

## Klasifikasi Teks Pengaduan Suara Warga Kabupaten Pasuruan menggunakan Metode *Maximum Entropy*

Mayang Panca Rini<sup>1</sup>, Putra Pandu Adikara<sup>2</sup>, Sigit Adinugroho<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya  
Email: <sup>1</sup>panca.mayang@gmail.com, <sup>2</sup>adikara.putra@ub.ac.id, <sup>3</sup>sigit.adinu@ub.ac.id

### Abstrak

Suara Warga adalah *website* yang disediakan pemerintah Kabupaten Pasuruan untuk menampung pengaduan dari masyarakat. Dalam penerapannya, admin harus meneruskan pengaduan kepada dinas terkait secara manual. Untuk meningkatkan efisiensi waktu dibutuhkan adanya klasifikasi teks pengaduan, *Maximum Entropy* merupakan metode yang digunakan dalam penelitian ini dengan metode evaluasi *Confusion Matrix* yang akan menghitung evaluasi dari data latih seimbang dan data latih tidak seimbang dengan pengaduan sebanyak 200 data. Sebelum melakukan proses klasifikasi, tahap pertama yang dilakukan adalah *pre-processing* selanjutnya dilakukan proses pembobotan kata. Klasifikasi dilakukan dengan cara mencari peluang setiap kata yang ada pada setiap dokumen dan hasil klasifikasi didapatkan berdasarkan hasil peluang tertinggi dari kelas dokumen. Hasil evaluasi data latih seimbang menghasilkan hasil lebih baik daripada hasil evaluasi dengan data latih tidak seimbang dengan akurasi sebesar 89,27%, *precision* sebesar 92,49%, *recall* sebesar 89,27%, dan *f-measure* sebesar 89,44 %.

**Kata kunci:** klasifikasi, pengaduan Suara Warga, *Maximum Entropy*.

### Abstract

*Suara Warga is a website that it provided by the government of Pasuruan Residence to receive complain from the society. In the application, the admin must pass on the complain to related official manually. To increase the efficient of time, it is needed a classification of the text, Maximum Entropy is a method that is used in this research with Confusion Matrix evaluation method which will count the evaluation from the equal-wont data and the unequal-wont data, with the complain as much 200 data. Before doing the classification, the first step which is done is pre-processing and the next is process of word quality. Classification is done through looking for the opportunity of every word in every document and the result of classification is got based on the higher opportunity result from document class. The result of equal-wont data evaluation produce better result than the result of the unequal-wont data evaluation with the accuration: 89,27%, precision: 92,49%, recall: 89,27% and f-measure: 89,44%.*

**Keywords:** classification, complain of Suara Warga, *Maximum Entropy*.

## 1. PENDAHULUAN

Seiring berjalannya perkembangan akan teknologi semua pelayanan publik bisa dilakukan dengan menggunakan teknologi yang dapat memudahkan dalam hal pelayanan publik, salah satunya adalah pengaduan masyarakat. Hal ini menandakan bahwa informasi yang akurat dan terbaru sangat dibutuhkan. Dengan adanya perkembangan teknologi dapat membantu masyarakat untuk mendapatkan dan membagikan informasi ke masyarakat lain dengan cepat sehingga proses pengaduan bisa dilakukan secara efektif dan efisien (Prasetya,

Domai and Mindarti 2013).

Masyarakat diminta untuk dapat menyesuaikan diri terhadap perkembangan teknologi yang ada salah satunya yaitu pemanfaatan teknologi dalam bentuk *E-Government* (Prasanti, Fauzi and Furqon 2018). Peningkatan pelayanan publik adalah salah satu hal yang pasti diinginkan oleh masyarakat, mengingat pentingnya informasi bagi masyarakat dan peranan teknologi informasi dalam kehidupan. Dikutip dari *website* Suara Warga Kabupaten Pasuruan, salah seorang warga mengatakan bahwa selama ini kualitas pelayanan publik di Pasuruan masih sangat

kurang. Hal ini dibuktikan dengan banyaknya pengaduan masyarakat terhadap kualitas pelayanan publik. Di pemerintah Kabupaten Pasuruan sendiri pelayanan publik yang ada masih memiliki kekurangan, masyarakat kurang puas dengan informasi yang didapatkan karena informasi yang diterima kurang mutakhir.

Suara Warga merupakan *website* pengaduan *online* yang di sediakan pemerintah Kabupaten Pasuruan untuk masyarakat untuk mengirimkan kritik, saran, dan keluhan. Dalam penerapan Suara Warga, admin yang bertugas untuk menerima pengaduan yang masuk adalah Dinas Komunikasi dan Informatika (Diskominfo). Admin harus menentukan Satuan Kerja Perangkat Daerah (SKPD) yang sesuai dengan pengaduan juga meneruskan pengaduan ke SKPD yang bersangkutan secara manual. Hal ini kurang efisien karena jika jumlah pengaduan yang masuk terlalu banyak maka pengaduan yang penting tidak bisa ditangani secara cepat.

Untuk meningkatkan efisiensi waktu maka diperlukan sebuah metode klasifikasi teks, *Maximum Entropy* merupakan metode yang digunakan dalam penelitian ini. *Maximum Entropy* merupakan metode klasifikasi dengan mengutamakan nilai dari setiap bobot kata di setiap dokumen. Penelitian oleh (Masithoh and Utami 2016) mengangkat masalah tentang analisis klasifikasi topik dengan metode *Naïve Bayes Classifier*, *Naïve Bayes Multinomial Classifier* dan *Maximum Entropy*. Hasilnya membuktikan bahwa urutan hasil tingkat akurasi terbaik adalah menggunakan metode *Maximum Entropy* sebesar 99,31%, *Naïve Bayes Classifier* sebesar 98,82% dan yang terendah adalah *Naïve Bayes Multinomial Classifier* dengan tingkat akurasi sebesar 97,39%.

## 2. KAJIAN PUSTAKA

### 2.1 Klasifikasi Text

Klasifikasi merupakan pengelompokan dokumen berdasarkan data latih yang sudah dilabeli. Pada klasifikasi, kelas atau kategori sebuah data sudah ditentukan diawal. Tiap dokumen menunjuk pada suatu kelas tertentu berdasarkan dari hasil nilai klaifikasi tertinggi yang diperoleh. Untuk mengklasifikasikan teks, kita perlu melakukan tahapan *pre-processing* dan pembobotan kata terlebih dahulu diantaranya *cleaning*, *case folding*, *filtering*, *stemming*, dan *tokenisasi*. Proses pembobotan kata dengan metode *Term Frequency* (TF),

metode ini merupakan faktor untuk menentukan bobot dari kata pada suatu dokumen berdasarkan jumlah kemunculannya pada dokumen tersebut.

### 2.2 Maximum Entropy

*Maximum Entropy* adalah teknik umum untuk mengestimasi probabilitas distribusi data dan ketika tidak ada yang diketahui distribusinya *diuniformkan* yaitu memiliki *Maximum Entropy*, dalam mengklasifikasikan teks metode ini mengestimasi distribusi label dalam dokumen (Susilo, T. H. & Rochimah Siti, 2013). Untuk menghitung besaran *entropy* dinyatakan dalam Persamaan 1.

$$Entropy(X) = - \sum_{x \in X} p(x) \log_2 p(x) \quad (1)$$

Keterangan:

$Entropy(X)$  = *entropy* dari  $X$

$P(x)$  = peluang dari  $x$

Untuk mengklasifikasikan teks dengan metode *Maximum Entropy* digunakan rumus yang dinyatakan dalam Persamaan 2.

$$p(b|a) = \frac{\exp(\sum \lambda_x f_x(a,b))}{\sum \exp(\sum \lambda_x f_x(a,b))} \quad (2)$$

Keterangan:

$\lambda_x$  = parameter yang akan diestimasi

$f_x(a,b)$  = peluang dari *term a* yang ada di kelas *b*

$a$  = setiap *term* yang ada pada dokumen

$b$  = kelas dari sebuah dokumen

## 3. METODOLOGI

### 3.1. Data Penelitian

Data yang digunakan pada penelitian ini adalah data pengaduan warga Kabupaten Pasuruan dimana total data berjumlah 200 data pengaduan. Data yang ada terbagi dari Dinas Pekerjaan Umum dan Bina Marga (DPUBM) sebanyak 100 data, Dinas Kependudukan dan Catatan Sipil (Dispendukcapil) sebanyak 40 data, Dinas Komunikasi dan Informatika (Diskominfo) sebanyak 32 data, dan Badan Kepegawaian Daerah sebanyak 28 data. Data yang ada dibagi menjadi data latih yang terdiri dari 150 pengaduan dan data uji yang terdiri dari 50 pengaduan. Data latih terbagi atas 75 data dari DPUBM, 30 data dari Dispendukcapil, 24 data dari Diskominfo, dan 21 data dari BKD. Data uji terbagi atas 25 data dari DPUBM, 10 data dari

Dispendukcabil, 8 data dari Diskominfo, dan 7 data dari BKD.

### 3.2 Implementasi Algoritme

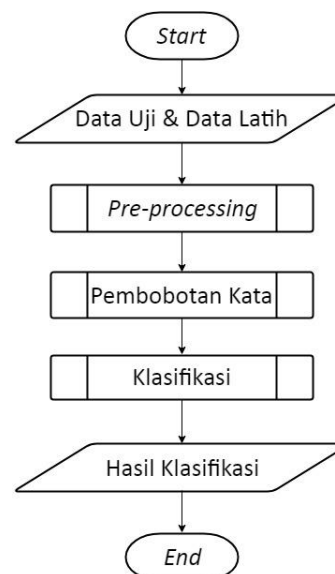
Tahap pertama yang dilakukan pada penelitian ini adalah melakukan tahapan *pre-processing* pada teks pengaduan Suara Warga. Selanjutnya dilakukan proses *term weighting* atau biasa disebut dengan proses pembobotan kata. Setelah didapatkan hasil pembobotan dari tiap pengaduan maka akan dilakukan proses klasifikasi untuk mendapatkan hasil kelas yang sesuai dengan pengaduan tersebut.

### 3.3 Teknik Analisis

Pada penelitian klasifikasi teks pengaduan Suara Warga ini dilakukan teknik pengujian dengan metode *Confusion Matrix* sebagai metode evaluasi. Pengujian dilakukan dengan menghitung *accuracy*, *recall*, *precision*, dan *f-measure*.

## 4. PERANCANGAN

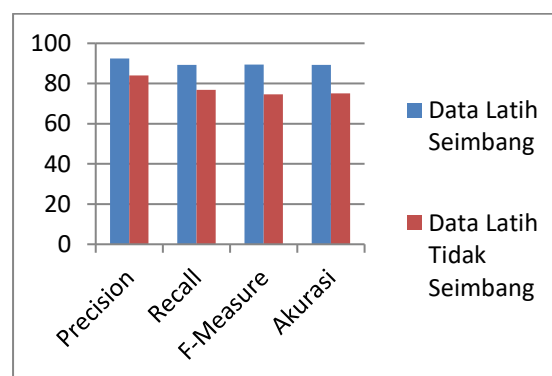
Dalam mengklasifikasikan pengaduan *online* Suara Warga di implementasikan dengan metode Maximum Entropy. Prinsip kerja secara umum dari sistem ini akan menghasilkan hasil akhir berupa klasifikasi pengaduan *online* Suara Warga. Gambaran umum sistem dijelaskan oleh Gambar 1 agar lebih mudah dimengerti bagaimana sistem dapat bekerja sampai menghasilkan klasifikasi pengaduan *online* Suara Warga. Sebelum melakukan proses klasifikasi terlebih dahulu sistem melakukan proses *pre-processing* dan pembobotan kata untuk *term* yang telah terpisah dengan metode *Term Frequency* (TF). Dataset yang digunakan berupa pengaduan *online* yang diambil dari web Suara Warga (<https://www.pasuruankab.go.id/suara.html>).



Gambar 1. Gambaran Umum Sistem

## 5. HASIL PENGUJIAN DAN ANALISIS

### 5.1 Pengujian *Maximum Entropy*



Gambar 2. Grafik Hasil Pengujian Data Latih Tidak Seimbang dan Data Latih Seimbang

Pada Gambar 2 menunjukkan hasil rata-rata dari nilai *precision* untuk data latih seimbang sebesar 92,49% dan data latih tidak seimbang sebesar 84,07%, nilai *recall* untuk data latih seimbang sebesar 89,27% dan data latih tidak seimbang sebesar 76,77%, nilai *f-measure* untuk data latih seimbang sebesar 89,44% dan data latih tidak seimbang sebesar 74,53%, dan nilai akurasi untuk data latih seimbang sebesar 89,27% dan data latih tidak seimbang sebesar 76,77%. Setelah dilakukan analisis menunjukkan bahwa hasil dari pengujian data latih seimbang lebih baik dalam mengkategorikan pengaduan daripada pengujian dengan data latih tidak seimbang. Hasil pengujian menunjukkan nilai *precision*, *recall*, *f-measure*, dan akurasi yang kurang maksimal karena terdapat beberapa data uji yang hasil klasifikasinya tidak sesuai. Setelah dilakukan

analisis hal itu bisa disebabkan karena adanya bobot dari fitur kata yang memiliki nilai lebih besar pada salah satu kelas sehingga dapat mempengaruhi hasil dari klaifikasinya.

### 5.2 Pengujian K-Fold Cross Validation

Tabel 1. Hasil Evaluasi K-Fold Cross Validation

	Akurasi	Precision	Recall	F-Measure
Fold-1	37,5%	52,68%	37,5%	34,5%
Fold-2	80,35%	94,55%	80,35%	85,16%
Fold-3	63,53%	94,12%	63,53%	69,97%
Fold-4	83,32%	93,41%	83,32%	83,92%
Fold-5	60,35%	92,7%	60,35%	66,53%
Average	65,01%	85,49%	65,01%	68,02%

Pada Tabel 1 dapat dilihat *fold-1* menghasilkan nilai *precision* sebesar 52,68%, nilai *recall* sebesar 37,5%, nilai *f-measure* sebesar 34,5% dan nilai akurasi sebesar 37,5%. Hasil tersebut merupakan hasil yang paling rendah dari semua *fold* yang ada. Setelah dilakukan analisis hal ini bisa terjadi karena adanya pembagian kelas pada data yang tidak seimbang sehingga mempengaruhi hasil pembobotan yang akan berpengaruh juga pada hasil klasifikasi. *Precision*, *recall*, dan akurasi dalam *fold-1* memiliki nilai yang lebih rendah daripada *fold* lain dikarenakan jumlah *true positif* yang dihasilkan lebih kecil daripada *fold* lain. Hasil *precision* dan *recall* yang rendah juga mempengaruhi nilai *f-measure* sehingga nilai *f-measure* yang dihasilkan juga rendah. Pengujian dengan metode *k-fold cross validation* menghasilkan rata-rata akurasi sebesar 65,01%, *precision* sebesar 85,49%, *recall* sebesar 65,01%, dan *f-measure* 68,02%. Hasil pengujian yang didapat tidak selalu sama dengan hasil pengujian selanjutnya karena metode pengambilan data dari setiap *fold* dilakukan secara acak. Hasil pengujian selanjutnya bisa mendapatkan hasil akurasi, *precision*, *recall*, dan *f-measure* yang lebih tinggi ataupun lebih rendah.

## 6. PENUTUP

### 6.1 Kesimpulan

1. Algoritme *Maximum Entropy* dapat digunakan untuk mengklasifikasikan teks

pengaduan Suara Warga Kabupaten Pasuruan. Sebelum melakukan proses klasifikasi, tahap pertama yang dilakukan adalah *pre-processing* selanjutnya dilakukan proses pembobotan kata. Hasil klasifikasi didapatkan berdasarkan hasil peluang tertinggi dari kelas dokumen.

2. Data uji seimbang menghasilkan nilai akurasi sebesar 89,27%, *precision* sebesar 92,49%, *recall* sebesar 89,27%, dan *f-measure* sebesar 89,44%. Data uji tidak seimbang menghasilkan nilai akurasi sebesar 76,77%, *precision* sebesar 84,07%, *recall* sebesar 76,77%, dan *f-measure* sebesar 74,53%. Hasil evaluasi metode *Maximum Entropy* menghasilkan nilai yang lebih baik dengan menggunakan data uji seimbang.

### 6.2 Saran

1. Data latih dan data uji pada penelitian ini perlu ditambahkan untuk mendapatkan nilai yang lebih baik.
2. Perlu ditambahkan beberapa kata ke dalam *stopword list* dan harus dilakukan penyaringan kata yang tidak memiliki makna penting dalam klasifikasi yang akan mempengaruhi keakuratan dari sistem.

### DAFTAR PUSTAKA

- Lipton, Zachary C., Elkan, C. and Naryanaswamy, B., 2014. Thresholding Classifiers to Maximize F1 Score. Joint European Conference on Machine Learning and Knowledge Discovery in Databases pp.225–239. Available at: <[https://link.springer.com/chapter/10.1007/978-3-662-44851-9\\_15](https://link.springer.com/chapter/10.1007/978-3-662-44851-9_15)>.
- Masithoh, N., 2016. Analisis Klasifikasi Topik Menggunakan Metode Naive Bayes Classifier, Naive Bayes Multinomial Classifier, dan Maximum Entropy pada Artikel Berita. S1. Universitas Gadjah Mada. Tersedia di <<http://etd.repository.ugm.ac.id>> [Diakses 16 Mei 2019]
- Niasita, Ananda Fitri, Putra Pandu Adikara, and Sigit Adinugroho. "Analisis Sentimen Pembangunan Infrastruktur di Indonesia dengan Automated Lexicon Word2Vec dan Naïve-Bayes." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2019: 2673-2679.

- Nigam, K., Lafferty, J. and Mccallum, A., 1999. Using Maximum Entropy for Text Classification. IJCAI-99 Workshop on Machine Learning for Information Filtering, [online] pp.61–67. Available at: <<http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.43.7517>>.
- Prasanti, Annisya Aprilia , M. Ali Fauzi, and M. Tanzil Furqon. "Klasifikasi Teks Pengaduan Pada Sambat Online Menggunakan Metode N-Gram dan Neighbor Weighted K-Nearest Neighbor (NW-KNN)." *JPTIHK* 2, no. 2 (2018): 594-601.
- Prasetya, D R, T Domai, and L I Mindarti. "Analisis Pengelolaan Pengaduan Masyarakat Dalam Rangka Pelayanan Publik." *Jurnal Administrasi Publik*, 2013: 1151-1158.
- Pratiwi, Y R, and Edy Widodo. "Comparison Of Maximum Entropy and Support Vector Machine Methods for Sentiment Analysis of Peralite Product Through Twitter Social Network." *International Journal of Advances in Electronics and Computer Science*, 2017: 2393-2835.
- Sabily, Alvandi Fadhil , Putra Pandu Adikara, and Mochammad Ali Fauzi. "Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan Metode Maximum Entropy." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* , 2019: 4204-4209 .
- Sari, Nur Hijriani Ayuning, Mochammad Ali Fauzi, and Putra Pandu Adikara. "Klasifikasi Dokumen Sambat Online Menggunakan Metode K-Nearest ." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* , 2018: 2449-2454 .
- Suharno, Claudio Fresta, M. Ali Fauzi, and Rizal Setya Perdana. "Klasifikasi Teks Bahasa Indonesia pada Dokumen Pengaduan Sambat." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2017: 1000-1007.
- Susilo, T.H., 2013. Pengklasifikasian Topik dan Analisis Sentimen Dalam Media Sosial. S2. Institut Teknologi Sepuluh Nopember. Tersedia di < <http://digilib.its.ac.id>> [Diakses 16 Mei 2019]
- Tika Grace, and Adiwijaya. "Klasifikasi Topik BeritaBerbahasa Indonesia Menggunakan Multilayer Perceptron." e-Proceeding of Engineering, 2019: 2355-9365.