

Pembentukan Daftar *Stopword* menggunakan *Zipf Law* dan Pembobotan *Augmented TF - Probability IDF* pada Klasifikasi Dokumen Ulasan Produk

Destin Eva Dila Purnama Sari¹, Yuita Arum Sari², Muhammad Tanzil Furqon³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹destindila@gmail.com, ²yuita@ub.ac.id, ³m.tanzil.furqon@ub.ac.id

Abstrak

Stopword merupakan suatu kata tidak penting yang terdapat di dalam kalimat. *Stopword* digunakan untuk membantu tahap *preprocessing* teks terutama dalam tahap *stopword removal*. *Digital library* sering digunakan dalam tahap ini untuk memperoleh daftar *stopword*. Namun tidak semua daftar *stopword* yang ada dalam *digital library* merupakan kata yang tidak penting dalam suatu data. Fokus utama dalam penelitian ini adalah mencari tahu pembentukan daftar *stopword* dan pembobotan kata dalam klasifikasi dokumen ulasan produk dengan menggunakan metode *Zipf Law*. *Augmented Term Frequency – Probability Term Frequency* digunakan untuk proses pembobotan kata. Proses klasifikasi dokumen bertujuan untuk mencari tahu pengaruh dari pembentukan daftar *stopword* dan pembobotan kata. Penggunaan metode *Support Vector Machine* dan *Polynomial Kernel* untuk memperoleh hasil klasifikasi. *Output* yang dihasilkan berupa hasil akurasi klasifikasi. Berdasarkan hasil akurasi klasifikasi, terdapat pengaruh pembentukan daftar *stopword* dan pembobotan kata terhadap hasil klasifikasi. Akurasi terbaik dihasilkan pada saat persentase 15% untuk pembentukan daftar *stopword* yang diambil dari *term* yang memiliki hasil konstanta rendah. Akurasi yang dihasilkan terdiri dari *precision* 0,73, *recall* 0,7 dan *f-measure* 0,63.

Kata kunci: Pembentukan *Stopword*, *Zipf Law*, *Augmented Term Frequency – Probability Term Frequency*

Abstract

Stopword is an insignificant word contained in a sentence. *Stopword* was used to help the text preprocessing stage, especially in the *stopword removal* stage. *Digital library* was often used at this stage to get a *stopword* list. However, not all *stopword* lists in the *digital library* were words that were not important in the data. The main focus in this research was to find out forming *stopword* list and word weighting on the document classification of product review using the *Zipf Law* method. The method used for word weighting was *Augmented Term Frequency - Probability Inverse Document Frequency*. The document classification process aimed to find out the effect of forming *stopword* list and word weighting. Document classification using the *Support Vector Machine* algorithm and *Polynomial Kernel*. The output of the research was the result of classification accuracy. Based on the result of classification accuracy, there was an effect of forming a *stopword* list and weighting of words on the classification result. The best accuracy result of the document classification was found at a percentage of 15% for forming *stopword* list taken from *term* that has low constant result. The resulting accuracy consisted of a *precision* value of 0.73, a *recall* value of 0.7 and a *f-measure* value of 0.63.

Keywords: Forming a *Stopword*, *Zipf Law*, *Augmented Term Frequency - Probability Term Frequency*

1. PENDAHULUAN

Stopword removal merupakan salah satu tahap pemrosesan teks yang terdapat dalam *preprocessing* teks. Kata yang tidak penting di sebuah dokumen dihilangkan di proses *stopword removal* ini. Proses *stopword removal* membantu pemrosesan teks lebih ringan dan

lebih cepat. Biasanya proses *stopword removal* ini memanfaatkan *digital library* yang khususnya berisi kata-kata yang tergolong *stopword*. Namun ada kumpulan *stopword* pada *digital library* bukan merupakan termasuk *stopword* pada dokumen yang digunakan untuk pemrosesan teks. Hal itu membuat kualitas proses *stopword removal* menjadi berkurang.

Pembentukan daftar *stopword* secara otomatis dapat membantu permasalahan pada pemrosesan teks ini. Dengan membuat daftar *stopword* secara otomatis, dapat meningkatkan kualitas pemrosesan teks terutama pada proses *stopword removal*. Penelitian ini membentuk daftar *stopword* yang digunakan untuk proses klasifikasi dokumen teks.

Klasifikasi dokumen teks merupakan suatu proses yang mengelompokkan dokumen berdasarkan ciri suatu dokumen. Dokumen yang digunakan untuk klasifikasi berupa data ulasan produk. Alasan menggunakan data ulasan produk karena data ulasan ini memuat kalimat-kalimat yang tidak baku dari segi penulisannya. Dengan membangun daftar *stopword* secara otomatis dapat membantu proses klasifikasi dokumen ulasan produk

Salah satu metode untuk pembentukan daftar *stopword* adalah *Zipf Law*. Seperti penelitian sebelumnya yang telah dilakukan oleh (Daniel dan Adebukola, 2017) tentang penggunaan metode *Zipf Law* sebagai dasar untuk pembentukan *stopword*. Penelitian ini menggunakan bahasa *Yoruba* dan menghasilkan 255 daftar *stopword*. Hal ini membuat teks dikompres menjadi 63% lebih sedikit dibandingkan teks awal sebelum menggunakan daftar *stopword*.

Selain pembentukan daftar *stopword*, proses pembobotan kata mempunyai pengaruh terhadap proses klasifikasi. Seperti penelitian yang telah dilakukan oleh (Haibing dan Xiaodong, 2016) mengenai penyeimbangan bobot *term* dengan untuk kategorisasi teks. Pada penelitian ini, dilakukan penyeimbangan bobot dengan menggunakan metode pembobotan *Term Frequency – Inverse Document Frequency* terhadap pengkategorian dokumen teks. Hasil dari penelitian ini, pembobotan TF-IDF yang seimbang terdapat saat penambahan *Add-One Smoothing* menghasilkan bobot yang sangat optimal dan seimbang. Hal ini dapat dilakukan dengan menggunakan *Augmented* untuk perhitungan TF dan *Probability* untuk perhitungan IDF.

Berdasarkan penjabaran sebelumnya, penelitian ini menggunakan metode *Zipf Law* untuk pembentukan daftar *stopword*. *Support Vector Machine* dan *polynomial kernel* berderajat dua yang telah disediakan oleh *digital library sklearn python* digunakan untuk proses klasifikasi. Untuk meningkatkan hasil akurasi,

penelitian ini juga akan dipadukan dengan proses pembobotan kata menggunakan *Augmented TF – Probability IDF*.

2. DASAR TEORI

2.1. Ulasan Produk

Ulasan produk adalah kegiatan yang dilakukan untuk memberikan komentar positif atau negatif terhadap suatu produk (Priansa, 2017). Semakin lama semakin banyak penggunaan fitur ulasan produk terutama untuk situs *online shop*. Kegiatan ulasan ini dapat membantu masyarakat dalam mendapatkan informasi yang lebih mengenai suatu produk. Selain itu kegiatan ulasan produk ini membuat masyarakat untuk lebih berhati-hati untuk memutuskan pembelian suatu produk. Salah satu situs *online shop* yang menggunakan fitur ulasan produk yaitu “*Sociolla*”. *Sociolla* merupakan situs *online shop* yang memiliki member cukup besar. Fitur ulasan produk dapat dilihat langsung pada katalog produk. Data ulasan produk untuk penelitian ini diambil dari situs *Sociolla* untuk dokumen klasifikasi. Hasil klasifikasi dalam penelitian akan mengelompokkan ulasan negatif maupun positif dari suatu produk.

2.2. Text Mining

Text mining adalah suatu proses yang digunakan dalam pengekstrasian pola yang menarik dan signifikan. Tujuan *text mining* adalah mendapatkan informasi yang cepat dan mudah dari kumpulan teks yang tidak terstruktur maupun semi terstruktur. *Text mining* biasanya sering dimanfaatkan untuk peringkasan teks, *information retrieval*, analisis *sentiment*, klasifikasi maupun pengelompokkan teks. Proses pada *text mining* terdiri dari beberapa tahapan. Secara umum proses *text mining* sebagai berikut (Talib et al., 2016).

1. Pengumpulan data. Jenis data terbagi menjadi data tidak terstruktur dan semi terstruktur yang terdapat pada sumber data.
2. *Preprocessing* teks. Proses ini digunakan untuk membuat data yang telah dikumpulkan sebelumnya menjadi lebih terstruktur. Selain itu kata yang terdapat data akan dibersihkan agar dapat diproses ke tahap selanjutnya.
3. Melakukan pemrosesan teks yang bertujuan untuk mencari pola dari suatu teks maupun data.

4. Mengambil informasi yang sangat relevan dari data yang sudah diproses di tahap sebelumnya.

2.3. Preprocessing

Preprocessing merupakan salah satu tahapan dari proses *text mining*. *Preprocessing* berfungsi untuk menjadikan data lebih terstruktur dan nantinya dapat diproses di tahap selanjutnya. Selain membersihkan kata yang tidak penting dan tanda baca yang dapat mengganggu proses *text mining*, menurut (Gurusamy dan Kannan, 2014) *preprocessing* dapat membantu mengurangi ukuran dari dokumen.

2.4. Pembentukan Stopword

Stopword merupakan kumpulan kata yang tidak penting pada suatu dokumen. seringnya kemunculan *stopword* tidak memberikan dampak yang baik dalam pemrosesan teks. Biasanya daftar *stopword* yang digunakan untuk *preprocessing* teks memanfaatkan *digital library*. Namun hal ini juga menimbulkan proses *preprocessing* memakan waktu lebih untuk memuat *digital library*. Karena semakin lama volume teks yang terdapat pada *digital library* juga semakin banyak. Selain itu tidak semua kata di *digital library* termasuk *stopword*, karena kondisi ini juga menyesuaikan dari bentuk data yang digunakan dalam pemrosesan teks. Perlunya pembentukan daftar *stopword* secara otomatis dapat membantu permasalahan pada *preprocessing* teks seperti ini. Metode pembentukan *stopword* sangat beragam. Bentuk data dan fokus penelitian juga berpengaruh dalam penentuan metode yang akan digunakan dalam proses pembentukan *stopword* (Vikhers, Arif dan Firdaus, 2009).

2.4.1 Zipf Law

Salah satu metode pembentukan daftar *stopword* adalah *Zipf Law*. *Zipf Law* adalah metode yang menyatakan bahwa frekuensi kata atau term berbanding terbalik dengan peringkat dari sebuah *term*. Semakin sedikit frekuensi kemunculan suatu *term* pada dokumen maka semakin tinggi peringkat yang didapatkan oleh *term* tersebut. Berdasarkan penelitian yang telah dilakukan sebelumnya oleh (Ruby dan Lubiyal, 2018) bahwa metode *Zipf Law* cukup menghemat waktu pemrosesan dan hasil daftar *stopword* cukup umum seperti daftar *stopword* yang terdapat pada *library* yang lain. Persamaan

yang terdapat pada hukum *Zipf Law* terdapat pada Persamaan 1.

$$k = f \times r \quad (1)$$

Keterangan:

k = konstanta
 f = frekuensi kemunculan kata
 r = peringkat suatu kata

2.5. Pembobotan Kata

Pembobotan kata merupakan suatu proses pemberian bobot pada suatu kata atau fitur. Salah metode pembobotan kata adalah *Term Frequency - Inverse Document Frequency* atau yang biasa disebut dengan TF-IDF. TF-IDF adalah suatu proses yang digunakan untuk pemberian bobot untuk setiap kata yang terdapat pada dokumen. Menurut (Luqyana et al., 2018) pembobotan kata dengan metode TF-IDF menggunakan dua pendekatan yaitu TF dan IDF. Pendekatan TF berfungsi dalam perhitungan bobot suatu *term* berdasarkan kemunculannya di suatu dokumen. Sedangkan pendekatan IDF berfungsi dalam perhitungan bobot setiap *term* yang jarang muncul pada dokumen. Perhitungan IDF pada penelitian ini menggunakan metode Probabilitas *Inverse Document Frequency*. Menghitung TF-IDF dapat menggunakan Persamaan 2 berikut.

$$W_{t,d} = Tf_{t,d} \times idf_t \quad (2)$$

Keterangan:

$W_{t,d}$ = nilai bobot TF-IDF

2.5.1. Augmented Term Frequency

Augmented TF merupakan suatu pengembangan dari perhitungan frekuensi *term*. *Augmented Term Frequency* memperhatikan frekuensi *term* dengan maksimum semua *term* di sebuah dokumen. Penggunaan konstanta pada *augmented term frequency* berfungsi untuk menormalisasi frekuensi *term* dan mencegah frekuensi *term* yang kosong. Menurut (Salton dan Buckley, 1988) persamaan *Augmented Term Frequency* terdapat pada Persamaan 3.

$$atf_{t,d} = k + (1 - k) \frac{tf_{t,d}}{\max_t(tf_{t,d})} \quad (3)$$

Keterangan:

$atf_{t,d}$ = *augmented term frequency*
 $tf_{t,d}$ = frekuensi dokumen terhadap *term*
 k = konstanta, bernilai 0.5
 $\max_t(tf_{t,d})$ = frekuensi maksimum dari semua

term pada suatu data.

2.5.2. Probability Inverse Document Frequency

Probability IDF adalah salah satu perhitungan *inverse document frequency* yang membandingkan frekuensi dokumen dengan jumlah dokumen. Hal ini digunakan untuk mencari probabilitas terbaik dari hasil *inverse document frequency* untuk setiap *term*. Berikut persamaan dari *Probability Inverse Document Frequency* yang terdapat pada Persamaan 4.

$$pidf_t = \max(0, \log \frac{N-df_t}{df_t}) \quad (4)$$

Keterangan:

$pidf_t$ =probability idf

N =jumlah dokumen

df_t =frekuensi dokumen yang mengandung suatu *term*

2.6. Klasifikasi Support Vector Machine

Support Vector Machine merupakan suatu metode klasifikasi yang menggunakan konsep dalam mencari *hyperplane* terbaik. *Hyperplane* terbaik yang didapatkan dalam penggunaan metode *support machine* digunakan untuk memisahkan dua kelas data. Penggunaan trik *kernel* pada metode ini dapat membantu kinerja svm dalam meningkatkan dimensi data (Rofiqoh, et al., 2017). Menurut (Haryanto, et al., 2018) metode kernel adalah suatu cara yang digunakan untuk menghadapi masalah ketidaklinearan data.

Pada penelitian ini menggunakan metode klasifikasi *support vector machine* dengan *kernel polynomial* berderajat dua. Alasan penggunaan metode klasifikasi ini karena menurut penelitian sebelumnya yang telah dilakukan oleh (Luqyana et al., 2018) dan penelitian yang telah dilakukan oleh (Hussain et al., 2017) bahwa metode klasifikasi *support vector machine* dan penggunaan *kernel polynomial* derajat dua memiliki akurasi yang tinggi. Persamaan secara umum dari metode *support vector machine* terdapat pada Persamaan 5 dan *kernel polinomial* derajat dua yaitu sebagai berikut terdapat pada Persamaan 6.

$$f(x) = w \cdot x + b \quad (5)$$

$$Kx_i x_j = (x_i \cdot x_j + 1)^d \quad (6)$$

Keterangan:

x = titik data masukan SVM

w = parameter *hyperplane*

b = nilai bias

$Kx_i x_j$ = nilai *kernel*

x_i = fitur data i

x_j = fitur data j

d = nilai derajat

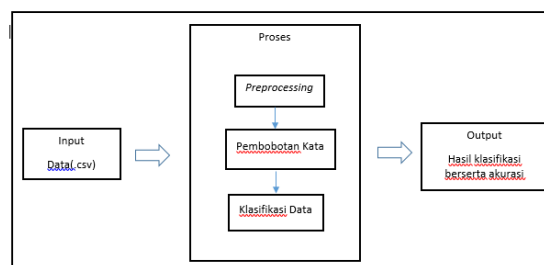
3. METODOLOGI

3.1. Pengumpulan Data

Proses pengumpulan data dilakukan secara langsung dengan cara manual berdasarkan produk. Data yang digunakan berupa postingan ulasan terhadap produk yang diposting oleh pengguna Sociolla. Jumlah data yang digunakan dalam penelitian ini sebesar 250, diantaranya 200 data latih dan 50 data uji.

3.2. Diagram Alir Sistem

Proses umum sistem merupakan gambaran seluruh proses dari sistem pada penelitian. Sistem ini digunakan untuk mengatasi masalah penelitian pembentukan daftar *stopword* menggunakan *Zipf Law* dan pembobotan *Augmented TF – Probability IDF* pada klasifikasi dokumen ulasan produk. Proses pertama yang terdapat pada sistem adalah melakukan *input* data dan menentukan jumlah data uji. Setelah proses *input*, data memasuki tahap *preprocessing*. Pada tahap *preprocessing* terjadi proses pembentukan daftar *stopword* dengan menggunakan *Zipf Law*. Selanjutnya terdapat tahap pembobotan menggunakan *Augmented TF – Probability IDF*. Tahap yang terakhir adalah tahap klasifikasi. *Output* dari sistem ini berupa hasil klasifikasi dari data uji yang telah ditentukan sebelumnya beserta kumpulan *stopword*. Alur proses sistem dapat dilihat pada Gambar 1 berikut.



Gambar 1. Alur Proses Sistem

4. PENGUJIAN DAN ANALISIS

4.1. Pengujian Sistem

Pengujian dilakukan berdasarkan perbandingan hasil akurasi dari proses klasifikasi dengan parameter. Pengujian terdiri dari pengujian *Augmented TF - Probability IDF* dan pengujian *Log TF-IDF*. Hal ini bertujuan untuk mencari tahu perbandingan hasil klasifikasi menggunakan dua metode tersebut. Persentase yang digunakan pada pengujian ini dihentikan sampai persentase 25%. Hal ini dilakukan karena pada pengujian 25% beberapa data mulai kosong yang disebabkan pengambilan daftar *stopword* yang terlalu banyak. Data kosong menyebabkan proses klasifikasi menjadi tidak efektif. Berikut hasil pengujian pembobotan kata menggunakan *Augmented TF - Probability IDF* yang terdapat pada Tabel 1.

Tabel 1. Pengujian *Augmented TF - Probability IDF*

Persentase		Hasil Akurasi		
		<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
<i>Lower</i>	10 %	0.22	0.47	0.3
	15 %	0.73	0.7	0.63
	20%	0.29	0.54	0.38
<i>Upper</i>	10 %	0.26	0.51	0.35
	15 %	0.73	0.5	0.38
	20%	0.31	0.55	0.4
<i>Upper + Lower</i>	10 %	0.7	0.69	0.63
	15 %	0.37	0.48	0.42
	20%	0.35	0.46	0.4

Pengujian pembobotan kata selanjutnya menggunakan metode *Log TF-IDF*. Hasil dari pengujian ini terdapat pada Tabel 2.

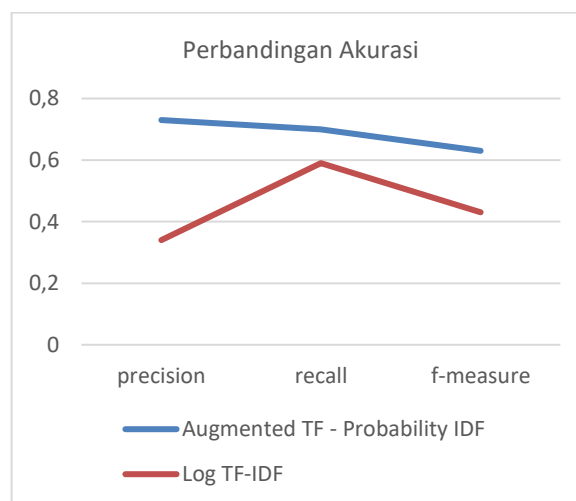
Tabel 2. Hasil Pengujian *Log TF-IDF*

Persentase		Hasil Akurasi		
		<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
<i>Lower</i>	10 %	0.24	0.49	0.32
	15 %	0.34	0.59	0.43
	20%	0.31	0.55	0.4
	10 %	0.21	0.45	0.29
	15 %	0.3	0.55	0.39

<i>Upper</i>	20%	0.29	0.54	0.38
<i>Upper + Lower</i>	10 %	0.28	0.53	0.37
	15 %	0.33	0.57	0.42
	20%	0.24	0.49	0.32

4.2. Analisis Pengujian Sistem

Berdasarkan hasil pengujian pembobotan kata menggunakan *Augmented TF - Probability IDF* dan *Log TF - IDF*, akurasi yang dihasilkan pembobotan *Augmented TF - Probability IDF* lebih baik. Nilai akurasi yang dihasilkan terdiri dari nilai *precision* 0,73, *recall* 0,7 dan *f-measure* 0,63 dengan menggunakan parameter konstanta untuk pembentukan daftar *stopwords* sebesar 15%. Berikut grafik dari perbandingan hasil akurasi pembobotan kata menggunakan *Augmented TF - Probability IDF* dan *Log TF - IDF* dengan menggunakan parameter *lower* 15% yang terdapat pada Gambar 2.



Gambar 2. Perbandingan Akurasi

5. KESIMPULAN DAN SARAN

Kesimpulan yang dihasilkan dari penelitian ini sebagai berikut.

1. Terdapat pengaruh dari penggunaan daftar *stopword* terhadap hasil klasifikasi dokumen ulasan produk. Berdasarkan hasil penelitian ini, akurasi hasil klasifikasi terbaik terdapat pada pembentukan daftar *stopword* dengan persentase 15% dan diambil dari hasil konstanta yang paling rendah atau *lower*. Akurasi yang dihasilkan terdiri dari *precision* 0.73, *recall* 0.7 dan *f-measure* 0.63. *Term* yang memiliki hasil konstanta rendah adalah *term* yang paling sering muncul pada data. Daftar *stopword* yang

terlalu sedikit membuat hasil klasifikasi tidak terlalu bagus. Namun jika daftar *stopword* yang dihasilkan terlalu banyak membuat hasil akurasi klasifikasi menurun. Jadi pengambilan *stopword* harus menggunakan persentase dan parameter sisi konstanta (*upper*, *lower*, dan *upper lower*) yang tepat sehingga hasil klasifikasi lebih efektif.

2. Terdapat pengaruh pembobotan kata menggunakan metode *Augmented TF – Probability IDF* terhadap hasil klasifikasi dokumen ulasan produk. Hasil akurasi klasifikasi menggunakan pembobotan *Augmented TF - Probability IDF* lebih bagus daripada pembobotan *Log TF-IDF*. Hal ini disebabkan karena penggunaan metode *Augmented TF* dapat membuat *term* yang memiliki frekuensi paling besar di dalam suatu dokumen dapat menjadi bobot yang dapat mewakili satu dokumen tersebut daripada *term* yang lain. Selain itu metode *Probability IDF* dapat membantu mencari probabilitas terbaik untuk data yang jarang muncul di suatu dokumen.

Adapun saran yang dapat digunakan untuk mengembangkan penelitian sebagai berikut.

1. Data ulasan produk terdapat banyak kata tidak baku secara penulisan dan singkatan kata. Perlu dilakukan proses perbaikan kata tidak baku dan singkatan kata untuk mencegah *term* yang mempunyai makna sama namun berbeda penulisannya.
2. Data ulasan produk terdapat kata yang tidak menggunakan bahasa Indonesia. Jadi perlu dilakukan penyesuaian kata sebelum memasuki tahap *preprocessing*.

6. DAFTAR PUSTAKA

- Buckley, C. 1993. The Importance of Proper Weighting Methods. Tersedia di: <https://www.researchgate.net/publication/234804406_The_importance_of_proper_weighting_methods> [Diakses 6 November 2019]
- Gurusamy, V. & Kannan, S. 2014. Preprocessing Techniques for Text Mining. Conference Paper. Tersedia di: <https://www.researchgate.net/publication/273127322_Preprocessing_Techniques_for_Text_Mining> [Diakses 29 Oktober 2019]
- Haibing, W. & Xiaodong, G. 2016. Balancing Between Over-Weighting and Under-Weighting in Supervised Term Weighting. Departement of Electronic Engineering: Fudan University.
- Haryanto, D., Muflikhah, L. & Fauzi M., 2018. Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion. Jurnal Pengembangan Teknologi dan Ilmu Komputer, [e-journal] Tersedia di: <<http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2464>> [Diakses 2 November 2019]
- Luqyana, W., Cholissodin, I. & Perdana, R., 2018. Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. Jurnal Pengembangan Teknologi dan Ilmu Komputer, [e-journal] 2(11) Tersedia di: <<http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3051>> [Diakses 29 Oktober 2019]
- Priansa, J. 2017. Komunikasi Pemasaran Terpadu. Bandung: Cv.Pustaka Setia.
- Rofiqoh, U., Perdana, R. & Fauzi, M., 2017. Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine. Jurnal Pengembangan Teknologi dan Ilmu Komputer, [e-journal] Tersedia di: <j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628> [Diakses 29 Oktober 2019]
- Ruby, R. & Lobiyal, D. K., 2018. Automatic Construction of Generic Stop Words List for Hindi Text. Tersedia di: <<https://www.sciencedirect.com/science/article/pii/S1877050918309311>> [Diakses 28 Oktober 2019]
- Salton, G. & Buckley, C. 1988. Term-Weighting Approaches in Automatic Text Retrieval. Tersedia di: <<https://www.sciencedirect.com/science/article/abs/pii/0306457388900210>> [Diakses 6 November 2019]
- Talib, R., Hanif, M. K., Ayesha, S., & Fatima, F., 2016. Text Mining: Techniques,

Applications and Issues. International Journal of Advanced Computer Science and Applications, [e-journal] 7(12) Tersedia di: <
https://thesai.org/Downloads/Volume7No11/Paper_53-Text_Mining_Techniques_Applications_and_Issues.pdf > [Diakses 29 Oktober 2019]