

ANALISA PENENTUAN DOSEN PEMBIMBING TUGAS AKHIR MAHASISWA MENGGUNAKAN *NAIVE BAYES CLASSIFIER*

Fatayat¹, Riki Ario Nugroho²

Jurusan Ilkom, Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA), Universitas Riau
Email: Fatayat79@gmail.com

Abstrak

Penentuan dosen pembimbing untuk Mahasiswa/i sangat berpengaruh dalam menyelesaikan Tugas Akhir/Skripsi. Dosen pembimbing yang kompeten dan sesuai dengan bidang tugas akhir mahasiswa mampu mengarahkan dan menentukan keberhasilan tugas akhir mahasiswa tersebut. Dalam skripsi ini text mining digunakan untuk mengklasifikasi dosen pembimbing yang tepat untuk mahasiswa/i dalam menyelesaikan tugas akhir. Klasifikasi diterapkan pada judul tugas akhir mahasiswa berdasarkan pengalaman, keahlian dan bidang dosen pembimbing. Algoritma yang digunakan adalah algoritma naive bayes classifier (NBC). Sebelum dilakukan klasifikasi data terlebih dahulu diproses pada tahap text preprocessing, yang terdiri dari case folding, stopcharacter removal, stopword, stemming dan pembobotan. Evaluasi untuk mengukur akurasi pada sistem ini menggunakan confusion matrix yang menghasilkan akurasi 78% pada dosen pembimbing dan menghasilkan akurasi 95% pada bidang keahlian.

Kata Kunci : *Dosen pembimbing, Text Mining, Naive Bayes Classifier (NBC), Text preprocessing, Tugas Akhir, Skripsi*

1. Pendahuluan

Salah satu syarat kelulusan mahasiswa untuk mendapatkan gelar sesuai dengan program studinya adalah dengan mengerjakan Skripsi atau Tugas Akhir (TA). Tugas Akhir (TA) adalah sebuah mata kuliah yang harus ditempuh oleh seorang mahasiswa menjelang akhir studinya. Mata kuliah ini berbentuk proyek mandiri yang dilakukan oleh mahasiswa dibawah bimbingan dosen pembimbing (Maulani dan Simbolon, 2019). Skripsi adalah bentuk pengalaman belajar yang meliputi penggalan kembali apa yang telah dipelajari, mencari dan mengumpulkan pengetahuan baru secara mandiri, melakukan analisis dan bimbingan, serta mengungkapkannya dengan bantuan pembimbing sehingga menghasilkan luaran berupa tulisan tentang suatu pengetahuan baru atau sesuatu yang bersifat pemecahan masalah (Patmi, 2016).

Tugas penentuan pembimbing dan penguji tugas akhir mahasiswa biasanya diberikan kepada ketua jurusan atau ketua prodi atau staf di jurusan, dalam penentuannya masih menggunakan cara konvensional dengan mengandalkan pengetahuan pribadi tentang spesifikasi keahlian dosen yang sesuai dengan judul tugas akhir mahasiswa. Dalam menentukan hal tersebut kadang terjadi keputusan yang kurang optimal dimana dosen yang ditunjuk kurang sesuai dengan topik tugas akhir mahasiswa, akibatnya dapat mengurangi kualitas tugas akhir atau skripsi mahasiswa dan dosen yang ditunjuk spesifikasinya masih kurang padahal masih banyak dosen yang lebih kompeten dengan judul tugas akhir mahasiswa. Untuk memecahkan masalah tersebut dibutuhkan suatu solusi yang dapat memperhitungkan spesifikasi keahlian dosen yang sesuai dengan kebutuhan dosen untuk membimbing mahasiswa.

2. Metode Penelitian

Tahap awal penelitian yaitu melakukan proses pengumpulan data dari berbagai sumber terkait dengan bidang yang sedang dikaji Untuk mendukung pembuatan sistem aplikasi ini diperlukan data-data yang akurat sehingga dapat dipertanggung jawabkan kebenarannya, adapun pengumpulan data diperoleh dari peneliti meliputi:

a. Studi Literatur

Pada tahap ini dilakukan pencarian beberapa literatur yang berkaitan dengan permasalahan dan bagaimana mencari solusi terbaik dari masalah tersebut. Studi literatur dilakukan untuk mendapatkan teori-teori tentang penelitian melalui jurnal, media online, buku atau penelitian lain yang sebelumnya berhubungan dengan penelitian yang akan dilakukan.

b. Pengumpulan Data

Metode pengumpulan data dilakukan dengan cara melakukan pengamatan langsung ke jurusan. Selain pengamatan, juga dilakukan wawancara kepada pihak-pihak yang terkait dengan penelitian ini.

c. Text Mining

Text mining adalah penggalian informasi dari teks oleh *user* menggunakan *tools* analisis. Secara umum *text mining* mengadopsi proses-proses di dalam *data mining* dan di dalam *text mining* juga menggunakan teknik *data mining* (Efendi, 2017). Prosedur utama dalam metode ini terkait dengan menemukan kata-kata yang dapat mewakili isi dari dokumen untuk selanjutnya dilakukan analisis keterhubungan antar dokumen dengan menggunakan metode statistik tertentu seperti analisis kelompok, klasifikasi, dan asosiasi.

d. Text Preprocessing

Pada tahap dilakukan *text preprocessing* yang berguna untuk menghilangkan bagian atau teks yang tidak diperlukan agar diperoleh data yang berkualitas untuk dilakukan tahap berikutnya. Beberapa tahapan yang dilakukan dari *text preprocessing* yaitu:

- 1) *Case folding* (perubahan kata menjadi huruf kecil).
- 2) *Tokenizing* (pemisahan rangkaian kata).
- 3) *Filtering/stop word* (penghapusan kata yang tidak penting).
- 4) *Stemming* (pembentukan kata dasar).
- 5) Pembobotan kata (tf idf).

e. Implementasi Algoritma Naive Bayes Classifier

Setelah *dataset* dipersiapkan untuk menyeimbangkan antara jumlah data *training* dan *testing* maka data akan diberikan label berdasarkan *rule*. Apabila data judul mahasiswa berisi kata yang sama terhadap judul yang telah dosen bimbing sebelumnya maka akan dilakukan proses perhitungan dan klasifikasi menggunakan algoritma *naive bayes classifier*.

f. Algoritma Naive Bayes Classifier

Naive Bayes Classifier merupakan sebuah metode klasifikasi yang berakar pada Teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris, Thomas Bayes, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes.

Jika X adalah vektor masukan yang berisi fitur dan Y adalah label kelas, *Naive Bayes* dituliskan dengan $P(X|Y)$. Notasi tersebut berarti probabilitas label kelas Y didapatkan setelah fitur-fitur X diamati, notasi ini disebut juga probabilitas akhir (*posterior probability*) untuk Y , sedangkan $P(Y)$ disebut probabilitas awal (*prior probability*). Selama proses pelatihan harus dilakukan pembelajaran probabilitas akhir $P(X|Y)$ pada model untuk setiap kombinasi X dan Y berdasarkan informasi yang didapat dari data latih.

Formulasi *Naive Bayes* untuk klasifikasi adalah :

$$P(W_k | V_j) = \frac{n_{k+1}}{(1) n + |vocab|}$$

Keterangan :

W_k : Kata dalam kategori j V_j : Kategori atau label data

n_k : frekuensi munculnya kata w_k pada dokumen berkategori

V_j n : Banyaknya seluruh kata dalam dokumen berkategori V_j

$|vocab|$: Banyaknya kata dalam contoh penelitian.

n_k+1 : 'add one' atau *laplace smoothing* yang menghentikan nilai 0 pada hasil perkalian.

g. Confusion Matrix

Confusion Matrix adalah evaluasi dari sebuah klasifikasi *data mining* yang direpresentasikan menjadi tabel (Gorunescu 2011). *Confusion matrix* berisi informasi perbandingan label hasil klasifikasi dengan label sebenarnya (Indriyanti, Sugianti and Karomi, 2017).

3. Hasil dan Pembahasan

a. Data

Data yang digunakan dalam penelitian ini adalah data judul skripsi mahasiswa dan *list* dosen pembimbing judul skripsi mahasiswa. Total judul skripsi yang didapatkan adalah sebanyak 361 judul skripsi dengan 28 dosen pembimbing,

b. Text Preprocessing

Pada tahapan ini akan dilakukan *text preprocessing* terhadap data yang akan diolah agar *dataset* bersih dari *noise*.

Tabel 1. Sampel Data Testing dan Data Training

Data Training	Judul Skripsi Mahasiswa	Kelas
1	Implementasi Fuzzy dan Metode Certainty Factor Untuk Mendiagnosa Gangguan Kepribadian	23
2	Perbandingan algoritma mean filter dalam mereduksi noise pada citra digital	
3	Klasifikasi Penentuan Subsidi Listrik Rumah Tangga Menerapkan Metode Learning Vector Quantization	25
4	Pencarian pola peminjaman buku pada perpustakaan UIN SUSKA Riau dengan Algoritma Prefixspan	26
5	Algoritma Stemming Bahasa Melayu	?

Tahap *preprocessing* terdiri dari beberapa tahapan, yaitu:

a. Case folding dan stop character removal

Tabel 2. Hasil Proses Case Folding dan Stop Character Removal

Data Training	Judul Skripsi Mahasiswa	Kelas
1	implementasi fuzzy dan metode certainty factor untuk mendiagnosa gangguan kepribadian	23
2	perbandingan algoritma mean filter dalam mereduksi noise pada citra digital	24

3	Klasifikasi penentuan subsidi listrik rumah tangga menerapkan metode learning vector quantization	25
4	pencarian pola peminjaman buku pada perpustakaan uin suska riau dengan algoritma prefixspan	26
5	algoritma stemming bahasa melayu	?

b. *Tokenizing*

Tokenizing adalah proses sebuah kalimat atau paragraf dipecah menjadi sebuah kata.

Tabel 3. Hasil *Tokenizing*

1	2	3	4	5
implementasi	perbandingan	klasifikasi	pencarian	algoritma
fuzzy	algoritma	klasifikasi	pola	stemming
dan	mean	penentuan	peminjaman	bahasa
Metode	filter	subsidi	minjam	melayu
certainty	dalam	listrik	buku	
Factor	mereduksi	rumah	pada	
Untuk	noise	tangga	perpustakaan	
mendiagnosa	pada	menerapkan	uin	
gangguan	citra	metode	suska	
kepribadian	digital	learning	riau	
		vector	dengan	
		quantization	algoritma	
			prefixspan	

c. *Stopword removal dan stemming*

Stopword removal adalah sebuah proses penghapusan kata/*term* yang tidak berhubungan. Tujuan *stopword* ini untuk meminimalisir kata yang dianggap tidak penting dalam memudahkan proses klasifikasi.

Tabel 4. Hasil *Stopword Removal dan Stemming*

	2	3	4	5
fuzzy	banding	klasifikasi	pola	stemming
metode	mean	penentu	Pinjam	bahasa
certainty	filter	subsidi	buku	melayu
factor	reduksi	listrik	pustaka	
diagnosa	noise	rumah	uin	
ganggu	citra	tangga	suska	
pribadi	digital	nerap	riau	
		learning	prefixspan	
		vector		
		quantization		

c. **Klasifikasi Algoritma *Naive Bayes Classifier***

Klasifikasi algoritma *Naive Bayes Classifier* dibagi menjadi 2 tahapan, yaitu tahap *training* (pelatihan) dan tahap *testing* (pengujian). Pada tahapan *training* ini dicari nilai probabilitas kelas dari data latih. Perhitungan nilai probabilitas setiap *term* atau token untuk setiap kelas dari data latih dapat dilihat sebagai berikut :

1. Hitung probabilitas setiap kategori menggunakan persamaan: $P(V_j) = \frac{|doc_j|}{|D|}$
2. Hitung probabilitas masing-masing kata (*stemming*, bahasa, melayu). Probabilitas masing- masing kata pada dosen dihitung menggunakan persamaan :

$$P(W_k | V_j) = \frac{n_k + 1}{n + |vocab|}$$

Probabilitas tiap *term id* pada dosen 23

- Probabilitas kata “ stemming”
 $P(\text{“fuzzy”} | \text{”23”}) = \frac{0 + 1}{139 + 1211} = 0.0007407407407407407$
- Probabilitas kata “ bahasa”
 $P(\text{“bahasa”} | \text{”23”}) = \frac{0 + 1}{139 + 1211} = 0.0007407407407407407$
- Probabilitas kata “ melayu”
 $P(\text{“melayu”} | \text{”23”}) = \frac{0 + 1}{139 + 1211} = 0.0007407407407407407$

Menghitung nilai probabilitas kata (*term*) pada tiap kelas $P(\text{sampel}|23) = p(\text{stemming} | 23) \times p(\text{bahasa} | 23) \times p(\text{melayu} | 23)$
 $P(\text{sampel}|23) = 0.0007407407407407407 \times 0.0007407407407407407 \times 0.0007407407407407407$
 $= 1.5762297793996056e-11$

Tabel 5. Probabilitas kata pada setiap dosen

	<i>Stemming</i>	Bahasa	Melayu		<i>Stemming</i>	Bahasa	Melayu
23	0,00074	0,00074	0,00074	38	0,00075	0,00075	0,00075
24	0,00077	0,00077	0,00077	39	0,00074	0,00074	0,00074
25	0,000776	0,000776	0,000776	40	0,00077	0,00077	0,00077
26	0,00069	0,00069	0,00069	41	0,00076	0,00076	0,00076
28	0,00080	0,00080	0,00080	42	0,00079	0,00079	0,00079
29	0,00068	0,00068	0,00068	43	0,00074	0,0022	0,00074
30	0,00076	0,00076	0,00076	44	0,00070	0,00070	0,00070
31	0,00082	0,00082	0,00082	45	0,00068	0,00137	0,00068
32	0,00068	0,00068	0,00068	46	0,00073	0,00073	0,00073
33	0,00075	0,00075	0,00075	47	0,00081	0,00081	0,00081
34	0,00075	0,00075	0,00075	48	0,00079	0,00079	0,00079
35	0,00077	0,00077	0,00077	49	0,00073	0,00073	0,00073
36	0,00407	0,00814	0,00271	50	0,00078	0,00078	0,00078
37	0,00069	0,00069	0,0013	117	0,00077	0,00077	0,00077

Tabel 6. Probabilitas kata pada setiap bidang keahlian

	<i>Stemming</i>	Bahasa	Melayu
RPL	0.00031	0.00187	0.00062
Pemrograman	0.00268	0.00537	0.00179
<i>Machine Learning</i>	0.00063	0.00126	0.000633
Android	0.000746	0.00224	0.000746

Tabel 7. Probabilitas rekomendasi dosen pembimbing

No	ID Dosen	Probabilitas
1	36	7.738178421521332e-09
2	40	5.90235521668236e-11
3	43	5.192345476252833e-11
4	45	4.398880572901416e-11
5	37	3.7497048646848675e-11
6	30	2.510093164661977e-11
7	26	2.3240765301572132e-11
8	29	2.149440286450708e-11
9	32	2.0769392380999937e-11
10	44	1.926743952692843e-11
11	49	1.7619600438206508e-11
12	46	1.7503512669277966e-11
13	23	1.5762297793996056e-11
14	39	1.4636419380139195e-11

No	ID Dosen	Probabilitas
15	34	1.4225294807284243e-11
16	41	1.3869328232782807e-11
17	35	1.3039853573343887e-11
18	35	1.2666852712731157e-11
19	25	1.1667766209850492e-11
20	24	1.1479580088101947e-11
21	38	1.0839606430366715e-11
22	117	1.0323183665609968e-11
23	48	8.230056688533218e-12
24	50	7.925271327475763e-12
25	42	7.0405986730619204e-12
26	28	4.306317022081572e-12
27	31	1.5292612400461116e-12
28	47	1.5255038431041996e-12

Tabel 8. Probabilitas rekomendasi bidang keahlian

No	Bidang	Probabilitas
1	Pemrograman	7.5949510981803e-09
2	RPL	2.0689857265128327e-10
3	<i>Machine Learning</i>	5.216791999998871e-11
4	Android	5.192345476252833e-11

Nilai probabilitas tertinggi yang diperoleh yaitu kategori atau kelas Id_Dosen 36 dengan nilai probabilitas sebesar 7.738178421521332e-09. Sehingga data uji judul atau skripsi mahasiswa yaitu “*Stemming Bahasa Melayu*” memiliki dosen pembimbing dengan Id_Dosen 36.

d. Evaluasi Klasifikasi Dosen Pembimbing dan Bidang Keahlian Evaluasi dilakukan dengan menggunakan *confusion matrix* untuk melihat seberapa baik *classifier* mengenali *class*.

- Dosen Pembimbing: Akurasi = 78.08 %
- Bidang Keahlian : Akurasi 90,41 %

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat diperoleh beberapa kesimpulan:

- Data mining* sangat berguna untuk penentuan dosen pembimbing dengan menggunakan kaidah klasifikasi.
- Dari hasil penelitian, peneliti berhasil mengimplementasikan algoritma *naive bayes* untuk penentuan dosen pembimbing dan bidang, hasil dari pengolahan 361 data menggunakan algoritma *naive bayes* dengan pembagian data 80:20 dimana data *training* sebanyak 80% dan data *testing* sebanyak 20%.
- Penelitian ini telah menguji kemampuan algoritma klasifikasi *naive bayes* dan diperoleh akurasi sebesar 78% untuk klasifikasi dosen pembimbing dan 90% untuk klasifikasi bidang keahlian.

Daftar Pustaka

- Asril, H. and Kamila, I. (2019) ‘Klasifikasi Dokumen Skripsi Berbasis Text Mining menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor’, (November), pp. 178–187.
- Berliana, G., Shaufiah, S.T., M. T. and Siti Sa’adah, S.T., M. T. (2018) ‘Klasifikasi Posting

- Tweet mengenai Kebijakan Pemerintah Menggunakan Naive Bayesian Classification', *e-Proceeding of Engineering*, 5(1), pp. 1562–1569.
- [3] Indriyanti, Sugianti, D. and Karomi, M. A. Al (2017) 'Peningkatan Akurasi Algoritma KNN dengan Seleksi Fitur Gain Ratio untuk Klasifikasi Penyakit Diabetes Mellitus', pp. 1-6.
- [4] Efendi, Z. (2017) 'Text Mining Classification Sebagai Rekomendasi Dosen Pembimbing Skripsi Program Studi Sistem Informasi', pp. 235–242. Hanik Mujiati, S. (2013) 'Analisis Dan Perancangan Sistem Informasi Stok Obat Pada Apotek Arjowinangun', *Indonesian Journal on Computer Science - Speed (IJCSS) FTI UNSA*, 9330(2), pp. 1–6