

## Klasifikasi Berita Kriminal Menggunakan *Naive Bayes Classifier* (NBC) dengan Pengujian *K-Fold Cross Validation*

Herfia Rhomadhona<sup>1)</sup>, Jaka Permadi<sup>2)</sup>

<sup>1)2)</sup> Program Studi Teknologi Informasi Politeknik Negeri Tanah Laut  
Jl. A. Yani Km. 6 Desa Panggung Pelaihari, Kalimantan Selatan 70815

<sup>1)</sup> herfia.rhomadhona@politala.ac.id

<sup>2)</sup> jakapermadi.88@politala.ac.id

### Abstrak

Berita kriminalitas merupakan berita yang selalu menjadi trending topik di setiap media massa, khususnya media massa *online*. Media massa *online* telah menyediakan beberapa fasilitas untuk mempermudah masyarakat dalam mencari sebuah berita berdasarkan topik. Media massa *online* melabeli suatu berita berdasarkan kategorinya. Namun, media massa *online* tidak memberikan sub kategori pada berita tersebut. Sebagai contoh jika seorang pengguna membuka kategori kriminal, maka yang ditampilkan adalah semua jenis berita kriminal tanpa memberikan informasi yang spesifik dari jenis kriminalitasnya. Permasalahan tersebut dapat diatasi dengan mengklasifikasikan berita kriminalitas berdasarkan subkategori. Penelitian ini menggunakan metode *Naive Bayes Classifier* (NBC) untuk mengklasifikasi berita berdasarkan sub kategorinya. Adapun subkategori terbagi kedalam 5 kategori yaitu korupsi, narkoba, pencurian, pemerkosaan dan pembunuhan. Penelitian ini bertujuan untuk mengetahui kemampuan NBC dalam mengklasifikasi berita dengan melakukan pengujian menggunakan teknik *K-Fold Cross Validation* dengan nilai *K* dari 3 sampai 10. Hasil pengujian menyatakan bahwa NBC memiliki kemampuan dalam klasifikasi berita kriminal dengan nilai *precision* sebesar 98,53 %, nilai *recall* sebesar 98,44 % dan nilai *accuracy* sebesar 99,38 %.

**Kata kunci:** *Confusion Matrix, K-Fold Cross Validation, Naive Bayes Classification*

### Abstract

*Crime news is news that has always been a trending topic in every newspaper, especially online newspaper. Online newspaper has provided several facilities to facilitate the community in finding a news based on the topic. Online newspaper label a news based on its category. However, online newspaper does not provide sub-categories for the news. For example if a user opens the crime category, then all types of criminal news are displayed without providing specific information on the type of crime. These problems can be overcome by classifying crime news by subcategory. This research use Naive Bayes Classifier (NBC) method to classify news based on its sub-categories. The subcategories are divided into 5 categories, that is corruption, drugs, theft, rape and murder. This research aims to determine the ability of NBC in classifying news by testing using the K-Fold Cross Validation technique with a K value of 3 to 10. The test results state that NBC has the ability to classify criminal news with a precision value of 98.53%, recall value of 98.44% and an accuracy value of 99.38%.*

**Keywords:** *Confusion Matrix, K-Fold Cross Validation, Naive Bayes Classification*

## 1. PENDAHULUAN

Surat kabar atau koran merupakan salah satu bentuk media massa yang memuat berita-berita terkini dengan harga yang terjangkau oleh masyarakat. Seiring perkembangan zaman, *trend* dari media ini juga berkembang. Perusahaan-perusahaan dari media massa surat kabar sudah

memanfaatkan teknologi internet sehingga dapat menyebarkan berita-berita terkini secara *online* melalui media massa *online*. Media massa *online* ini dapat diakses oleh masyarakat dengan lebih cepat dan lebih murah. Berbagai fasilitas disediakan pada media massa *online* tersebut sehingga lebih mempermudah masyarakat dalam penggunaannya. Salah satu fasilitas yang biasa diberikan adalah pencarian berita berdasarkan topik atau kategorinya.

Media massa *online* melabeli suatu berita berdasarkan kategorinya. Walaupun demikian sub kategori tidak diberikan pada berita tersebut. Sebagai contoh jika seorang pengguna membuka kategori kriminal, maka yang ditampilkan adalah semua jenis berita kriminal tanpa memberikan informasi yang spesifik dari jenis kriminalitasnya. Permasalahan tersebut dapat diatasi dengan menanamkan *classifier* yang dapat mengklasifikasi suatu berita berdasarkan sub kategorinya.

Penelitian mengenai klasifikasi teks berita telah dilakukan oleh Hamzah (2012) dimana peneliti menggunakan *Naïve Bayes Classifier* (NBC) untuk pengelompokan teks berita yang menghasilkan akurasi pengujian sebesar 91%. Sehingga dapat dikatakan bahwa NBC merupakan *classifier* yang baik untuk mengklasifikasi suatu teks berita. Selain itu, Rhomadhona dan Permadi (2019) juga melakukan klasifikasi berita kriminal menggunakan *Naïve Bayes Classifier* (NBC) dengan 180 data latih dan 45 data uji. Pengujian dilakukan dengan menghitung akurasi untuk mengukur kinerja dari NBC, akurasi yang didapatkan hanya sebesar 77,78 %.

Berdasarkan latar belakang tersebut dilakukan penelitian yang bertujuan untuk mengetahui kemampuan NBC dalam mengklasifikasi berita berdasarkan sub kategorinya. Penelitian ini dibatasi pada kategori kriminal dengan lima sub kategori, yaitu korupsi, narkoba, pemerkosaan, pencurian dan pembunuhan. Studi kasus menggunakan media massa *online* di Kalimantan Selatan, yaitu banjarmasin.tribunnews.com, antarakalsel.com, dan kalsel.prokal.co. Pengujian yang dilakukan menggunakan *K-Fold Cross Validation*.

## 2. TINJAUAN PUSTAKA

*Naïve Bayes Classifier* (NBC) adalah *classifier* yang banyak digunakan dalam klasifikasi teks. Penelitian dengan menggunakan NBC dilakukan oleh Herlambang dan Wijoyo (2019) yang memanfaatkan NBC untuk mengelompokkan dokumen pembelajaran berdasarkan kriteria atau ciri esensial setiap pembelajaran. Akurasi tertinggi yang dihasilkan adalah sebesar 81,48%. Utami dan Sari (2018) juga melakukan penelitian dengan memanfaatkan NBC sebagai *classifier* untuk mendeteksi *hoax* pada kolom komentar di forum *Female Day*. Akurasi yang didapatkan adalah sebesar 88% dari *testing set* sebanyak 50 komentar. Fanissa dkk (2018) melakukan penelitian dengan menggunakan NBC untuk menganalisa ulasan masyarakat tentang pariwisata kota Malang. Akurasi yang dihasilkan dari penelitian ini adalah sebesar 86,6%. Penelitian lainnya dilakukan oleh Putra dkk (2016) yang menggunakan NBC untuk klasifikasi teks bahasa Bali. Penelitiannya ini diuji dengan teknik *cross validation* dan didapatkan rata-rata akurasi sebesar 95,22%.

Metode NBC merupakan model sederhana dari algoritma *bayes* yang mengklasifikasikan suatu dokumen berdasarkan persamaan berikut:

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} P(v_j | a_1, a_2, \dots, a_n) \quad (1)$$

Dimana:

$V$  : Himpunan seluruh dokumen

$v_j$  : Dokumen pada kategori  $j$

$a_i$  : Kata  $i$  pada suatu dokumen

Persamaan (1) dapat juga ditulis menjadi persamaan berikut:

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} \frac{P(a_1, a_2, \dots, a_n | v_j) P(v_j)}{P(a_1, a_2, \dots, a_n)} \quad (2)$$

Karena  $P(a_1, a_2, \dots, a_n)$  konstan, maka persamaan (2) dapat disederhanakan menjadi:

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} P(a_1, a_2, \dots, a_n | v_j) P(v_j) \quad (3)$$

Karena  $P(a_1, a_2, \dots, a_n | v_j)$  susah dihitung, maka diasumsikan setiap kata pada suatu dokumen tidak memiliki keterkaitan, sehingga persamaan (3) dituliskan menjadi:

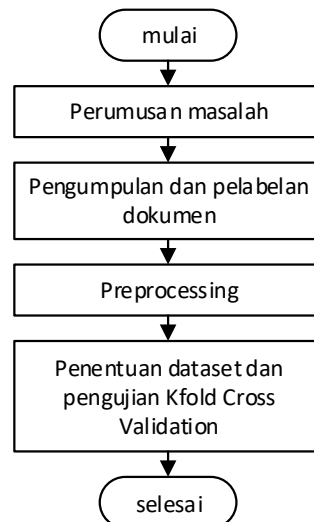
$$V_{MAP} = \operatorname{argmax}_{v_j \in V} P(v_j) \prod_i P(a_i | v_j) \quad (4)$$

Dengan  $P(v_j) = \frac{|v_j|}{|V|}$  dan  $P(a_i | v_j) = \frac{n_{ij}+1}{n_j+|kosakata|}$ , dimana:

- $P(v_j)$  : Probabilitas setiap dokumen  $j$  terhadap seluruh dokumen.
- $P(a_i | v_j)$  : Probabilitas kemunculan kata  $a_i$  pada dokumen  $v_j$
- $|v_j|$  : Frekuensi dokumen pada kategori  $j$
- $|V|$  : Jumlah seluruh dokumen
- $n_{ij}$  : Frekuensi kata ke- $i$  pada kategori  $j$
- $n_j$  : Frekuensi seluruh kata pada kategori  $j$
- $|kosakata|$  : Jumlah seluruh kata yang tersimpan

### 3. METODE PENELITIAN

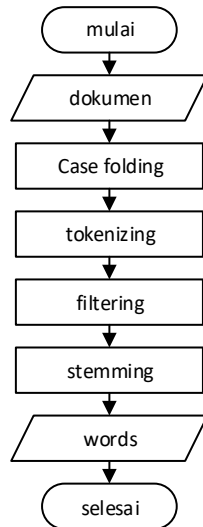
Langkah-langkah yang dilakukan dalam penelitian ini digambarkan pada kerangka penelitian berikut.



Gambar 1. Kerangka penelitian

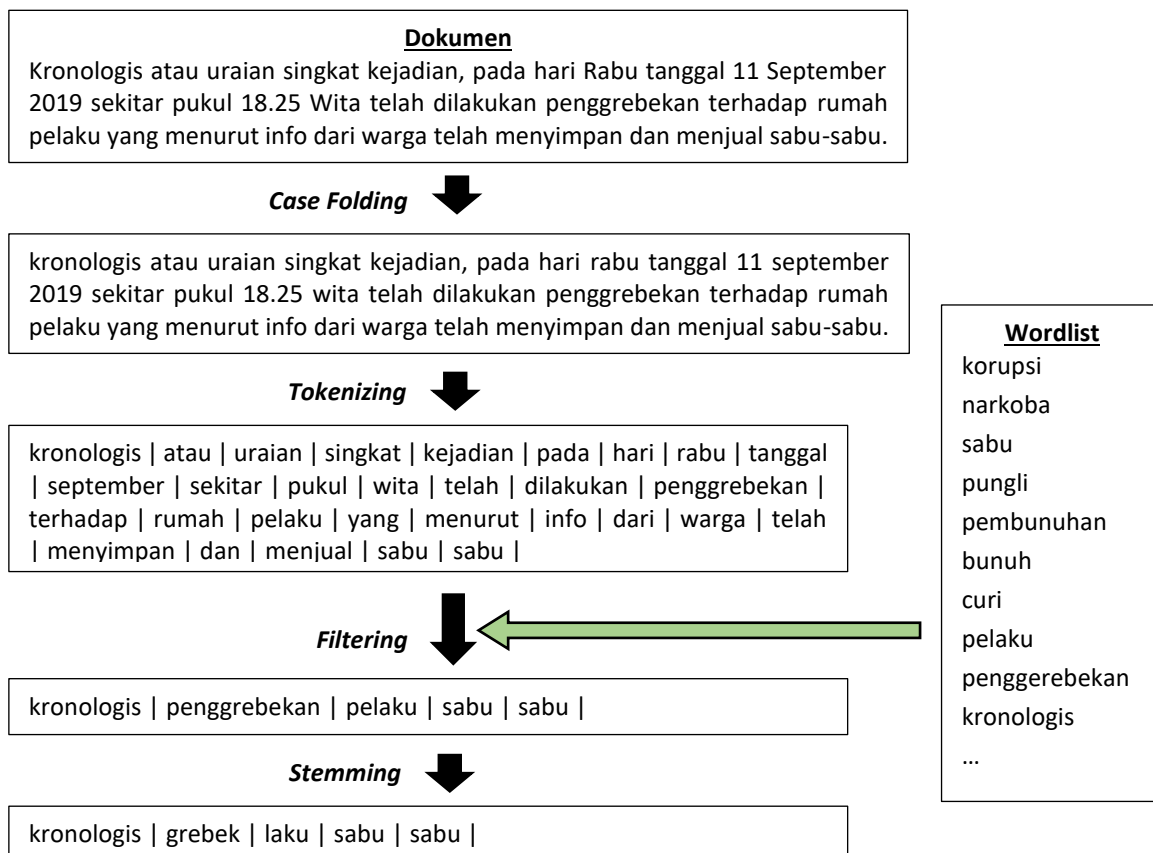
Pada penelitian ini dibatasi pada lima kategori dari berita kriminal, yaitu kategori korupsi, narkoba, pemerkosaan, pencurian dan pembunuhan. Dokumen yang digunakan dalam penelitian berupa artikel-artikel berita yang diambil dari tiga media massa *online* di Kalimantan Selatan, yaitu banjarmasinpost.tribunnews.com, antarakalsel.com dan kalsel.prokal.co. Sebanyak 100 dokumen per kategori dikumpulkan dengan persentase 34% dokumen dari banjarmasinpost.tribunnews.com, 33% dokumen dari antarakalsel.com dan 33% dokumen dari kalsel.prokal.co. Dengan demikian total data yang digunakan pada penelitian ini sebanyak 500 dokumen untuk semua kategori.

Tahapan selanjutnya adalah melakukan *preprocessing* terhadap seluruh dokumen yang dimiliki. Tahapan ini dilakukan untuk mendapatkan data berupa sekumpulan kata yang dibutuhkan pada proses klasifikasi dari setiap dokumen. Ada empat tahapan pada *preprocessing*, yaitu *case folding*, *tokenizing*, *filtering* dan *stemming*.



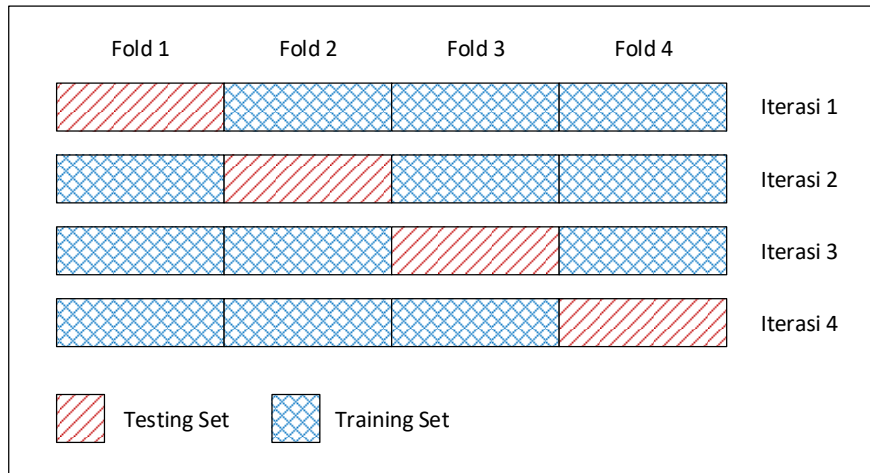
Gambar 2. Alur tahapan *preprocessing*

Pada tahap *case folding*, semua huruf pada dokumen akan diubah menjadi huruf kecil (*lowercase*). Kemudian pada tahap *tokenizing* setiap kata dikumpulkan dan membuang karakter-karakter bukan huruf karena bukan bagian dari penyusun kata. Tahapan selanjutnya dari *preprocessing* adalah *filtering*. Pada tahapan ini kata-kata penting dari hasil *tokenizing* diambil dengan berdasarkan pada *wordlist* yang disimpan dalam *database*. Selanjutnya pada *preprocessing* dilakukan tahapan *stemming* dimana pada tahapan ini setiap kata hasil *filtering* dijadikan kata dasar dengan membuang imbuhan pada setiap kata. Ilustrasi dari *preprocessing* dapat dilihat pada gambar berikut.



Gambar 3. Ilustrasi tahapan *preprocessing* pada suatu dokumen

Setelah *preprocessing*, tahapan selanjutnya pada penelitian yang dilakukan adalah menentukan dataset berdasarkan nilai  $k$  pada *kfold cross validation* yang ditentukan. Metode *cross validation* membagi semua dokumen ke dalam sekumpulan  $k$  dataset, dimana setiap dataset tersebut memiliki kesempatan untuk menjadi *testing set*. Saat dataset ke- $i$  menjadi *testing set*, maka dataset lainnya digunakan untuk menentukan frekuensi seluruh kata pada kategori  $j$  ( $n_j$ ), frekuensi kata ke- $i$  pada kategori  $j$  ( $n_{ij}$ ) dan probabilitas setiap dokumen pada kategori  $j$  terhadap seluruh dokumen ( $P(v_j)$ ). Dapat dikatakan dataset lainnya tersebut menjadi *training set* pada *classifier* NBC. Setiap pengujian yang dilakukan pada *testing set* diukur presisi, sensitivitas (*recall*) dan akurasinya dengan menggunakan *confusion matrix*.



Gambar 4. Ilustrasi *k-fold cross validation*

#### 4. HASIL DAN PEMBAHASAN

Sebanyak 500 dokumen digunakan sebagai dataset untuk pelatihan dan pengujian klasifikasi berita kriminal menggunakan NBC. Dataset tersebut terbagi ke dalam lima kategori, yaitu korupsi, narkoba, pencurian, pemerkosaan dan pembunuhan, sehingga terdapat 100 dokumen di setiap kategori. Pengujian dilakukan dengan teknik *K-Fold Cross Validation* dengan nilai  $K$  dari 3 sampai 10.

##### 4.1 Hasil Pengujian dengan $k = 3$

Dengan menggunakan 3-fold, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 66,67 : 33,33. Jumlah dokumen untuk setiap *fold* adalah sebanyak 170 dokumen. Pengujian dilakukan sebanyak 3 iterasi.

Tabel 1. Hasil pengujian dengan  $k = 3$

<i>Iteration</i>	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
1	94.38 %	98.29 %	98.48 %
2	100 %	98.65 %	99.73 %
3	100 %	98.37 %	99.67 %
rata-rata	98.13 %	98.44 %	99.29 %

Dapat dilihat pada Tabel 1, nilai rata-rata *precision* sebesar 98,13 %, nilai rata-rata *recall* sebesar 98,44 % dan nilai rata-rata *accuracy* sebesar 99,29 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai  $k = 3$  adalah sebesar 98,62 %.

#### 4.2 Hasil Pengujian dengan $k = 4$

Dengan menggunakan 4-fold, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 75 : 25. Jumlah dokumen untuk setiap *fold* adalah sebanyak 125 dokumen. Pengujian dilakukan sebanyak 4 iterasi.

Tabel 2. Hasil pengujian dengan  $k = 4$

<i>Iteration</i>	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
1	98.41 %	98.29 %	99.34 %
2	98.43 %	99.47 %	99.57 %
3	96.81 %	96.47 %	98.65 %
4	98.41 %	98.46 %	99.37 %
rata-rata	98.02 %	98.17 %	99.23 %

Dapat dilihat pada Tabel 2, nilai rata-rata *precision* sebesar 98,02 %, nilai rata-rata *recall* sebesar 98,17 % dan nilai rata-rata *accuracy* sebesar 99,23 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai  $k = 4$  adalah sebesar 98,47 %.

#### 4.3 Hasil Pengujian dengan $k = 5$

Dengan menggunakan 5-fold, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 80 : 20. Jumlah dokumen untuk setiap *fold* adalah sebanyak 100 dokumen. Pengujian dilakukan sebanyak 5 iterasi.

Tabel 3. Hasil pengujian dengan  $k = 5$

<i>Iteration</i>	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
1	98.05 %	99.62 %	99.52 %
2	96.09 %	97.41 %	98.68 %
3	100 %	98.42 %	99.68 %
4	98.02 %	98.06 %	99.21 %
5	96.03 %	95.92 %	98.38 %
rata-rata	97.64 %	97.89 %	99.09 %

Dapat dilihat pada Tabel 3, nilai rata-rata *precision* sebesar 97,64 %, nilai rata-rata *recall* sebesar 97,89 % dan nilai rata-rata *accuracy* sebesar 99,09 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai  $k = 5$  adalah sebesar 98,21 %.

#### 4.4 Hasil Pengujian dengan $k = 6$

Dengan menggunakan 6-fold, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 83,33 : 16,67. Jumlah dokumen untuk setiap *fold* adalah sebanyak 85 dokumen. Pengujian dilakukan sebanyak 6 iterasi.

Tabel 4. Hasil pengujian dengan  $k = 6$

<i>Iteration</i>	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
1	100 %	96.58 %	99.31 %
2	100 %	100 %	100 %
3	97.68 %	97.98 %	99.12 %
4	100 %	100 %	100 %
5	95.45 %	97.48 %	98.55 %
6	100 %	100 %	100 %

rata-rata	98.86 %	98.67 %	99.5 %
-----------	---------	---------	--------

Dapat dilihat pada Tabel 4, nilai rata-rata *precision* sebesar 98,86 %, nilai rata-rata *recall* sebesar 98,67 % dan nilai rata-rata *accuracy* sebesar 99,5 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai  $k = 6$  adalah sebesar 99,01 %.

#### 4.5 Hasil Pengujian dengan $k = 7$

Dengan menggunakan *7-fold*, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 85,71 : 14,29. Jumlah dokumen untuk setiap *fold* adalah sebanyak 75 dokumen. Pengujian dilakukan sebanyak 7 iterasi.

Tabel 5. Hasil pengujian dengan  $k = 7$

<i>Iteration</i>	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
1	97.42 %	99.23 %	99.31 %
2	100 %	98.28 %	99.65 %
3	97.43 %	99.93 %	99.45 %
4	100 %	96.46 %	99.28 %
5	100 %	96.28 %	99.25 %
6	100 %	99.09 %	99.82 %
7	100 %	100 %	100 %
rata-rata	99.26 %	98.47 %	99.54 %

Dapat dilihat pada Tabel 5, nilai rata-rata *precision* sebesar 99,26 %, nilai rata-rata *recall* sebesar 98,47 % dan nilai rata-rata *accuracy* sebesar 99,54 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai  $k = 7$  adalah sebesar 99,09 %.

#### 4.6 Hasil Pengujian dengan $k = 8$

Dengan menggunakan *8-fold*, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 87,5 : 12,5. Jumlah dokumen untuk setiap *fold* adalah sebanyak 65 dokumen. Pengujian dilakukan sebanyak 8 iterasi.

Tabel 6. Hasil pengujian dengan  $k = 8$

<i>Iteration</i>	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
1	100 %	96.97 %	99.39 %
2	97.05 %	99.72 %	99.33 %
3	100 %	96.32 %	99.25 %
4	100 %	96.69 %	99.33 %
5	97.05 %	99.72 %	99.33 %
6	97.04 %	99.21 %	99.23 %
7	100 %	100 %	100 %
8	97 %	98 %	98.98 %
rata-rata	98.52 %	98.33 %	99.36 %

Dapat dilihat pada Tabel 6, nilai rata-rata *precision* sebesar 98,52 %, nilai rata-rata *recall* sebesar 98,33 % dan nilai rata-rata *accuracy* sebesar 99,36 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai  $k = 8$  adalah sebesar 98,74 %.

#### 4.7 Hasil Pengujian dengan $k = 9$

Dengan menggunakan 9-fold, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 88,89 : 11,11. Jumlah dokumen untuk setiap *fold* adalah sebanyak 60 dokumen. Pengujian dilakukan sebanyak 9 iterasi.

Tabel 7. Hasil pengujian dengan  $k = 9$

<i>Iteration</i>	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
1	100 %	100 %	100 %
2	100 %	100 %	100 %
3	100 %	98.14 %	99.62 %
4	96.63 %	93.93 %	98.11 %
5	100 %	100 %	100 %
6	100 %	100 %	100 %
7	93.66 %	96.83 %	98.03 %
8	100 %	100 %	100 %
9	100 %	100 %	100 %
rata-rata	98.92 %	98.77 %	99.53 %

Dapat dilihat pada Tabel 7, nilai rata-rata *precision* sebesar 98,92 %, nilai rata-rata *recall* sebesar 98,77 % dan nilai rata-rata *accuracy* sebesar 99,53 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai  $k = 9$  adalah sebesar 99,07 %.

#### 4.8 Hasil Pengujian dengan $k = 10$

Dengan menggunakan 10-fold, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 90 : 10. Jumlah dokumen untuk setiap *fold* adalah sebanyak 50 dokumen. Pengujian dilakukan sebanyak 10 iterasi.

Tabel 8. Hasil pengujian dengan  $k = 10$

<i>Iteration</i>	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
1	100 %	100 %	100 %
2	100 %	100 %	100 %
3	100 %	100 %	100 %
4	100 %	100 %	100 %
5	96.21 %	99.45 %	99.09 %
6	100 %	94.59 %	98.9 %
7	92.42 %	95.61 %	97.51 %
8	100 %	100 %	100 %
9	100 %	98.12 %	99.62 %
10	100 %	100 %	100 %
rata-rata	98.86 %	98.78 %	99.51 %

Dapat dilihat pada Tabel 8, nilai rata-rata *precision* sebesar 98,86 %, nilai rata-rata *recall* sebesar 98,78 % dan nilai rata-rata *accuracy* sebesar 99,51 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai  $k = 10$  adalah sebesar 99,05 %.



#### 4.9 Rata-Rata Hasil Pengujian

Berikut disajikan tabel mengenai rata-rata nilai *precision*, *recall* dan *accuracy* dari hasil pengujian dengan *k* sama dengan 3 sampai 10.

Tabel 9. Rata-rata hasil pengujian

<i>K-Fold</i>	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
3	98.13 %	98.44 %	99.29 %
4	98.02 %	98.17 %	99.23 %
5	97.64 %	97.89 %	99.09 %
6	98.86 %	98.67 %	99.5 %
7	99.26 %	98.47 %	99.54 %
8	98.52 %	98.33 %	99.36 %
9	98.92 %	98.77 %	99.53 %
10	98.86 %	98.78 %	99.51 %
rata-rata	98.53 %	98.44 %	99.38 %

Dapat dilihat pada Tabel 9, dari seluruh pengujian menggunakan *K-Fold Cross Validation* dengan nilai *k* dari 3 sampai 10 menghasilkan rata-rata nilai *precision*, *recall* dan *accuracy* yang saling berdekatan dengan nilai yang sangat tinggi. Jika diambil nilai rata-rata, didapatkan kesimpulan bahwa klasifikasi berita kriminal dengan menggunakan NBC memiliki nilai *precision* sebesar 98,53 %, nilai *recall* sebesar 98,44 % dan nilai *accuracy* sebesar 99,38 %.

#### 5. KESIMPULAN

Penelitian ini membahas tentang kemampuan *Naïve Bayes Classifier* (NBC) dalam klasifikasi berita kriminal ke dalam lima kategori, yaitu korupsi, narkoba, pencurian, pemerkosaan dan pembunuhan. Dataset yang digunakan sebanyak 500 dokumen/artikel berita *online* yang diambil dari tiga media massa *online* Kalimantan Selatan, yaitu banjarmasin.tribunnews.com, antarakalsel.com dan kalsel.prokal.co. Pengujian dilakukan menggunakan *K-Fold Cross Validation* dengan nilai *k* diatur dari 3 sampai 10. Hasil pengujian menyatakan bahwa NBC memiliki kemampuan dalam klasifikasi berita kriminal dengan nilai *precision* sebesar 98,53 %, nilai *recall* sebesar 98,44 % dan nilai *accuracy* sebesar 99,38 %.

#### UCAPAN TERIMA KASIH

Direktorat Riset dan Pengabdian Masyarakat Kementerian Riset, Teknologi, dan Pendidikan Tinggi yang telah membiayai penelitian ini melalui skema Penelitian Dosen Pemula Tahun Anggaran 2019 dengan No. Kontrak 043/SP2H/LT/DRPM/2019.

#### DAFTAR PUSTAKA

- Fanissa, S., Fauzi, M. A., & Adinugroho, S. (2018). Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(8), 2766–2770.
- Hamzah, A. (2012). Klasifikasi Teks dengan Naive Bayes Classifier (NBC) untuk Pengelompokan Tekse Berita dan Abstract Akademis. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III*, 269–277.
- Herlambang, A. D., & Wijoyo, S. H. (2019). Algoritma Naive Bayes untuk Klasifikasi Sumber Belajar Berbasis Teks pada Mata Pelajaran Produktif di SMK Rumpun Teknologi Informasi dan Komunikasi. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 6(4), 430–435.
- Putra, I. B. G. W., Sudarma, M., & Kumara, I. N. S. (2016). Klasifikasi Teks Bahasa Bali dengan Metode Supervised Learning Naive Bayes Classifier. *Teknologi Elektro*, 15(2), 81–86.

- Rhomadhona, H., & Permadi, J. (2019). Penerapan Naïve Bayes Classifier untuk Klasifikasi Berita Kriminal di Kalimantan Selatan. *Seminastika Universitas Mulia Balikpapan*, 2(001), 18–23.
- Utami, P. D., & Sari, R. (2018). Filtering Hoax Menggunakan Naive Bayes Classifier. *MULTINETICS*, 4(1), 57–61.

#### **Biodata Penulis**



***Herfia Rhomadhona, S.Kom., M.Cs.*** lahir di Tanah Laut pada tanggal 21 April 1989. Penulis kedua memperoleh gelar Magister *Computer Science* Program Sarjana Magister Ilmu Komputer Universitas Gadjah Mada, Yogyakarta pada tahun 2016. Penulis saat ini menjadi dosen di Politeknik Negeri Tanah Laut.



***Jaka Permadi***, mendapatkan gelar S.Si. dari Universitas Tanjungpura Pontianak dalam bidang matematika pada tahun 2011 dan gelar M.Cs. dari Universitas Gadjah Mada Yogyakarta dalam bidang Ilmu Komputer pada tahun 2016. Pekerjaan penulis saat ini adalah sebagai dosen di Program Studi D3 Teknik Informatika, Politeknik Negeri Tanah Laut. Ketertarikan penelitian pada bidang kecerdasan buatan, pengolahan citra, data mining dan jaringan syaraf tiruan.