Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X
Vol. 2, No. 8, Agustus 2018, hlm. 2449-2454 http://j-ptiik.ub.ac.id

Klasifikasi Dokumen Sambat *Online* Menggunakan Metode *K-Nearest Neighbor* dan *Features Selection* Berbasis *Categorical Proportional Difference*

Nur Hijriani Ayuning Sari[1], Mochammad Ali Fauzi[2], Putra Pandu Adikara[3]

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: [1]nayuningsari@gmail.com, [2]moch.ali.fauzi@ub.ac.id, [3]adikara.putra@ub.ac.id

Abstrak

Sambat *Online* merupakan fasilitas yang berfungsi untuk menampung saran, kritik, keluhan atau pertanyaan dari masyarakat kota Malang seputar Pemerintah Kota Malang melalui situs *web* yang sudah disediakan atau melalui pesan singkat kepada nomor yang sudah disediakan. Suatu teks pengaduan yang masuk akan dikategorikan ke dalam berbagai bidang SKPD yang bertanggung jawab, untuk mempermudah mengorganisir teks pengaduan dan meningkatkan efisiensi waktu administrator dalam memilah dan menentukan bidang SKPD tujuan maka perlu dibuat sistem cerdas yang dapat mengklasifikasikan dokumen sesuai tujuan. *K-Nearest Neighbor* (K-NN) merupakan metode klasifikasi yang mana akan mencari dokumen yang memiliki kedekatan antara dokumen. Metode seleksi fitur yang digunakan adalah menggunakan metode *Categorical Proportional Difference* (CPD) untuk mengukur derajat kontribusi sebuah kata. Proses yang dilakukan adalah mengumpulkan dokumen latih dan dokumen uji, melakukan tahap *preprocessing* dan seleksi fitur, pembobotan, kemudian dilakukan klasifikasi, dan pada tahap akhir dilakukan pengujian dan analisis terhadap hasil klasifikasi oleh sistem terkait nilai *accuracy, precision, recall*, dan *F-Measure*. Hasilnya kinerja yang paling optimal adalah penggunaan *k*=1 dengan *feature* sebanyak 100% sebesar 91,84%, yang mana nilai akurasinya lebih baik dibandingkan dengan adanya seleksi fitur karena adanya penghapusan *term* yang memiliki nilai CPD yang rendah.

Kata Kunci: *Klasifikasi Dokumen, K-Nearest Neighbor, Categorical Proportional Difference*

Abstract

Sambat Online is a platform to facilitate the suggestions, criticisms, complaints or questions from public to the Government of Malang through provided websites or via short messages. Incoming complaints, will be categorized into various fields of SKPD. To make it easier to organize the text and increase the efficiency of the administrator in sorting out and define the field of SKPD, an intelligent systems that can classify documents according to its SKPD's field is needed. K-Nearest Neighbor (K-NN) is a method of classification that will be used to find similarities between documents. Feature selection method used in this research is Categorical Proportional Difference (CPD) to measure the degree of contribution of a word. Started from collecting the test documents and training documents, continue to the preprocessing stage and selection features, weighting, and then do the classification, and analysing the results of the classification system by value of accuracy, precision, recall, and F-Measure. The result is the most optimal performance is the use of k = 1 with featured as much as 100% of 91.84%, which shows better value compared to the featured selection due to the removal of the term with low CPD value.

Keywords: *Document Classification, K-Nearest Neighbor, Categorical Proportional Difference*

1. PENDAHULUAN

Pengumpulan informasi atau tanggapan dari pengguna jasa layanan sudah berubah seiring berkembangnya teknologi informasi. Pemerintah kota Malang adalah salah satu penyelenggara pemerintahan yang menggunakan teknologi informasi dalam menangani urusan-urusan pemerintahan. Cara yang digunakan untuk mengambil informasi atau tanggapan tidak lagi

secara manual melalui survei atau kotak saran. Cara yang sekarang digunakan sudah modern dan praktis yaitu menggunakan situs *web* Sambat Online, Sambat Online adalah sistem berbasis situs *web* yang dapat diakses oleh pengguna jasa layanan selama terhubung dengan jaringan internet.

Sambat Online akronim dari Sistem Aplikasi Masyarakat Terpadu Online Kota Malang. Sambat Online merupakan fasilitas yang berfungsi untuk menampung saran, kritik, keluhan atau pertanyaan dari masyarakat kota Malang seputar Pemerintah Kota Malang melalui situs *web* yang sudah disediakan atau atau melalui pesan singkat kepada nomor yang sudah disediakan. Penyampaian pengaduan pada Sambat Online harus sesuai dengan peraturan yang berlaku. Pengelola sistem berhak untuk tidak menanggapi atau menayangkan pengaduan jika tidak sesuai dari prosedur.

Pada kenyataannya instansi yang menangani Sambat Online kurang dapat memberi respon yang cepat dalam menangani pengaduan yang masuk. Hal ini dikarenakan lembaga yang menangani Sambat Online masih menggunakan sumber daya manusia untuk menyortir dokumen secara manual, untuk meningkatkan kualitas pelayanan maka perlu dibuat sistem cerdas yang dapat mengklasifikasikan dokumen tersebut berdasarkan tujuannya.

Pengklasifikasian yang dilakukan pada penelitian ini menggunakan metode *K-Nearest Neighbor* karena memiliki kesederhanaan yang mana prosesnya berdasarkan pada pendekatan pembobotan yang sederhana dan kemudahan dalam implementasi, adaptasi dan proses *learning* serta memiliki nilai akurasi yang tinggi. *K-Nearest Neighbor* digunakan untuk memecahkan teks dengan kategori statis, yang mana jumlah kategori tidak berubah dalam jangka waktu lama. Pada algoritme ini juga mudah untuk melakukan perubahan dan bisa disesuaikan dengan permasalahan yang berbeda (Toker, dkk). Metode *K-Nearest Neighbor* mengkategorikan sebuah sampel data yang tidak memiliki label dengan menggunakan label mayoritas dari sampel data tetangga yang paling terdekat (paling dekat) dalam data latih (Hassanat, et al., 2014).

Feature selection adalah salah satu teknik yang sering digunakan dan penting dalam *preprocessing* data. Metode ini digunakan untuk mengurangi dimensi sehingga jumlah fitur bisa berkurang. Pada penelitian ini menggunakan *feature selection* berbasis *Categorical Proprotional Difference*. Pada penelitian sebelumnya yang berkaitan dengan *Categorical Proportional Difference* (Allotey, 2011), membahas tentang penggunaan CPD sebagai seleksi fitur pada beberapa metode seperti Naïve bayes dan SVM dan menggunakan dua dataset dalam percobaannya. Terdapat juga seleksi fitur lain yang digunakan pada penelitian tersebut yaitu IG dan X2 untuk dibandingkan dengan CPD. Pada hasil pengujian dari penelitian tersebut menunjukkan bahwa pengguanaan CPD dapat bekerja dengan baik daripada seleksi fitur yang lain.

2. STUDI PUSTAKA

2.1. Sambat Online

Sistem Aplikasi Masyarakat Bertanya Terpadu Online atau yang biasa disingkat Sambat Online adalah fasilitas yang disediakan oleh Dinas Komunikasi dan Informatika untuk memfasilitasi pengaduan melalui jalur Online. Pengaduan yang ingin disampaikan melalui Sambat Online dapat dikirimkan melalui situs *web* secara langsung maupun melalui pesan singkat yang ditujukan kepada nomor yang telah disediakan.

2.2. *Text Mining*

Text mining dapat didefinisikan sebagai proses untuk mendapatkan informasi dari sumber data melalui identifikasi dan eksplorasi pola menarik menggunakan seperangkat alat analisis (Feldman & Sanger, 2007).

Proses text mining dimulai dengan pengumpulan atau koleksi dokumen dari berbagai sumber, kemudian diambil sebuah dokumen tertentu untuk dilakukan preprocessing dengan memeriksa format dan data set karakter. Setelah itu dilakukan analisa teks yang merupakan analisis semantik guna memperoleh informasi yang berkualitas tinggi dari teks. Informasi yang didapatkan ini disimpan dalam sistem informasi (Gaikwad et al., 2014).

Tujuan dari text mining yaitu untuk mendapatkan informasi dari sekumpulan dokumen. Data yang digunakan dalam text mining adalah kumpulan text yang memiliki format yang tidak terstruktur dan melakukan pengkategorian text dan pengelompokan text. Berdasarkan data yang tidak struktur pada text, maka proses text mining memerlukan beberapa tahap yang intinya adalah untuk mempersiapkan

agar text dapat diubah menjadi data yang lebih terstruktur.

2.3. *Text Preprocessing*

Preprocessing merupakan teknik *text mining* yang melibatkan perubahan data mentah menjadi data yang terstruktur dan dimengerti. Metode yang digunakan dalam text preprocessing dokumen antara lain, *case folding, tokenizing , filtering, dan stemming*.

1. *Case Folding*

Case folding adalah tahapan untuk merubah semua huruf yang berada dalam dokumen menjadi huruf kecil. Perubahan dilakukan dari huruf ‘a’ sampai dengan ‘z’.

2. *Tokenizing*

Tokenizing adalah proses pemisahan setiap kata yang menyusun suatu dokumen. Pada proses *tokenizing* dilakukan penghilangan tanda baca, karakter dan angka selain huruf alfabet, karena karakter-karakter tersebut dianggap sebagai pemisah kata atau delimiter dan tidak memiliki pengaruh terhadap pemrosesan text (Feldman & Sanger, 2007). Tokenizing juga sering disebut sebagai istilah (*term*) atau kata, sebagai contoh sebuah token merupakan suatu urutan karakter dari suatu dokumen yang dikelompokkan sebagai unit kerja semantic yang bisa untuk diproses.

3. *Filtering*

Filtering adalah tahap pengambilan kata-kata penting dari hasil token. Bisa menggunakan algoritme *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist* atau *stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Eliminasi *stopword* memiliki banyak kelebihan,yaitu akan mengurangi *space*. Pada *filtering* menggunakan daftar *stopword* yang menggunakan bahasa Indonesia Contohnya adalah “yang”, “dan”, “di”, “dari”, dan seterusnya. Sedangkan *wordlist* adalah kata kata yang dianggap sebagai kata kunci.

4. *Stemming*

Stemming adalah proses mengubah bentuk kata menjadi kata dasar atau mencari *base* kata dari tiap kata hasil *filtering*. Proses *stemming* pada teks berbahasa Indonesia berbeda dengan *stemming* pada teks berbahasa Inggris. Pada teks berbahasa Inggris hanya diperlukan proses menghilangkan akhiran (*suffix*), sedangkan untuk teks berbahasa Indonesia diperlukan proses menghilangkan awalan (*prefix*), sisipan (*infix*), akhiran (*suffix*) dan kombinasi dari awalan dan akhiran (*confix*) (Agusta, 2009).

2.4. *K-Nearest Neighbor*

K-Nearest Neighbor adalah salah satu metode pengenalan pola yang umum dan sering digunakan untuk proses klasifikasi sekelompok data karena tekniknya yang sederhana. Klasifikasi *K-Nearest Neighbor* mengkategorikan sebuah sampel data tidak berlabel dengan menggunakan label mayoritas dari sampel data tetangga terdekat (paling mirip) dalam data training (Hassanat, et al., 2014).

Semakin mirip suatu dokumen maka semakin tinggi peluang untuk dikelompokan menjadi satu dokumen. Sebaliknya semakin tidak mirip maka semakin rendah peluang untuk dikelompokkan menjadi satu dokumen. Untuk mengukur tingkat kemiripan bisa menggunakan *cosine similarity*, kovarian, dan kolerasi.

Cosine similarity merupakan metode yang digunakan untuk mencari kemiripan antara vektor dokumen dan vektor *query*. Semakin mirip antara vector dokumen dan vector *query*, maka dokumen tersebut dipandang semakin sesuai dengan *query*. Rumus untuk menghitung nilai *similarity* di antara dua vektor adalah sebagai berikut

$$\text{CosSim(Di, Dj)} = \frac{\sum_{k=1}^{t}(W_{ik} \cdot W_{jk})}{\sqrt{\sum_{k=1}^{t} {W_{ik}}^2 \cdot \sum_{k=1}^{t} {W_{jk}}^2}} \quad (1)$$

t = Banyaknya kata unik pada dokumen pada kategori
D_i= Dokumen uji
D_j= Dokumen latih
W_{ik} = Bobot nilai dari elemen ke-k dari vektor kata
W_{jk} = Bobot nilai dari elemen ke-k dari vektor kata

2.5. *Categorical Proportional Difference*

Categorical proportional difference adalah f*eature selection* untuk mengukur derajat kontribusi sebuah kata guna membedakan apakah kata tersebut termasuk pada suatu kategori tertentu dari beberapa kategori yang ada. Setiap kelas yang ada akan dihitung berapa banyak kata yang dicari pada suatu dokumen tertentu dan menghitung juga kata selain yang

dicari, sehingga akan mendapat jumlah kata tersebut pada setiap-setiap kelas.

Tabel 1. Tabel *Contingency*

	C	¬C	∑ Row
W	A	B	A+B
¬W	C	D	C+D
∑ Column	A+C	A+B	N

Categorical proportional difference mengukur sejauh mana kata yang berkontribusi untuk membedakan kategori tertentu dari kategori lain dalam korpus. Nilai yang mungkin untuk *Categorical proportional difference* dibatasi pada interval [-1, 1], yang mana nilai yang dekat -1 menunjukkan bahwa kata terjadi pada sekitar jumlah yang sama pada dokumen di semua kategori dan 1 menunjukkan bahwa kata terjadi di dokumen dari satu kategori. Lebih formal, perbedaan proporsional kategori untuk kata w_i di Kategori c_j didefinisikan sebagai berikut.

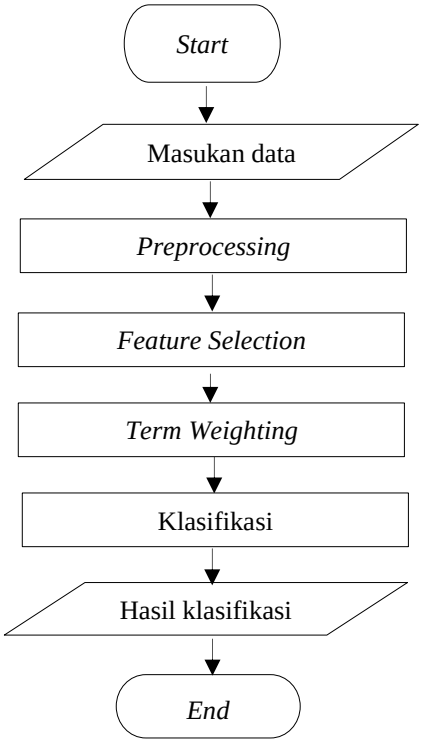
$$CPD(w_i, c_j) = \frac{A-B}{A+B} \quad (2)$$

Pada Persamaan 2 A adalah jumlah berapa kali kata w_i dan kategori c_j terjadi bersama sama, B adalah berapa kali kata w_i terjadi tanpa kategori c_j, C adalah jumlah berapa kali kategori c terjadi tanpa kata w, D adalah jumlah kali kata bukan w_i atau kategori c_j terjadi. Sehingga setelah dilakukan perhitungan di masing-masing kategori atau *class* maka akan dipilih nilai *ratio* yang tertinggi untuk menentukan masuk pada suatu kategori/class tertentu. Untuk persamaanya adalah sebagai berikut.

$$CPD(w_i) = max_j\{CPD(w_i, C_j)\} \quad (3)$$

3. PERANCANGAN SISTEM

Pada penelitian ini terdapat empat proses utama yang akan dilakukan oleh sistem. Langkah awal yaitu melakukan perancangan proses pada *preprocessing text* antara lain *case folding, tokenizing, filtering, dan stemming*. Langkah selanjutnya adalah melakukan tahap perancangan proses pada *features selection* berbasis *Categorical Proportional Difference*. Langkah selanjutnya adalah *term weighting*. Langkah yang terakhir adalah melakukan perancangan proses pada klasifikasi *K-Nearest Neighbor*. Pada Gambar 1 akan menunjukkan diagram alir perancangan proses.



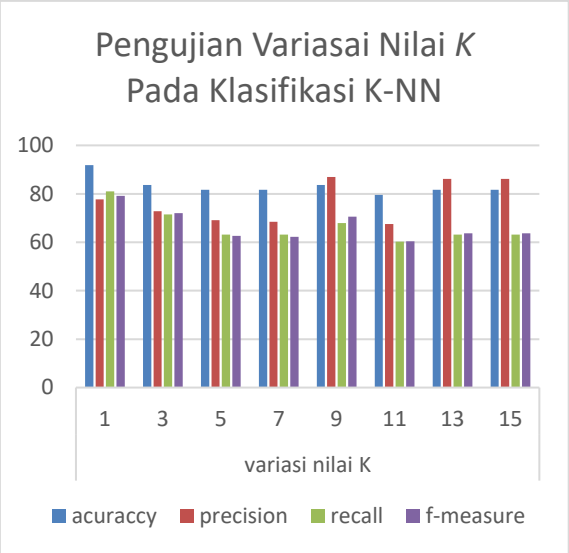
Gambar 1. Alur Kerja Sistem

4. PENGUJIAN DAN ANALISIS

Pada penelitian ini dilakukan beberapa pengujian terhadap hasil penerapan metode gabungan *Categorical Proportional Difference* dan *K-Nearest Neighbor*.

4.1 Pengujian Variasi Nilai *K* Pada Klasifikasi *K-Nearest Neighbor*

Pengujian ini bertujuan untuk mengetahui nilai *k* tetangga terdekat yang paling optimal untuk digunakan pada saat proses klasifikasi tanpa mengunakan variasi rasio *feature*. Nilai *K* berfungsi untuk menentukan klasifikasi data uji terhadap data latih yang sudah memiliki label. Pengujian nilai *k* pada klasifikasi tanpa rasio *feature* dilakukan menggunakan nilai *k* yang bervariasi untuk mengetahui nilai *k* yang paling optimal sehingga dengan variasi nilai *k* yang beragam akan dapat menentukan apakah semakin tinggi nilai *k* akan memiliki nilai keakuratan yang tinggi. Berikut merupakan variasi penggunaan nilai *k* antara lain 1, 3, 5, 7, 9, 11, 13 dan 15:

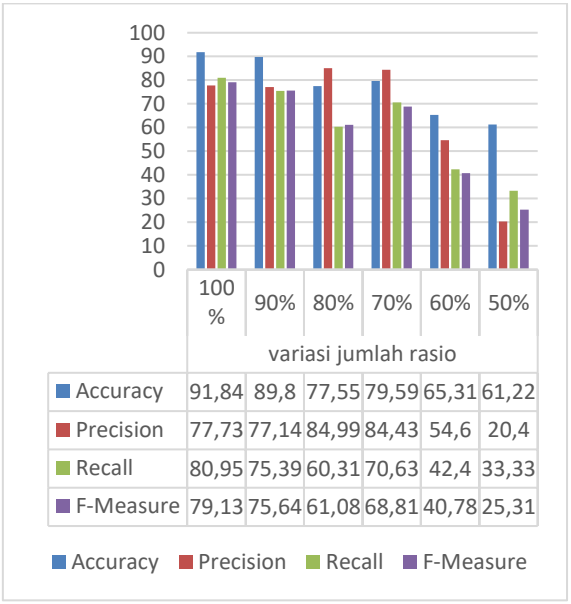


Gambar 2. Grafik Variasi Nilai *K*

Berdasarkan Gambar 2 nilai *k* yang paling optimal adalah ketika *k*=1 karena memiliki nilai *accuracy* sebesar 91,84% yang mana nilai *accuracy*-nya merupakan nilai *accuracy* yang paling tinggi dibandingkan dengan nilai *accuracy* pada nilai *k* yang lain. Kinerja pada Gambar 2 akan cenderung menurun dari *k*=3 hingga *k*=15. Hal ini terjadi karena terdapat perbedaan data pada tiap kelas data latih yang tidak seimbang yang mana pada kelas Dishub terdapat 150 data, pada kelas DPUPPB terdapat 60 data, dan pada kelas DKP terdapat 30 data. Pada proses klasifikasi yang digunakan perhitungan nilai *cosine similarity* pada perhitungan ini juga penting untuk dipertimbangkan. Perhitungan *cosine similarity* ini dikatan baik jika nilai *cosine similarity*-nya tinggi, semakin tinggi nilai pada *cosine similarity* semakin dekat juga kemiripan antar data. Pada penelitian ini kemiripan antara data latih dan data uji sangat dekat, karena isi dokumen yang terdiri dari *term* pada data uji dan data latih memiliki kemiripan, hal ini dapat dibuktikan dengan besarnya *accuracy* pada *k*=1 yaitu sebesar 91.84%.

4.2 Pengujian Variasi Penggunaan Jumlah Rasio *Feature* untuk Klasifikasi K-NN

Pengujian ini berfungsi untuk mengetahui pengaruh variasi banyaknya jumlah rasio *feature* yang digunakan terhadap hasil klasifikasi menggunakan *K-Nearest Neighbor*. Pemilihan banyaknya *term* yang digunakan untuk proses klasifikasi berdasarkan hasil perhitungan *Categorical Proportional Difference* pada tiap-tiap *term* kemudian nilai *Categorical Proportional Difference* tersebut diurutkan secara *descending*. *Term* dengan nilai *Categorical Proportional Difference* tertinggi memiliki peluang untuk digunakan pada saat proses klasifikasi. Pengujian ini dilakukan dengan memilih *term* dengan nilai *Categorical Proportional Difference* tertinggi dari seluruh *term* pada data latih. *Term* pada data uji akan dicocokan dengan *term* pada latih jika *term* yang tidak ada pada latih maka *term* tersebut tidak akan masuk pada proses klasifikasi. Variasi banyaknya rasio *feature* yang digunakan pada saat proses klasifikasi adalah 100%, 90%, 80%, 70%, 60% dan 50% rasio dari seluruh jumlah *term*. Hasil pengujian penggunaan feature 90%, 80%, 70%, 60% dan 50% seperti pada Gambar 3.



Gambar 3. Pengujian Variasi Jumlah *Feature*

Berdasarkan Gambar 3 Penggunaan variasi banyaknya *feature* yang digunakan memiliki pengaruh terhadap hasil *accuracy*, *precision*, *recall*, *dan f-measure*. Pada saat penggunaan *feature* sebesar 90% tingkat *accuracy* sebesar 85%, *precision* 85%, *recall* 69% dan *f-measure* sebesar 68%. Pada saat penggunaan *feature* sebesar 80% tigkat *accuracy* sebesar 79%, *precision* 67%, tingkat sebesar *recall* 62% dan tingkat *f-measure* sebesar 59%. Pada saat penggunaan *feature* sebesar 70% tingkat *accuracy* sebesar 75%, *precision* sebesar 67%, *recall* sebesar 62% dan *f-measure* sebesar 59%. Hal ini menunjukkan bahwa semakin kecil rasio *feature* yang digunakan maka tingkat keakuratannya akan semakin rendah, hal ini dapat dibuktikan pada penggunaan *feature*

sebesar 50% dengan *accuracy* sebesar 61%, *precision* sebesar 20%, *recall* sebesar 33% dan *f-measure* sebesar 25%. Peneliti menganalisis bahwa semakin rendah jumlah rasio yang digunakan maka semakin rendah tingkat keakuratannya disebabkan sebaran datanya tidak seimbang data yang paling banyak adalah pada kelas Dishub sehingga ketika penggunaan *Feature* sebanyak 50% semua kelas pada data training dikenali oleh sistem sebagai kelas Dishub. Faktor lain yang mempengaruhi adalah karena pengguna jumlah *feature* yang terlalu sedikit sehingga informasi yang diperlukan oleh sistem terlalu sedikit untuk proses klasifikasi dan bisa saja *term* yang seharusnya diprioritaskan untuk dilakukan proses klasifikasi namun terbuang atau tidak dilakukan proses klasifikasi, sehingga menyebabkan kurang maksimimal dalam menghasilkan hasil klasifikasi.

5. KESIMPULAN

Metode *K-Nearest Neighbor* dan *Categorical Proprortional difference* dapat diimplementasikan pada klasifikasi dokumen sambat *Online*, untuk mengimplementasikan Klasifikasi Dokumen Sambat *Online* Menggunakan Metode *K-Nearest Neighbor* dan *feature selection* Berbasis *Categorical Proportional Difference* yang harus dilakukan adalah melakukan *preprocessing text* antara lain *tokenizing, filtering* dan *stemming*. Tahap selanjutnya adalah proses *feature selection* menggunakan *Categorical Proportional Difference* tahap *feature selection* dilakukan untuk mengurangi dimensi *feature*. Tahap selanjutnya proses *term weighting* antara lain *term frequency, document frequency, inverse document frequency, term frequency weight*, TF-IDF, setelah pembobotan menghitung nilai *cosine similarity* tiap kategori pada masing-masing dokumen data uji terhadap data latih. Proses selanjutnya adalah klasifikasi yaitu dengan memasukkan nilai *k* yang akan digunakan, kemudian jumlahkan nilai *cosine similarity* dari dokumen pada tiap kategori dari kumpilan *k* dokumen latih, kemudian cari nilai *cosine similarity* terbesar dari hasil tersebut.

Berdasarkan hasil pengujian nilai *accuracy, precision, recall dan f-measure* yang didapatkan dari hasil implementasi kinerja yang paling optimal adalah penggunaan *k*=1 dengan *feature* sebanyak 100% *accuracy* sebesar 91,84%, *precision* sebesar 77,73%, *recall* sebesar 80,95% dan *f-measure* sebesar 79,13%. Namun, pengurangan jumlah fitur berpengaruh negatif terhadap kenerja sistem, semakin sedikit fitur yang dipakai semakin rendah kinerjanya. Hasilnya perpaduan metode *Categorical Proportional Difference* dengan metode *K-Nearest Neighbor* menghasilkan nilai keakuratan yang rendah dibandingkan dengan metode *K-Nearest Neighbor*.

REFERENSI

Agusta, L., 2009. Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia. Bali, Universitas Kristen Satya Wacana.

Allotey, D. A. (2011). *Sentiment Analysis and Classification of Online Reviews Using Categorical Proportional Difference.* Regina: Departement of Computer Science.

Barlow, J. & Moller, C., 2008. A Complaint Is a Gift. Second Edition, Revised and Expanded ed. s.l.:Barret-Koehler.

Buana, P. W., Jannet, S. & Putra, I. K. G. D., 2012. Combination of *K-Nearest Neighbor* and K-Means based on *Term* Re-Weighting for Classfy Indonesian News. *International Jurnal of Computer Application,* 50(11), pp. 37-42.

Feldman, R. & Sanger, J., 2007. *The Text Mining Handbook: Advanced Approaches in Anlayzing Unstructured Data*. Cambridge: Cambridge University Press.

Gaikwad, S.V., Chaugule, A. & Patil, P., 2014. Text Mining Methods and Techniques. *International Journal of Computer Applications*, LXXXV, pp.42-45.

Hassanat, A. B., Abbadi, M. A. & Altarawneh, G. A., 2014. Solving the problem of the K parameter in the K-NN Classifier using an Ensemble learning Approach. *Computer Science and Informatic,* 12(8), pp. 33-39.

Simeon, M. & Hilderman, R., n.d. Categotical Proportional Difference: A Features Selection Method for Text Categorization.

Toker, G., Kirnemis, O., Text Categorization Using *K-Nearest Neighbor* Classification.