

**SKRIPSI**  
**PENERAPAN ALGORITMA *K-MEANS* DAN ALGORITMA**  
***NAÏVE BAYES* DALAM PEMILIHAN KONSENTRASI**  
**JURUSAN SISWA DI SMA ISLAM TA'ALUMUL**  
**HUDA BUMIAYU**



**Oleh:**  
**Muhamad Ghozi Ghozali**  
**42415012**

**JURUSAN INFORMATIKA**  
**FAKULTAS SAINS DAN TEKNOLOGI**  
**UNIVERSITAS PERADABAN**  
**2021**

# PENGESAHAN SKRIPSI

## PENGESAHAN SKRIPSI

JUDUL PENERAPAN ALGORITMA K-MEANS DAN ALGORITMA  
NAIVE BAYES DALAM PEMILIHAN KONSENTRASI  
JURUSAN SISWA DI SMA ISLAM TA'ALUMUL HUDA  
BUMILAYU  
NAMA M. GHOZI GHOZALI  
NIM 42415012

Skrripsi ini telah diujikan dan dipertahankan dihadapan Dewan Penguji Skripsi,  
pada tanggal. Untuk Memenuhi Salah Satu Syarat Guna Memperoleh Gelar  
Sarjana Komputer (S.Kom)

Paguyangan, 21 Januari 2021

### Nama Penguji

1. Nurul Mega Saraswati M.kom
2. Rito Cipta Sigitta H, M.Kom
3. Fathulloh, M.Kom
4. Tezhar Rayendra TPN., M.Kom

### Tanda Tangan



# PERSETUJUAN SKRIPSI

## PERSETUJUAN SKRIPSI

JUDUL : PENERAPAN ALGORITMA *K-MEANS* DAN ALGORITMA  
*NAIVE BAYES* DALAM PEMILIHAN KONSENTRASI  
JURUSAN SISWA DI SMA ISLAM TA'ALUMUL HUDA  
BUMIAYU

NAMA : M. GHOZI GHOZALI

NIM : 42415012


Skripsi ini telah diperiksa dan disetujui

Bumiayu, 15 Oktober 2020

Pembimbing 1,

  
Nurul Mega Sariwati, M.com  
NIDN.0606069102

Pembimbing 2,

  
Rito Cipta Sigita Haryono, M.com  
NIDN.0619128301

Ketua Jurusan,  
  
Rito Cipta Sigita Haryono, M.com  
NIDN.0619128301



## PERNYATAAN KEASLIAN SKRIPSI

### PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini:

Nama : M. Ghozi Ghozali

NIM : 42415012

Program Studi : Informatika

Menyatakan dengan sebenarnya bahwa skripsi ini benar-benar merupakan hasil karya saya dan bukan merupakan pengambilan tulisan atau pikiran orang lain yang saya akui sebagai tulisan dan pikiran saya, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa skripsi ini hasil plagiat maka saya bersedia menerima sanksi berupa pencabutan gelar akademik yang saya peroleh terkait dengan skripsi ini.

Paguyangan, 21 Januari 2021

Yang membuat pernyataan



M.GHOZI GHOZALI

42415012

## ABSTRAK

Penentuan jurusan pada sekolah menengah atas, sekolah menengah kejuruan maupun madrasah Aliyah dilakukan pada awal masuk sekolah. SMA Islam Ta'alumul Huda Bumiayu menerapkan kurikulum tingkat satuan pendidikan 2013 sebagai landasan dalam proses penjurusan. Pemilihan konsentrasi jurusan harus mempertimbangkan minat dan kemampuan akademik siswa agar tidak kesulitan dalam menyerap dan memahami materi, untuk mengatasi masalah diatas diperlukan metode yang dapat memberikan acuan bagi siswa dalam memilih konsentrasi jurusan. Metode yang digunakan adalah algoritma *K-Means clustering* dan algoritma *Naïve Bayes*. Hasil klasifikasi menggunakan algoritma *K-Means clustering* dan *Naïve Bayes* memperoleh akurasi sebesar 80%.

Kata kunci : *K-Means*, *Naïve Bayes*, konsentrasi jurusan, prediksi, *Confusion Matrix*

## **ABSTRACT**

*The determination of majors in high school, vocational high school and Madrasah Aliyah is done at the beginning of school. Sma Islam Ta'alumul Huda Bumiayu implemented the 2013 education unit level curriculum as a foundation in the cornering process. The selection of the concentration of majors should take into account the interests and academic ability of students so as not to have difficulty in absorbing and understanding the material, to overcome the above problems is required methods that can provide reference for students in selecting the concentration of majors. The methods used are the K-Means clustering algorithm and the Naïve Bayes algorithm. Classification results using the K-Means clustering algorithm and Naïve Bayes obtained 80% accuracy.*

*Keywords : K-Means, Naïve Bayes, concentration majors, prediction, Confusion Matrix*

## KATA PENGANTAR

Alhamdulillah segala puji kehadiran ALLOH SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis bisa menyelesaikan skripsi dengan judul “Penerapan Algoritma *K-Means* dan Algoritma *Naïve Bayes* Dala Pemilihan Konsentrasi Jurusan Siswa Di SMA Islam Ta’alumul Huda Bumiayu”

Penyusunan skripsi ini ditujukan untuk memnuhi syarat kelulusan program studi strata satu (S1) jurusan Informatika di Universitas Peradaban, dalam penyusunan skripsi ini mendapat bantuan dari berbagai pihak, oleh karena itu pada kesempatan ini penulis menyampaikan terimakasih yang sebesar-besarnya kepada

1. Orang tua tercinta, kakak serta adik yang telah memberikan dukungan materil dan moril kepada penulis
2. Prof. Dr. H. Yahya A. Muhaimin selaku Rektor Universitas Peradaban
3. Bapak Dr. Pudjono, Su., Apt. selaku Dekan Fakultas Sains dan Teknologi Universitas Peradaban yang selalu memberikan support kepada penulis
4. Bapak Rito Cipta Sigitta H, M.Kom dan Ibu Nurul Mega Saraswati M.Kom selaku pembimbing yang selalu sabra membimbing dan memotivasi penulis serta memberikan sumbangsih ilmu dalam dunia penelitian.
5. Kawan seperjuangan yang sudah berkontribusi dalam penelitian ini.

Penulis menyadari dalam penelitian skripsi ini masih banyak kekurangan, maka dari itu kritik dan saran selalu penulis terima sebagai bahan penyempurnaan.

Bumiayu, 4 November 2020

Penulis

M.Ghozi Ghozali

## DAFTAR ISI

PENGESAHAN SKRIPSI .....	i
PERSETUJUAN SKRIPSI .....	<b>Error! Bookmark not defined.</b>
PERNYATAAN KEASLIAN SKRIPSI.....	iv
ABSTRAK .....	v
<i>ABSTRACT</i> .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI .....	viii
DAFTAR TABEL .....	x
DAFTAR GAMBAR .....	xi
DAFTAR LAMPIRAN.....	xii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian .....	4
1.5 Ruang Lingkup .....	5
1.6 Sistematika Penulisan .....	5
BAB II TINJAUAN PUSTAKA .....	6
2.1 Penelitian Terkait.....	6
2.6 Landasan Teori .....	9
2.2.1 Kelas konsentrasi.....	9
2.2.2 <i>Data Mining</i> .....	9
2.2.3 <i>K-Means</i> .....	11
2.2.4 <i>Naïve Bayes</i> .....	12
2.2.5 <i>Android</i> .....	13
2.2.6 <i>MySQL</i> .....	14
2.2.7 <i>Xampp</i> .....	15
2.2.8 <i>HYPERTEXT PREPROCESSOR (PHP)</i> .....	15



2.2.9	<i>CodeIgniter</i> .....	15
2.2.10	<i>Confusion Matrix</i> .....	16
2.2.11	<i>Root Mean Square Error</i> .....	17
2.3	Kerangka Pemikiran .....	17
BAB III METODE PENELITIAN .....		19
3.1	Rancangan Penelitian.....	19
3.2	Peralatan yang digunakan .....	20
3.3	Jadwal penelitian .....	20
BAB IV HASIL DAN PEMBAHASAN .....		22
4.1	Identifikasi Masalah.....	22
4.2	Pengumpulan Data.....	22
4.3	Pengolahan Data .....	22
4.3.1	<i>Data Selection</i> .....	22
4.3.2	<i>Preprocessing / Data cleaning</i> .....	23
4.3.3	Transformasi Data .....	24
4.3.4	`Data Mining .....	24
BAB V KESIMPULAN DAN SARAN .....		44
5.1	Kesimpulan .....	44
5.2	Saran .....	44
DAFTAR PUSTAKA .....		45
LAMPIRAN .....		48

## DAFTAR TABEL

Tabel 2.1 Penelitian Terkait .....	8
Tabel 2.2 Versi Android.....	13
Tabel 2.3 Confusion Matrix .....	16
Tabel 3.1 Jadwal Penelitian.....	21
Tabel 4.1 Hasil seleksi nilai Pretest .....	23
Tabel 4.2 Tabel Inisialisasi.....	23
Tabel 4.3 Data Transformasi .....	24
Tabel 4.4 <i>Centroid</i> Awal.....	25
Tabel 4.5 Hasil Perhitungan Iterasi 1.....	25
Tabel 4.6 <i>Centroid</i> ke 2.....	26
Tabel 4.7 Hasil perhitungan iterasi 2 .....	26
Tabel 4.8 <i>Centroid</i> ke 3.....	27
Tabel 4.9 Hasil Iterasi ke 3.....	27
Tabel 4.10 Hasil Akhir Iterasi .....	27
Tabel 4.11 Transformasi .....	28
Tabel 4.12 Hasil Transformasi .....	28
Tabel 4.13 Hasil perhitungan Naive Bayes.....	35
Tabel 4.14 Hasil pengujian Confusion Matrix .....	38

## DAFTAR GAMBAR

Gambar 2.1 Proses KDD.....	10
Gambar 2.2 Tahapan MVC[18].....	16
Gambar 2.3 Kerangka Pemikiran .....	17
Gambar 3.1 Rancangan Penelitian.....	19
Gambar 4.1 Perhitungan label pada tiap data.....	29
Gambar 4.2 Perhitungan kasus pada <i>atribut</i> minat 2.....	29
Gambar 4.3 Perhitungan kasus pada <i>atribut</i> minat 1.....	30
Gambar 4.4 Perhitungan kasus pada <i>atribut</i> Bahasa .....	30
Gambar 4.5 Perhitungan kasus pada <i>atribut</i> IPS.....	31
Gambar 4.6 Perhitungan kasus pada <i>atribut</i> IPA .....	31
Gambar 4.7 Perhitungan kasus pada <i>atribut</i> MTK.....	32
Gambar 4.8 Root Mean Square error.....	36
Gambar 4.9 Confusion Matrix.....	36
Gambar 4.10 Rancangan tampilan awal .....	38
Gambar 4. 11 Rancangan halaman login .....	39
Gambar 4. 12 Rancangan halaman utama.....	39
Gambar 4.13 Rancangan halaman input .....	40
Gambar 4.14 Rancangan halaman hasil.....	40
Gambar 4.15 Tampilan awal .....	41
Gambar 4.16 Tampilan Login .....	41
Gambar 4.17 Tampilan Utama Aplikasi .....	42
Gambar 4.18 Tampilan Input .....	42
Gambar 4.19 Tampilan Hasil .....	43

## DAFTAR LAMPIRAN

Lampiran 1 data nilai siswa.....	48
Lampiran 2 data selection .....	51
Lampiran 3 Data cleaning .....	55
Lampiran 4 Hasil K-Means <i>Clustering</i> .....	59
Lampiran 5 Data transformasi .....	63
Lampiran 6 Data testing.....	66
Lampiran 7 Scan Surat Izin Pengambilan Data SMA Islam T.H.....	67
Lampiran 8 Scan Surat Keterangan penelitian .....	68
Lampiran 9 Source Code web .....	69
Lampiran 10 Source Code Android.....	72