



Jurnal Politeknik Caltex Riau  
<http://jurnal.pcr.ac.id>

## Rancang Bangun Aplikasi *Data Mining* untuk Memprediksi Hasil Belajar Siswa Sekolah Menengah Atas Berbasis Web dengan Algoritma K-NN (Studi Kasus: SMKN 2 Pekanbaru)

Okta Riveranda<sup>1</sup>, Maksum Rois Adin Saf<sup>2</sup> dan Muhammad Ihsan Zul<sup>3</sup>

<sup>1</sup>Politeknik Caltex Riau, email: okta12si@mahasiswa.pcr.ac.id

<sup>2</sup>Politeknik Caltex Riau, email: maksum@pcr.ac.id

<sup>3</sup>Politeknik Caltex Riau, email: ihsan@pcr.ac.id

### Abstrak

*Kurikulum 2013 (K-13) pertama kali diumumkan pada tahun 2014 dengan penerapan ke beberapa sekolah percobaan. Penyusunan kurikulum baru ini oleh pemerintah ditujukan agar pendidikan di Indonesia tidak hanya menitikberatkan pada aspek kognitif atau keterampilan yang dimiliki saja, melainkan juga melihat minat dan motivasi belajar siswa. Namun dibalik tujuan tersebut, terdapat kendala yang dialami oleh sekolah selama penerapan K-13. Salah satunya adalah proses input dan konversi nilai yang membutuhkan waktu yang relatif lama. Hal itu dikarenakan standar dan skala penilaian yang berbeda daripada kurikulum sebelumnya. Sementara itu, sistem akademik yang berjalan di sekolah-sekolah percobaan tersebut masih relatif konvensional. Maka dibangunlah sebuah model aplikasi yang mampu menangani masalah tersebut. Di samping tambahan yang ada, aplikasi ini dibangun dalam rangka menerapkan data mining dengan algoritma k-NN untuk memprediksi hasil belajar siswa berdasarkan mata pelajaran tertentu. Pemilihan algoritma k-NN didasari pada karakteristik data latih yang sesuai untuk diterapkannya algoritma tersebut. Sumber data yang digunakan berjumlah 500 sumber data latih yang mencakup semua label yang ada, yang atribut-atributnya adalah nilai harian 1, nilai harian 2, nilai harian 3, nilai harian 4, dan nilai ujian tengah semester. Metode yang digunakan dalam pengujian ini adalah black box testing dan confusion matrix. Ada 3 teknik black box testing yang diterapkan untuk menguji fungsionalitas aplikasi berdasarkan nilai inputnya. Teknik-teknik tersebut ialah equivalence class partitioning, boundary value analysis dan decision table based testing. Sementara confusion matrix, telah dilakukan 3 kali pengujian berdasarkan nilai k. Dengan k-5 diperoleh persentase akurasi sebesar 79.34%, k-10 dengan persentase akurasi sebesar 62.67% dan dengan k-15 diperoleh sebesar 64%. Sehingga kesimpulan yang bisa diambil dari pengujian tersebut ialah bahwa dengan menerapkan k-5 untuk algoritma k-NN, persentase akurasi lebih akurat.*

**Kata kunci:** Aplikasi Web, Data Mining, k-NN, Klasifikasi, Prediksi, Prestasi Akademik.

### Abstract

*Curriculum 2013 (K-13) was first announced in 2014 which has been applied to number of schools. Preparation of this new curriculum by the government aimed at making education in Indonesia is not only focused on cognitive aspects or skills possessed, but also at students' interest and motivation. Unfortunately, behind the goal, there are issues occurred in the school during the application of K-13. Those are input process and values conversion that takes relatively much time. The things are caused by the dissimilarity of the standards and the*

assessment scale between current curriculum with the previous one. Meanwhile, the academic system running in schools is still pretty conventional. Therefore, this research will construct an application which have capability to handle the things. Beside those additional features, this research is build an application in order to apply the data mining with  $k$ -NN algorithm to predict students learning outcomes based on certain subjects. Reason why choose the  $k$ -NN algorithm is according to the suitability of the characteristics of training data with the implementation of the algorithm. Data source that used in this research are consisted into 500 data training that covered up all classes or labels, whose attributes are daily assignment 1, daily assignment 2, daily assignment 3, daily assesment 4, and a mid term test assignment. Testing methods which have been applied are black box testing and confusion matrix. There are 3 techniques of black box testing that applied in order to test the system functionality according to its input values. Those are equivalence class partitioning, boundary value analysis and decision table based testing. Meanwhile in confusion matrix, it has been done 3 times testing according by variety of  $k$  value in  $k$ -NN algorithm. With  $k$ -5 acquired accurate rate 79.34%,  $k$ -10 with accurate rate 62.67%, then  $k$ -15 with accurate rate 64%. Therefore, information that can conluded from those testing methods is the algorithm with  $k$ -5 is more accurate than any others.

**Keywords:** Web Application, Data Mining,  $k$ -NN, Classification, Prediction, Academic Achievement

---

## 1. Pendahuluan

Teknik-teknik *data mining* telah banyak diaplikasikan dalam berbagai bidang penerapan, seperti dalam perbankan, militer, kesehatan, keuangan, kedokteran, pendidikan, dan lain sebagainya. Tujuan dari masing-masing penerapan tersebut ialah untuk memperoleh informasi yang berguna dalam rangka meningkatkan kualitas yang lebih baik di bidang terapan itu.

Salah satu tujuan penerapan *data mining* dalam bidang pendidikan adalah untuk meningkatkan performa akademik suatu institusi pendidikan, seperti membantu untuk memrediksi hasil belajar siswa berdasarkan mata pelajaran tertentu. Hasil belajar adalah hasil yang telah dicapai seseorang setelah mengikuti suatu program pendidikan dan pelatihan [1]. Sehingga dapat disimpulkan bahwa hasil belajar merupakan suatu capaian dari hasil kerja keras peserta didik selama proses pembelajaran yang ia ikuti untuk masing-masing mata pelajaran yang ada.

Kurikulum 2013 pertama kali diterapkan di berbagai sekolah percobaan pada tahun 2014. Kurikulum ini memiliki standar dan skala penilaian yang berbeda dibanding kurikulum sebelumnya [2]. Hal itu menimbulkan dampak tersendiri dalam proses input dan konversi nilai bagi sekolah yang masih menggunakan file excel dalam pengoperasiannya. Penerapan aplikasi berbasis web yang akan dibangun dalam penelitian ini diharapkan mampu untuk menangani kendala tersebut. Selain itu, aplikasi ini memiliki tujuan utama dalam implementasi *data mining* dengan memrediksi hasil belajar siswa pada mata pelajaran tertentu.

Algoritma yang digunakan dalam penelitian ini ialah  $k$ -NN, salah satu algoritma dalam *data mining* yang dapat digunakan untuk proses prediksi. Algoritma  $k$ -NN termasuk ke dalam golongan algoritma *supervised learning*. Dimana tiap data dalam data latihnya sudah memiliki label masing-masing yang merujuk pada baris data tersebut. Cara kerja algoritma ini adalah dengan menghitung jarak kedekatan antara data uji dengan masing-masing data latih dan memilih sejumlah  $k$  data latih terdekat, untuk kemudian menentukan label untuk data uji berdasarkan label terbanyak yang muncul dari data latih terdekat yang telah dipilih [3].

Berdasarkan standar penilaian K-13, predikat suatu nilai mata pelajaran dibagi menjadi 4, sangat baik (SB), baik (B), cukup (C) dan kurang (K). Hasil yang diharapkan dari penerapan ini berupa prediksi nilai akhir suatu mata pelajaran yang disesuaikan dengan standar penilaian K-13 tersebut.

Sumber data yang digunakan dalam penelitian ini berjumlah 500 baris data. Terdiri dari 208 baris data nilai siswa SMKN 2 Pekanbaru dan 292 baris data yang ditambahkan karena

label yang ada dari data nilai siswa tidak menampung semua label yang dibutuhkan dalam proses prediksi. Berdasarkan total data tersebut, diharapkan aplikasi *data mining* yang dibangun dapat digunakan untuk memprediksi hasil belajar siswa dengan baik, dan membantu para wali kelas maupun guru dalam proses input dan konversi nilai siswa serta membantu dalam memberi peringatan bagi siswa yang diprediksi mendapat nilai dengan predikat kurang.

## 2. Penelitian Terdahulu

Penelitian mengenai *data mining* di bidang pendidikan (*educational data mining*) sudah banyak diterapkan oleh para akademisi di berbagai institusi perguruan tinggi di dunia. Salah satu dari penelitian tersebut ialah penelitian yang dilakukan dalam penelitian [4]. Penelitian tersebut berfokus pada pemrediksian performa akademik para mahasiswa dengan menggunakan tiga algoritma *data mining*, k-Means, *Artificial Neural Network* dan *Decision Tree*. Algoritma k-Means digunakan untuk *clustering* data mahasiswa, membaginya ke dalam dua *cluster* atau kelompok berdasarkan tingkat kemiripan antar data tersebut. Sementara itu, dua algoritma klasifikasi yang digunakan dalam penelitian itu diterapkan untuk mengklasifikasikan data yang sudah di-*cluster*.

Sumber data yang digunakan dalam penelitian [4] berasal dari data mahasiswa yang disediakan oleh National Defence University of Malaysia (NDUM). Terdapat 85 data mahasiswa untuk satu semester dalam periode 2008/2009 yang menjadi fokus dalam penelitian tersebut. Sumber data tersebut dibagi menjadi dua bagian, data primer dan data sekunder.

Penelitian lain yang juga menerapkan *data mining* dalam bidang pendidikan ialah penelitian yang telah dilakukan dalam penelitian [5]. Sumber data yang digunakan dalam penelitian tersebut ialah data akademik di kampus *Science and Technology Khanyounis* yang menampung data dalam periode lima belas tahun (1993-2007). Sumber data tersebut diolah untuk mendapatkan jenis-jenis *knowledge* yang mungkin bisa dihasilkan dari teknik-teknik *data mining*. Beberapa dari teknik tersebut yang difokuskan pada penelitian itu ialah *association rules*, *classification*, *clustering*, dan *outlier detection*. Tujuan dari penelitian itu ialah untuk memberikan bantuan kepada pihak manajemen akademik pada kampus tersebut untuk mengatasi masalah pada mahasiswa yang memiliki nilai akademik yang rendah, dan juga untuk meningkatkan performa akademik para mahasiswa.

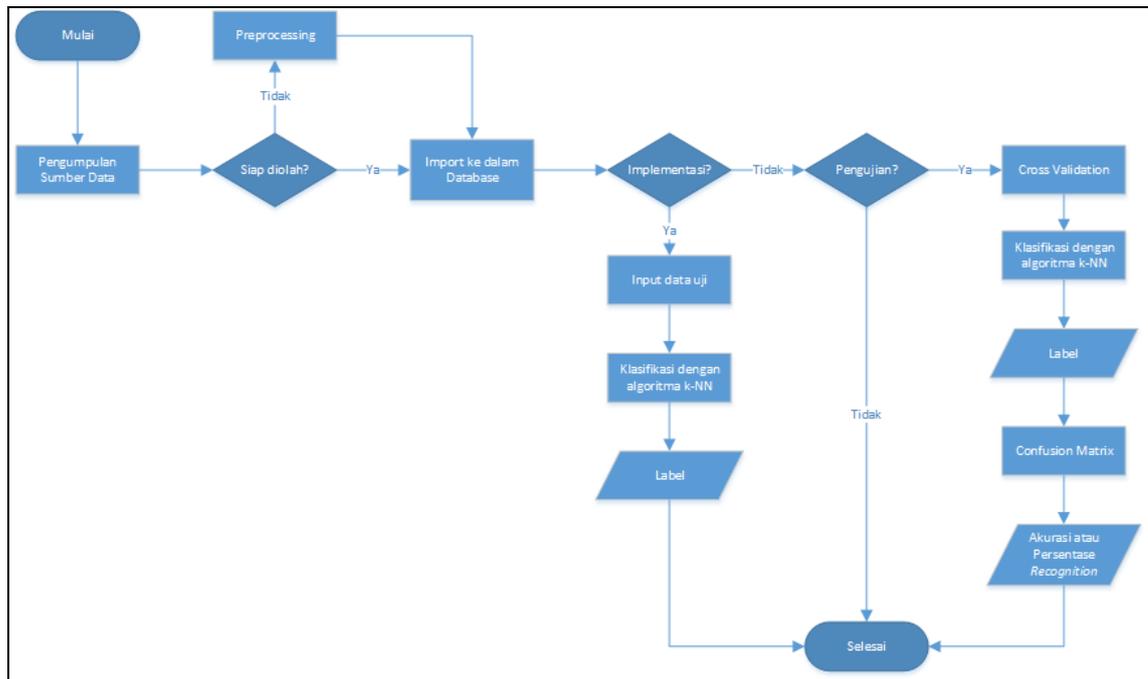
Penelitian lain yang juga menerapkan *educational data mining* ialah yang dilakukan dalam penelitian [6]. Para peneliti tersebut menerapkan teknik *classification* untuk memrediksi performa mahasiswa. Sumber data dikumpulkan dari berbagai program studi yang ada di kampus Dr. R. M. L. Awadh University periode 2009-2010. Jumlah data yang digunakan mencapai 300 baris data yang terdiri dari 226 laki-laki dan 74 perempuan. Sementara itu, algoritma yang digunakan dalam penelitian tersebut ialah Naive Bayes. Tujuan dari penelitian tersebut ialah untuk meningkatkan performa mahasiswa tiap divisi dengan mengidentifikasi mahasiswa mana yang harus diberi perhatian lebih untuk mengurangi rasio kegagalan dan mengambil langkah tepat di waktu yang tepat.

Tabel 1. Penelitian Terdahulu

Peneliti	Teknik Data Mining	Algoritma	Sumber data yang digunakan	Keluaran
[4]	<i>Clustering dan Classification</i>	<i>k-Means, Artificial Neural Network, dan Decision Tree.</i>	85 data mahasiswa untuk satu semester dalam periode 2008/2009	Kesimpulan terkait kemampuan akademik mahasiswa
[5]	<i>Association Rules, Classification, Clustering dan Outlier Detection</i>	<i>Rule Induction, k-NN, k-Means, Distance-based Approach dan Density-based Approach.</i>	3360 baris dan 18 atribut (kolom) untuk 24 program studi komputer	Pembagian kelompok mahasiswa berdasarkan kemampuan akademisnya
[6]	<i>Classification</i>	<i>Naive Bayes</i>	300 baris data yang terdiri dari 226 laki-laki dan 74 perempuan	Peningkatan performa akademis mahasiswa tiap divisi
Penelitian saat ini	<i>Classification</i>	<i>k-NN</i>	Rincian nilai tiap mata pelajaran dengan jumlah 208 siswa SMKN 2 Pekanbaru dengan tambahan 292 data latih baru	Prediksi nilai akhir dalam bentuk angka dan huruf pada masing-masing mata pelajaran

### 3. Metodologi Penelitian

Penelitian proyek akhir ini diawali dengan mengumpulkan sumber data untuk dijadikan sebagai data latih sistem aplikasi. Data latih yang diperoleh tidak langsung diimport ke dalam database, melainkan dilakukan tahap *preprocessing* dimana variabel-variabel yang tidak digunakan dalam proses *data mining* dihilangkan. Setelah data latih siap diimport ke dalam database, sistem aplikasi dikembangkan sedemikian rupa agar algoritma k-NN dapat berjalan dengan semestinya. Tahap implementasi memungkinkan guru untuk melihat prediksi hasil belajar siswa dengan variabel-variabel yang telah disiapkan untuk data uji. Adapun pada tahap pengujian, dilakukan *cross validation* dengan mengambil beberapa sampel dari data latih untuk dijadikan data uji dalam perhitungan akurasi algoritma dengan *confusion matrix*. Gambaran umum dari deskripsi yang telah dijelaskan dapat dilihat pada Gambar 1.



Gambar 1. Gambaran Metodologi Penelitian Proyek Akhir

### 3.1 Pengumpulan Sumber Data

Sumber data diperoleh dari bagian akademik SMKN 2 Pekanbaru dalam bentuk *flat file* yang terdiri dari hasil rekapitulasi nilai sejumlah 6 kelas pada jurusan Teknik Informatika. Tipe data adalah numerik, sehingga tepat untuk menggunakan algoritma k-NN dalam rangka prediksi hasil belajar para siswa.

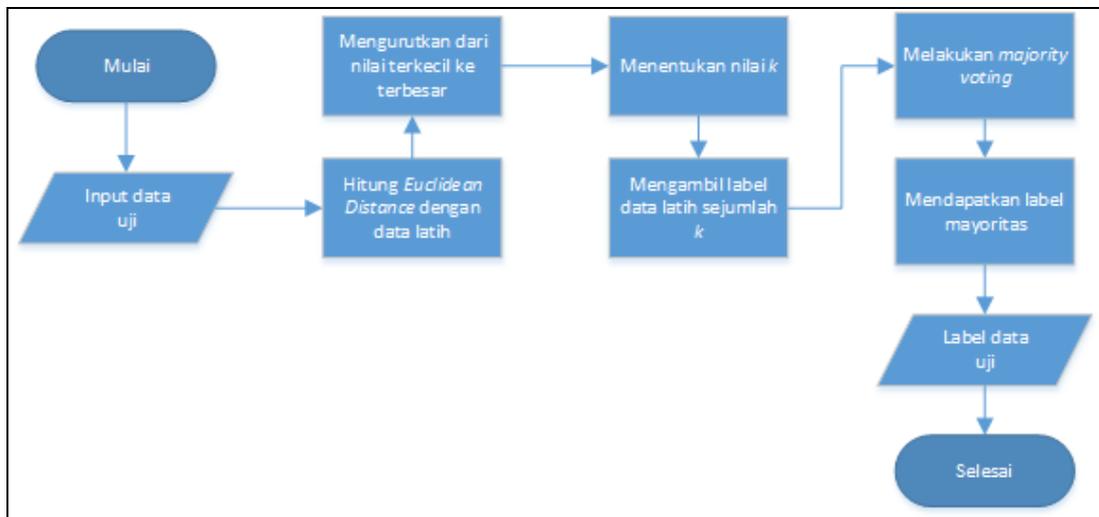
Tabel 2. Sumber Data

No.	Kelas	Jumlah
1	X RPL	36 baris data
2	X TKJ 1	37 baris data
3	X TKJ 2	36 baris data
4	XI RPL	31 baris data
5	XI TKJ 1	33 baris data
6	XI TKJ 2	35 baris data
7	Data latih baru	292 baris data
Jumlah Total		500 baris data

Penambahan data latih baru disebabkan karena data latih yang didapatkan dari pihak SMKN 2 Pekanbaru tidak mampu memenuhi kebutuhan akan label yang ada. Terdapat 10 label yang diambil berdasarkan variasi yang ada di dalam standar penilaian Kurikulum 2013. Label-label tersebut ialah nilai 'A', 'A-', 'B+', 'B', 'B-', 'C+', 'C', 'C-', 'D+', 'D'. Sementara label yang diperoleh hanya untuk nilai 'A-', 'B+', dan 'B' saja. Oleh karena itu, diputuskanlah untuk menambah data latih sebanyak 292 baris data yang menampung semua label tersebut.

### 3.2 Implementasi Algoritma K-NN

Algoritma k-NN adalah algoritma *data mining* yang digunakan untuk mengklasifikasikan suatu data uji berdasarkan data latih yang ada untuk diketahui label atau kelas dari data uji tersebut. Untuk mendapatkan label atau kelas yang dimaksud, algoritma ini menggunakan *euclidean distance* untuk menghitung jarak dengan data latih atau tetangga terdekat sejumlah  $k$  yang ditentukan [3]. Berikut ini adalah *flowchart* algoritma k-NN yang menjelaskan bagaimana itu bekerja.



Gambar 2. Flowchart k-NN

### 3.3 Pengujian Confusion Matrix

Setelah dilakukan implementasi algoritma k-NN, penelitian dilanjutkan dengan menguji tingkat akurasi algoritma berdasarkan *recognition* dengan tabel *confusion matrix*. Pengujian dilakukan untuk mengetahui variasi akurasi algoritma berdasarkan perubahan pada nilai  $k$  yaitu, dengan  $k=5$ ,  $k=10$  dan  $k=15$ . Oleh karena itu, pengujian ini melibatkan teknik *cross validation*, dimana dari 500 data latih yang tersedia, diambil sebesar 30% atau 150 data untuk dijadikan sebagai data uji. Setelah dilakukan pengujian tersebut, maka akan diperoleh hasil akurasi dan kesimpulan berdasarkan variasi nilai akurasi yang didapat.

## 4. Hasil Perancangan dan Pengujian

### 4.1 Hasil Perancangan

#### 4.1.1 Dasbor Kelas

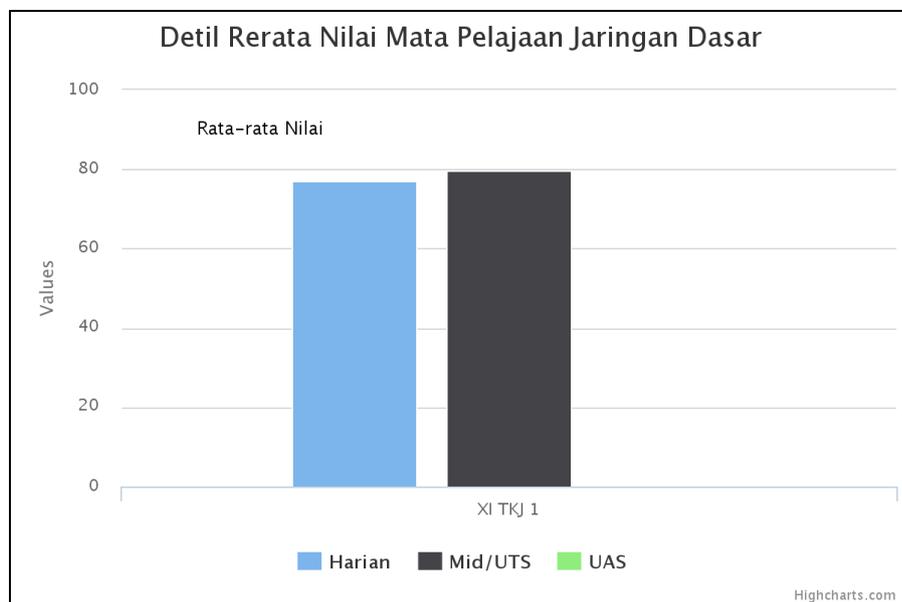
Dasbor kelas adalah *chart* yang menampilkan informasi kelas yang diampu oleh seorang guru berdasarkan mata pelajaran yang diampunya.



Gambar 3. Dasbor Detil Kelas

#### 4.1.2 Dasbor Nilai

Dasbor nilai adalah sebuah *chart* yang di dalamnya terdapat informasi rata-rata nilai harian, UTS, dan UAS pada masing-masing kelas untuk tiap mata pelajaran yang diampu oleh seorang guru.



Gambar 4. Dasbor Nilai

#### 4.1.3 Tabel Nilai Siswa

Tabel nilai menampilkan data nilai siswa per kelas pada satu mata pelajaran tertentu. Tabel ini bisa dilihat oleh guru mata pelajaran maupun wali kelas. Di tabel ini juga guru dapat melihat hasil prediksi nilai akhir per siswa atau per kelas.

#	NIS	Nama Siswa	Harian 1	Harian 2	Harian 3	Harian 4	Mid	UAS	Aksi
1	83184	ADAM PERMANA	75.00	78.00	80.00	80.00	75.00	0.00	 
2	43283	AHMAD ZAKI	80.00	82.00	84.00	80.00	80.00	0.00	 
3	86228	ANDREW ALFAREZ	74.00	80.00	88.00	76.00	78.00	0.00	 
4	75687	CHINDY ARISKA	90.00	82.00	88.00	84.00	82.00	0.00	 
5	59493	DIMITIA ANJARWATI	78.00	72.00	70.00	80.00	82.00	0.00	 
6	74120	FAISAL CANIAGO	86.00	82.00	90.00	82.00	75.00	0.00	 
7	72209	FEBRI WARDHANA	60.00	65.00	65.00	62.00	68.00	0.00	 
8	77899	FUAD LATIF	70.00	80.00	80.00	75.00	82.00	0.00	 
9	39903	IRVAN MULYADI HAMRYSON	90.00	90.00	85.00	88.00	87.00	0.00	 
10	93557	JAKA ANUGERAH	65.00	70.00	75.00	77.00	78.00	0.00	 

Showing 1 to 10 of 33 entries

Previous 1 2 3 4 Next

**Gambar 5. Tabel Nilai Siswa**

#### 4.1.4 Form Input Nilai Siswa

*Form* ini digunakan untuk mengisi nilai tiap siswa atau per kelas. Di *form* ini juga ditampilkan hasil prediksi nilai akhir per siswa. Sementara untuk mendapatkan hasil prediksi per kelas, terdapat tombol di atas di halaman guru mata pelajaran dan wali kelas untuk *generate* hasilnya ke dalam bentuk *excel file*.

NIS		Nama Lengkap	
<input type="text" value="NIS"/>		<input type="text" value="Nama Lengkap"/>	
Nilai Harian 1	Nilai Harian 2	Nilai Harian 3	Nilai Harian 4
<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Nilai Mid Semester			
<input type="text" value="0"/>			
Input via file (Nilai harian dan UTS)		Input via file (Nilai UAS)	
<input type="button" value="Choose File"/> No file chosen		<input type="button" value="Choose File"/> No file chosen	
<input type="button" value="Submit"/>	<input type="button" value="Download Template"/>	<input type="button" value="Download Template UAS"/>	

**Gambar 6. Form Input Nilai Siswa**

#### 4.1.5 Hasil Prediksi Nilai Akhir

Tujuan utama dari proyek akhir ini adalah untuk melakukan prediksi nilai akhir siswa berdasarkan variabel nilai yang telah ditentukan. Berikut ini adalah gambar dari hasil perancangan sistem aplikasi berdasarkan tujuan dari proyek akhir ini dengan mengambil satu contoh data siswa.



NIS ""		Nama Lengkap	
83184		ADAM PERMANA	
Nilai Harian 1	Nilai Harian 2	Nilai Harian 3	Nilai Harian 4
75.00	78.00	80.00	80.00
Nilai Mid Semester		Prediksi Hasil Belajar (Nilai Akhir)	
75.00		B	
Nilai UAS minimal yang dibutuhkan		Nilai UAS maksimal yang dibutuhkan	
53.5		85.5	

Gambar 7. Hasil Prediksi Nilai Akhir Per Siswa

## 4.2 Pengujian

### 4.2.1 Black Box Testing

Berikut ini adalah hasil pengujian dengan menggunakan teknik-teknik dalam metode pengujian *black box testing* [7]:

#### 1. *Equivalence Class Partitioning*

Objek yang diuji adalah *form input* nilai pada masing-masing siswa yang ada di halaman guru di menu **Akademik**. Berikut ini adalah partisi-partisi kelas *valid* dan *invalid* dari pilihan jenis *input* yang ada.

Tabel 3. Partisi Kelas

No.	Nama Kelas/Partisi	Validitas	Item
1	Angka I	Valid	0, 1, 2 ... 100
2	Angka II	Invalid	-1, -2, -3 ...
3	Angka III	Invalid	101, 102, ...
4	Huruf I	Invalid	a, b, c ... z
5	Huruf II	Invalid	A, B, C ... Z
6	Alpanumerik I	Invalid	a1, a2, a3 ...
7	Alpanumerik II	Invalid	A1, A2, A3 ...
8	Karakter Spesial	Invalid	&, ^, * ...
9	Angka dengan koma	Invalid	80,5;90,7;...
10	Angka dengan titik	Valid	90.5; 95.8; ...

Dengan menggunakan teknik ini, pengujian *input* akan jauh lebih mudah dan menghemat *cost* (waktu dan tenaga). Hal ini dikarenakan adanya ekivalensi nilai intra partisi. Hanya dengan mengambil salah satu dari *item* di masing-masing partisi, hasil dari pengujian akan selalu sama untuk tiap *item* yang ada pada kelas/partisi tersebut. Berikut ini adalah contoh tes *input* yang mewakili masing-masing kelas/partisi.

Tabel 4. Tes Input Nilai

No	Tes Input	Hasil	Kelas/Partisi
1	85	T	1
2	-85	F	2
3	850	F	3
4	a	F	4
5	A	F	5
6	a85	F	6
7	A85	F	7
8	85*/	F	8
9	85,5	F	9
10	85.5	T	10

### 2. Boundary Value Analysis

Berdasarkan kelas partisi yang telah dibuat di atas, didapatkan nilai batas atas dan batas bawah dari tiap partisi sebagai berikut:

Tabel 5. Boundary Value Analysis

No.	Nama Kelas/Partisi	Tes input batas bawah	Tes input batas atas	Validitas
1	Angka I	0	100	Valid
2	Angka II	Tak terhingga	-1	Invalid
3	Angka III	101	Tak Terhingga	Invalid
4	Huruf I	a	z	Invalid
5	Huruf II	A	Z	Invalid
6	Alpanumerik I	a1	z[tak terhingga]	Invalid
7	Alpanumerik II	A1	Z[tak terhingga]	Invalid
8	Karakter Spesial	Tak ada batasan	Tak ada batasan	Invalid
9	Angka dengan koma	0,0	100,0	Invalid
10	Angka dengan titik	0.0	100.0	Valid

### 3. Decision Table Based-Testing

Di bawah ini adalah isi dari tabel yang dihasilkan oleh pengujian *decision table based-testing* yang objek ujinya adalah *form* nilai di halaman guru dan pengembangan aplikasi di halaman admin:

Tabel 6. Decision Table Based-Testing

Kondisi	Rule											
	1	2	3	4	5	6	7	8	9	10	11	12
Nilai harian 1 sudah diinput	T	T	T	T	T	T	F	F	F	F	F	F
Nilai harian 2 sudah diinput	T	T	T	T	T	F	T	F	F	F	F	F
Nilai harian 3 sudah diinput	T	T	T	T	F	F	T	T	F	F	F	F
Nilai harian 4 sudah diinput	T	T	T	F	F	F	T	T	T	F	F	F
Nilai mid sudah diinput	T	T	F	F	F	F	T	T	T	T	F	F
Nilai UAS sudah diinput	T	F	F	F	F	F	T	T	T	T	T	F
<b>Hasil/Aksi yang diharapkan</b>												
Prediksi hasil belajar (Guru dan Wali Kelas)	F	T	F	F	F	F	F	F	F	F	F	F
Upload nilai UAS (Guru Mata Pelajaran)	F	T	F	F	F	F	F	F	F	F	F	F

### 4.3 Confusion Matrix

Berikut ini adalah hasil pengujian akurasi algoritma k-NN dengan menggunakan metode *confusion matrix* dengan nilai  $k = 5$ .

Tabel 7. Confusion Matrix dengan  $k = 5$ 

	true B	true B-	true B+	true A-	true A	true C+	true C	true C-	true D+	true D	class precision
pred. B	110	20	23	0	0	0	0	0	0	0	70.51%
pred. B-	4	11	0	0	0	0	0	0	0	0	73.33%
pred. B+	11	0	25	0	0	0	0	0	0	0	69.44%
pred. A-	0	0	0	26	7	0	0	0	0	0	78.79%
pred. A	0	0	0	14	26	0	0	0	0	0	65%
pred. C+	0	5	0	0	0	44	2	0	0	0	86.27%
pred. C	0	0	0	0	0	0	22	8	0	0	73.33%
pred. C-	0	0	0	0	0	0	10	19	4	1	55.88%
pred. D+	0	0	0	0	0	0	2	10	52	17	64.20%
pred. D	0	0	0	0	0	0	0	0	6	18	75%
Average Precision											71.175%

Dari Tabel 7 di atas, dapat disimpulkan bahwa label dengan akurasi tertinggi adalah label untuk nilai huruf C+ dengan akurasi mencapai 86.27%. Sementara label dengan akurasi terendah adalah label untuk nilai huruf C- dengan persentase 55.88%. Rata – rata presisi label dengan pengujian  $k = 5$  adalah sebesar 71.175%. Pengujian dilanjutkan dengan menguji akurasi algoritma k-NN dengan nilai  $k = 10$ . Berikut penjelasannya pada Tabel 8.

Tabel 8. Confusion Matrix dengan k = 10

	true B	true B-	true B+	true A-	true A	true C+	true C	true C-	true D+	true D	class precision
pred. B	107	21	4	3	0	0	0	0	0	0	64.85%
pred. B-	2	10	0	0	0	0	0	0	0	0	83.33%
pred. B+	16	0	14	0	0	0	0	0	0	0	46.67%
pred. A-	0	0	0	28	6	0	0	0	0	0	82.35%
pred. A	0	0	0	2	27	0	0	0	0	0	69.23%
pred. C+	0	5	0	0	0	44	3	0	0	0	84.62%
pred. C	0	0	0	0	0	0	23	7	0	0	76.67%
pred. C-	0	0	0	0	0	0	8	20	3	0	64.52%
pred. D+	0	0	0	0	0	0	2	10	55	23	61.11%
pred. D	0	0	0	0	0	0	0	0	4	13	76.47%
Average Precision											70.982%

Berdasarkan hasil dari pengujian pada Tabel 8, diketahui bahwa akurasi tertinggi dicapai oleh label untuk nilai huruf C+ dengan persentase 84.62%. Sementara label dengan akurasi terendah didapat oleh label untuk nilai huruf B+ dengan persentase 46.67%. Rata – rata presisi label dengan k = 10 adalah sebesar 70.982%. Pengujian dilanjutkan lagi dengan menggunakan nilai  $k = 15$ . Berikut ini adalah hasil dari pengujian ketiga.

Tabel 9. Confusion Matrix dengan k =15

	true B	true B-	true B+	true A-	true A	true C+	true C	true C-	true D+	true D	class precision
pred. B	107	1	5	2	0	0	0	0	0	0	64.85%
pred. B-	6	10	0	0	0	0	0	0	0	0	62.50%
pred. B+	12	0	12	1	0	0	0	0	0	0	48.00%
pred. A-	0	0	1	29	4	0	0	0	0	0	85.29%
pred. A	0	0	0	11	29	0	0	0	0	0	72.50%
pred. C+	0	5	0	0	0	44	4	0	0	0	83.02%
pred. C	0	0	0	0	0	0	20	7	0	0	74.07%
pred. C-	0	0	0	0	0	0	10	19	4	0	57.58%
pred. D+	0	0	0	0	0	0	2	11	53	20	61.63%
pred. D	0	0	0	0	0	0	0	0	5	16	76.19%
Average Precision											68.563%

Melihat hasil pengujian pada Tabel 9 diperoleh informasi sebagai berikut. Akurasi tertinggi dimiliki oleh label untuk nilai huruf A- dengan akurasi mencapai 85.29%. Sementara itu, akurasi terendah dimiliki oleh label untuk nilai huruf B+ dengan akurasi 48%. Rata – rata presisi label dengan k = 15 adalah sebesar 68.563%.

## 5. Kesimpulan

Berdasarkan hasil pengujian yang telah dilakukan, didapatkan kesimpulan bahwa dengan jumlah data latih sebanyak 500 baris data, penentuan nilai  $k = 5$  adalah pilihan yang tepat untuk

memperoleh akurasi yang lebih tinggi dengan rata – rata presisi label sebesar 71.175%. Selain itu, pengujian dengan menggunakan teknik *black box testing* juga dapat membantu para pengembang dalam menganalisis kesalahan yang mungkin muncul dari sistem aplikasi yang dibangun. Kemudian, setelah dilakukan pengisian kuesioner kepada 10 responden, diperoleh hasil berdasarkan *rating scale* sebesar 82%, dimana nilai tersebut berada diantara skala setuju dan sangat setuju. Sehingga dapat disimpulkan bahwa aplikasi yang telah dibangun dapat mempermudah guru maupun wali kelas dalam proses input dan konversi nilai.

### Daftar Pustaka

- [1]. I.L. Pasaribu dan B. Simandjuntak. 1983. Metode Belajar dan Kesulitan Belajar. Bandung. Tarsito.
- [2]. Kementerian Pendidikan dan Kebudayaan R.I Bidang Pendidikan Periode 2009-2014. 2013. *Konsep dan Implementasi Kurikulum 2013*. Jakarta.
- [3]. Han, Jiawei & Kamber, Micheline. 2006. *Data Mining: Concepts and Techniques Second Edition*. San Fransisco. Morgan Kauffman Publisher.
- [4]. Wook dkk. 2009. *Predicting NDUM Student's Academic Performance Using Data Mining Techniques*. *IEEE Computer Society Journal*
- [5]. M. Abu Tair, Mohammed & M. El-Halees Alaa. 2012. *Mining Educational Data to Improve Student's Performance*. *International Journal of Information and Communication Technology Research, Vol. 2 No. 2*.
- [6]. K. Baradwaj, Brijesh & Pal, Saurabh. 2011. *Mining Educational Data to Analyze Students' Performance*. *International Journal of Advanced Computer Science and Applications, Vol 2, No. 6.Tabel 0.1.1*
- [7]. Burnstein, Ilene. 2002. *Practical Software Testing*. New York. Springer.

