

Klasifikasi Target Keluhan Untuk Lembaga Terkait Pada Instansi Pemerintah Menggunakan Support Vector Machine

Dwi Putri Kartini

Program Studi Teknik Elektro, Fakultas Teknik, Universitas Panca Marga

Jl. Yos Sudarso No. 107 Pabean Dringu Probolinggo 67271

Telp : (0335) 422715 , Fax : (0335) 427923

E-mail : dwiputri@upm.ac.id

ABSTRAK

Pengaduan merupakan bentuk penerapan dari pengawasan masyarakat terhadap kinerja aparatur pemerintah yang dapat disampaikan secara lisan maupun tertulis kepada lembaga atau instansi terkait. Media sosial seperti facebook sering digunakan untuk menuliskan keluhan melihat kinerja aparatur pemerintah. Keluhan ini sebagian besar tidak menyertakan alamat atau kepada lembaga atau instansi yang dituju. Padahal keluhan tersebut disampaikan dengan harapan segera mendapatkan jawaban yang cepat dan tepat dari lembaga atau instansi terkait.

Penelitian ini menggunakan metode Support Vector Machine (SVM). Metode SVM merupakan sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (feature space) berdimensi tinggi. Input yang digunakan dalam penelitian ini adalah dokumen keluhan masyarakat. Tahapan praproses meliputi case folding, tokenisasi, filtering/ stopword removal, stemming dan tf-idf. Output yang dihasilkan berupa nama lembaga atau instansi.

Terdapat enam rasio dalam uji coba penelitian ini, nilai akurasi tertinggi sebesar 85% yaitu pada rasio 80:20 sedangkan nilai akurasi terendah sebesar 68,8% yaitu pada rasio 50:50. Hal ini menunjukkan bahwa sistem yang dibangun sudah sesuai dengan yang diharapkan.

Kata Kunci: keluhan, klasifikasi, SVM

ABSTRACT

Complaints is an application of community supervision on the performance of government officials. The complaints can be delivered in spoken or written to the institutions or relevant agencies. Social networks like facebook are often used to write down complaints on the performances of government officials, most of them do not include the addresses of the institutions or agencies, in fact the complaints are delivered with an expectation to be answered quickly and accurately from the related institution.

This research uses the Support Vector Machine (SVM) method. This method is a system learning that uses space hypothesis of linear functions in a feature space with high dimension. The input used in this research is the document of complaints from the community. Stage preprocessing covers case folding, tokenization, filtering/stopword removal, stemming and tf-idf. The output result is the name of the institution.

There are six ratios in the trial of this research, the highest value of accuracy of 85% is obtained from the ratio of 80:20, while the lowest value of accuracy of 68.8% is from ratio 50:50. This shows that the system established is in accordance with the expectations.

Keywords: complaint, classifications, SVM

PENDAHULUAN

Pelayanan publik yang baik adalah pelayanan yang tidak menghasilkan kesenjangan antara apa yang diliat dan diharapkan oleh masyarakat dengan apa yang diberikan oleh pemerintah. Rendahnya tingkat produktivitas aparatur pemerintah dalam memberikan pelayanan kepada masyarakat, serta munculnya praktek korupsi, kolusi dan nepotisme (KKN) dalam penyelenggaraan pelayanan publik mampu memunculkan pelayanan yang bersifat diskriminatif. Hal ini bisa menjadi salah satu penyebab banyaknya masyarakat menyampaikan pengaduan.

Pengaduan biasanya berupa keluhan yang sifatnya membangun. Pengaduan juga merupakan bentuk penerapan dari pengawasan masyarakat terhadap kinerja aparatur pemerintah yang dapat disampaikan secara lisan maupun tertulis kepada lembaga atau instansi terkait.

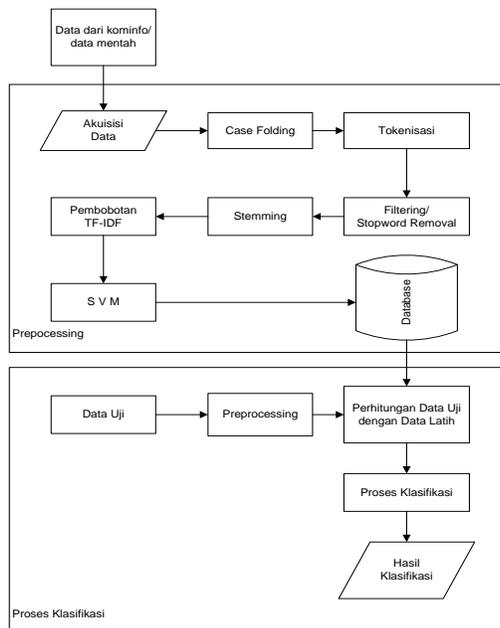
Dewasa ini, media sosial seperti facebook sering digunakan masyarakat untuk menuliskan keluhan mereka melihat kinerja aparat pemerintah. Keluhan ini sebagian besar tidak menyertakan alamat atau kepada lembaga atau instansi yang di tuju. Padahal keluhan tersebut disampaikan dengan harapan segera mendapatkan jawaban yang cepat dan tepat dari lembaga atau instansi terkait.

Sebuah disiplin ilmu *text mining* dapat digunakan untuk mengatasi permasalahan tersebut. *Text mining* merupakan proses ekstraksi pola berupa informasi dan pengetahuan yang berguna dari sejumlah besar sumber data teks. Jenis masukan untuk penambangan teks ini disebut data tak terstruktur dan merupakan pembeda utama dengan penambangan data yang menggunakan data terstruktur atau basis data sebagai masukan.

Metode yang digunakan adalah Support Vector Machine (SVM). SVM merupakan sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (*feature space*) berdimensi tinggi, dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan *learning bias* yang berasal dari teori pembelajaran statistik.

METODE

Sistem klasifikasi target keluhan ini melewati dua tahap. Diawali dengan tahap preprocessing dan dilanjutkan dengan tahap klasifikasi.



Gambar 1. Arsitektur Sistem

1) Case Folding

Proses penyamaan case dalam sebuah dokumen, hal ini dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu peran case folding dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar.

2) Tokenisasi

Proses pemotongan sebuah dokumen menjadi bagian-bagian, yang disebut dengan token. Pada saat bersamaan token juga membuang beberapa karakter tertentu yang dianggap sebagai tanda baca.

3) Filtering/ Stopword Removal

Tahapan yang berguna untuk menghilangkan karakter-karakter yang dianggap tidak valid (karakter angka dan symbol) dan Stopping atau stoplist removal adalah proses penghilangan kata-kata yang tidak berkontribusi banyak pada isi dokumen (Yates dan Nito, 1999). Kata-kata yang termasuk stopwords dihilangkan karena memberikan pengaruh tidak baik dalam pencarian dokumen yang dikehendaki user, seperti kata-kata dan, dari, kemudian, yang, di, ke, aku, kamu, dia dan lain-lain.

4) Stemming

Salah satu cara yang digunakan untuk meningkatkan performa IR dengan cara mentransformasi kata-kata dalam sebuah dokumen teks ke bentuk kata dasarnya. Algoritma *stemming* untuk bahasa yang satu berbeda dengan

algoritma stemming untuk bahasa lainnya. Sebagai contoh bahasa Inggris memiliki morfologi yang berbeda dengan bahasa Indonesia sehingga algoritma *stemming* untuk kedua bahasa tersebut juga berbeda.

5) Pembobotan TF-IDF

Cara pemberian bobot hubungan suatu kata (term) terhadap dokumen. Untuk dokumen tunggal tiap kalimat dianggap sebagai dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot yaitu *Term Frequency* (TF) yang merupakan frekuensi kemunculan kata (t) pada kalimat (d). *Document Frequency* (DF) adalah banyaknya kalimat dimana suatu kata muncul. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Bobot kata akan semakin besar jika sering muncul dalam suatu dokumen dan semakin kecil jika muncul dalam banyak dokumen (Robertson, 2005).

6) SVM

Suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. SVM memiliki prinsip dasar linier *classifier* yaitu kasus klasifikasi yang secara linier dapat dipisahkan, namun SVM telah dikembangkan agar dapat bekerja pada problem non-linier dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi. Pada ruang berdimensi tinggi, akan dicari hyperplane yang akan memaksimalkan jarak (margin) antara kelas data. Untuk mendapatkan hyperplane terbaik adalah dengan mencari hyperplane yang terletak di tengah-tengah antara dua bidang pembatas kelas dan untuk mendapatkan hyperplane terbaik itu, sama dengan memaksimalkan margin atau jarak antara dua set objek dari kelas yang berbeda. Margin dapat dihitung dengan $\frac{2}{\|w\|}$. (Santosa, 2007).

Data input yang digunakan adalah dokumen keluhan berbahasa Indonesia yang diambil dari Facebook beberapa dinas dan dari penelitian terdahulu yang jumlah keseluruhan adalah 500 data. Sedangkan Outputnya adalah salah satu nama dinas.

Kelas target dalam sistem klasifikasi ini adalah 10, yaitu Dinas Kependudukan dan Pencatatan Sipil (Disdukpil), Dinas Pendidikan (Diknas), Dinas Sosial (Dinsos), Dinas Perhubungan (Dishub), Dinas Pekerjaan Umum dan Penataan Ruang (Dispu), Perusahaan Listrik Negara (PLN), Perusahaan Daerah Air Minum (PDAM), Badan Penyelenggara Jaminan Kesehatan (BPJS), Satuan Administrasi Manunggal Satu Atap (SAMSAT) dan Dinas Lingkungan Hidup (DLH)

HASIL DAN PEMBAHASAN

Pembahasan terhadap hasil penelitian dan pengujian yang diperoleh disajikan dalam bentuk uraian teoritik, baik secara kualitatif maupun kuantitatif. Hasil percobaan sebaiknya ditampilkan dalam bentuk grafik atau pun tabel. Contoh penulisan gambar dan tabel :

Pada bagian ini akan dijelaskan tentang tahapan pemrosesan data dari sistem klasifikasi target keluhan masyarakat menggunakan SVM yang terdiri dari case folding, tokenisasi, filtering/stopword removal, stemming dan pembobotan tf-idf.

Berikut ini adalah contoh data keluhan yang digunakan

pak tolong saya sedang buat kartu keluarga baru, KK asli dan master KK di tahan di disdukcapil, sekarang di minta susah banyak alasan!

Case folding

pak tolong saya sedang buat kartu keluarga baru, kk asli dan master kk di tahan di disdukcapil, sekarang di minta susah banyak alasan!

Tokenisasi

pak tolong saya sedang buat kartu keluarga baru, , kk
 asli dan dan master kk di tahan di di minta susah banyak
 alasan !

Filtering

pak tolong saya sedang buat kartu keluarga baru kk asli
 dan dan master kk di tahan di di minta susah banyak alasan

Stemming

pak tolong saya sedang buat kartu keluarga baru kk asli dan master kk di tahan di disdukcapil sekarang di minta susah banyak alas

Pembobotan tf-idf

(0, 174) 1

(0, 224) 1
 (0, 230) 1
 (0, 396) 1
 (0, 465) 1
 (0, 497) 3
 (0, 950) 1
 (0, 1004) 1
 (0, 1079) 2
 (0, 1448) 1
 (0, 1569) 1
 (0, 1953) 1
 (0, 1977) 1
 (0, 2000) 1
 (0, 2162) 1
 (0, 2175) 1
 (0, 2321) 1

Dalam uji coba sistem klasifikasi target keluhan, data dibagi menjadi data latih dan data uji. Uji coba dilakukan terhadap program satu dan program dua. Program satu untuk tahapan preprocessing dan program dua untuk nilai akurasi. Uji coba nilai akurasi dilakukan sebanyak tujuh kali dengan rasio perbandingan yang bervariasi. Uji coba juga menampilkan *screenshot* program. Program dalam penelitian ini dibangun dengan menggunakan bahasa pemrograman Python.

Tabel berikut ini menjelaskan dokumen keluhan beserta target kelas sesuai dengan data input yang digunakan.

Tabel 1. Dokumen Keluhan Dan Target Kelas

No.	Dokumen Keluhan	Target Kelas
1.	katanya jaman canggih kapan mau maju bikin Kartu Keluarga saja lama. Ibulan masih belum beres.	Disdukpil
2.	mohon agar di awasi adanya praktek jual beli nilai karena nantinya apabila di loloskan anak yang seperti itu akan mencoreng nama sekolah juga dan prestasi sekolah akan menurun	Diknas
3.	tolong amankan pengamen yg bawa bayi di alun-alun ... kasian bgt bayinya.. tersiksa sampe kotor gt badannya. Mohon ditindaklanjuti, terima kasih	Dinsos
4.	ada lobang mengangah di jalan dekat pasar. Sudah banyak pengendara motor yg hampir terjatuh karena lobang ini. Mohon dg segera lakukan perbaikan agar tidak ada korban yg berjatuhan	Dishub
5.	Jalan yang menuju tiris tepatnya di desa Segaran ... sejak aku masih lajang sampai sekarang anak sudah dua belum diperbaiki jugak.	Dispu
6.	Bikin BPJS nya ank baru lahir sulit banget. Harus bikin surat pindah dulu buat kk n akte kelahiran. Dari BPJS d kasik waktu 2 minggu sedangkan buat kk n akte lama bnget. Gmn dengan BPJS nya ank saya kalou wktu berlakux sdh habis. Apakah anak saya tidak bisa mempunyain BPJS sterusx??? Mohon infox bagi yang tau.	BPJS
7.	didesa sumberkerang dlm bbrpa bulan ne..sering mati lampu dan voltase mnim shingga bxx peralatan rumah tdk bs trpaka..aplgi yg listrik fulsaan 10 menit mati 10 menit hidup..msyrkat sudah sgt resah..tolong bantuanx dan tinjauanx trmkasih.	PLN
8.	Air mati lagi pak,buk jangan suruh bayar aja pelayanan juga donk yang baik	PDAM
9.	gmn nih pelayanan samsat kraksan...hari kamis tgl 12/10/2017 petugase jm 8 blm datang .. payahhhh	SAMSAT
10.	Tiba-tiba muncul pabrik tahu, limbah dan polusinya sangat mengganggu warga. Mohon bantuannya..	DLH

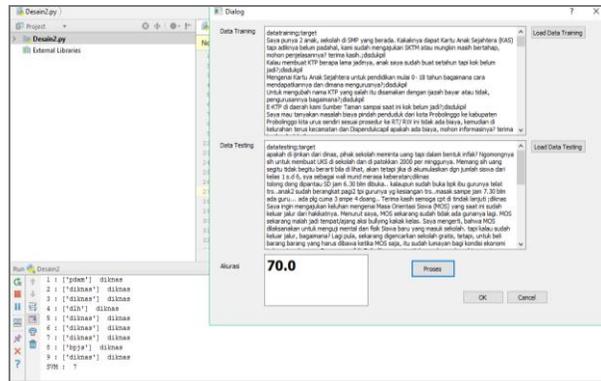
Selanjutnya data dalam tabel 1. akan diimplementasikan satu per satu ke dalam program yang dibangun dengan bahasa pemrograman Python yang selanjutnya akan dihitung nilai akurasinya. Dari hasil implementasi program, didapatkan 8 data yang diprediksi secara benar dan 2 data yang diprediksi salah dari total 10 data yang di uji coba. Selanjutnya akan dihitung nilai akurasinya.

$$\begin{aligned} \text{Nilai Akurasi} &= \frac{\text{jumlah data yang diprediksi benar}}{\text{jumlah data yang diprediksi}} \times 100\% \\ &= \frac{8}{10} \times 100\% \end{aligned}$$

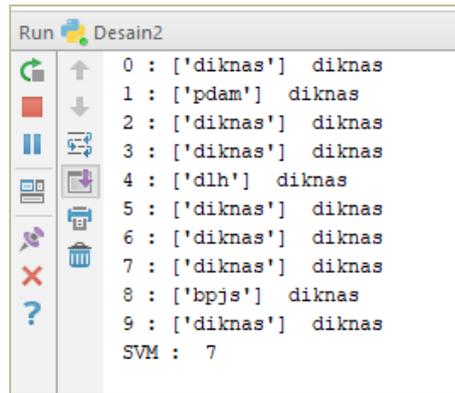
= 80%

Jadi, nilai akurasi yang didapatkan adalah sebesar 80%.

Berikut ini adalah Tampilan Uji Coba Akurasi yang menggunakan program dua, melibatkan 400 data training dan 10 data testing, menghasilkan akurasi sebesar 70%.



Gambar 2 . Tampilan Uji Coba Akurasi Dengan Program 2



Gambar 3. Hasil Tracing

Gambar 3 merupakan hasil *tracing* untuk 10 datatesting. Angka 0 sampai 9 menunjukkan nomor data, sedangkan yang berada didalam kurung siku merupakan nama dinas yang diprediksi oleh system, sedangkan nama dinas di sisi kanannya merupakan kelas target. Dari gambar tersebut didapatkan 7 data yang diprediksi secara benar dan 3 data yang diprediksi salah sehingga nilai akurasinya mecapai 70%.

SIMPULAN

Dapat disimpulkan bahwa algoritma SVM berhasil diterapkan untuk melakukan klasifikasi target keluhan dan berhasil mendapatkan nilai akurasi tertinggi sebesar 85%. Diharapkan agar penelitian selanjutnya dapat dikembangkan dengan menggunakan data yang dilakukan secara otomatis sehingga keluhan yang ditulis langsung tertuju pada lembaga atau instansi yang dimaksud. Harapan lainnya adalah agar penelitian selanjutnya dapat dikembangkan untuk menangani komentar yang menyangkut lebih dari dua lembaga atau instansi.

DAFTAR PUSTAKA

- [1] Anonymous, *Python* . [Online] Available at: <https://www.tuliskode.com/kelebihan-dan-kekurangan-python/> [Accessed 16 Mei 2019]
- [2] Budi Santosa, *Data Mining Terapan Dengan MATLAB*, Graha Ilmu, 2007
- [3] Dwiyanto, Agus.. *Reorientasi Ilmu Administrasi Publik: Dari Government ke Governance*. Pidato Pengukuhan Jabatan Guru Besar pada FISIPOL UGM. Yogyakarta, 2004
- [4] Eko Prasetyo, *DATA MINING-Konsep dan Aplikasi Menggunakan MATLAB*, Penerbit ANDI, 2012

- [5] Imam Cholissodin, Maya Kurniawati, Indriati, Issa Arwani . *Classification of Campus E-Complaint Documents using Direct Acyclic Graph Multi-Class SVM Based on Analytic Hierarchy Process*. Informatics Department, PTIIK, Brawijaya University, Malang, Indonesia. ICACSSIS 2014
- [6] Kusrini & Emha Taufiq Luthfi, *Algoritma Data Mining*, Penerbit ANDI Yogyakarta, 2009
- [7] Petrix Nomleni, Mochamad Hariadi dan I Ketut Eddy, *Sentiment Analysis Berbasis Big Data*. Jurusan Teknik Elektro, Institut Teknologi Sepuluh Nopember
- [8] Ratminto & Atik Septi Winarsih. 2006. *Manajemen Pelayanan*. Yogyakarta:PUSTAKA PELAJAR
- [9] Seyyed Mohammad Hossein Dadgar, Mohammad Shirzad Araghi, Morteza Mastery Farahani. *A Novel Text Mining Approach Based on TF-IDF and Support Vector Machine for News Classification*. Department of Computer Engineering, Islamic Azad University, Central Branch Tehran, Iran
- [10] Santoso, Budi. 2015. *Tutorial SVM*. [Online]
Available at: <https://bsantosa.files.wordpress.com/2015/03/tutorial-svm-2015.pdf/>
[Accessed 15 Juni 2019]
- [11] Shweta Joshi, Bhawna Nigam, *Categorizing the Document using Multi Class Classification in Data Mining*, Department of Computer Engineering, Institute of Engineering & Technology, Devi Ahilya Vishwavidyalaya, Indore (M.P.), India, 2011
- [12] Suryadi. 2010. Penanganan Keluhan Publik Pada Birokrasi Dinas Perijinan. Vol, 23. No, 4. pp, 293-303
- [13] Thorsten Joachims, *Text Categorization with Support Vector Machines: Learning with Many Relevant Features*, Universitas Dortmund Informatic LS8, Dortmund-Germany
- [14] Turban, E.; et.al. (2011). *Decision Support and Business Intelligence Systems* (9 ed.). New Jersey: Pearson Education, Inc.
- [15] Wibawa, Samodra (Ed.). 2009. *Administrasi Negara:Isu-Isu Kontemporer*. Yogyakarta: Graha Ilmu