

## Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor

Yusra<sup>1</sup>, Dhita Olivita<sup>2</sup>, Yelfi Vitriani<sup>3</sup>

<sup>1,2,3</sup> Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau  
Jl. HR. Soebrantas No. 155 Simpang Baru, Panam, Pekanbaru, 28293  
Email: usera84@yahoo.com, dhita.olivita@uin-suska.ac.id

(Received: 10 Desember 2016; Revised: 30 Desember 2016; Accepted: 30 Desember 2016)

### ABSTRAK

Klasifikasi merupakan salah satu cara untuk mengorganisasikan teks sehingga teks dengan isi yang sama akan dikelompokkan ke dalam kategori yang sama. *Naïve Bayes Classifier* dan *K-Nearest Neighbor* merupakan metode populer dalam pengklasifikasian teks. Di Jurusan Teknik Informatika UIN Suska Riau, tugas akhir dikelompokkan dalam dua topik umum yaitu Ilmu Komputer dan Teknologi Informasi. Dalam penelitian ini, dilakukan klasifikasi terhadap tugas akhir mahasiswa Jurusan Teknik Informatika dengan membandingkan kedua metode tersebut. Klasifikasi yang biasanya dilakukan berdasarkan abstrak bahasa Indonesia, ditambahkan dengan daftar isi. Hal ini dilakukan karena abstrak dan daftar isi mengandung istilah-istilah yang merincikan dan menjelaskan topik penelitian dari sebuah tugas akhir. Proses metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* dilakukan dengan fungsi metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* yang ada pada WEKA. Pengujian akurasi metode pada penelitian ini dilakukan dengan *test option 10-fold cross validation* dan evaluasi data uji menggunakan *confusion matrix*. Dari penelitian yang telah dilakukan, didapatkan hasil pada seratus data tugas akhir dengan jumlah kelas acak, metode *Naïve Bayes* menghasilkan nilai akurasi lebih baik, yaitu sebesar 87%. Pengujian pada metode *K-Nearest Neighbor* menghasilkan nilai akurasi 84% dengan nilai  $k=3$ , 85% dengan nilai  $k=5$ , 86% dengan nilai  $k=7$  dan 84% dengan nilai  $k=9$ .

**Kata Kunci:** *10-fold cross validation*, *Confusion matrix*, klasifikasi, *K-Nearest Neighbor*, *Naïve Bayes Classifier*, tugas akhir, WEKA

### ABSTRACT

*Classification is a way to organize texts, which of texts that have a same content will be grouped in the same category. Naïve Bayes Classifier and K-Nearest Neighbor are popular methods for classifying texts. In Informatic Engineering Department at UIN Suska Riau, final tasks are grouped into two general topics, which is Computer Science and Information Technology. Therefore the research will be conducted to classify the student's final task by comparing the Naïve Bayes Classifier and K-Nearest Neighbor methods. There will be the comparison of the classification based on the abstract in Bahasa and table of contents of the final task in this research. WEKA tools was used in Naïve Bayes Classifier and K-Nearest Neighbor process. Accuration test of the method in this research using test option 10-fold cross validation and evaluation of data testing using confusion matrix. The result of this research is Naïve Bayes method get accuration value 87%. Meanwhile K-Nearest Neighbor method get accuration value 84% if k value = 3, 85% if k value = 5, 86% if k value = 7 and 84% if k value = 9 of one hundred data testing with random class.*

**Keyword :** *10-fold cross validation*, *Confusion Matrix*, *Classification*, *Final task*, *K-Nearest Neighbor*, *Naïve Bayes Classifier*, *WEKA*

---

#### Corresponding Author:

**Yusra**

Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau  
Email: usera84@yahoo.com

---

## Pendahuluan

Tugas akhir merupakan salah satu syarat yang harus dipenuhi oleh mahasiswa untuk menyelesaikan pendidikan di perguruan tinggi. Ilmu yang diperoleh mahasiswa dituangkan ke dalam suatu penelitian yang nantinya akan menghasilkan keluaran berupa dokumen tugas akhir. Untuk menghasilkan penelitian yang baik dan bermanfaat, maka dibutuhkan literatur yang mendukung penelitian serta harus sesuai dengan topik penelitian yang akan diangkat.

Universitas Islam Negeri Sultan Syarif Kasim (UIN SUSKA) Riau merupakan salah satu perguruan tinggi negeri yang berada di Riau. UIN SUSKA memiliki delapan fakultas, salah satunya adalah Fakultas Sains dan Teknologi yang di dalamnya terdapat lima jurusan yaitu Teknik Informatika, Sistem Informasi, Teknik Industri, Teknik Elektro dan Matematika Terapan. Jurusan Teknik Informatika menerima lebih dari 250 mahasiswa baru setiap tahunnya. Banyaknya lulusan sarjana IT yang akan dihasilkan tentunya berbanding lurus dengan banyaknya penelitian tugas akhir mahasiswa yang ada di jurusan Teknik Informatika.

Tugas akhir mahasiswa di jurusan Teknik Informatika dikelompokkan dalam dua topik umum yaitu Ilmu Komputer dan Teknologi Informasi. Topik Ilmu Komputer diantaranya membahas tentang sistem pakar, sistem pendukung keputusan, *machine learning*, *data mining*, dan *information retrieval*. Sementara itu, topik Teknologi Informasi diantaranya membahas tentang aplikasi *mobile*, *arsitektur enterprise*, perencanaan strategis teknologi informasi serta *networking* (jaringan).

Berdasarkan penelitian sebelumnya, yaitu *Text Mining Dengan Metode Naïve Bayes Classifier dan Support Vector Machine untuk sentiment analysis* [3], membahas tentang klasifikasi opini positif dan opini negatif pada data berbahasa Inggris dan Bahasa Indonesia. Didapatkan hasil dengan Metode NBC memberikan hasil dengan akurasi hingga 80,18% untuk data uji opini positif berbahasa Inggris, dan memberikan hasil dengan akurasi hingga 83,86% untuk data uji opini negatif berbahasa Inggris. Untuk data berbahasa Indonesia, metode NBC memberikan hasil dengan akurasi hingga 74,29% pada data uji opini positif dan hingga 87,14% pada data uji opini negatif. Untuk Metode SVM memberikan hasil dengan akurasi hingga 80,15% untuk data uji opini positif berbahasa Inggris, dan memberikan hasil dengan akurasi hingga 98,95% untuk data uji opini negatif berbahasa Inggris. Untuk data berbahasa Indonesia metode SVM memberikan hasil dengan akurasi hingga

78,20% pada data uji opini positif dan hingga 78,14% pada data uji opini negatif.

*Klasifikasi Konten Berita Dengan Metode Text Mining* [2] membahas tentang klasifikasi berita berbasis *web* dengan menggunakan *text mining* dan algoritma *naive bayes classifier*. *Perbandingan Kinerja Metode Naïve Bayes, K-Nearest Neighbor, dan Metode Gabungan K-Means dan LVQ dalam Pengkategorian Buku Komputer Berbahasa Indonesia Berdasarkan Judul dan Sinopsis* [4] yang membahas tentang perbandingan kinerja metode *naive bayes*, *k-nearest neighbor*, dan metode gabungan *k-means* dan *LVQ* dalam pengkategorian buku komputer berbahasa Indonesia didapatkan hasil bahwa rata-rata akurasi metode KNN adalah 96% sedangkan metode *Naïve Bayes* adalah 98%, dan metode gabungan *k-means* dan *LVQ* adalah 92,2%.

Penelitian selanjutnya yaitu *Review on Comparison Between Text Classification Algorithm* [1] membahas tentang perbandingan metode SVM, *Naïve Bayes* dan *K-Nearest Neighbor* dalam pengkategorian dokumen teks berbahasa Inggris, didapatkan hasil pada tiga kategori yang telah ditentukan, *Naive Bayes* mendapatkan rata-rata akurasi yaitu sebesar 72,3%, *K-Nearest Neighbor* mendapatkan rata-rata akurasi sebesar 84,53% dan *SVM* mendapatkan rata-rata akurasi sebesar 86,06%.

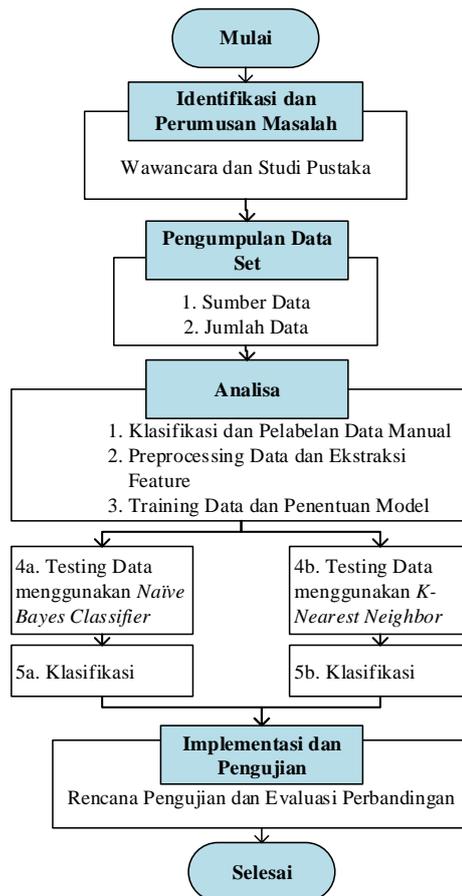
Berdasarkan penelitian sebelumnya, dalam penelitian ini akan dilakukan perbandingan metode yang akan diimplementasikan pada klasifikasi bidang keahlian tugas akhir yang ada di jurusan Teknik Informatika. Metode yang akan dibandingkan pada penelitian ini adalah metode *naive bayes classifier* dan *k-nearest neighbor*. Metode ini dipilih karena *Naïve Bayes Classifier* dan *K-Nearest Neighbor* ini merupakan algoritma yang populer digunakan dalam proses pengklasifikasian teks. Selanjutnya, dilakukan perbandingan terhadap kedua metode dan selanjutnya dapat diketahui perbandingan keefektifan dari kedua metode ini terhadap pengklasifikasian tugas akhir.

Mahasiswa umumnya akan melihat kepada judul, abstrak dan daftar isi ketika membaca laporan tugas akhir. Sejauh ini belum ada penelitian yang melibatkan daftar isi untuk proses klasifikasi. Oleh karena itu, dalam penelitian ini dilibatkan daftar isi selain abstrak Bahasa Indonesia. Abstrak dan daftar isi dipilih karena pada laporan tugas akhir, abstrak dan daftar isi mendefinisikan istilah secara lebih rinci dan jelas pada setiap bidang penelitian sebuah tugas akhir. Hal ini dimaksudkan untuk mengetahui perbandingan dari kedua metode yang digunakan dalam mengklasifikasikan bidang keahlian mahasiswa Teknik Informatika.

Proses dari penelitian ini nantinya akan dilakukan dengan menggunakan 100 data tugas akhir mahasiswa dan dilabel secara manual sesuai kelas yang telah ditentukan yaitu Ilmu Komputer dan Teknologi Informasi. Data tugas akhir selanjutnya akan dilakukan *preprocessing* dan *stemming*, dimana data tersebut akan melalui tahapan; penyeragaman huruf menjadi huruf kecil (*case folding*), penghapusan tanda baca serta angka (*stop character removal*), penguraian kalimat menjadi kata (*token*), penghapusan kata penghubung (*stopword removal*) dan juga penguraian kata menjadi kata dasar (*stemming*). Setelah tahapan *preprocessing* dan *stemming* selesai, akan dilakukan tahapan *ekstraksi feature* dan pembobotan. Tahapan *ekstraksi feature* yang akan digunakan yaitu fitur *unigram* dan pembobotan kata akan menggunakan *augmented term frequency* yang nantinya akan dilakukan didalam *tools* WEKA. Proses selanjutnya yaitu tahapan klasifikasi menggunakan WEKA, dimana proses klasifikasi ini akan membandingkan hasil dari metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor*. Hasil klasifikasi dari kedua metode ini selanjutnya akan dibandingkan sehingga dapat diketahui metode yang paling baik dalam melakukan proses klasifikasi terhadap tugas akhir.

### Metodologi Penelitian

Dalam melakukan sebuah penelitian, dibutuhkan suatu acuan pelaksanaan yang dinamakan metodologi penelitian. Metodologi penelitian terdiri dari beberapa tahapan kerangka kerja penelitian secara terstruktur / sistematis mulai dari tahap awal penelitian hingga mendapatkan hasil yang ingin dicapai. Berikut adalah gambaran tahapan yang dilakukan dalam penelitian yang akan disajikan pada Gambar 1 berikut.



Gambar 1. Tahapan pengerjaan penelitian

### Pengumpulan Data

Untuk membandingkan hasil klasifikasi tugas akhir jurusan Teknik Informatika dengan menggunakan metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* berdasarkan bidang keahlian yang ada di jurusan Teknik Informatika. Jenis bidang keahlian dari tugas akhir mahasiswa di jurusan Teknik Informatika diketahui setelah melakukan proses wawancara. Wawancara dilakukan dengan Ibu Fadhilah Syafria, ST., M.Kom selaku koordinator Tugas Akhir di jurusan Teknik Informatika.

Selanjutnya dikumpulkan data berupa laporan tugas akhir mahasiswa jurusan Teknik Informatika di Universitas Islam Negeri Sultan Syarif Kasim (UIN SUSKA) Riau. Pengambilan data bertujuan untuk mendapatkan informasi tentang analisa klasifikasi tugas akhir menggunakan algoritma *Naïve Bayes Classifier* dan *K-Nearest Neighbor*. Adapun sumber data dan jumlah data sebagai berikut :

1. Sumber data berasal dari abstrak Bahasa Indonesia dan daftar isi laporan tugas akhir Teknik Informatika
2. Data diambil secara acak dalam rentang tahun 2012-2015, dimana pada proses pengumpulan data, jumlah masing –

masing kelasnya tidak ditentukan terlebih dahulu.

3. Jumlah data yang akan diklasifikasi sebanyak 100 tugas akhir.

### Pengklasifikasian Data

Pada tahap ini ada beberapa aktifitas yang akan dilakukan, yaitu :

- a. Klasifikasi dan Pelabelan Data

Pada tahap ini akan dianalisa kategori tugas akhir yang akan diklasifikasi. Kategori dibagi menjadi dua yaitu Ilmu Komputer dan Teknologi Informasi. Selanjutnya data yang ada akan diberi label manual sesuai dengan kategori yang telah ditetapkan.

Penentuan label dari masing-masing data dilakukan setelah proses wawancara dengan koordinator TA mengenai jenis-jenis dari tugas akhir dan dengan membaca abstrak beserta *keyword* dan juga daftar isi dari tugas akhir tersebut.

- b. *Preprocessing* Data, *Ekstraksi Feature* dan *Pembobotan*

*Preprocessing Data* merupakan langkah untuk membersihkan data dengan tujuan untuk menyeragamkan bentuk kata dan mengurangi volume kata. Dalam *preprocessing* data akan dilakukan sejumlah proses yaitu :

1. *Case folding*. Pada tahap ini semua karakter yang ada diubah menjadi huruf kecil, umumnya juga disebut *lowercase*.
2. *Stop Character Removal*. Pada tahap ini semua tanda baca seperti titik (.), koma (,), serta angka dihilangkan.
3. *Token*. Dimana pada proses ini akan dilakukan penguraian deskripsi yang semula berupa kalimat – kalimat menjadi kata – kata
4. *StopwordRemoval*. Pada tahap ini akan dilakukan penghapusan kata penghubung (konjungsi) dalam sebuah kalimat seperti kata atau, yang, dengan ini, pada dan lainnya.
5. *Stemming*. Pada proses ini akan dilakukan sebuah proses pemetaan dan penguraian berbagai bentuk dari suatu kata menjadi kata dasarnya (*stem*). Proses stemming dilakukan dengan bantuan KBBI dan algoritma yang digunakan yaitu algoritma *stemming* Nazief Andriani.

Tahap selanjutnya yaitu melakukan ekstraksi feature *unigram* dan pembobotan dengan *Augmented Term Frequency*. Berdasarkan metode ekstraksi feature, dilihat urutan peringkat kata berdasarkan frekuensi kemunculan kata. Minimal kemunculan kata dipilih sebanyak delapan kali setelah melakukan sortir terhadap

pengaruh kemunculan kata pada tugas akhir. Pembobotan kata akan di proses menggunakan WEKA.

- c. Training Data dan Penentuan Model

Pada tahap ini data akan dikelompokkan menjadi dua kelompok yaitu data training dan data testing. Data *training* digunakan untuk proses pengolahan data awal dimana proses ini melatih mesin (komputer) yang diberikan pengetahuan dengan sekumpulan *dataset* sampai mesin dinilai cukup cerdas. Hasil dari data *training* ini adalah sebuah model yang digunakan untuk melakukan proses *testing*.

- d. Testing Data

Tahap ini adalah bagian pembelajaran pola yang akan menggunakan metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor*.

### *Naïve Bayes Classifier*

Pada tahapan ini akan dilakukan testing data dengan menggunakan metode *Naïve Bayes Classifier*. Metode ini akan dipakai dalam tools WEKA. Hasil dari klasifikasi dapat dilihat pada nilai probabilitas yang paling tinggi.

Ringkasan untuk algoritma *Naïve Bayes Classifier* dapat dilihat sebagai berikut :

- A. Proses pelatihan. Input adalah dokumen-dokumen contoh yang telah diketahui kategorinya:

1. Kosakata  $\leftarrow$  himpunan semua kata yang unik dari dokumen-dokumen contoh
2. Untuk setiap kategori  $v_j$  lakukan:
  - a.  $Docs_j \leftarrow$  Himpunan dokumen-dokumen yang berada pada kategori  $v_j$
  - b. Hitung  $P(v_j)$  dengan persamaan

$$P(V_j) = \frac{|doc\ j|}{Contoh} \quad (1)$$

- c. Untuk setiap kata  $w_k$  pada kosakata lakukan:
  - (i). Hitung  $P(w_k | v_j)$  dengan persamaan

$$P(W_k | V_j) = \frac{nk+1}{n+|vocabulary|} \quad (2)$$

- B. Proses klasifikasi. Input adalah dokumen yang belum diketahui kategorinya:

1. Hasilkan  $v_{map}$  sesuai dengan menggunakan  $P(v_j)$  dan  $P(w_k | v_j)$  yang telah diperoleh dari pelatihan

$$V_{map} = \underset{v_j \in V}{\operatorname{argmax}} P(V_j) \prod_i P(a_i | V_j) \quad (3)$$

2. Probabilitas kelas maksimum adalah kelas terpilih hasil klasifikasi.

**K-Nearest Neighbor**

Pada tahapan ini juga akan dilakukan pengujian data dengan menggunakan metode *K-Nearest Neighbor* yang ada di dalam WEKA. WEKA adalah aplikasi data mining open source berbasis Java yang dikembangkan pertama kali oleh Universitas Waikato di Selandia Baru. Hasil dari klasifikasi dapat dilihat dari kelas mayoritas yang muncul berdasarkan jumlah tetangga K yang telah ditentukan.

Prinsip kerja dari *K-Nearest Neighbor* adalah mencari jarak antara dua titik yaitu titik *training* dan titik *testing*. Selanjutnya dilakukan evaluasi dengan k tetangga terdekatnya dalam data training. Jarak yang digunakan adalah persamaan jarak *Euclidean distance*, dengan rumus :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

Dimana :

- d : jarak antara titik pada data *training* x dan titik data *testing* y yang akan diklasifikasi, dimana  $x=x_1, x_2, \dots, x_i$  dan  $y=y_1, y_2, \dots, y_i$
- i : merepresentasikan nilai atribut
- n : merupakan dimensi atribut.

e. Klasifikasi

Pada tahap klasifikasi, akan dilakukan proses dengan menggunakan tools WEKA, tahap yang dilakukan yaitu:

1. Persiapan data dalam format file yang dapat diidentifikasi oleh WEKA yaitu format CSV dan ARFF.
2. Perancangan *dataset*, dalam perancangan *test option* yang digunakan adalah *10-fold validation*, dengan demikian penelitian ini membagi data *training* menjadi 90% data *training* dan 10% data *testing*

**Implementasi dan Pengujian**

1. Rencana Pengujian

Pengujian dataset dilakukan dengan alat bantu WEKA. Proses pengujian ini dibagi dalam dua sesi yaitu pengujian perhitungan manual dan pengujian *testoptioncrossvalidation*. Rencana pengujian akan dijelaskan sebagai berikut :

a. Pengujian Perhitungan Manual

Pengujian ini dilakukan untuk menguji perhitungan manual yang telah dilakukan pada bab 4. Untuk pengujian ini dilakukan dengan menggunakan *test option use training set* dan *supplied test*. Setelah dilakukan proses klasifikasi oleh *naïve bayes*

dan *K-NN*, WEKA akan menampilkan tingkat akurasi dari masing – masing pengujian.

- b. Pengujian Test Option Cross Validation
  - i. Pengujian dilakukan dengan dua jenis pengujian yaitu data dengan posisi kelas acak dan data dengan posisi kelas terurut. Tujuannya adalah untuk melihat bagaimana pengaruh posisi data terhadap akurasi.
  - ii. Pengujian ini akan menggunakan *testoptioncrossvalidation* dengan melibatkan seluruh total *dataset* sebagai data latih dan data uji. Untuk pembagian data latih dan data uji, digunakan *10-fold-crossvalidation*.
  - iii. Data yang diuji dibagi menjadi dua kelompok yaitu tugas akhir yang berjumlah seratus data dengan jumlah kelas 79 data Ilmu Komputer dan 21 data Teknologi Informasi, serta tugas akhir berjumlah dua puluh data dengan jumlah kelas 10 data Ilmu Komputer dan 10 data Teknologi Informasi.

2. Evaluasi Perbandingan

Pada tahap evaluasi, ukuran atau parameter yang digunakan adalah *confusion matrix*. Tujuan dari tahapan evaluasi adalah untuk melihat efektivitas dari kedua metode pengujian yang dilakukan dalam mengklasifikasikan tugas akhir.

**Hasil dan Pembahasan**

Setelah proses analisa data dilakukan, tahapan selanjutnya adalah proses implementasi analisa dan pengujian terhadap *dataset*. Dengan melakukan implementasi dan pengujian ini, maka akan diketahui akurasi klasifikasi dari dataset dengan menggunakan *naïvebayes* dan *k-nearest neighbor*.

Berdasarkan hasil pengujian dari data yang telah dilakukan sebelumnya, *classifier* output WEKA menunjukkan *confusion matrix* sehingga nilai akurasi dapat dihitung dapat dilihat hasilnya pada tabel 1 dan 2.

Tabel 1. Evaluasi pengujian *Naïve Bayes*

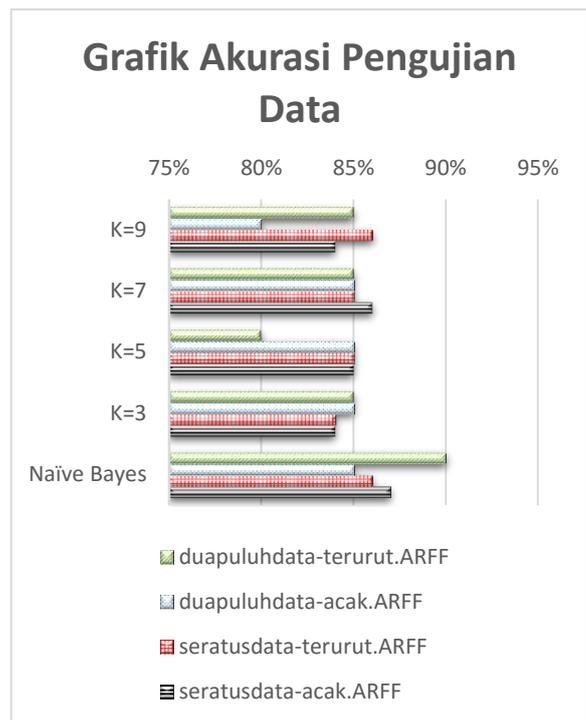
No	Nama Data	Confusion Matrix		Akurasi
		TP	FP	
1	seratusdata-acak.ARFF	74	5	87%
		8	13	
2	seratusdata-terurut.ARFF	73	6	86%
		8	13	
3	duapuluhdata-acak.ARFF	9	1	85%
		8	2	
4	duapuluhdata-terurut.ARFF	10	0	90%
		2	8	

Tabel 2. Evaluasi pengujian *K-Nearest Neighbor*

No	Nama Data	Confusion Matrix		Aku rasi	K	Rata - Rata Akur asi
		TP	FP			
1	seratusdata-acak.ARFF	75	4	84%	3	84,75 %
		12	9			
		75	4	85%	5	
		11	10			
76	3	86%	7			
11	10					
76	3	84%	9			
13	8					
2	seratusdata-terurut.ARFF	75	4	84%	3	85%
		12	9			
		75	4	85%	5	
		11	10			
		76	3	85%	7	
12	9					
77	2	86%	9			
12	9					
3	duapuluhdata-acak.ARFF	10	0	85%	3	83,75 %
		3	7			
		10	0	85%	5	
		3	7			
10	0	85%	7			
3	7					
9	1	80%	9			
3	7					
4	duapuluhdata-terurut.ARFF	10	0	85%	3	83,75 %
		3	7			
		10	0	80%	5	
		4	6			
10	0	85%	7			
3	7					
10	0	85%	9			
3	7					

Untuk visualisasi hasil dari pengujian pada penelitian ini, dapat dilihat pada grafik Gambar 2

Berdasarkan hasil pengujian, dapat dilihat bahwa pada metode *naivebayes*, dataset *seratusdata-acak.ARFF* yaitu sebesar 87% dan *duapuluhdata-terurut.ARFF* sebesar 90% mampu mengklasifikasi data dengan baik. Sementara itu, pada metode *k-nearest neighbor*, dataset *seratusdata-terurut.ARFF* mampu diklasifikasikan dengan baik, dimana rata-rata akurasi 85% pada masing-masing akurasi; K=3 yaitu 84%, pada K=5 yaitu 85%, pada K=7 yaitu 85% dan pada K=9 yaitu 86%.



Gambar 2. Grafik akurasi pengujian data

Pada pengujian menggunakan *test option cross validation*, proses ini membagi data menjadi data *training* dan data *test* secara berulang. Dalam penelitian ini, penulis menggunakan *10 fold cross validation*, dimana data dikelompokkan menjadi 9 data sebagai data latih dan 1 sebagai data untuk *testing*. Begitupun selanjutnya hingga iterasi sebanyak k. Akurasi yang didapat pada pengujian *10-fold cross validation* untuk seratus data tugas akhir, metode *Naïve Bayes Classifier* dapat melakukan klasifikasi terhadap tugas akhir lebih baik daripada *K-Nearest Neighbor* dengan akurasi sebesar 87%.

### Kesimpulan

Pada hasil analisa dan implementasi klasifikasi tugas akhir dengan menerapkan metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* pada WEKA, maka dapat diambil beberapa kesimpulan yaitu :

1. Penggunaan abstrak Bahasa Indonesia dan daftar isi dapat mengklasifikasikan data tugas akhir sesuai kelas yang diprediksi. Hal ini disebabkan karena abstrak dan daftar isi telah menjelaskan secara lebih detail mengenai garis besar topik tugas akhir tersebut.
2. Metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* dapat mengklasifikasikan data tugas akhir

mahasiswa jurusan teknik informatika. Berdasarkan hasil yang diperoleh, dapat diketahui perbandingan akurasi dari masing-masing metode.

3. Berdasarkan akurasi yang didapat pada pengujian *10-fold cross validation* untuk seratus data tugas akhir, metode *Naïve Bayes Classifier* dapat melakukan klasifikasi terhadap tugas akhir lebih baik daripada *K-Nearest Neighbor* dengan akurasi sebesar 87%.
4. Akurasi data antara metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* pada seratus data tugas akhir memiliki selisih data yang kecil. Hal ini disebabkan karena jumlah data dari masing-masing kelasnya tidak seimbang, sehingga mengakibatkan kurangnya keefektifan data latih pada salah satu kelas.

### Daftar Pustaka

- [1] Gandhi, V. C. dan Jignesh A .P., Review on Comparison between Text Classification Algorithms. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 1, Issue 3* September-October, 2012.
- [2] Kurniawan, B., Klasifikasi Konten Berita Dengan Metode Text Mining. *Jurnal Dunia Teknologi Informasi* Vol.1, No.1, 2012: 14-19, 2012.
- [3] Saraswati, N.W. dan Sumartini. Text Mining Dengan Metode Naïve Bayes Classifier dan Support Vector Machine untuk Sentimen Analysis. Thesis Teknik Elektro Universitas Udayana Denpasar, 2011.
- [4] Santoso, D., Perbandingan Kinerja Metode Naïve Bayes, K-Nearest Neighbor, dan Metode Gabungan K-Means dan LVQ dalam pengkategorian Buku Komputer Berbahasa Indonesia Berdasarkan Judul dan Sinopsis. *Jurnal Ilmu Komputer Universitas Brawijaya*, 2014.