

PEMANFAATAN ALGORITMA FUZZY NAIVE BAYES DALAM PEMILIHAN BIDANG KEAHLIAN MAHASISWA TEKNIK INFORMATIKA UNIVERSITAS ISLAM LAMONGAN

Nurul Fuad¹

¹ Program Studi Teknik Informatika, Universitas Islam Lamongan
Jln. Veteran No.53 A Lamongan
Telp +0856 4611 1939
nurulfuad2@gmail.com

ABSTRACT

One of the factors that can determine the smoothness of students in taking lectures well is the selection of areas of interest in accordance with the abilities of students, meaning that if students determine the choice of majors in accordance with their field of expertise then certainly there will be no difficulty to develop themselves in the lecture process. This study by utilizing two algorithms, namely Fuzzy and Naive Bayes to determine the areas of expertise of students. Informatics Engineering students have the opportunity to choose a field of expertise, namely: the field of network expertise, programming, and design. Not all students will master the area of expertise, so they must choose to accord to their skills, hobbies and favorite subjects, and grades. This research will create a system for classifying the expertise of UNISLA informatics engineering students. The algorithm is implemented, namely Fuzzy to categorize the values or input numbers, while Naive Bayes is used for classification

Keyword : Naive Bayes, Science Expertise

1. PENDAHULUAN

Pembelajaran di dunia pendidikan yang paling tinggi adalah di perguruan, Pada tingkat perguruan tinggi seseorang belajar sesuai bidang keahlian tertentu, sehingga *output* dari sebuah perguruan tinggi diharapkan mampu menghasilkan sumber daya manusia yang berkualitas sesuai dengan bidang keahlian mahasiswa masing-masing berdasarkan bidang minat yang telah di pilih. Oleh sebab itu pengelompokan mahasiswa berdasarkan minat bidang keahlian sangat dibutuhkan di fakultas guna mendukung dan mempermudah kebijakan yang dikeluarkan oleh fakultas untuk mendapatkan *output* sumber daya manusia yang sesuai kebutuhan masyarakat umum.

Bidang keahlian merupakan bagian dari kurikulum berbasis kompetensi. Bidang keahlian adalah kumpulan dari beberapa matakuliah pendukung yang akan mengantarkan Mahasiswa menuju proses penyelesaian skripsi. Adapun bidang peminatan yang ada pada program studi Teknik Informatika Unisla adalah Jaringan, Pemrograman dan Desain. Mahasiswa harus mampu menjadi tenaga yang profesional. Sehingga perlu disesuaikan dengan bidang keahlian, untuk menentukan bidang keahlian tersebut dapat diambil dari beberapa variabel diantaranya skill kesukaan mata kuliah, hoby dan IPK agar dapat memilih bidang sesuai dengan keahlian karena tidak jarang juga mahasiswa dalam menentukan bidang minat secara asal - asalan yang mereka minati tidak berdasarkan kriteria yang telah dianjurkan oleh jurusan Universitas Islam Lamongan. seiring dengan

perkembangan teknologi permasalahan tersebut dapat diantisipasi dengan cara klasifikasi atau pengelompokan bidang minat sesuai dengan kriteria yang telah ditentukan.

Dari permasalahan di atas, maka akan dibangun system untuk mengklasifikasikan atau menemukan bidang keahlian mahasiswa berdasarkan hoby, skill, mata kuliah favorit, IPK dengan algoritma Fuzzy Naive Bayes. Data inputan angka (IPK) dilakukan konversi ke bentuk kategori menggunakan algoritma Fuzzy sedangkan untuk menentukan bidang keahlian mahasiswa menggunakan algoritma Naive Bayes.

Algoritma Naive Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema tersebut dikombinasikan dengan Naive dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi Naive Bayes diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Persamaan dari teorema Bayes adalah :

$$P(H | E) = \frac{P(E | H) * P(H)}{P(E)} \dots\dots\dots(1)$$

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik
 $P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posterior probability)
 $P(H)$: Probabilitas hipotesis H (prior probability)
 $P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H
 $P(X)$: Probabilitas X

Dalam penelitian Nur Nafi'iyah, Siti Mujilawati, 2018 menyebutkan bahwa algoritma Naïve Bayes dapat digunakan dalam proses identifikasi jenis kelamin manusia dengan nilai akurasi 80%, dan inputan dataset dari naïve bayes lebih baik dalam bentuk kategori.

Penelitian lain dalam klasifikasi, yaitu penelitian Nur Nafi'iyah, Chastine Faticah, 2018 algoritma ID3 dataset inputan lebih baik dibuat dalam bentuk kategori, sehingga algoritma fuzzy digunakan sebagai cara mengkonversi inputan angka ke kategori.

Penelitian Darmawan, 2017 proses penentuan dosen pembimbing seminar proposal menggunakan data 100 mahasiswa yang berbeda judul beserta dosen pembimbing. Dari data lama tersebut digunakan untuk menentukan mahasiswa baru yang akan mengajukan proposal skripsi menggunakan algoritma naïve bayes.

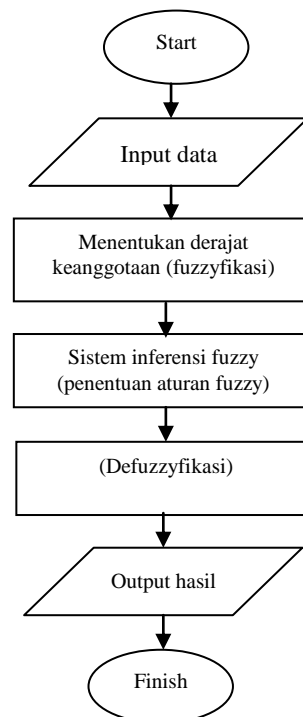
Algoritma Naïve Bayes harus menggunakan dataset terlebih dahulu sebagai sumber belajar dalam menentukan keputusan. Dataset dari algoritma Naïve Bayes lebih baik dalam bentuk kategori, agar dapat mengubah bentuk kategori diterapkan algoritma Fuzzy. Di mana Fuzzy terlebih dahulu diset himpunan keanggotaannya, selanjutnya dilakukan tahap inferensi menggunakan rule, dan dihasilkannya value kategori.

2. METODE

Proses pertama yang dilakukan adalah mengumpulkan data informasi mahasiswa semester 5 dan 7 melalui quisioner online kemudian melakukan preprocessing dengan cara melakukan penapisan data untuk selanjutnya disimpan pada database yang nantinya dijadikan data training dan testing pada uji coba sbidang minat, sistem pemilihan bidang keahlian ini berdasakan skill, hoby dan mata kuliah kesukaan, dan nilai. Agar dapat membantu mahasiswa dalam memilih bidang keahlian, maka peneliti akan membuat system untuk mengklasifikasi bidang keahlian mahasiswa teknik informatika UNISLA. Algoritma yang digunakan, yaitu Fuzzy untuk mengkategorikan dari nilai atau inputan angka, sedangkan naïve bayes digunakan untuk klasifikasi

2.1 Proses Konversi Ipk Menggunakan Fuzzy

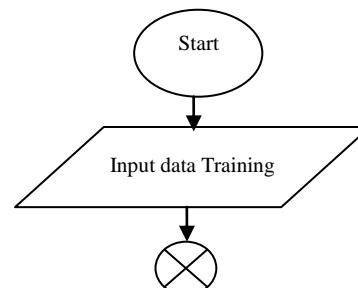
1. Inputan sistem ada data masasiswa semester 5 & 7
2. Pre processing ini adalah menyiapkan data dan mealkukan filtering pada data yang kurang layak
3. proses traning adalah..pembelajaran pada sistem yang nantinya ada testing
4. pada bagian ini ada proses konversi angka ke bentuk kategori (Rendah, sedang, tinggi) menggunakan fuzzy
5. pada bagian ini Proses Uji coba system untuk menentukan bidang keahlian mahasiswa dengan menggunakan algoritma Naïve Bayes
6. kemudian proses yang terakhir melakukan analisa mengenai hasil dari sistem.

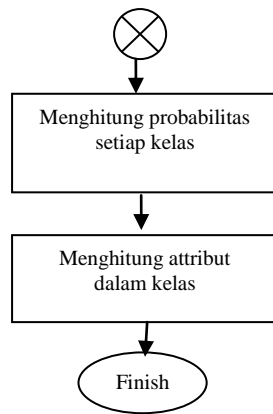


Gambar 1. Alur Konversi IPK dengan Fuzzy

2.2 Proses Pemilihan Jurusan Menggunakan Naive Bayes

a. Proses Training

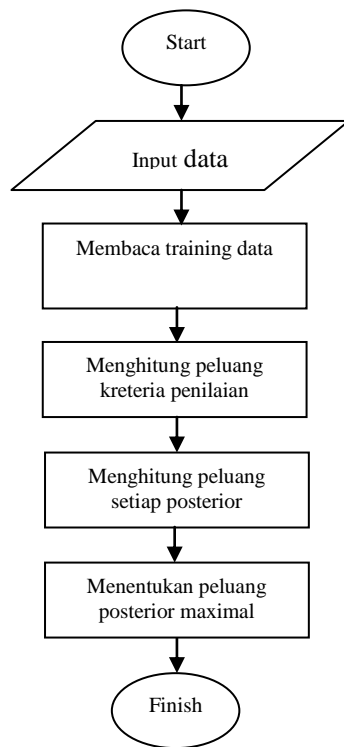




Gambar 2. Pemilihan Jurusan Dengan Naive Bayes

Pada proses training ini data mahasiswa yang telah melalui preprosesing dan pengolahan fuzzy akan di masukkan data base kemudian dilakukan penghitungan probabilitas menggunakan naive bayes selanjutnya menentukan kelas berdasarkan hasil perhitungan

b. Proses Testing



Gambar 3. Proses Testing

Pada proses data testing data dimputkan atau di load dari database kemudian dilakukan pembelajaran dari data traning dan selanjutnya dilakukan perhitungan peluang berdasarkan kreteria secara sistem dan terakhir menghitung peluang posterior dan hasil

3. PEMBAHASAN

3.1 Data Inputan

Berikut contoh data inputan yang akan diolah untuk dilakukan klasifikasi seperti Tabel 1:

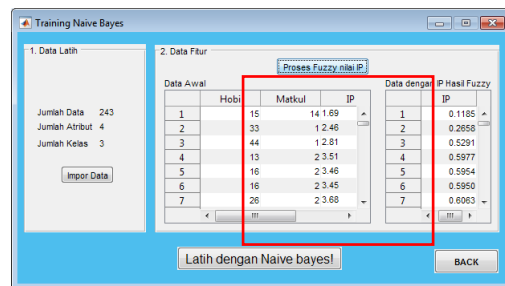
Table 1. Contoh Data Inputan

Nim	Nama	Hoby	Mata kuliah (favorit)	Skill indidividu	Ipk
111810135	Firnanda	Membaca	Multimedia	Desain	03.36
111810043	Khoirul abidin	Menulis	Alpro	Desain	03.04
111810111	Ansori	Makan	Islam nusantara	Desain	03.05
111510046	Prasetyo	Futsal	Indonesia	Desain	03.41

Dari data diatas kemudian diolah dengan cara dilakukan preprocessing, prosesnya yaitu menghilangkan data yang kurang lengkap dalam mengisi questioner, menghapus data ganda dan di simpan dengan format *.CSV. Data tersebut selanjutnya dijadikan untuk data latih atau training pada proses klasifikasi pemilihan bidang minat.

3.2. Menu Perhitungan Fuzzy

Untuk prosesnya data di load atau diimport kemudian ditampung didalam table kemudian disimpan dalam file *.CSV.mat.berikut hasil dari import data pada Gambar 4 untuk proses pengolahan IPK menggunakan fuzzy seperti gambar berikut ini:



Gambar 4. Menu Training

Adapun segmen program pada proses fuzzy tersebut seperti pada segment 1 .

Segmen 1. Program Proses Perhitungan Nilai fuzzy

```

1: set(handles.figure1, 'pointer',
'watch')
drawnow
2: set(handles.figure1, 'pointer',
'arrow')
3: fis = readfis('fuzzy_ipk2');
4: output = evalfis([3.4],fis); %
5: fitur_ip = {};
6: fitur_ip = handles.vektorFitur(:,
3)
7: assignin('base', 'fitur_ip',
fitur_ip);
8: jumlah_baris_data =
sum(cellfun('size', fitur_ip, 1))
9: for i = 1:jumlah_baris_data
fitur_ip_fuzzy{i, 1} =
evalfis(str2double(fitur_ip{i, 1}),
fis);
end
handles.fitur_ip_fuzzy =
fitur_ip_fuzzy;
guidata(hObject, handles);

[alamatFitur, namaFitur, ekstensi] =
fileparts(strcat(alamat,
namaBerkas));
[~, ~, raw] = xlsread(strcat(alamat,
namaBerkas));
cellVectors = raw(:,4);
data = raw(:, [1,2,3]);
    
```

Implementasi algoritma naive bayes untuk menentukan bidang minat menggunakan data training Tabel 2 contoh data training berikut ini :

Tabel 2. Contoh Data Training

N o	Nama	Hobb y	Mata Kulia h (Vavo rit)	IP K	Bidan g Minat
1	MIRZ A	olara ga	alpro	3,68	pemro graman
2	HASA N	mem baca	islamo logi	3,86	desain
3	HIMA M	editin g	b.ingg ris	3,23	desain
4	ALVI N	editin g	alpro	3,14	desain
5	YAFI	mem baca	pancas ila	3,09	jaringa n
6	FILA	editin g	alpro	3,86	desain
7	QORN I	olahr aga	b.ingg ris	2,77	design
8	RIZK Y	editin g	islamo logi	3,36	jaringa n
9	ROSI	olahr aga	islamo logi	3,36	pemro graman
10	ROSI DA	olahr aga	pancas ila	3,36	pemog raman

Pada Tabel 3 berikut ini adalah contoh perhitungan testing

Tabel 3. Contoh Data Testing

N o	Nama	Hobb y	Mata Kulia h (Vavo rit)	IP K	Bidang Minat
1	xxxx	editin g	alpro	3, 14	?

Dalam proses perhitungan menggunakan algoritma naive bayes, ada beberapa langkah yang dilakukan, berikut ini adalah langkah-langkahnya:

- Tahap 1 menghitung jumlah class/label
 $P(Y = \text{DESAIN}) = 5/10$ ‘ jumlah data “DESAIN” pada komom ‘BIDANG MINAT’ dibagi jumlah data
 $P(Y = \text{PEMOGRAMAN}) = 3/10$ ‘ jumlah data “PEMOGRAMAN” pada komom ‘BIDANG MINAT’ dibagi jumlah data
 $P(Y = \text{JARINGAN}) = 2/10$ ‘ jumlah data “JARINGAN” pada komom ‘BIDANG MINAT’ dibagi jumlah data
- Tahap 2 menghitung jumlah kasus yang sama dengan class yang sama
 $P(\text{HOBBY} = \text{EDITING} | Y = \text{DESAIN}) = 3/5$
 ‘jumlah data Hobby “Editing” dengan keterangan “DESAIN” dibagi jumlah data DESAIN
 $P(\text{HOBBY} = \text{EDITING} | Y = \text{PEMOGRAMAN}) = 2/3$
 ‘jumlah data Hobby “Editing” dengan keterangan “PEMOGRAMAN” dibagi jumlah data PEMOGRAMAN
 $P(\text{HOBBY} = \text{EDITING} | Y = \text{JARINGAN}) = 1/2$
 jumlah data Hobby “Editing” dengan keterangan “JARINGAN” dibagi jumlah data JARINGAN

 $P(\text{MATA KULIAH} = \text{ALPRO} | Y = \text{DESAIN}) = 2/5$
 ‘jumlah data MK “ALPRO” dengan keterangan “DESAIN” dibagi jumlah data DESAIN
 $P(\text{KULIAH} = \text{ALPRO} | Y = \text{PEMOGRAMAN}) = 1/3$
 jumlah data MK “ALPRO” dengan keterangan “PEMOGRAMAN” dibagi jumlah data PEMOGRAMAN
 $P(\text{KULIAH} = \text{ALPRO} | Y = \text{JARINGAN}) = 0/2$
 jumlah data MK “ALPRO” dengan keterangan “JARINGAN” dibagi jumlah data JARINGAN

 $P(\text{IPK} = 3.23 | Y = \text{DESAIN}) = 1/5$
 ‘jumlah data IPK “3.23” dengan keterangan “DESAIN” dibagi jumlah data DESAIN
 $P(\text{IPK} = 3.23 | Y = \text{PEMOGRAMAN}) = 0/3$

- ‘jumlah data IPK “3.23” dengan keterangan “PEMOGRAMAN” dibagi jumlah data PEMOGRAMAN
 $P(IPK = 3.23 | Y= JARINGAN) = 0/2$
 ‘jumlah data IPK “3.23” dengan keterangan “JARINGAN” dibagi jumlah data JARINGAN
- Tahap 3 kalikan semua hasil variable DESAIN,PEMOGRAMAN & JARINGAN
 $P(HOBBY = EDITING),(MATA KULIAH = ALPRO),(IPK = 3.23)DESAIN)$
 $=\{P(P(HOBBY = EDITING | Y= DESAIN). P(MATA KULIAH = ALPRO | Y= DESAIN). P(IPK = 3.23 | Y= DESAIN)$
 $=3/5.2/5.1/5 = 0.048$
 $P(HOBBY = EDITING),(MATA KULIAH = ALPRO),(IPK = 3.23)PEMOGRAMAN)$
 $=\{P(P(HOBBY = EDITING | Y= DESAIN). P(KULIAH = ALPRO | Y= PEMOGRAMAN). P(IPK = 3.23 | Y= DESAIN)$
 $=2/3.1/3.0/3 = 0$
 $P(HOBBY = EDITING),(MATA KULIAH = ALPRO),(IPK = 3.23)JARINGAN)$
 $=\{P(P(HOBBY = EDITING | Y= DESAIN). P(KULIAH = ALPRO | Y= JARINGAN). P(IPK = 3.23 | Y= DESAIN)$
 $=1/2.0/2.0/2 = 0$

- Tahap 4 Bandingkan hasil class DESAIN,PEMOGRAMAN & JARINGAN
 Karena hasil (PIDESAIN) lebih besar dari (PIPEMOGRAMAN) lebih besar dari (PIJARINGAN) maka keputusanya adalah “DESAIN”

Tabel 4. Contoh Hasil Data Testing

No	Nama	Hobby	Mata	IPK	Bidang Minat
			Kuliah (Vavorit)		
1	xxxx	editing	alpro	3,14	DESAIN

3.3 Menu Testing

Pada Gambar 5 berikut ini akan dilakukan proses testing, Pada proses testing, performa algoritma akan diuji menggunakan testing set, dimana testing set dan training set merupakan data yang berbeda, ada beberapa uji coba yang kami lakukan seperti dibawa ini:

1. Fitur1.csv

Jumlah data = 81 baris
 Confusion Matrix = 0%
 Cross Validation
 - kfold (10) = 78,2809%
 - leaveout = 78,2609%
 - holdout = 0%

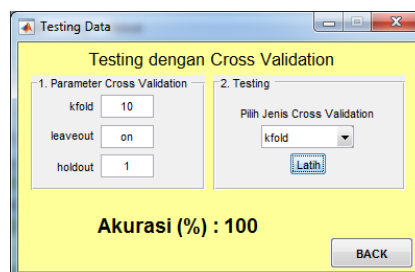
2. Fitur2.csv

Jumlah data = 243 baris
 Confusion Matrix = 10% (data uji baris

234-243)

Cross Validation

- kfold (10) = 100%
- leaveout = 100%
- holdout = 100%



Gambar 5. Proses Testing

Proses tetsing adalah langkah pertama menentukan parameter cross validasi kemudian memberikan nilai besaran validasi setelah itu menentukan pilihan pada testing kemudian menekan tombol latihan untuk melakukan validasi data dengan menggunakan data hasil training di awal.adapun segment program tersebut diatas seperti pada Segmen 2 berikut ini

Segmen 2 Proses Testing

```
1:if strcmpi('kfold', PilihanCV) nilaiCV = 10;
2:elseif strcmpi('leaveout', PilihanCV) nilaiCV = 'on'
3:else strcmpi('holdout', PilihanCV) nilaiCV = 1
4: end
assignin('base', 'CVMdl', CVMdl); Loss= kfoldLoss(CVMdl)
assignin('base', 'Loss', Loss); set(handles.lblAkurasi CV, 'String', 100-Loss*100);
save(strcat('data\Mdl.mat'), 'Mdl'); [cdata, map]=imread('aset/fff.jpg'); cdata, map, ... 'modal');
```

4. KESIMPULAN

Dari analisa dan uji coba diatas maka dalam penelitian ini dapat disimpulkan sebagai berikut :

- Aplikasi yang dibuat pada penelitian ini adalah membuat progam aplikasi klasifikasi bidang minat dengan menggunakan program matlab dan Algoritma Fuzzy Naive Bayes menghasilkan akurasi 78.2% dari 81 data dan 100% dari 243 data yang olah pada proses training dengan menggunakan kfold (10).
- algoritma fuzzy digunakan untuk konversi variabel ke dalam bentuk kategori dan algoritma

naive bayes digunakan untuk menentukan pemilihan bidang minat

PUSTAKA

- Anief Fauzan Rozi, Agus Sidiq Purnomo (2017) *.Rekomendasi Pemilihan Minat Studi Menggunakan Metode Mamdani.*
- Basri, Muhammad Assidiq (2017) *Klasifikasi Data pada Sistem Penjurusan dengan Preferensi Standar Simple Additive Weighting (PS-SAW).*
- Darmawan, R. (2017). *Perancangan Implementasi Sistem Pendukung Keputusan Penentuan Dosen Pembimbing Seminar Proposal Menggunakan Metode Naive Bayes.*
- Dina Yulina Heriyani, Agus Sidiq Purnomo (2018). *Rekomendasi Pemilihan Jurusan SMK Menggunakan Inferensi Fuzzy (Sugeno).*
- Eza Budi Perkasa (2017), *Logika Fuzzy Untuk Penentuan Peminatan Pada SMAN 1 Pangkalpinang.*
- Hartatik, (2017). *Klasifikasi konsentrasi penjurusan mahasiswa universitas amikom Yogyakarta.*
- Husni Naparin, (2016). *Klasifikasi Peminatan Siswa Sma Menggunakan Metode Naive Bayes.*
- Nur Nafi'iyah, Chastine Fatichah. (2018). *Metode Fuzzy ID3 untuk Klasifikasi Bentuk Wajah Manusia Menggunakan Dental Panoramic.*
- Nur Nafi'iyah, Siti Mujilawati. (2018, Agustus). *Analisi Algoritma Backpropagation dan Naive Bayes dalam identifikasi Jenis Kelamin Manusia Berdasarkan Foto Panoramik Gigi.*
- Teddy Pratama, Yulmainis, *Implementasi Algoritma Naive Bayes Dalam Menentukan Konsentrasi Skripsi Dan Rekomendasi Bahasa Pemrograman .*