

KLASIFIKASI PENGADUAN LARAS ONLINE BERBASIS TEXT MINING MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DAN NAÏVE BAYES

Muhammad Azis Suprayogi
Universitas Gunadarma
Jalan Margonda Raya No. 100, Depok, Jawa Barat
azissuprayogi.gunadarma@gmail.com

Abstrak

Laras Online adalah fasilitas pada Pemkab Bogor yang diperuntukkan bagi masyarakat terutama warga Kabupaten Bogor sebagai wadah aspirasi dan pengaduan. Seiring dengan jumlah pengaduan yang masuk sangat banyak sehingga mengakibatkan waktu yang lebih lama yang digunakan oleh petugas admin dalam memilah kemudian menentukan unit tujuan pengaduan. Penelitian ini bertujuan untuk membandingkan performansi klasifikasi dokumen pengaduan pada situs Laras Online menggunakan algoritma K-Nearest Neighbor (KNN) dan Naïve Bayes Classifier (NBC). Penelitian dilakukan dengan cara mengumpulkan dokumen pengaduan, melakukan preprocessing, pembobotan kata, klasifikasi, dan pengujian. Pengujian menggunakan cross validation dengan parameter $k\text{-fold}=10$ dan confusion matrix berdasarkan nilai accuracy, precision, recall, dan score-f1. Hasil pengujian terhadap 360 dataset menunjukkan bahwa algoritma NBC lebih baik dari algoritma KNN dengan nilai $k=3$, $k=5$, $k=7$, dan $k=9$ untuk mengklasifikasikan dokumen pengaduan ke dalam 6 kategori. Hasil klasifikasi menggunakan algoritma NBC memberikan nilai accuracy sebesar 79,16% dengan nilai precision tertinggi pada 2 kategori yaitu Dinsos 91,30% dan SatpolPP 66,80%, nilai recall tertinggi pada 4 kategori yaitu Disdukcapil 89,90%, Dislinghidup 88,40%, Dispupr 93,20%, dan Dishub 76,50%, serta nilai score-f1 tertinggi pada 4 kategori yaitu Disdukcapil sebesar 82,10%, Dislinghidup 82,90%, Dinsos 88,90%, dan Dishub 81,20%.

Kata Kunci: k-nearest neighbor, klasifikasi, naïve bayes classifier, pengaduan masyarakat

Abstract

Laras Online is a facility at the Bogor Regency Government that is intended for the community, especially residents of Bogor Regency for aspirations and complaints. Along with the large number of incoming complaints, it results in a longer time used by admin in sorting and determining the complaint unit. This study aims to compare the performance of the classification of public complaints on the Laras Online site using the K-Nearest Neighbor (KNN) and Naïve Bayes Classifier (NBC). The research is carried out by collecting documents, preprocessing, terms weighting, classifying and testing. The test used the $k\text{-fold}$ cross validation ($k\text{-fold}=10$) and confusion matrix based on the accuracy, precision, and f1-score values. The test results of 360 datasets show that the NBC algorithm is better than the KNN ($k=3$, $k=5$, $k=7$, and $k=9$) to classify documents into 6 categories. The NBC classification provide accuracy rate of 79,16% with highest score of precision in 2 categories, namely Dinsos 91,30% and SatpolPP 66,80%, the highest score of recall in 4 categories namely Dukcapil 89,90%, Dislinghidup 88,40%, Dispupr 93,20%, and Dishub 76,50%, also the highest score of score-f1 in 4 categories namely Disdukcapil 82,10%, Dislinghidup 82,90%, Dinsos 88,90% and Dishub 81,20%.

Keywords: classification, complaints, k-nearest neighbor, naïve bayes classifier

PENDAHULUAN

Pemkab Bogor memiliki visi mewujudkan Kabupaten Bogor yang termaju, aman, nyaman, dan berkeadaban. Untuk mencapai visi tersebut, salah satu upaya yang dilakukan Pemerintah Kabupaten Bogor adalah menyediakan layanan aspirasi dan pengaduan masyarakat dalam rangka untuk memenuhi hak bertanya masyarakat serta untuk menyampaikan aspirasi atau pengaduan masyarakat terkait penyelenggaraan pemerintahan Kabupaten Bogor.

Pemkab Bogor telah menyediakan fasilitas pengaduan masyarakat melalui aplikasi Layanan Resmi Aspirasi Online (Laras Online). Fasilitas pengaduan tersebut dapat diakses oleh masyarakat melalui situs Laras Online <http://pengaduan.bogorkab.go.id>, melalui SMS yang dikirim ke no. **08111184399** dengan format pesan **teman#nama#alamat#pesan**, melalui aplikasi *mobile android* Laras Online, melalui *email* Satuan Kerja Perangkat Daerah, melalui surat kabar serta dengan cara datang langsung ke kantor yang dituju. Seluruh sumber aspirasi dan pengaduan tersebut terintegrasi dalam satu sistem dan dapat diakses pada situs Laras Online. Dalam hal ini, Dinas Komunikasi dan Informatika memiliki peran sebagai *admin* yang bertugas memonitor, mengelola, serta mengingatkan Satuan Kerja Perangkat Daerah (SKPD) atau Kecamatan untuk segera merespon dan menindaklanjuti aspirasi pengaduan masyarakat tersebut[1]. Jumlah

petugas *admin* yang terbatas serta jumlah pengaduan masuk yang cukup banyak sehingga menyebabkan proses penyampaian aduan masyarakat oleh petugas ke unit tujuan menjadi lambat. Oleh karena itu, perlu adanya solusi berupa sistem yang dapat melakukan klasifikasi penentuan unit tujuan pengaduan masyarakat. Aplikasi pengaduan Laras Online dapat digunakan sebagai objek penelitian dengan tema klasifikasi dokumen secara otomatis.

Penelitian serupa yang pernah dilakukan oleh beberapa peneliti tentang klasifikasi dokumen teks, yaitu Kurnia Iswardani dan Dyah Ariyanti[2] meneliti tentang klasifikasi keluhan masyarakat menggunakan *Naïve Bayes Classifier* untuk mengelompokkan keluhan masyarakat yang berasal dari SMS, telepon, datang ke kantor secara langsung maupun melalui media sosial ke Satuan Kerja Perangkat Desa (SKPD) tujuan. Dari hasil penelitian menunjukkan rata-rata akurasi yang diperoleh sebesar 95%. Penelitian selanjutnya oleh Titien S. Sukanto dan Heru Pramono Hadi[3] tentang klasifikasi jenis laporan masyarakat menggunakan algoritma *K-Nearest Neighbour*. Hasil penelitian ini menunjukkan algoritma *K-Nearest Neighbour* memberikan akurasi yang baik untuk klasifikasi laporan masyarakat menjadi tiga kategori yaitu permintaan informasi, pengaduan, dan aspirasi. Berdasarkan evaluasi dan validasi dengan *Confusion Matrix* diperoleh akurasi tertinggi sebesar 82% menggunakan $k=11$. Penelitian

Selanjutnya oleh Aida Indriani[4] yaitu melakukan analisa perbandingan algoritma *K-Nearest Neighbour* dan *Naïve Bayes Classifier* untuk klasifikasi data forum diskusi mahasiswa menjadi kategori topik berdasarkan isi materi forum. Penelitian tersebut menunjukkan bahwa algoritma *K-Nearest Neighbour* dan *Naïve Bayes Classifier* dapat digunakan untuk klasifikasi data forum, hasil pengukuran terhadap efektifitas klasifikasi terhadap 15 data uji menunjukkan nilai 73% untuk *Naïve Bayes Classifier* dan nilai 80% untuk *K-Nearest Neighbour* dengan menggunakan *Confusion Matrix*, sehingga disimpulkan bahwa *K-Nearest Neighbour* lebih baik dari *Naïve Bayes Classifier*. Penelitian berikutnya dilakukan oleh Hamdani Asril, Insanul Kamila, dan Mustakim[5] terhadap klasifikasi dokumen tugas akhir menggunakan algoritma *K-Nearest Neighbor* dan *Naïve Bayes Classifier*. Penelitian tersebut menghasilkan akurasi terbaik pada percobaan tiga kelas sebesar 91,67% untuk *K-Nearest Neighbor* nilai $k=6$, $k=7$, $k=8$ dan $k=9$ dan sebesar 86,11% untuk *Naïve Bayes Classifier*. Berdasarkan penelitian-penelitian sebelumnya tersebut, kami mencoba melakukan perbandingan akurasi klasifikasi antara *K-Nearest Neighbor (KNN)* dengan *Naïve Bayes Classifier (NBC)* menggunakan pengujian *K-fold Cross Validation* pada objek penelitian yang berbeda yaitu pengaduan masyarakat

Laras Online pada Pemerintah Kabupaten Bogor.

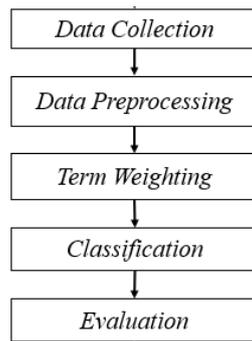
Tujuan penelitian ini untuk membandingkan dan mendapatkan model klasifikasi pengaduan dengan akurasi terbaik antara algoritma *K-Nearest Neighbor* dan *Naïve Bayes Classifier*. Selain itu dapat memberi saran kepada admin Laras Online untuk peningkatan fungsi sistem Laras Online agar dapat memberikan pelayanan yang lebih cepat.

METODE PENELITIAN

Kerangka kerja secara umum dapat dilihat pada Gambar 1, proses terdiri dari 5 tahapan yaitu proses pengumpulan data (*Data Collection*), praproses teks (*Data Preprocessing*), proses pembobotan kata (*Term Weighting*), proses klasifikasi (*Classification*) dan proses evaluasi (*Evaluation*).

Pengumpulan Data

Dataset yang digunakan pada penelitian adalah data aduan masyarakat pada sistem Laras Online yang bersumber dari SMS, *website* Laras Online <http://pengaduan.bogorkab.go.id> dan aplikasi *mobile* mulai bulan Januari 2019 sampai dengan Desember 2020. Data-data tersebut dapat diakses pada situs pengaduan Kabupaten Bogor secara bebas.



Gambar 1. Kerangka Kerja Umum Penelitian

Hasil *webscrapping* data *website* Laras Online seluruhnya sebanyak 829 data yang terbagi menjadi 53 instansi tujuan aduan. Proses pengumpulan data dilakukan dengan cara *webscrapping* halaman *website* Laras Online menggunakan *library* berbasis Python yaitu *BeautifulSoup*. Selanjutnya data teks tersebut diubah ke dalam *format CSV* dengan tujuan supaya lebih mudah diolah menggunakan *tools* pemrograman *Jupyter Notebook*.

Data yang digunakan dalam penelitian perlu dibatasi jumlahnya supaya waktu yang digunakan lebih efisien sehingga proses pengujian dan evaluasi dapat dilakukan secara efektif. Dengan demikian, perlu ditentukan kriteria data yang digunakan untuk penelitian yaitu: 1) Data aduan yang masuk mulai Januari 2019 s.d. Desember 2020, 2) Data aduan masuk melalui *website*, SMS, serta aplikasi *mobile*, 3) Data aduan sudah mendapat respon petugas *admin* Laras Online, 4) Kategori atau kelas ditentukan berjumlah 6 kategori berdasarkan unit yang memiliki

jumlah aduan terbanyak yaitu Dinas Kependudukan dan Pencatatan Sipil, Dinas Lingkungan Hidup, Dinas Sosial, Dinas Pekerjaan Umum dan Penataan Ruang, Dinas Perhubungan, dan Satuan Polisi Pamong Praja. Berdasarkan kriteria data tersebut, ditentukan dataset penelitian berjumlah 360 aduan. Tabel 1 memperlihatkan contoh data yang digunakan untuk penelitian.

Selanjutnya dipilih 6 kategori dari data tersebut yang merupakan unit tujuan terpopuler berdasarkan jumlah data aduan terbanyak, selanjutnya diambil 60 data dari masing-masing unit tersebut. Tujuan pengambilan data seimbang pada masing-masing unit dalam rangka mengeliminasi faktor data tidak seimbang yang dapat mempengaruhi tingkat akurasi klasifikasi kategori tertentu. Kami membuat kategori/label **0,1,2,3,4,5** yang mencerminkan unit/kategori klasifikasi untuk mempermudah proses penelitian. Jumlah data masing-masing kategori yang digunakan pada penelitian dapat dilihat pada Tabel 2.

Tabel 1. Contoh Data Aduan pada Laras Online

Aduan	Kategori
Lampu merah perempatan pemda mati sejak pagi hari tadi. Mohon ditindak lanjuti, karena sangat berbahaya bagi pengguna jalan. Terimakasih	Dinas Perhubungan
Mohon dibantu perbaikan jalan raya bojonggede menuju citayam terutama depan rumah sakit citama dan suzuki citayam, saya lihat ini masih bagian dari kabupaten bogor, dan setiap kali hujan pasti jalan tergenang ber hari hari	Dinas Pekerjaan Umum dan Penataan Ruang
Saya urus surat pindah dari ciomas ke jakarta dari tanggal 27 oktober samapai skrg belum Selesai dan belum ada jawaban dari dukcapil	Dinas Kependudukan dan Pencatatan Sipil
Truck pengangkut sampah menjatuhkan sampah didepan restoran kami sehingga berserakan dan berantakan, mengganggu dan beraroma busuk, TOLONG SEGERA DIBERSIHKAN	Dinas Lingkungan Hidup
Assalamu'alaikum maaf bapak/ibu, keluarga saya belum mendapat bantuan selama masa pandemi ini padahal tetangga-tetangga sudah mendapat bantuan lebih dari satu kali, bagaimana agar keluarga saya bisa dapat bantuan? Lokasi desa ciomas, kecamatan ciomas	Dinas Sosial
Kepada pemerintahan kabupaten bogor mohon kira nya dapat ditertibkan atau ditutup warung minum beralkohol dan karaokean ditempat terbuka dijalan utama narogong kp cikerewis desa limusnunggal sehingga mengganggu waktu istirahat warga sekitar...mohon kira nya ditindak demi kenyamanan warga sekitar ...terima kasih untuk perhatiannya	Satuan Polisi Pamong Praja

Tabel 2. Jumlah Dataset Penelitian

Kategori	Label	Jumlah Data
Dinas Kependudukan dan Pencatatan Sipil	0	60
Dinas Lingkungan Hidup	1	60
Dinas Sosial	2	60
Dinas Pekerjaan Umum dan Penataan Ruang	3	60
Dinas Perhubungan	4	60
Satuan Polisi Pamong Praja	5	60
Total		360

Preprocessing

Preprocessing data dilakukan melalui beberapa tahapan, dimulai dari tahap *case folding* yaitu mengubah teks dokumen menjadi huruf kecil seluruhnya, kemudian tahap *tokenization* yaitu membersihkan teks dari tanda baca, spasi berulang, mengubah baris baru menjadi spasi, dan memisahkan kata per kata dari kalimatnya, *stopword* yaitu menghapus kata yang tidak dibutuhkan, dan

tahap *stemming* yaitu menghilangkan imbuhan kata pada teks dokumen. Adapun prosesnya menggunakan *tools* berbasis Python berupa *library Numpy, Pandas, Sastrawi, dan NLTK*.

Pembobotan, Klasifikasi dan Evaluasi

Proses *preprocessing* dokumen hanya menghasilkan dokumen *bag of words* yang perlu diubah menjadi *term vector* yang bisa dilakukan dengan cara proses pembobotan[6].

Proses pembobotan dilakukan setelah *preprocessing* dengan menggunakan metode *Term Frequency – Inverse Document Frequency (TF-IDF)* yaitu menghitung bobot *term* pada sebuah dokumen berdasarkan seringnya kata tersebut muncul dimana bobot tersebut mengindikasikan pentingnya sebuah *term* terhadap dokumen, semakin banyak *term* tersebut muncul pada dokumen maka semakin tinggi nilai *term* tersebut[7]. Teknik TF-IDF juga mengeliminasi *term* yang bersifat sangat umum dan mengekstrak *term* yang memiliki relevansi yang tinggi dari *corpus*[8]. Tahap pertama menentukan nilai *Term Frequency (TF)* yaitu jumlah *term* yang terdapat pada setiap dokumen, tahap selanjutnya menentukan nilai *Inverse Document Frequency (IDF)* yang berfungsi untuk mengurangi bobot *term* yang jumlah kemunculannya banyak di seluruh dokumen menggunakan Persamaan (1). Selanjutnya menentukan bobot *term* dengan cara mengalikan nilai TF dengan IDF menggunakan Persamaan (2).

$$idf_i = \log\left(\frac{N}{df_i}\right) \quad (1)$$

$$w_{i,j} = tf_{i,j} \times idf_i \quad (2)$$

Dimana $w_{i,j}$ adalah bobot *term* i pada dokumen j , $tf_{i,j}$ adalah jumlah kemunculan *term* i pada dokumen j , idf_i adalah *inverse* df_i , df_i adalah banyaknya dokumen yang memuat *term* ke- i . N adalah jumlah dokumen. Proses klasifikasi menggunakan algoritma Naïve bayes Classifier dan *K-Nearest*

Neighbor serta memanfaatkan modul klasifikasi pada *scikit-learn python* yaitu *KNeighborsClassifier* dan *MultinomialNB*. Model algoritma *K-Nearest Neighbor* dibuat menggunakan ukuran jarak *Euclidean Distance* dan nilai k yang umum digunakan yaitu $k=3,5,7$, dan 9 untuk kemudian dipilih nilai k dengan performansi terbaik. Adapun model algoritma Naïve Bayes merupakan algoritma klasifikasi sederhana yang tidak mempunyai banyak pengaturan.

Selanjutnya dilakukan uji *K-fold Cross Validation* dengan nilai $k=10$. Pemilihan nilai $k=10$ tersebut karena pengujian *Cross Validation* pada nilai $k=10$ memiliki bias terkecil pada estimasi *error* prediksi serta memiliki nilai *mean square error* terendah[9][10]. Berdasarkan nilai $k=10$, maka jumlah keseluruhan 360 data pada penelitian ini akan dibagi ke dalam 10 bagian data yang masing-masing bagian memiliki 36 data dan pada setiap bagiannya terdiri dari 6 data dari masing-masing kategori. Proses pengujian dilakukan sebanyak k kali, jika salah satu bagian digunakan sebagai data *testing*, maka sembilan bagian sisanya akan menjadi data *training*. Tahapan selanjutnya adalah evaluasi klasifikasi menggunakan pengukuran nilai akurasi, recall, precision dan score-f1 dalam satuan persen berdasarkan rumusan menggunakan *Confusion Matrix* yang ditampilkan pada Tabel 3. True Positive (TP) adalah data yang diklasifikasikan berkaitan dengan kategori yang benar, False Positive (FP) adalah data yang

diklasifikasikan berkaitan dengan kategori yang salah, True Negative (TN) adalah data yang diklasifikasikan tidak berkaitan dengan kategori yang benar, False Negative (FN) adalah data yang diklasifikasikan tidak berkaitan dengan kategori yang salah[11]. Terdapat beberapa pengukuran untuk menghitung performansi klasifikasi yaitu *accuracy*, *precision*, *recall*, dan *score-f1*. Accuracy adalah jumlah proporsi prediksi yang benar[12], didefinisikan dalam Persamaan (3)

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (3)$$

Precision adalah proporsi jumlah dokumen yang relevan di antara seluruh dokumen yang terpilih oleh sistem[12], didefinisikan dalam Persamaan (4)

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Recall adalah proporsi jumlah dokumen yang relevan di antara seluruh dokumen relevan yang terdapat pada koleksi dokumen[12], didefinisikan dalam Persamaan (5)

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

Adapun *score-f1* (*f-measure*) didefinisikan dalam Persamaan (6)

$$score - f1 = \frac{TP+TN}{TP+FN+FP+TN} \quad (6)$$

HASIL DAN PEMBAHASAN

Hasil Preprocessing

Beberapa tahapan *preprocessing* yang dilakukan, yaitu *case folding*, *tokenization*, *stopword*, dan *stemming*. Contoh tahapan *preprocessing* pada teks aduan dapat dilihat seperti Tabel 4.

Tabel 3. Confusion Matrix

Terprediksi	Aktual	
	Positif	Negatif
Positif	True Positive (TP)	False Negative (FN)
Negatif	False Positive (FP)	True Negative (TN)

Tabel 4. Contoh Teks Aduan Hasil Preprocessing

Tahapan	Hasil
<i>Raw Data</i>	Lampu merah perempatan pemda mati sejak pagi hari tadi. Mohon ditindak lanjuti, karena sangat berbahaya bagi pengguna jalan. Terimakasih
<i>Case Folding</i>	lampu merah perempatan pemda mati sejak pagi hari tadi. mohon ditindak lanjuti, karena sangat berbahaya bagi pengguna jalan. terimakasih
<i>Tokenization</i>	lampu merah perempatan pemda mati sejak pagi hari tadi mohon ditindak lanjuti karena sangat berbahaya bagi pengguna jalan terimakasih
<i>Stopword</i>	lampu merah perempatan pemda mati ditindak lanjuti berbahaya pengguna jalan
<i>Stemming</i>	lampu merah empat mati bahaya guna jalan

Hasil Pembobotan

Proses ini dilakukan dengan menghitung bobot tiap kata (*term weighting*) pada sebuah dokumen berdasarkan seringnya kata tersebut muncul pada dokumen dengan cara mengalikan nilai TF dengan nilai IDF. Hasil dari pembobotan berupa nilai TF-IDF yang akan digunakan untuk perhitungan klasifikasi KNN dan NBC. Contoh nilai TF-IDF dapat dilihat pada Tabel 5.

Hasil Klasifikasi KNN dan NBC

Klasifikasi menggunakan data sebanyak 360 dokumen yang merupakan hasil

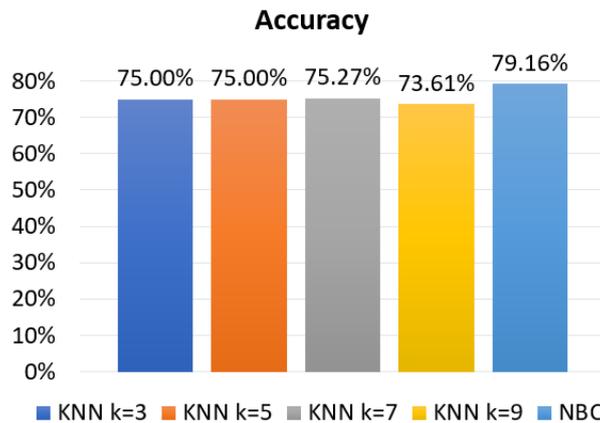
tahap *preprocessing* dan tahap pembobotan TF-IDF. Klasifikasi dilakukan menggunakan algoritma KNN dengan variasi nilai k $k=3$, $k=5$, $k=7$, dan $k=9$ dan algoritma NBC kemudian dilakukan uji *Cross Validation* dengan nilai $k\text{-fold}=10$ terhadap seluruh data dan hasilnya diukur dengan menghitung nilai rata-rata *accuracy*. Pengukuran juga dilakukan terhadap nilai *precision*, *recall*, dan *score-f1* untuk masing-masing kategori yaitu Disdukcapil, Dislinghidup, Dinsos, DisPupr, Dishub, dan SatpolPP. Perbandingan nilai *accuracy* hasil klasifikasi dapat dilihat pada Tabel 6 dan grafik pada Gambar 2.

Tabel 5. Contoh Perhitungan Nilai TF dan IDF

Term	tf pada d1	df untuk N=360	idf = $\text{Log}(N/df)$	w = tf x idf
lampu	1	32	1,0511	1,0511
merah	1	8	1,6532	1,6532
empat	1	2	2,2552	2,2552
mati	1	23	1,1945	1,1945
bahaya	1	22	1,2138	1,2138
guna	1	18	1,3010	1,3010
jalan	1	118	0,4844	0,4844

Tabel 6. Perbandingan Nilai Accuracy

Algoritma	Accuracy
KNN $k=3$	75,00 %
KNN $k=5$	75,00 %
KNN $k=7$	75,27 %
KNN $k=9$	73,61 %
NBC	79,16 %



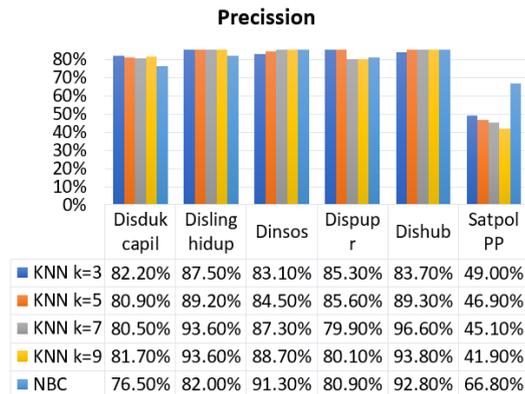
Gambar 2. Perbandingan Accuracy Klasifikasi dengan Cross Validation

Tabel 7. Perbandingan Nilai Precision, Recall, Dan Score-F1

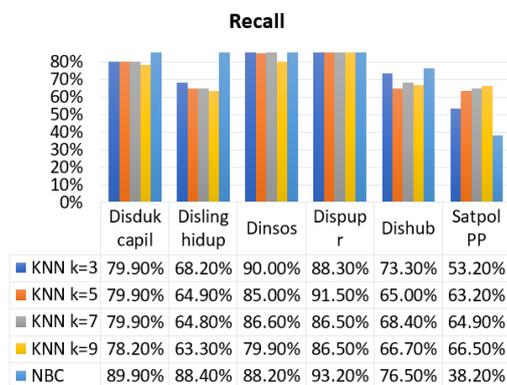
Algoritma	Pengukuran	Disduk capil	Disling hidup	Dinsos	Dis pupr	Dishub	Satpol PP
KNN k=3	precision	82,20%	87,50%	83,10%	85,30%	83,70%	49,00%
KNN k=3	recall	79,90%	68,20%	90,00%	88,30%	73,30%	53,20%
KNN k=3	score-f1	80,20%	73,20%	85,50%	82,10%	74,50%	49,30%
KNN k=5	precision	80,90%	89,20%	84,50%	85,60%	89,30%	46,90%
KNN k=5	recall	79,90%	64,90%	85,00%	91,50%	65,00%	63,20%
KNN k=5	score-f1	79,60%	71,60%	83,40%	87,90%	71,50%	53,00%
KNN k=7	precision	80,50%	93,60%	87,30%	79,90%	96,60%	45,10%
KNN k=7	recall	79,90%	64,80%	86,60%	86,50%	68,40%	64,90%
KNN k=7	score-f1	79,40%	72,90%	85,90%	82,50%	76,20%	52,10%
KNN k=9	precision	81,70%	93,60%	88,70%	80,10%	93,80%	41,90%
KNN k=9	recall	78,20%	63,30%	79,90%	86,50%	66,70%	66,50%
KNN k=9	score-f1	78,70%	72,40%	83,20%	82,90%	74,60%	50,30%
NBC	precision	76,50%	82,00%	91,30%	80,90%	92,80%	66,80%
NBC	recall	89,90%	88,40%	88,20%	93,20%	76,50%	38,20%
NBC	score-f1	82,10%	82,90%	88,90%	85,80%	81,20%	44,60%

Berdasarkan hasil perbandingan pada Tabel 6 dapat disimpulkan bahwa hasil accuracy tertinggi klasifikasi dokumen pengaduan diperoleh algoritma NBC sebesar 79,16 %. Adapun hasil akurasi algoritma KNN k=3 sebesar 75,00 %, k=5 sebesar 75,00

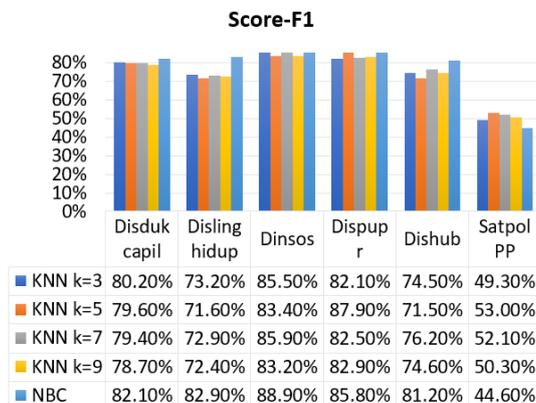
%, k=7 sebesar 75,27 %, dan k=9 sebesar 73,61%. Adapun untuk perbandingan pengukuran nilai precision, recall, dan score-f1 masing-masing kategori pada 6 kategori yang tersedia dapat dilihat pada Tabel 7 serta grafik perbandingan pada Gambar 3,4, dan 5.



Gambar 3. Perbandingan Nilai *Precision* terhadap Kategori Klasifikasi



Gambar 4. Perbandingan Nilai *Recall* terhadap Kategori Klasifikasi



Gambar 5. Perbandingan Nilai *Score-F1* terhadap Kategori Klasifikasi

Hasil pengukuran berdasarkan Tabel 7 dan grafik pada Gambar 3,4, dan 5 menunjukkan bahwa algoritma NBC memberikan nilai tertinggi untuk pengukuran *precision* pada 2 kategori yaitu Dinsos sebesar 91,30% dan SatpolPP 66,80%, sedangkan untuk 4 kategori lainnya yaitu Disdukcapil, Dislinghidup,

Dispupr dan Dishub nilai tertinggi diperoleh algoritma KNN dengan nilai *k* yang bervariasi yaitu secara berurutan sebesar 82,20% (KNN $k=3$), 93,60% (KNN $k=7, k=9$), 85,60% (KNN $k=5$), dan 96,60% (KNN $k=7$). Algoritma NBC memberikan nilai tertinggi untuk pengukuran *recall* pada 4 kategori yaitu

Disdukcapil sebesar 89,90%, Dislinghidup 88,40%, Dispupr 93,20%, dan Dishub 76,50%, sedangkan 2 kategori lainnya diperoleh algoritma KNN dengan nilai k yang bervariasi yaitu Dinsos 90,00% (KNN $k=3$), dan SatpolPP 66,50% (KNN $k=9$).

Adapun untuk pengukuran nilai *score-f1*, nilai tertinggi diperoleh algoritma NBC untuk 4 kategori yaitu Disdukcapil sebesar 82,10%, Dislinghidup 82,90%, Dinsos 88,90%, dan Dishub 81,20%. Untuk 2 kategori lainnya diperoleh algoritma KNN dengan nilai k yang bervariasi yaitu Dispupr 87,90% (KNN $k=5$) dan SatpolPP 53,00% (KNN $k=5$). Hasil perbandingan pengukuran nilai *precision*, *recall*, dan *score-f1* tersebut menunjukkan bahwa algoritma NBC mengungguli algoritma KNN dengan nilai k yang bervariasi ($k=3, k=5, k=7, k=9$) pada sebagian besar kategori klasifikasi.

Nilai *precision*, *recall* dan *score-f1* pada kategori SatpolPP memiliki nilai yang lebih rendah dibandingkan dengan kategori yang lain. Hal ini disebabkan karena pada kategori tersebut banyak terdapat kata yang memiliki nilai pembobotan kata yang rendah sehingga berpengaruh pada klasifikasi.

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pengujian dapat disimpulkan bahwa klasifikasi pengaduan masyarakat pada situs Laras Online dapat dilakukan menggunakan

algoritma KNN dan NBC didahului dengan proses *preprocessing* serta pembobotan kata.

Hasil pengujian *k-fold cross validation* dengan parameter *k-fold=10* terhadap klasifikasi menggunakan algoritma KNN dengan nilai $k=3, k=5, k=7$, dan $k=9$ dan algoritma NBC menunjukkan bahwa nilai *accuracy* tertinggi diperoleh algoritma NBC sebesar 79,16%.

Algoritma NBC memperoleh nilai *precision* tertinggi pada 2 kategori dari total 6 kategori yaitu kategori Dinsos 91,30% dan SatpolPP 66,80%, nilai *recall* tertinggi pada 4 kategori dari 6 kategori yaitu Disdukcapil 89,90%, Dislinghidup 88,40%, Dispupr 93,20%, dan Dishub 76,50% serta nilai *score-f1* tertinggi pada 4 kategori dari 6 kategori yaitu Disdukcapil 82,10%, Dislinghidup 82,90%, Dinsos 88,90%, dan Dishub 81,20%. Dengan demikian dapat disimpulkan bahwa algoritma NBC lebih baik dari algoritma KNN dalam mengklasifikasikan pengaduan masyarakat pada situs Laras Online.

Saran untuk hasil klasifikasi yang lebih akurat agar dilakukan proses *preprocessing* yang lebih baik yaitu dengan menambahkan fungsi perbaikan penulisan kata yang salah (*spelling correction*) maupun fungsi perbaikan penulisan spasi antar kata pada data aduan karena dalam data aduan ditemukan penulisan beberapa kata yang digabung tanpa menggunakan spasi. Diharapkan dengan melakukan proses *preprocessing* yang lebih baik akan meningkatkan akurasi klasifikasi.

DAFTAR PUSTAKA

- [1] Dinas Komunikasi dan Informatika Kabupaten Bogor, "Layanan Aspirasi Online." <https://diskominfo.bogorkab.go.id/laras-online/> (accessed Jan. 02, 2021).
- [2] D. Ariyanti, K. Iswardani, U. Panca, and M. Probolinggo, "Teks Mining untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma Naïve Bayes," *J. IKRA-ITH Inform.*, vol. 4, no. 3, pp. 125–132, 2020.
- [3] H. P. Hadi and T. S. Sukamto, "Klasifikasi Jenis Laporan Masyarakat Dengan K-Nearest Neighbor Algorithm," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 77–85, 2020, doi: 10.33633/joins.v5i1.3355.
- [4] A. Indriani, "Analisa Perbandingan Metode Naïve Bayes Classifier Dan K-Nearest Neighbor Terhadap Klasifikasi Data," *Sebatik*, vol. 24, no. 1, pp. 1–7, 2020, doi: 10.46984/sebatik.v24i1.909.
- [5] I. K. Hamdani Asril, Mustakim, "Klasifikasi Dokumen Tugas Akhir Berbasis Text Mining menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor," *Semin. Nas. Teknol. Inf. Komun. dan Ind.*, vol. 0, no. 0, pp. 2579–5406, 2019, [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/7995>.
- [6] V. Kalra and R. Aggarwal, "Importance of Text Data Preprocessing & Implementation in RapidMiner," *Proc. First Int. Conf. Inf. Technol. Knowl. Manag.*, vol. 14, pp. 71–75, 2018, doi: 10.15439/2017km46.
- [7] N. G. Yudiarta, M. Sudarma, and W. G. Ariastina, "Penerapan Metode Clustering Text Mining Untuk Pengelompokan Berita Pada Unstructured Textual Data," *Maj. Ilm. Teknol. Elektro*, vol. 17, no. 3, p. 339, 2018, doi: 10.24843/mite.2018.v17i03.p06.
- [8] P. Bafna, D. Pramod, and A. Vaidya, "Document clustering: TF-IDF approach," *Int. Conf. Electr. Electron. Optim. Tech. ICEEOT 2016*, no. November, pp. 61–66, 2016, doi: 10.1109/ICEEOT.2016.7754750.
- [9] D. Berrar, "Cross-validation," *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, no. April, pp. 542–545, 2018, doi: 10.1016/B978-0-12-809633-8.20349-X.
- [10] A. M. Molinaro, R. Simon, and R. M. Pfeiffer, "Prediction error estimation: A comparison of resampling methods," *Bioinformatics*, vol. 21, no. 15, pp. 3301–3307, 2005, doi: 10.1093/bioinformatics/bti499.
- [11] D. Yuliana and C. Supriyanto, "Klasifikasi Teks Pengaduan Masyarakat Dengan Menggunakan Algoritma Neural Network," *UPI*

- YPTK J. KomTekInfo*, vol. 5, no. 3, pp. 92–116, 2019.
- [12] L. A. Andika, P. A. N. Azizah, and R. Respatiwan, “Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier,” *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: 10.13057/ijas.v2i1.29998.