

NASKAH PUBLIKASI

**KLASIFIKASI PEMINATAN SISWA DENGAN METODE
K-NEAREST NEIGHBOR**
(Studi kasus SMP Negeri 1 Pabuaran)



Disusun oleh:

WILIANA ARIE MUNANDA

5150411239

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN ELEKTRO
UNIVERSITAS TEKNOLOGI YOGYAKARTA
2020**

NASKAH PUBLIKASI

**KLASIFIKASI PEMINATAN SISWA DENGAN METODE
K-NEAREST NEIGHBOR**
(Studi kasus SMP Negeri 1 Pabuaran)



Pembimbing



Saucha Diwandari, S.Kom., M.Eng

Tanggal: 5/12/2019

KLASIFIKASI PEMINATAN SISWA DENGAN METODE K-NEAREST NEIGHBOR

(Studi kasus SMP Negeri 1 Pabuaran)

WILIANA ARIE MUNANDA

*Program Studi Informatika, Fakultas Teknologi Informasi dan Elektro
Universitas Teknologi Yogyakarta*

Jl. Ringroad Utara Jombor Sleman Yogyakarta

E-mail : wilianaariemunanda@gmail.com

ABSTRAK

Salah satu parameter utama tercapainya efektivitas pembelajaran adalah dengan tercapainya tujuan pembelajaran siswa secara optimal sesuai standar kompetensi kelulusan. Hasil evaluasi yang sangat beragam berbagai cara dilakukan untuk melakukan evaluasi agar hasil evaluasi sesuai dengan kompetensi minat dan bakat siswa, namun kemampuan siswa yang berbeda-beda menjadi salah satu masalah tersendiri dalam menentukan kebijakan kegiatan belajar yang sesuai dengan kompetensi minat dan bakat siswa. Minat dan bakat siswa yang berbeda-beda menjadi salah satu masalah tersendiri dalam menentukan kegiatan belajar dibutuhkan siswa. Oleh sebab itu dibuatlah sistem Klasifikasi Peminatan Siswa Berdasarkan Ujian Menggunakan Metode k Nearest Neighbor (kNN). Algoritma k Nearest Neighbor adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dengan dengan objek tersebut. Implementasi k Nearest Neighbor (kNN) telah dilakukan dengan menggunakan 243 data dengan perbandingan data latih dan data uji 70%:30% yaitu sebanyak 170 data latih dan 73 data uji dengan $k = 5$ memberikan hasil prediksi besar sebanyak 61 data dan prediksi salah sebanyak 12 data dengan akurasi sebesar 83.56%.

Kata kunci : Minat dan Bakat, Nilai Siswa, k Nearest Neighbor

1. PENDAHULUAN

Undang-undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional menegaskan bahwa, pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya masyarakat, bangsa dan negara. Pendidikan tidak lepas dari interaksi yang dilakukan antara siswa dengan guru di sekolah. Sederhananya, sekolah menjadi salah satu tempat sarana untuk guru memberikan ilmu pengetahuan, keterampilan maupun teladan sikap terpuji, sedangkan siswa menerima ilmu pengetahuan dan keterampilan guna menjadi pribadi yang baik. Sekolah akan berusaha memberikan yang terbaik kepada siswanya guna meningkatkan kualitas pendidikannya. Salah satu aspek sebagai indikator kualitas di sekolah adalah nilai ujian siswa, selain itu kemajuan teknologi memberi dampak yang besar

terhadap dunia pendidikan dimana masyarakat dituntut untuk selalu berkembang setiap tahunnya supaya tidak ketinggalan informasi..

Hasil evaluasi yang sangat beragam berbagai cara dilakukan untuk melakukan evaluasi agar hasil evaluasi sesuai dengan kompetensi minat dan bakat siswa, namun kemampuan siswa yang berbeda-beda menjadi salah satu masalah tersendiri dalam menentukan kebijakan kegiatan belajar yang sesuai dengan kompetensi minat dan bakat siswa. Selain itu keterlibatan faktor luar, seperti keluarga dan lingkungan yang mempengaruhi psikologis siswa yang mencakup minat, bakat, dan motivasi siswa untuk belajar diperlukan adanya bimbingan yang tepat dari wali di sekolah.

Metode klasifikasi yang digunakan adalah metode k-nearest neighbor untuk menghasilkan klasifikasi siswa dimana hasil dari sample uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori knn yang bertujuan untuk mengetahui kemampuan siswa sehingga diharapkan hasil dari penelitian ini dapat memberikan rujukan atau acuan

bagi guru atau bagian konseling dalam memberikan bimbingan yang dibutuhkan siswa. Pemilihan metode k-Nearest Neighbor (KNN) dalam penelitian ini karena metode ini merupakan salah satu metode klasifikasi yang mengklasifikasikan objek uji berdasarkan klasifikasi mayoritas, dimana dengan mencari kelompok K objek dalam data training yang paling dekat (mirip) dengan objek data uji baru.

Sulitnya menentukan kegiatan belajar serta bimbingan yang sesuai dengan kompetensi minat dan bakat siswa sehingga diperlukan suatu cara tersendiri yang dapat membantu menentukan kegiatan belajar dan bimbingan yang dibutuhkan, oleh sebab itu dibuatlah sistem Klasifikasi Peminatan Siswa Berdasarkan Ujian yang dapat mengklasifikasi siswa sesuai kemampuan apakah siswa tersebut sudah cukup atau kurang dalam suatu pelajaran sehingga guru di sekolah dapat menentukan porsi belajar dan bimbingan yang dibutuhkan siswa sesuai klasifikasi yang dihasilkan.

2. LANDASAN TEORI

2.1. Algoritma K-Means

Menurut Agustin, F, E, M ., dkk. (2015) Algoritma k-means merupakan metode pengklasteran secara partitioning yang memisahkan data ke dalam kelompok yang berbeda. Dengan partitioning secara iteratif, k-means mampu meminimalkan rata-rata jarak setiap data ke klasternya. Dalam algoritma K-Means setiap data harus termasuk ke cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya dapat berpindah ke cluster yang lain. Berdasarkan jumlah cluster yang perlu di inputkan serta atribut yang dimiliki merupakan nilai mata pelajaran Bahasa Indonesia, Bahasa Inggris, Matematika, dan IPA yang bertipe numerik maka dipilihkan K-Means sebagai metode kelasterisasi data.

Algoritma K-Means pada dasarnya melakukan 2 proses yakni proses pendekatan lokasi pusat cluster dan proses pencarian anggota tiap-tiap cluster, seperti sebagai berikut:

1. Tentukan jumlah cluster yang ingin di bentuk. Tetapkan pusat cluster.
2. Hitung jarak setiap data ke pusat cluster menggunakan Euclidean.

$$D(x,y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (1)$$

3. Kelompokkan data kedalam cluster dengan jarak yang paling pendek menggunakan persamaan berikut:

$$\text{Min} \sum_k^n D(x,y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (2)$$

4. Hitung pusat cluster yang baru menggunakan persamaan berikut:

$$C(x,y) = \frac{\sum_{k=1}^n x_{ij}}{p} \quad (3)$$

Dimana

$$x_{ij} \in \text{Kluster ke } - k$$

5. Ulangi langkah 2 sampai 4 hingga sudah tidak ada lagi data yang berpindah ke cluster lain.

2.2. Algoritma K-Nearest Neighbor

Menurut Ndaumanu, R, I ., dkk. (2014) Algoritma k-Nearest Neighbor (kNN) adalah algoritma yang termasuk dalam kelompok instance based learning. Instance based learning adalah algoritma pembelajaran yang melakukan generalisasi eksplisit membandingkan masalah baru dengan pola yang ada yang terlihat dalam pelatihan yang telah di simpan dalam memori. Algoritma ini juga merupakan salah satu teknik lazy learning. kNN dilakukan dengan mencari kelompok k objek data training yang paling dekat (mirip) dengan objek pada data baru atau data testing. Algoritma k-Nearest Neighbor adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dengan dengan objek tersebut. Nearest Neighbor sendiri adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dan kasus lama yaitu berdasarkan pencocokan bobot dari sejumlah fitur yang ada.

Untuk mendefinisikan jarak antara dua titik yaitu titik pada data training dan titik pada data testing maka digunakan rumus euclidean seperti yang ditunjukkan pada persamaan.

$$D(x,y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (4)$$

Dengan D adalah jarak antara titik pada data training x dan titik data testing y yang akan di klasifikasi.

Langkah-langkah untuk menghitung metode Algoritma K-Nearest Neighbor:

1. Menentukan Parameter K (Jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak *Euclidean (query instance)* masing-masing objek terhadap data sampel yang diberikan.
3. Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak Euclid terkecil.
4. Mengumpulkan kategori Y (*Klasifikasi Nearest Neighbor*)
5. Dengan menggunakan kategori Nearest Neighbor yang paling mayoritas maka dapat diprediksi nilai *query instance* yang telah dihitung.

2.3. Confusion Matrix

Menurut Rahman, M, F., dkk (2017) Confusion matrix adalah suatu metode yang biasanya di gunakan untuk melakukan perhitungan akurasi pada konsep data mining. Confusion Matrix digambarkan dengan tabel yang menyatakan jumlah data uji yang benar dan jumlah data uji yang salah diklasifikasikan. Tabel *Confusion Matrix* dapat dilihat pada Tabel 1

Tabel 1 Tabel Confusion Matrix

Correct Classification	Classified as	
	Predicted "+"	Predicted "-"
Actual "+"	True Positives	False Negatives
Actual "-"	False Positives	True Negatives

Berdasarkan Tabel *Confusion Matrix* diatas:

1. True Positives (TP) adalah jumlah record data positif yang di klasifikasikan sebagai nilai positif.
2. False Positives (FP) adalah jumlah record data negatif yang diklasifikasikan sebagai nilai positif.
3. False Negatives (FN) adalah jumlah record data negatif yang diklasifikasikan sebagai nilai negatif.
4. True Negatives (TN) adalah jumlah record data positif yang diklasifikasikan sebagai nilai negatif.

Nilai yang dihasilkan melalui metode *Confusion Matrix* adalah berupa evaluasi sebagai berikut:

1. *Accuracy*, Presentase jumlah record data yang diklasifikasikan (Prediksi) secara benar oleh algoritma.

$$Accuracy = \frac{TP + TN}{total\ data} \quad (5)$$

2. *Misclassificationfunction (Error) Rate*, presentasi jumlah record data yang diklasifikasikan (prediksi) secara salah oleh algoritma.

$$Missfunction\ Rate = \frac{FP + FN}{total\ data} \quad (6)$$

3. METODOLOGI PENELITIAN

3.1 Pengumpulan Data

Pengumpulan data adalah suatu metode dan prosedur yang digunakan untuk mendapatkan suatu informasi tentang apa saja yang harus dikerjakan pada saat pembangunan sistem informasi pengelolaan persediaan barang. Pada metode ini terdapat beberapa kegiatan yang dilakukan, diantaranya sebagai berikut:

1. Melakukan observasi ke SMP N 1 Pabuaran khususnya bagian kurikulum, karena bagian

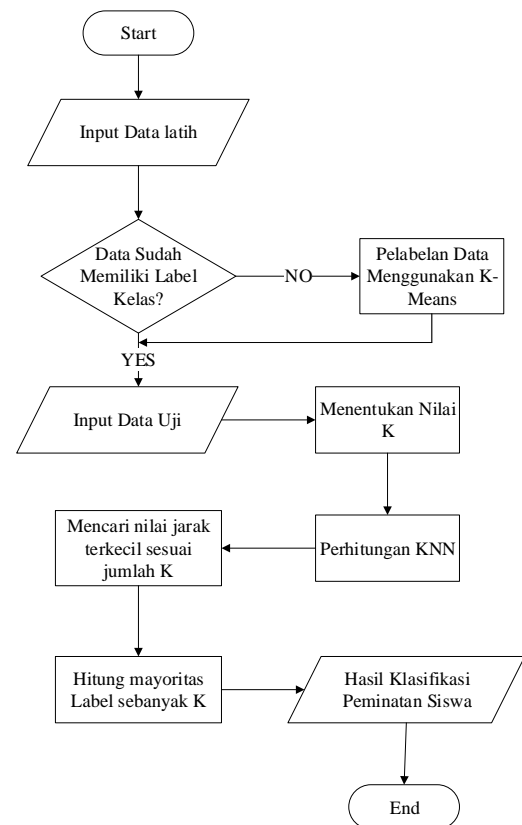
kurikulumlah yang membawa data nilai siswa yang dibutuhkan untuk penelitian ini.

2. Wawancara Pak Eris Heriyono M.Pd. selaku petugas kurikulum di SMP N 1 Pabuaran dan Pak Drs. H. Yusep Riyadi M.Pd. selaku kepala sekolah SMP N 1 Pabuaran.
3. Pengambilan data, proses mendapatkan data dari seorang atau sebuah instansi. Pada penelitian ini keperluan data diambil dengan meminta data dari Bapak Eris Heriyono M.Pd selaku kurikulum di SMP N 1 Pabuaran sekaligus yang memegang data nilai tryout siswa.

3.2 Perancangan Sistem

3.2.1 Flowchart Sistem

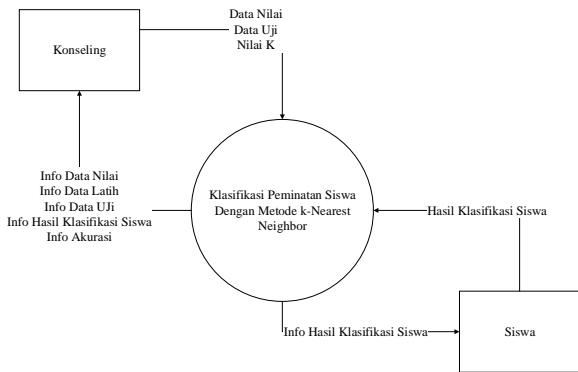
Flowchart Klasifikasi Peminatan Siswa dengan Metode K-Nearest Neighbor dapat dilihat pada Gambar 1.



Gambar 1 Flowchart Klasifikasi Peminatan

3.2.2 Diagram Konteks

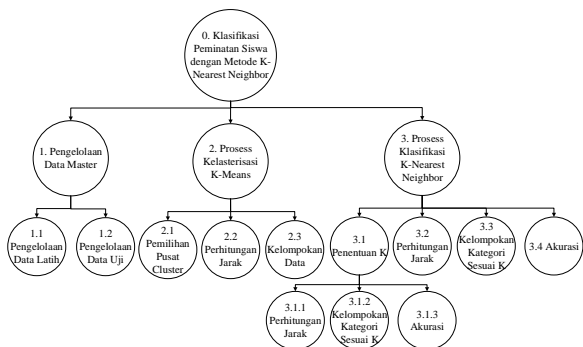
Diagram Konteks adalah gambaran sistem secara garis besar. Dalam diagram konteks ini memperlihatkan bahwa pengguna harus memasukan data latih terlebih dahulu sebelum menguji data uji. Rancangan DAD pada diagram konteks Sistem Klasifikasi Peminatan Siswa Dengan Metode k-Nearest Neighbor dapat dilihat pada Gambar 2.



Gambar 2 Diagram Konteks

3.2.3 Diagram Jenjang

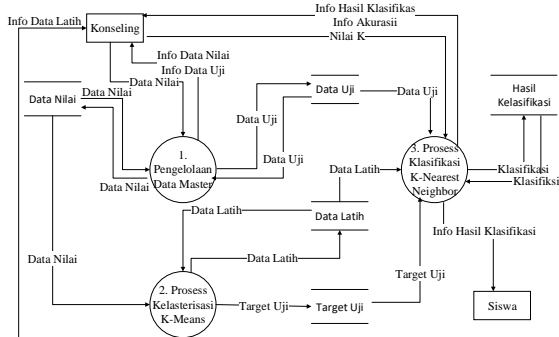
Diagram ini menggambarkan hubungan dari seluruh fungsi fungsi didalam sistem secara berjenjang. Rancangan diagram jenjang Sistem Klasifikasi Peminatan Siswa Dengan Metode k-Nearest Neighbor dapat dilihat pada Gambar 3.



Gambar 3 Diagram Jenjang

3.2.4 DAD Level 1

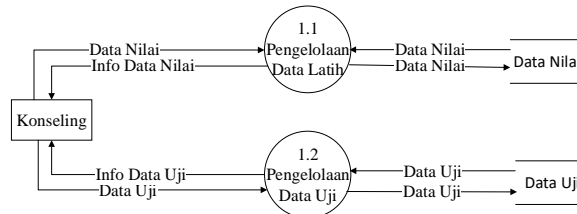
DAD Level 1 merupakan suatu proses yang di buat untuk menggambarkan asal dan tujuan data yang keluar dari sistem serta proses ang terjadi didalam sistem. Rancangan DAD Level 1 Sistem Klasifikasi Peminatan Siswa Dengan Metode kNearest Neighbor dapat dilihat pada Gambar 4.



Gambar 4 DAD Level 1

3.2.5 DAD Level 2 Prosess 1

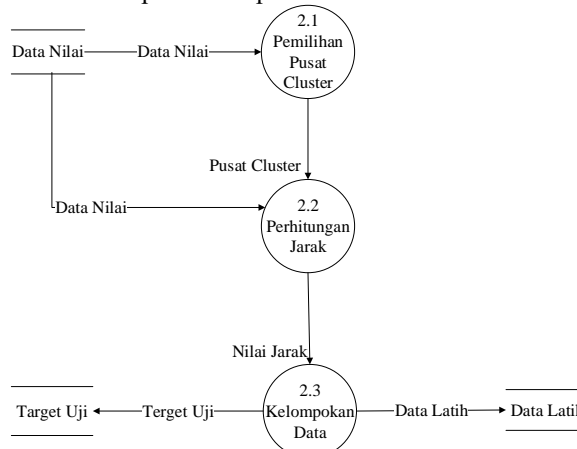
DAD Level 2 Prosess 1 adalah proses pengelolaan master data yang akan digunakan sebagai dataset. Rancangan Diagram Arus Data Level 2 Prosess 1 Sistem Klasifikasi Peminatan Siswa Dengan Metode k-Nearest Neighbor dapat dilihat pada Gambar 5.



Gambar 5 DAD Level 2 Prosess 1

3.2.6 DAD Level 2 Prosess 2

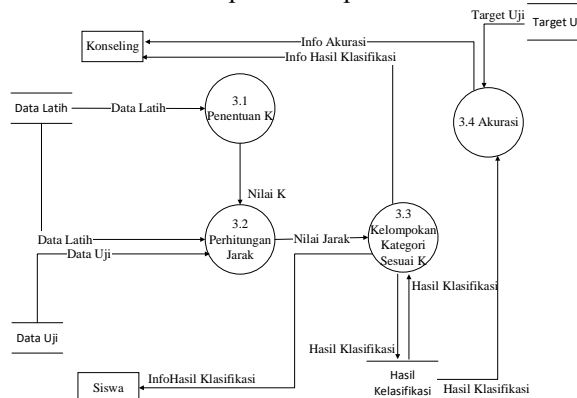
DAD Level 2 Prosess 2 adalah proses pelabalan terhadap data yang akan digunakan sebagai data latih. Rancangan Diagram Arus Data Level 2 Prosess 2 dapat dilihat pada Gambar 6.



Gambar 6 DAD Level 2 Prosess 2

3.2.7 DAD Level 2 Prosess 3

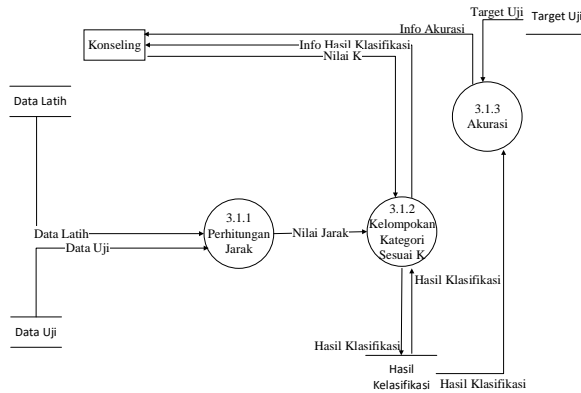
DAD Level 2 Prosess 3 adalah proses pengklasifikasian data uji menggunakan metode k-Nearest Neighbor. Rancangan Diagram Arus Data Level 2 Prosess 3 dapat dilihat pada Gambar 7.



Gambar 7 DAD Level 2 Prosess 3

3.2.8 DAD Level 3 Proses 1

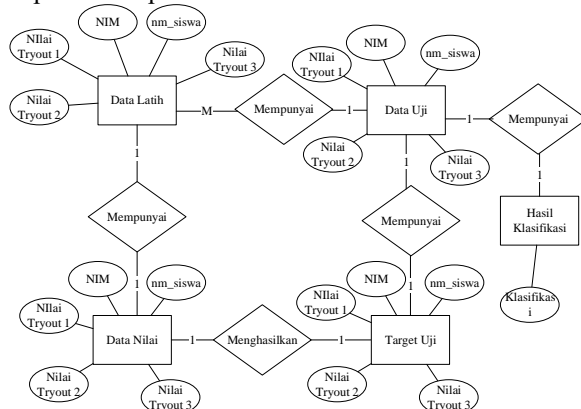
DAD Level 3 Proses 1 adalah alur proses penentuan nilai k. Rancangan Diagram Arus Data Level 3 Proses 1 dapat dilihat pada Gambar 8.



Gambar 8 DAD Level 3 Proses 1

3.2.9 Entity Relationship Diagram (ERD)

Entitas (entity) yang ada dalam Klasifikasi Peminatan Siswa ini adalah sebagai berikut: Entitas Data Nilai, Entitas Data Lath, Entitas Data Uji, Entitas Data Target Uji, Hasil ERD pada Klasifikasi Peminatan Siswa Dengan Metode k-Nearest Neighbor dapat dilihat pada Gambar 9.



Gambar 9 ERD Klasifikasi Peminatan Siswa

4. HASIL DAN PEMBAHASAN

4.1. Perhitungan Manual

Pada perhitungan manual ini, data yang digunakan dalam proses manualisasi data adalah dataset sebanyak 243 dataset dengan masing-masing mata pelajaran Bahasa Indonesia, Bahasa Inggris, Matematika, dan IPA. Data nilai yang digunakan adalah data nilai yang belum memiliki label sehingga sebelum dilakukan perhitungan klasifikasi menggunakan Metode k-Nearest Neighbor data terlebih dahulu di kelasterisasi menggunakan Metode K-Means.

4.1.1 Kelasterisasi Data

Kelasterisasi data sebagai bentuk pelabelan terhadap data latih di lakukan dengan menggunakan parameter-parameter berikut: Jumlah Cluster: 3 Jumlah Data: 243 Jumlah Atribut: 4 (Bahasa Indonesia, Bahasa Inggris, Matematika dan IPA). Untuk dataset yang digunakan dapat dilihat pada Tabel 2.

Tabel 2 Tabel Datasets

No.	Nama Siswa	Rata-Rata Bahasa Indonesia	Rata-Rata Bahasa Inggris	Rata-Rata MTK	Rata-Rata IPA
1	ABDUL FATAH	59,3	50,0	43,8	52,5
2	ABDUL MUNIP	52,7	33,8	42,2	53,3
3	ADAM PRATAMA	64,7	47,2	39,7	51,7
4	ALFINA LISTIANI	64,7	43,0	42,3	52,5
...
...
240	WINDA NUREHA	62,0	54,7	49,7	54,2
241	YUGIS GAMARA	64,7	50,7	38,5	44,2
242	YUNENSIH	62,7	50,0	38,8	34,2
243	ZAMRONI USMAN	58,7	42,0	36,3	35,0

4.1.1.1 Penentuan Pusat Awal Cluster

Untuk penentuan awal diasumsikan dengan menggunakan nilai tertinggi pada masing-masing mata pelajaran. Nilai Centroid Awal dapat dilihat pada Tabel 3.

Tabel 3 Tabel Centroid Awal

29	SUSI JUNIYANTI	77,3	58,0	49,7	54,2
11	ELA AMELIA	68,7	68,0	58,8	51,7
9	DESI KURNIASIH	61,3	56,7	63,0	60,8
90	WAHENDRA FAROMADONI	64,0	58,7	49,7	63,3

4.1.1.2 Perhitungan Jarak Pusat Cluster

Untuk mengukur jarak antara data dengan pusat cluster digunakan euclidean distance. Sebagai contoh perhitungan jarak dari data ke -1 terhadap pusat cluster:

$$C0 = \sqrt{(77,3 - 59,3)^2 + (58 - 50)^2 + (49,7 - 43,8)^2 + (54,2 - 52,5)^2}$$

$$C1 = \sqrt{(68,7 - 59,3)^2 + (68 - 50)^2 + (58,8 - 43,8)^2 + (51,7 - 52,5)^2}$$

$$C2 = \sqrt{(61,3 - 59,3)^2 + (56,7 - 50)^2 + (63 - 43,8)^2 + (60,8 - 52,5)^2}$$

$$C3 = \sqrt{(64 - 59,3)^2 + (58,7 - 50)^2 + (49,7 - 43,8)^2 + (63,3 - 52,5)^2}$$

Dan seterusnya di lanjutkan untuk data ke 2,...N kemudian akan di dapatkan perhitungan jarak seperti terlihat pada Tabel 4.

Tabel 4 Tabel Jarak Tiap Cluster

No.	Nama Siswa	C0	C1	C2	C3
1	ABDUL FATAH	20,61	25,24	22,03	15,76
2	ABDUL MUNIP	35,35	41,28	32,97	30,02
3	ADAM PRATAMA	19,60	28,59	27,02	19,20
....
...
239	UMROH	39,62	44,28	43,93	38,91
240	WINDA NUREHA	15,69	17,68	15,06	10,20
241	YUGIS GAMARA	20,95	28,04	30,42	23,59
242	YUNENSIH	28,22	32,65	36,62	32,33
243	ZAMRONI USMAN	33,91	39,50	40,01	35,87

Kolom C0, C1, C2, C3 menunjukkan nilai jarak data terhadap pusat cluster. Baris pertama pada tabel menunjukkan nilai jarak data terhadap titik pada pusat cluster pertama, baris kedua pada tabel menunjukkan nilai jarak data terhadap titik pusat cluster kedua dan seterusnya.

4.1.1.3 Pengelompokan Data

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat data dengan pusat cluster, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat cluster terdekat. Berikut ini ditampilkan data pengelompokan data pada Tabel 5.

Tabel 5 Tabel Anggota Cluster

	1	2	3	4	241	242	243
C0	1	1	1	1				0	0	0
C1	1	1	1	1				1	1	1
C2	1	1	1	1				1	1	1
C3	0	0	0	0				1	1	1

Keterangan:

- Jika nilai jarak lebih dekat jaraknya dari nilai hasil perhitungan pusat centroid ke-1(C0) atau ke-2(C1) atau ke-3(C2) atau data ke-4(C4) maka nilai G1 = 1 dan tidak termasuk kelompok cluster baru.
- Jika nilai jarak lebih jauh jaraknya dari nilai hasil perhitungan pusat centroid ke-1(C0) atau ke-2(C1) atau data ke-3(C2) atau ke-4(C4) maka nilai G1 = 0 dan termasuk grup atau kelompok cluster baru.

Berdasarkan perhitungan Iterasi 1 didapatkan: Bahasa = 112 data Matematika = 12 data IPA = 119 data.

4.1.1.3 Penentuan Pusat Cluter Baru

Setelah diketahui anggota tiap-tiap cluster kemudian pusat cluster baru di hitung berdasarkan data anggota tiap-tiap cluster yang sesuai. sehingga didapatkan perhitungan sebagai berikut:

$$C0 = \frac{6391}{105}, \frac{4884,67}{105}, \frac{3567,33}{105}, \frac{3989}{105}$$

Lakukan lah hal yang sama pada C1, C2, dan C3 sehingga didapat nilai cluster baru seperti pada Tabel 6.

Tabel 6 Tabel Cluster Baru

C0	61,45714	46,90159	34,20159	38,4873
C1	60,95238	63,71429	47,64286	46,90476
C2	61,55556	53,38889	57,23611	55,76389
C3	55,81793	47,7563	37,23109	45,43838

Ulangi langkah ke 2(kedua) hingga posisi data tidak mengalami perubahan. Iterasi akan terus di lakukan hingga nilai G pada tiap-tiap iterasi memiliki nilai yang sama. Pada saat nilai G yang sama telah terjadi maka clustering telah mencapai stabil dan konvergen. Dan pada proses manual ini nilai yang konvergen terjadi pada iterasi ke-14 seperti pada Tabel 7.

Tabel 7 Tabel Anggota Cluster

	1	2	3	4	241	242	243
C0	1	1	1	1				1	1	1
C1	0	1	0	0				0	1	1
C2	1	1	1	1				1	1	1
C3	1	0	1	1				1	0	0

Diketahui data kelasterisasi yang didapatkan Bahasa = 102 data Matematika = 41 data IPA = 100 data.

4.1.2 Klasifikasi Data

Setelah dilakukan pelabelan data, barulah proses klasifikasi dengan menggunakan Metode k-Nearest Neighbor dapat dilakukan. Pada saat klasifikasi data testing ditentukan parameter K=5 dengan cara menghitung jarak euclidean masing-masing objek data uji terhadap data training. Tabel data training tryout 1 dapat dilihat pada Tabel 8. Tabel data training tryout 2 dapat dilihat pada Tabel 9. Tabel data training tryout 3 dapat dilihat pada Tabel 10. Tabel hasilkelasterisasi data training dapat dilihat pada Tabel 11.

Tabel 8 Tabel data training tryout 1

Nama Siswa	Tryout 1			
	Bindo	BIng	MTK	IPA
ABDUL FATAH	64	40	30	32,5
ABDUL MUNIP	48	26	17,5	47,5
ADAM PRATAMA	70	36	22,5	35
ALFINA LISTIANI	68	36	37,5	37,5
....
...
MUHAMAD ARIF ARFIANSYAH	70	30	15	15
MUHAMMAD ROHIT	80	52	35	40
NATASYA NURAINY	76	32	35	35
NURLAELA	70	44	25	42,5

Tabel 9 Tabel data training tryout 2

Nama Siswa	Tryout 2			
	Bindo	B Ing	MTK	IPA
ABDUL FATAH	60	44	45	60
ABDUL MUNIP	52	35,5	55	45
ADAM PRATAMA	66	50	65	62,5
ALFINA LISTIANI	60	48	37,5	60
....
...
MUHAMAD ARIF ARFIANSYAH	60	54	35	37,5
MUHAMMAD ROHIT	60	66	35	52,5
NATASYA NURAINY	66	66	32,5	37,5
NURLAELA	52	38	15	35

Tabel 10 Tabel data training tryout 3

Nama Siswa	Tryout 3			
	Bindo	B Ing	MTK	IPA
ABDUL FATAH	54	66	56,5	65
ABDUL MUNIP	58	40	54	67,5

ADAM PRATAMA	58	55,5	31,5	57,5
ALFINA LISTIANI	66	45	52	60
....
...
MUHAMAD ARIF ARFIANSYAH	56	72	54	60
MUHAMMAD ROHIT	64	68	49	47,5
NATASYA NURAINY	68	80	51,5	45
NURLAELA	56	40	29	30

Tabel 11 Tabel hasil klasterisasi data

Nama Siswa	Kelaster isasi
ABDUL FATAH	Bahasa
ABDUL MUNIP	IPA
ADAM PRATAMA	Bahasa
ALFINA LISTIANI	Bahasa
....
...	...
MUHAMAD ARIF ARFIANSYAH	IPA
MUHAMMAD ROHIT	Bahasa
NATASYA NURAINY	Bahasa
NURLAELA	Bahasa

Untuk tabel data uji tryout 1 dapat dilihat pada Tabel 12. Data uji tryout 2 dapat dilihat pada Tabel 13. Data uji tryout 3 dapat dilihat pada Tabel 14.

Tabel 12 Tabel data uji tryout 1

Nama Siswa	Tryout 1			
	B indo	B Ing	MTK	IPA
PIPIT RAMDANI	40	38	22,5	32,5
RAHAYU	76	46	25	32,5
RAMA PERMANA	58	32	30	35
....
....
YUGIS GAMARA	62	52	27,5	35
YUNENSIH	74	54	55	42,5
ZAMRONI USMAN	58	44	25	35

Tabel 13 Tabel data uji tryout 2

Nama Siswa	Tryout 2			
	B indo	B Ing	MTK	IPA
PIPIT RAMDANI	58	42	22,5	55
RAHAYU	48	34	57,5	35
RAMA PERMANA	52	40	20	52,5
....
....
YUGIS GAMARA	60	42	34	45
YUNENGSIH	56	48	30	30
ZAMRONI USMAN	68	54	40	45

Tabel 14 Tabel data uji tryout 3

Nama Siswa	Tryout 3			
	B indo	B Ing	MTK	IPA
PIPIT RAMDANI	44	56	64	52,5
RAHAYU	62	48	36,5	15
RAMA PERMANA	46	32	36,5	37,5
....
....
YUGIS GAMARA	72	58	54	52,5
YUNENGSIH	58	48	31,5	30
ZAMRONI USMAN	50	28	44	25

4.2 Pengujian Data

Pengujian dilakukan terhadap 73 data uji sampel nilai Tryout siswa SMP N1 Pabuaran. Hasil yang diperoleh akan dicocokkan dengan nilai K yang dipilih. Pengujian dilakukan terhadap jumlah data latihan yang berbeda, dalam pengujian ini diambil 4 set data latihan yang masing-masing berjumlah 40, 80, 120 dan 170 dengan nilai K yang digunakan adalah K = 5. Pengujian Jumlah data latihan dapat dilihat pada Tabel 15.

Tabel 15 Tabel Pengujian Jumlah Data Latihan

No	Jml Data Latihan	Indeks yg Sesuai	Akurasi
1	40	52	71.23%
2	80	58	79.45%

3	120	58	79.45%
4	170	61	83.56%

Dari hasil pengujian diatas, dapat dilihat data latihan yang berjumlah 170 memiliki nilai akurasi lebih tinggi dari ke empat jumlah data latihan yang berbeda yaitu sebesar 83.56% dan akurasi terkecil berada didata latihan yang berjumlah 40 data dengan akurasi sebesar 71.23%.

4.3 Mencari Nilai K Optimal

Pada tahap ini pengujian dilakukan terhadap nilai K yang dimana ada beberapa nilai K yang akan digunakan. Data latihan yang digunakan adalah sebanyak 170 karena memiliki akurasi tertinggi. Adapun nilai K yang digunakan adalah dari 1 sampai 9. Pengujian nilai K dapat dilihat pada Tabel 16.

Tabel 16 Tabel Pengujian Nilai K

Nilai K	Jumlah data sesuai	Akurasi
1	55	75.34%
2	42	57.53%
3	57	78.08%
4	52	71.23%
5	61	83.56%
6	56	76.71%
7	61	83.56%
8	57	78.08%
9	61	83.56%

Dari hasil pengujian diatas di temukan nilai K dengan akurasi tinggi adalah K= 5, K = 7, dan K = 9 yaitu dengan jumlah data yang sesuai 61 data dan hasil akurasi sebanyak 83.56%. Dapat di simpulkan K yang optimal yang dapat digunakan adalah K = 5 atau K = 7 K = 9. Pada penelitian ini di tetapkan K = 5 karena nilai k yang besar akan membuat batas antara setiap klasifikasi menjadi lebih kabur, jadi dipilih nilai k terkecil dengan presentasi akurasi peling besar yaitu k = 5.

4.4. Tampilan Program

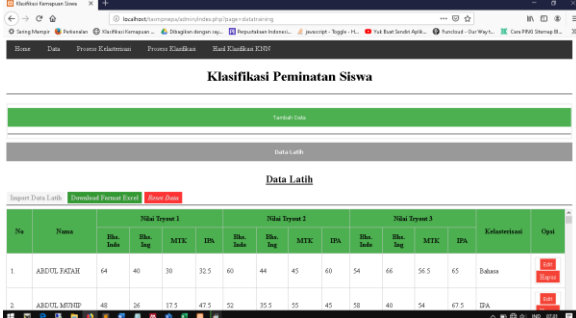
1. Tampilan Halaman Home (Dashboard)

Halaman Home merupakan halaman awal dari program yang dibuat karena sebelum memulai menjalankan sistem maka akan ditampilkan halaman home (dashbard) ini. Tampilan halaman home dapat dilihat pada Gambar 10.



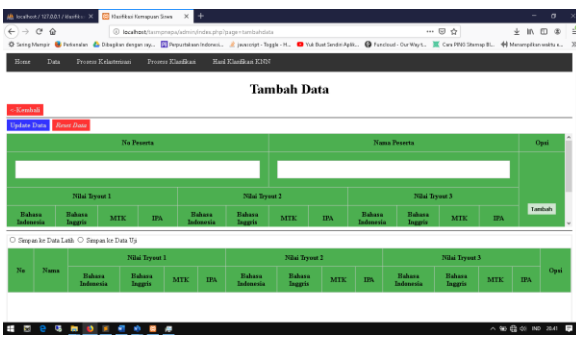
Gambar 10 Tampilan Halaman Home (Dashboard)

2. Tampilan Halaman Data
 Halaman ini dibuat untuk menampilkan data training dan data latih yang digunakan pada penelitian ini. Tampilan halaman data dapat dilihat pada Gambar 11.



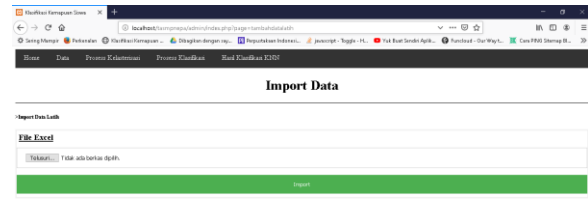
Gambar 11 Tampilan Halaman Data

3. Tampilan Halaman Tambah Data
 Halaman ini dibuat untuk menambahkan data yang akan digunakan dalam penelitian. Tampilan halaman data latih dapat dilihat pada Gambar 12.



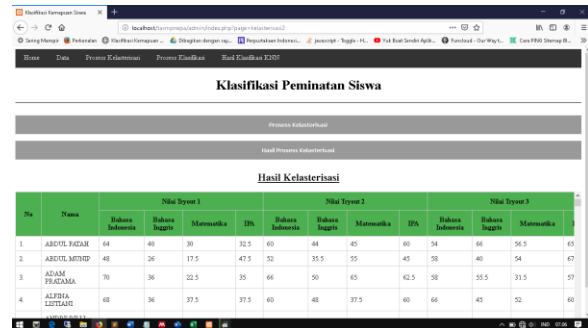
Gambar 12 Tampilan Halaman Tambah Data

4. Tampilan Halaman Import Data
 Halaman data uji dibuat untuk mengimport dataset dalam bentuk excel agar tidak perlu menginputkan data secara manual. Implementasi Import data dapat dilihat pada Gambar 13.



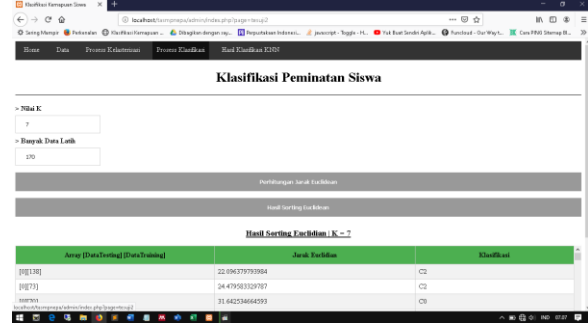
Gambar 13 Tampilan Halaman Import Data

5. Tampilan Halaman Proses Kelasterisasi
 Halaman kelasterisasi dibuat untuk menampilkan data hasil kelasterisasi dengan ditampilkan label kelas dari masing-masing data. Tampilan halaman hasil kelasterisasi dapat dilihat pada Gambar 14.



Gambar 14 Tampilan Halaman Hasil Kelasterisasi

6. Tampilan Halaman Proses Klasifikasi
 Halaman ini dibuat untuk menampilkan hasil proses klasifikasi dengan kNN. Tampilan halaman proses kNN dapat dilihat pada Gambar 15.



Gambar 15 Tampilan Halaman Proses Klasifikasi

7. Tampilan Halaman Hasil Klasifikasi
 Halaman Hasil Klasifikasi kNN dibuat untuk menampilkan hasil pengujian dari data uji terhadap data latih. Tampilan hasil klasifikasi kNN dapat dilihat pada Gambar 16.

No	No. Peserta	Nama	Klasifikasi
1	16-013-171-4	PITRI KAMARIE	IPA
2	16-013-172-5	PITRI DEAR BATHINYA	IPA
3	16-013-173-4	KAMUFP	IPA
4	16-013-174-3	BABKA PERMATA	Bahasa
5	16-013-175-2	KEA WEBATIA	IPA
6	16-013-176-9	RICTY ALIB KIRAH	Bahasa
7	16-013-175-2	MELASTRI	Matematika

Gambar 16 Tampilan Halaman Hasil Klasifikasi

4.5 Pengujian Akurasi KNN

Pengujian kNN dilakukan dengan menggunakan 73 data uji dan 170 data latih. Dengan $K = 5$ didapatkan hasil perhitungan seperti pada Tabel 17.

Tabel 17 Tabel Akurasi Perhitungan KNN

Prediksi Benar	Prediksi Salah	Total
61	12	73

Dengan akurasi yang didapat sebesar:

$$\text{Accuracy} = \frac{\text{Prediksi Benar}}{\text{total data}} = \frac{61}{73} \times 100\% = 83,56\%$$

$$\text{Missfunction Rate} = \frac{\text{Prediksi Salah}}{\text{total data}} = \frac{12}{73} \times 100\% = 16,44\%$$

5. PENUTUP

5.1. Kesimpulan

Berdasarkan keseluruhan proses analisis, perancangan dan implementasi pada pembuatan sistem diperoleh kesimpulan sebagai berikut:

1. Penentuan minat mata pelajaran siswa kelas 9 SMP N 1 Pabuaran menggunakan metode k-Nearest Neighbor dengan PHP dan Mysql, dengan atribut yang digunakan adalah nilai tryout 1, 2, dan 3 mata pelajaran Bahasa indonesia, Bahasa Inggris, Matematika, dan IPA.
2. Pelabelan terhadap data dilakukan dengan menggunakan k Means dengan didapatkan 3 kelasterisasi, yaitu Bahasa sebanyak 102 data, Matematika sebanyak 41 data dan IPA sebanyak 100 data dari jumlah data sebanyak 243 data.
3. Pembagian data latih dan data uji adalah 70%:30% yaitu 170 data latih dan 73 data uji dari total data 243 data. Hasil kNN yang didapatkan adalah total prediksi benar sebanyak 61 data, dan prediksi salah sebanyak 12 data. Akurasi yang didapat sebesar 83.56%, dapat disimpulkan bahwa klasifikasi kemampuan siswa dengan metode k-Nearest Neighbor memiliki hasil yang baik.

5.2. Saran

Berdasarkan kesimpulan diatas, penulis mencantumkan beberapa saran antara lain:

1. Untuk mempermudah penggunaan sistem sebaiknya dibuat tampilan yang lebih user friendly.
2. Menggunakan Data Latih yang memiliki nilai atribut yang beragam untuk bisa meningkatkan nilai akurasinya.
3. Sangat bergantung terhadap data latih, disarankan untuk menggunakan data latih dengan nilai yang beragam.
4. Disarankan untuk menggunakan data latih yang sudah memiliki label sehingga tidak perlu melakukan kelasterisasi sebelum perhitungan kNN.

DAFTAR PUSTAKA

- [1] Aditama, K. (2018), Mendeteksi Objek Dengan Algoritma k-Nearest Neighbor Menggunakan Perangkat Bluetooth Low Energy, Skripsi, S.T., Universitas Islam Indonesia, Yogyakarta
- [2] Agustin, F, E, M., dkk. (2015), Implementasi Algoritma k-Means untuk Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus: SMP Negeri 101 Jakarta), Journal, Journal Teknik Informatika Vol. 8 Universitas Islam Syarif Hidayatullah
- [3] Agusvianto, H. (2017), Sistem Informasi Inventori Gudang Untuk Mengontrol Persediaan Barang Pada Gudang Studi Kasus: PT . Alaisys Sidoarjo, Jurnal, Journal Information Engineering and Educational Technology, Vol.1, 40–46
- [4] Alghifari, M, R. dan Wibowo, A, P., (2019), Penerapan Metode K-Nearest Neighbor Untuk Klasifikasi Kinerja Satpam, Jurnal, Jurnal Teknologi & Manajemen Informatika, Vol.5 No.1 2019.
- [5] Arifudzaki, B., Somantri, M. dan Adian. (2010), Aplikasi Sistem Informasi Persediaan Barang pada Perusahaan Ekspor Hasil Laut Berbasis Web, Jurnal, Transmisi, 12(4), 138–144.
- [6] Aswar, M. (2019), Penerapan Metode k-Nearest Neighbor Untuk Identifikasi Kareis Gigi Pada Manusia, Tugas Akhir, S.Kom., Universitas Teknologi Yogyakarta, Yogyakarta
- [7] Harjoko, A.(2017), Metode Klasifikasi Mutu Jambu Biji Menggunakan k-NN Berdasarkan Fitur Warna dan Tekstur, Journal, Universitas Gajah Mada Yogyakarta
- [8] Mastuti, Y. F. (2007), Pengaruh Nilai Ujian Akhir Nasional SLTA dan Nilai Tes Masuk Perguruan Tinggi Terhadap Prestasi Akademik Mahasiswa, Skripsi, S.Pd., Universitas Sanata Dharma, Yogyakarta.
- [9] Mukminin, M, A. (2018), Klasifikasi Film Berdasarkan Kata Kunci Alur Menggunakan

- Multi-label k-Nearest Neighbor, Skripsi, S.Kom., Universitas Gajah Mada, Yogyakarta
- [10] Ndaumanu, R, I., dkk. (2014), Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode k-Nearest Neighbor, Journal, Jatisi vol 1 STMIK AMIKOM Yogyakarta.
- [11] Nugrahaeni, R, A. (2017), Identifikasi Morfologi Telur dan Larva Nyamuk Pembawa Vektor Penyakit Zoonosis Berbasis Citra Mikroskopis, Tugas Akhir, S.Kom., Universitas Islam Indonesia, Yogyakarta.
- [12] Purnama, C. (2016), Sistem Informasi Manajemen, Mojokerto: Insan Global.
- [13] Rahmah, M, F., dkk. (2017), Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN), Jurnal, Jurnal Informatika vol 11 Universitas Padjajaran
- [14] Rahmahtillah, Q. (2017). Identifikasi Morfologi Telur Cacing Diagnosis Penyakit Kecacingan Berbasis Pencitraan, Tugas Akhir, S.Kom., Universitas Islam Indonesia, Yogyakarta
- [15] Riantika, I. (2018), Aplikasi Klasifikasi k-Nearest Neighbor (k-NN) Pada Pasien Kanker Paru-Paru, Tugas Akhir, S.Kom., Universitas Islam Indonesia, Yogyakarta
- [16] Rizal, M, A. (2019), Klasifikasi Mutu Biji Kopi Menggunakan Metode K-Nearest Neighbor Berdasarkan Warna dan Tekstur, Tugas Akhir, S.Kom., Universitas Teknologi Yogyakarta, Yogyakarta
- [17] Setiawati, L. (2013), Pembuatan Aplikasi Stok Barang Berbasis Web Di Gudang Spare Parts Pada Pt Arwana Citramulia Tbk, Tangerang, Journal of Chemical Information and Modeling
- [18] Utami, R, F. (2016), Penerapan Algoritma k-Nearest Neighbor pada E-Nose untuk Identifikasi Tahu Berformalin, Skripsi, S.Si., Universitas Gajah Mada, Yogyakarta
- [19] Zahro, A.A. (2016), Pengembangan Dan Analisis Kualitas Sistem Informasi Laporan Hasil Belajar Siswa Berbasis Web Di Smk Negeri 1 Banyumas, Skripsi, S.Pd., Universitas Negeri Yogyakarta.