diharapkan mampu memberikan informasi secara tepat dan membantu dalam kegiatan prediksi kelulusan Ujian Nasional
- Algoritma *K-Nearest Neighbor* dan metode *Euclidean Distance* dapat diimplementasikan kedalam aplikasi prediksi kelulusan Ujian Nasional

REFERENSI

Destriana, R., Permana, A, A., Legawa, S, D., dan Irawan, H. 2019, *Security system development for vehicle using the method of "mail notification" at villa Rizki Ilhami Tangerang residential*, IOP Conference Series: Materials Science and Engineering 012124, Vol 508 No 1, IOP Publishing doi:10.1088/1757-899X/508/1/012124

Gorunescu, F. 2011. *Data Mining Concept Model and Techniques*. Berlin: Springer. ISBN 978-3-642-19720-8

Han J, Dkk 2012 *Data Mining: Concepts And Techniques*, 3rd Edition

Lukman, K. U. S. bin and Labellapansa, A. 2015 Analisa Nilai Limda Model Jarak Minkowsky Untuk Penentuan Jurusan SMA (Studi Kasus di SMA Negeri 2 Tualang)', Jurnal Teknik Informatika dan Sistem Informasi

Mustakim, Giantika Oktaviani F, 2016 Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa. Jurnal Sains, Teknologi dan Industri, Vol. 13, No.2, Juni 2016, pp.195 - 202 ISSN 1693-2390

Pressman, Ph.D.Roger S. 2010. *Pendekatan Praktisi Rekayasa Perangkat Lunak*. Edisi 7. Penerbit Andi. Yogyakarta.

R. Taufiq dan A. A. Permana, 2018, "Sistem Pendukung Keputusan Penerimaan Karyawan Menggunakan Simple Additive Weighting Studi Kasus PT. Trafoindo Prima Perkasa," J. Al-AZHAR Indonesia. SERI SAINS DAN Teknol., vol. 4, no. 4, hlm. 186–194.

Siburian Beni, 2014 Aplikasi Data Mining Untuk Menampilkan Tingkat Kelulusan Mahasiswa Dengan Algoritma Apriori. Pelita Informatika Budi Darma, Vol. 8, No. 2, Agustus 2014, ISSN : 2301-9425 Pressman, Ph.D.Roger S. 2010. *Pendekatan Praktisi Rekayasa Perangkat Lunak*. Edisi 7. Penerbit Andi. Yogyakarta.