

Analisis Rating Sentimen pada Video di Media Sosial Youtube Menggunakan *STRUCT-SVM*

Kadek Ary Budi Permana¹, Made Sudarma², Wayan Gede Ariastina³

[Submission: 18-02-2019, Accepted: 15-04-2019]

Abstract— Sentiment analysis on comments can be used to determine sentiment rating. The comments used are comments on Youtube. The type of video used is the official trailer video Indonesian movie. This paper contains steps to determine sentiment rating by notice the structure of comments. The structure of comments is needed because not all comments are relevant to the topic. Classes on comments are divided into seven classes including positive films, neutral films, negative films, positive not films, neutral not film, negative not film, and spam / off-topic. Comments that have a positive film or film negative class are used to determine sentiment rating. The number of likes in comments also determines the sentiment rating. Comment classification using *STRUCT-SVM*. The results of *STRUCT-SVM* show accuracy of 70% for linear kernels and 71% for RBF kernels.

Key Words— Sentiment Analysis, Rating Sentimen, *STRUCT-SVM*, Linear Kernel, RBF Kernel, Youtube

Intisari— Analisis sentimen pada komentar dapat digunakan untuk menentukan rating sentimen. Komentar yang digunakan adalah komentar yang terdapat pada Youtube. Jenis video yang digunakan adalah video official trailer film Indonesia. Pada paper ini memuat tentang langkah-langkah dalam menentukan rating sentimen dengan memperhatikan struktur komentar. Pengenalan struktur komentar diperlukan karena tidak semua komentar relevan dengan topik yang bersangkutan. Kelas pada komentar dibagi menjadi tujuh kelas diantaranya positif film, netral film, negatif film, positif bukan film, netral bukan film, negatif bukan film, dan spam / diluar topik. Komentar dengan hasil klasifikasi positif film dan negatif film yang digunakan dalam menentukan rating sentimen. Jumlah like pada komentar juga ikut menentukan rating sentimen. Klasifikasi komentar menggunakan *STRUCT-SVM*. Hasil dari *STRUCT-SVM* menunjukkan akurasi mencapai 70% untuk linear kernel dan 71% untuk RBF kernel.

Kata Kunci— Analisis Sentimen, Rating Sentimen, *STRUCT-SVM*, Linear Kernel, RBF Kernel, Youtube

I. PENDAHULUAN

Media sosial menjadi tempat berbagi berbagai hal, diantaranya seperti berbagi berita, gambar, video, dan sebagainya. Salah satu hal yang menarik pada media sosial saat ini adalah bagaimana pengguna dapat memberikan suatu opini atau komentar pada suatu topik. Melakukan analisis terhadap komentar pengguna tentunya akan bermanfaat untuk

¹Mahasiswa Pascasarjana, Jurusan Teknik Elektro dan Komputer Fakultas Teknik Universitas Udayana, Jln. Jalan Kampus Bukit Jimbaran 80361 INDONESIA (tel: 0361-703315; fax: 0361-4321; e-mail: arybudi1@gmail.com)

^{2, 3}Dosen, Jurusan Teknik Elektro dan Komputer Fakultas Teknik Universitas Udayana, Jln. Jalan Kampus Bukit Jimbaran 80361 INDONESIA (tel: 0361-703315; fax: 0361-4321; e-mail: imasudarma@gmail.com, w.ariastina@gmail.com)

Kadek Ary Budi Permana: Analisis Rating Sentimen pada ...

mengetahui apakah penilaian yang terdapat pada komentar bersifat baik (positif) atau tidak baik (negatif) bahkan tidak keduanya (netral). Hasil dari analisis sentimen dapat dikembangkan untuk memberikan score atau rating terhadap suatu topik. Hal ini dapat diistilahkan sebagai rating sentimen. Rating sentimen merupakan persentase sentimen dari keseluruhan komentar. Rating sentimen ini juga dapat dijadikan sebagai pembanding antara penilaian pada suatu topik dengan topik lainnya.

Media sosial yang digunakan sebagai sumber data adalah media sosial Youtube. Youtube adalah salah satu sumber informasi video yang memiliki pengguna aktif terbesar, dimana pengguna dapat berinteraksi dengan berbagi video, memberikan like atau dislike, menambah viewer terhadap suatu video dan berlangganan (subscribe) pada suatu channel. YouTube juga memfasilitasi pengguna untuk menanggapi video dengan cara memberikan komentar. Terkadang video yang tidak relevan dan memiliki kualitas rendah berhasil menduduki peringkat yang lebih tinggi dalam hasil pencarian karena jumlah viewer atau like yang lebih banyak [1]. Dalam permasalahan tersebut, rating sentimen dapat dijadikan salah satu parameter untuk hasil pencarian atau rekomendasi video.

Teknik klasifikasi sentimen dapat digunakan untuk menentukan polaritas setiap komentar tunggal dan kemudian digabungkan menjadi rating sentimen [2]. Dalam melakukan analisis sentimen untuk menghasilkan rating sentimen memiliki permasalahan dimana tidak semua komentar relevan kepada topik yang bersangkutan [3]. Sehingga tidak semua komentar dapat dijadikan parameter dalam menentukan rating sentimen. Oleh karena itu, penelitian ini menambahkan pengenalan struktur komentar sebagai parameter dalam melakukan klasifikasi komentar. Jumlah like yang ada pada masing-masing komentar juga ikut dalam menentukan hasil dari rating sentimen. Jumlah like pada komentar dapat diasumsikan sebagai jumlah pengguna yang setuju pada komentar tersebut.

Penelitian ini menggunakan pendekatan struktural (*STRUCT*) dan menggunakan Support Vector Machine (*SVM*). Model *STRUCT* mengkodekan setiap komentar ke dalam shallow syntactic tree [4]. Penggunaan algoritma *SVM* untuk klasifikasi dapat bekerja lebih baik dengan dataset yang besar [5]. Penggunaan *STRUCT-SVM* diharapkan mampu mengklasifikasikan sentimen serta mengklasifikasikan apa yang dituju oleh komentar berdasarkan struktur komentar. Hasil dari klasifikasi sentimen dengan *STRUCT-SVM* ini yang menentukan rating sentimen

II. STUDI LITERATUR

A. Analisis Sentimen

Analisis sentimen atau bisa disebut juga *opinion mining*

p-ISSN:1693 – 2951; e-ISSN: 2503-2372



adalah riset komputasional dari opini, sentimen, dan emosi yang diekspresikan secara tekstual. Jika diberikan satu set dokumen teks yang berisi opini atau sentimen mengenai suatu objek, maka analisis sentimen bertujuan untuk mengekstrak atribut dan komponen dari objek yang telah dikomentari pada setiap dokumen dan menentukan apakah komentar tersebut positif atau negatif [6].

Sentimen memiliki polaritas (positif atau negatif), sumber (orang atau kelompok orang yang memiliki sentimen), dan target (hal yang menjadi arah tujuan sentimen). Dalam teks, sentimen dapat ditangkap pada berbagai tingkatan diantaranya pada tingkat dokumen, paragraf, kalimat, atau klausa. Pada setiap tingkatan tertentu memiliki berbagai komponen sentimen (polaritas, sumber, dan target) yang berlaku dan berbagai teknik yang berbeda dapat digunakan untuk mengidentifikasi sentimen tersebut [7].

B. Rating Sentimen

Rating sentimen adalah salah satu metode yang digunakan untuk mengetahui tingkat sentiment terhadap suatu topik. *Rating* sentimen ditentukan oleh semua kelas yang terdapat pada setiap komentar yang terkait dengan topik. Metode yang digunakan pada penelitian ini untuk menghitung *rating* sentimen adalah *weighted average* [2]. Persamaan *weighted average* dapat dilihat pada persamaan (1).

$$Score_{W_{Avg}}(m_i) = \frac{1}{n} \times \sum_{c \in em_i} Polarity(c) * Likes(c) \quad (1)$$

C. STRUCT-SVM

Klasifikasi *SVM* dengan input x dapat dihitung dengan persamaan (2). a_i adalah parameter model yang diperkirakan dari data pelatihan, y_i adalah variabel target yang merupakan daftar kelas yang sudah didefinisikan sebelumnya, x_i adalah vektor pendukung yang berasal dari data yang direpresentasikan menjadi numerik atau vektor, dan K adalah fungsi kernel.

$$h(x) = \sum_i a_i y_i K(x, x_i) \quad (2)$$

Shallow syntactic Tree Kernel (SHTK) digunakan untuk menangani rekayasa fitur atas representasi struktural dari model *STRUCT*. Dalam menggabungkan model *STRUCT* dan vektor maka memperlakukan setiap komentar dengan *shallow syntactic tree T* dan vektor v [4]. Oleh karena itu, untuk setiap pasangan komentar x_1 dan x_2 didefinisikan kernel seperti pada persamaan (3).

$$K(x_1, x_2) = K_{TK}(T_1, T_2) + K_v(v_1, v_2) \quad (3)$$

K_{TK} adalah *SHTK* dan K_v adalah *kernel* vektor. *Kernel* vektor pada *SVM* misalnya seperti *linear kernel* dan *RBF (Radial Basis Function) kernel*. Persamaan *kernel* vektor dapat dilihat pada persamaan (4) dan (5).

$$\text{Linear } K_v(v_1, v_2) = (v_1 \cdot v_2) \quad (4)$$

$$\text{RBF } K_v(v_1, v_2) = \exp\left(-\frac{\|v_1 - v_2\|^2}{2\sigma^2}\right) \quad (5)$$

Untuk meningkatkan perhitungan kecepatan *TK*, maka dipertimbangkan pasangan *node* (n_1, n_2) yang termasuk dalam tingkat pohon yang sama. Dengan demikian maka diberikan H untuk ketinggian pohon *STRUCT*. $N_{T_1}^h$ dan $N_{T_2}^h$ adalah set *node*

pada ketinggian h [4]. Persamaan (6) mendefinisikan persamaan *SHTK*.

$$SHTK(T_1, T_2) = \sum_{h=1}^H \sum_{n_1 \in N_{T_1}^h} \sum_{n_2 \in N_{T_2}^h} \Delta(n_1, n_2) \quad (6)$$

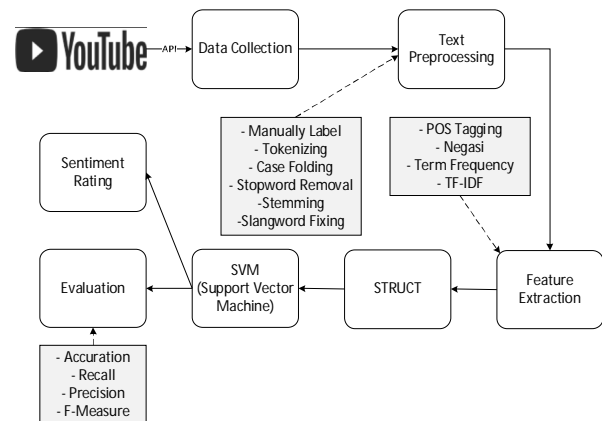
Untuk mendapatkan nilai diantara 0 dan 1, normalisasi *SHTK* dapat dilakukan dengan persamaan (7) [4].

$$\frac{SHTK(T_1, T_2)}{\sqrt{SHTK(T_1, T_1) \times SHTK(T_2, T_2)}} \quad (7)$$

III. METODE PENELITIAN

Data komentar pengguna dikumpulkan dengan memanfaatkan *API Youtube*. Data yang digunakan adalah data komentar pada konten video *official trailer*. Pengelolaan data diawali dengan memberi label atau kelas pada komentar pengguna ke dalam tujuh kelas secara manual berdasarkan teori penelitian [8] untuk proses pelatihan. Kelas tersebut terdiri dari positif film, netral film, negatif film, positif bukan film, netral bukan film, negatif bukan film, dan komentar *spam* / di luar topik. Selanjutnya terdapat proses *text preprocessing* diantaranya *tokenizing*, *case folding*, *stopword removal*, *stemming* dan *slangword fixing*. Setelah proses *text preprocessing* selesai maka dilanjutkan dengan tahap *feature extraction* dimana terdapat empat fitur diantaranya *POS Tagging* dengan pendekatan *STRUCT*, *negasi*, *TF* dan *TF-IDF*.

Hasil klasifikasi dengan *STRUCT-SVM* digunakan sebagai dasar dalam perhitungan *rating* sentimen. Kelas yang digunakan dalam perhitungan *rating* sentimen adalah film positif dan film negatif serta mempertimbangkan jumlah *like* pada setiap komentar pada kedua kelas tersebut. Untuk mengetahui performa *rating sentimen* maka dilakukan pengujian dari klasifikasi yang dihasilkan. Pengujian dilakukan dengan menghitung *recall*, *precision*, *f-measure*, dan akurasi.



Gambar 1: Gambaran Umum Sistem

A. Data

Jenis video yang difokuskan pada penelitian ini adalah *official trailer* film Indonesia pada *Youtube* dan komentar dalam Bahasa Indonesia. Secara umum *official trailer* adalah istilah yang sering digunakan dalam dunia perfilman yang merujuk pada upaya promosi sebuah film yang akan segera tayang dengan bentuk video berdurasi singkat. Untuk contoh komentar dengan kelasnya masing-masing dapat dilihat pada tabel 1.

TABEL I
 CONTOH KOMENTAR DENGAN KELASNYA

Komentar	Kelas
Sumpah senyum2 sendiri. Ini novel dan film romantis terhebat sepanjang masa. Mewakili seluruh perasaan wanita dan perjuangan seorang dilan untuk milea	Positif Film
Film Keren tapi jelek	Netral Film
Film ni Gak Bener, gak mendidik sma sekali nii film	Negatif Film
Iqballl ganteng banget sih	Positif Bukan Film
Siapa yang tau film "REFRAIN" yg pemeran nya afgan sama maudy ayunda... Sama kan ceritanya hahaha	Netral Bukan Film
uh guru gblk, masa tkut ama muridnya DONG00000 bat anyingggg	Negatif Bukan Film
mampir guy's.. https://www.wattpad.com/534934710-rindu-sendiri	Spam / di luar topik

Penelitian ini menggunakan tiga film dengan total jumlah komentar 4199. Film tersebut terdiri dari Dilan 1990 (Film 1), Pengabdian Setan (Film 2), dan Warkop DKI Reborn : Jangkrik Boss Part 1 (Film 3). Komentar dibagi menjadi data *train* dan data *test*. Tabel 2 menunjukkan jumlah dari kedua data tersebut. Data *train* digunakan untuk melatih metode yang digunakan dalam melakukan klasifikasi sedangkan data *test* digunakan untuk menguji performa dan kebenaran dari hasil klasifikasi.

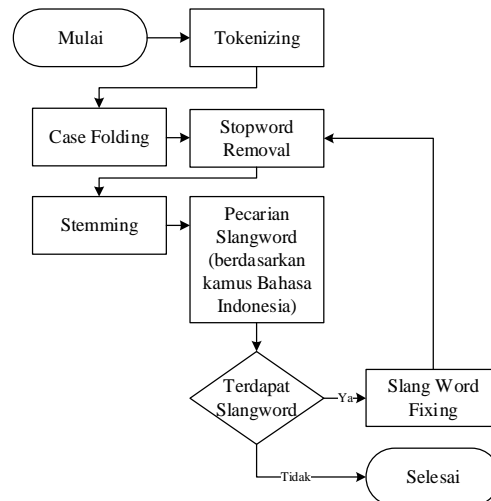
TABEL II
 JUMLAH DATA TRAIN DAN DATA TEST

Kelas	Train	Test
Positif Film	1098	753
Netral Film	144	81
Negatif Film	246	166
Positif Bukan Film	320	207
Netral Bukan Film	503	351
Negatif Bukan Film	69	49
Spam / di luar topik	130	81
Total	2510	1689

B. Text Preprocessing

Tahap *text preprocessing* diawali dengan pemecahan komentar menjadi beberapa *token* yang disebut *tokenizing*. *Token* yang dimaksud adalah per-kata. Kemudian dilakukan *case folding* atau merubah text menjadi *lowercase*/huruf kecil. Setelah itu *stopword* pada komentar dihilangkan dan semua kata yang memiliki imbuhan diubah ke dalam bentuk kata dasar yang dilakukan pada tahapan *stemming*. *Stemming* berhasil meningkatkan hasil akurasi namun tidak terlalu tinggi [9]. Setelah *stemming*, semua kata dibandingkan dengan kamus kata dasar Bahasa Indonesia yang dimiliki sistem. Jika terdapat kata yang tidak sesuai (*slangword*) dengan kamus, maka kata tersebut perlu didefinisikan sesuai dengan kata yang ada pada kamus. *Slangword fixing* dapat meningkatkan keakuratan klasifikasi dan memperkecil ukuran data [10]. Penggantian *slangword* digunakan untuk menghasilkan data yang masuk akal [11]. Setelah kata-kata *slangword* didefinisikan maka akan diperiksa kembali pada proses

penghapusan *stopword* dan *stemming*. Jika masih terdapat kata *slangword* maka proses *slangword fixing* akan berulang kembali sedangkan jika tidak maka *text preprocessing* selesai.



Gambar 2: Flowchart Text Preprocessing pada Komentar

C. Pendekatan STRUCT

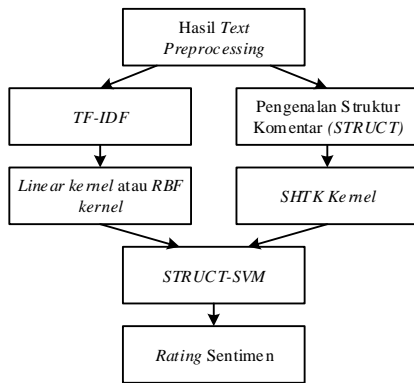
Pengenalan struktur komentar digunakan pada penelitian ini untuk mengetahui apa yang dituju dari suatu komentar. Pengenalan struktur komentar dengan metode *STRUCT* dilakukan dengan memberikan *tag* pada kata-kata tertentu. Pemberian *tag* digunakan untuk membantu mengatasi kata yang tidak dapat didefinisikan atau kata ambigu [12]. Pendekatan pada penelitian ini menggunakan 3 *tag* diantaranya kata yang berhubungan dengan film ([FILM]), kata sentimen positif atau negatif ([POSITIF] atau [NEGATIF]), dan kata negasi seperti kata “tidak”, “bukan”, dan sebagainya ([NEGASI]).

Sebagai contoh pada komentar “dilan peran iqbal bagus tidak suka milea” menghasilkan struktur komentar yaitu “[FILM] [FILM[dilan]] [FILM] [FILM[peran]] [FILM] [FILM[iqbal]] [POSITIF] [POSITIF[bagus]] [NEGASI] [NEGASI[tidak]] [POSITIF] [POSITIF[suka]] [FILM] [FILM[milea]]”.

D. STRUCT-SVM dalam menentukan Rating Sentimen

Hasil dari *text preprocessing* mengalami proses ekstraksi menggunakan *TF-IDF* dan pengenalan struktur dengan pendekatan *STRUCT*. Hasil dari *TF-IDF* diproses lebih lanjut dengan *linear kernel* atau *RBF kernel* dan hasil dari *STRUCT* menggunakan *SHTK*. Gabungan *kernel* inilah yang mempengaruhi perhitungan pada *STRUCT-SVM*. Hasil klasifikasi kemudian yang menentukan *rating* sentimen. Gambaran dari proses *STRUCT-SVM* dalam menentukan *rating* sentimen dapat dilihat pada gambar 3.





Gambar 3: STRUCT-SVM dalam menentukan Rating Sentimen

IV. HASIL DAN PEMBAHASAN

Hasil klasifikasi STRUCT-SVM menggunakan linear kernel dan RBF kernel. Tabel 3 menunjukkan komentar pada masing-masing kelas yang telah berhasil diklasifikasikan dengan menggunakan STRUCT-SVM dengan linear kernel.

TABEL III
HASIL KLASIFIKASI STRUCT-SVM DENGAN LINEAR KERNEL

Kelas	Film 1	Film 2	Film 3
Positif Film	177	396	393
Netral Film	4	10	7
Negatif Film	21	39	36
Positif Bukan Film	12	17	37
Netral Bukan Film	147	155	136
Negatif Bukan Film	8	3	10
Spam / di luar topik	33	31	17
Total	402	651	636

Untuk hasil dari perhitungan STRUCT-SVM dengan RBF kernel dapat dilihat pada tabel 4. Terdapat beberapa perbedaan hasil klasifikasi antara STRUCT-SVM yang menggunakan linear kernel dan RBF Kernel.

TABEL IV
HASIL KLASIFIKASI STRUCT-SVM DENGAN RBF KERNEL

Kelas	Film 1	Film 2	Film 3
Positif Film	175	385	366
Netral Film	4	16	5
Negatif Film	22	43	51
Positif Bukan Film	18	24	52
Netral Bukan Film	142	149	133
Negatif Bukan Film	8	3	12
Spam / di luar topik	33	31	17
Total	402	651	636



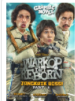
Dalam menentukan rating sentimen menggunakan kelas yang berhubungan dengan film diantaranya positif film dan negatif film. Kelas netral film adalah kelas yang berhubungan dengan film tetapi memiliki polaritas 0 sehingga tidak diikutsertakan dalam menentukan rating sentimen. Komentar dengan kelas yang tidak berhubungan dengan film seperti positif bukan film, netral bukan film, negatif bukan film, dan spam juga tidak digunakan dalam menentukan rating sentimen. Hal ini sejalan dengan tujuan penggunaan STRUCT-SVM sebagai metode klasifikasi yang melakukan klasifikasi berdasarkan sentimen dan struktur komentar. Perhitungan rating sentimen atau score menggunakan persamaan (1)

dimana komentar pada kelas film positif bernilai positif (+1) dan film negatif bernilai negatif (-1). Jumlah like pada setiap komentar berarti setiap nilai komentar yang bernilai positif (+1) ataupun negatif (-1) dikalikan dengan jumlah like yang terdapat pada masing-masing komentar. Pada tabel 5 menunjukkan score dari setiap film pada kedua kernel.

TABEL V
SCORE PADA SETIAP FILM

No	Film	Linear Kernel			RBF Kernel		
		Nilai Positif Film	Nilai Negatif Film	Score	Nilai Positif Film	Nilai Negatif Film	Score
1	Film 1	3422	37	0.98	3426	37	0.98
2	Film 2	338	37	0.80	334	39	0.79
3	Film 3	271	24	0.84	267	30	0.80

Nilai positif film adalah jumlah dari seluruh komentar pada kelas positif film yang dikalikan dengan jumlah like pada masing-masing komentar. Nilai negatif film didapat dengan cara yang sama dengan nilai positif film tetapi menggunakan kelas yang berbeda yaitu kelas negatif film. Score didapat dari nilai positif film dikurang dengan nilai negatif film yang selanjutnya dibagi dengan jumlah dari nilai positif film dan nilai negatif film. Score memiliki nilai minimal -1 sampai dengan nilai maksimal +1. Penelitian ini menggambarkan rating sentimen dengan nilai berbentuk bintang diantara 0 sampai dengan 5 maka nilai score tersebut perlu dikalikan 5. Rating sentimen dapat bernilai negatif jika score bernilai negatif. Jika score bernilai negatif maka rating sentimen adalah 0. Tampilan rating sentimen dapat dilihat pada gambar 4. Semakin besar rating sentimen menunjukkan komentar positif lebih mendominasi dari komentar negatif.

No.	Film	Rating (Linear Kernel)	Rating (RBF Kernel)
1	 Dilan 1990	★★★★★ 4.9	★★★★★ 4.9
2	 Pengabdian Setan	★★★★☆ 4.0	★★★★☆ 3.95
3	 Warkop DKI Reborn: Jangkrik Boss Part 1	★★★★☆ 4.2	★★★★☆ 4.0

Gambar 4: Tampilan Rating Sentimen

Gambar 4 menunjukkan beberapa perbedaan rating sentimen yang dihasilkan antara linear kernel dan RBF kernel pada masing-masing film. Perbedaan klasifikasi antara dua kernel tersebut tentunya yang mempengaruhi perbedaan rating sentimen. Untuk mengetahui kernel yang memiliki akurasi terbaik dalam menentukan rating sentimen maka perlu dilakukan pengujian terhadap hasil klasifikasi pada masing-masing kernel. Pada hasil klasifikasi dilakukan evaluasi pengujian menggunakan metode confusion matrix dengan membandingkan hasil klasifikasi (prediksi) dengan hasil

sebenarnya. Metode ini dapat menentukan *recall*, *precision*, *f-measure*, dan akurasi. Pada tabel 6 menunjukkan nilai *precision*, *recall*, dan *f-measure* pada masing-masing kelas menggunakan *STRUCT-SVM* dengan *linear kernel*.

TABEL VI
HASIL PENGUJIAN *STRUCT-SVM* DENGAN *LINEAR KERNEL*

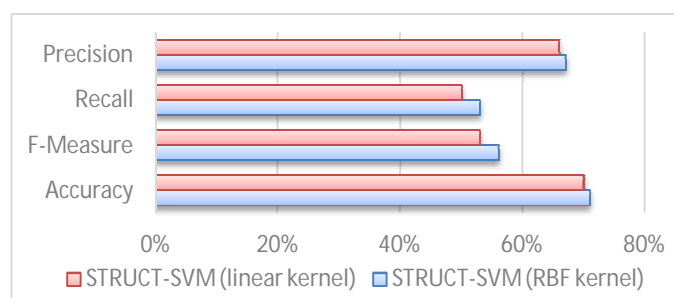
Kelas	Precision	Recall	F-Measure
Positif Film	0.72	0.89	0.8
Netral Film	0.36	0.2	0.26
Negatif Film	0.63	0.4	0.49
Positif Bukan Film	0.56	0.29	0.38
Netral Bukan Film	0.58	0.75	0.65
Negatif Bukan Film	0.43	0.18	0.26
Spam	0.74	0.32	0.44

Dari hasil pengujian pada tabel 6 maka didapatkan rata-rata nilai *precision* 66%, *recall* 50%, dan *f-measure* 53%. Nilai akurasi klasifikasi secara keseluruhan yaitu sebesar 70%. Pada Tabel 7 menunjukkan nilai *precision*, *recall*, dan *f-measure* pada masing-masing kelas menggunakan *STRUCT-SVM* dengan *RBF kernel*.

TABEL VII
HASIL PENGUJIAN *STRUCT-SVM* DENGAN *RBF KERNEL*

Kelas	Precision	Recall	F-Measure
Positif Film	0.7	0.87	0.78
Netral Film	0.46	0.25	0.32
Negatif Film	0.58	0.44	0.5
Positif Bukan Film	0.58	0.35	0.44
Netral Bukan Film	0.57	0.75	0.65
Negatif Bukan Film	0.48	0.19	0.27
Spam	0.76	0.32	0.45

Berbeda dengan hasil menggunakan *linear kernel*, *RBF kernel* menghasilkan hasil yang lebih baik. Ini dapat dilihat dari hasil pengujian pada tabel 7 dimana didapatkan rata-rata nilai *precision* 67%, *recall* 53%, dan *f-measure* 56%. Nilai akurasi klasifikasi secara keseluruhan yaitu sebesar 71%. Pada gambar 5 dapat dilihat grafik dari nilai rata-rata *precision*, *recall*, *f-measure*, dan akurasi dari *STRUCT-SVM* dengan *linear kernel* maupun *RBF kernel*.



Gambar 5: Grafik Pengujian *STRUCT-SVM* (*linear kernel* dan *RBF kernel*)

Grafik menunjukkan *STRUCT-SVM* dengan *RBF kernel* memiliki hasil klasifikasi yang lebih baik dari *linear kernel*. Hal ini juga menunjukkan *rating* sentimen dengan menggunakan hasil klasifikasi dari *STRUCT-SVM* dengan *RBF kernel* memiliki performa yang lebih baik.

V. KESIMPULAN

STRUCT-SVM berhasil melakukan klasifikasi dengan mempertimbangkan sentimen dan struktur komentar. *Rating* sentimen berhasil didapatkan dari hasil klasifikasi tersebut dan memanfaatkan jumlah *like* pada setiap komentar di kelas yang berhubungan dengan film. Nilai rata-rata dari *STRUCT-SVM* dengan *linear kernel* adalah *precision* 66%, *recall* 50%, *f-measure* 53%, dan akurasi sebesar 70%. Sedangkan untuk nilai rata-rata *STRUCT-SVM* dengan *RBF kernel* adalah *precision* 67%, *recall* 53%, dan *f-measure* 56% dan akurasi sebesar 71%. Nilai dari pengujian menunjukkan bahwa *STRUCT-SVM* dengan *RBF kernel* lebih baik daripada *linear kernel* dalam melakukan klasifikasi dan menentukan *rating* sentimen.

Selain memperhatikan struktur komentar dalam menentukan *rating* sentimen, hal lain juga dapat dikembangkan untuk lebih baik lagi yaitu dengan penggunaan sinonim kata. Penggunaan sinonim kata dapat memperkecil ukuran data pada proses klasifikasi. Misalnya kata "apik", "bagus", dan "keren" dapat dijadikan satu kata yaitu "baik".

REFERENSI

- [1] H. Bhuiyan, J. Ara, R. Bardhan, and M. R. Islam, "Retrieving YouTube video by sentiment analysis on user comment," *Proc. 2017 IEEE Int. Conf. Signal Image Process. Appl. ICSIPA 2017*, no. 1, pp. 474–478, 2017.
- [2] J. H. Wang and T. W. Liu, "Improving sentiment rating of movie review comments for recommendation," *2017 IEEE Int. Conf. Consum. Electron. - Taiwan, ICCE-TW 2017*, pp. 433–434, 2017.
- [3] E. Rinaldi and A. Musdholifah, "FVEC-SVM for Opinion Mining on Indonesian Comments of YouTube Video," 2017.
- [4] A. Severyn, A. Moschitti, O. Uryupina, B. Plank, and K. Filippova, "Opinion Mining on YouTube," pp. 1252–1261, 2014.
- [5] S. Anastasia and I. Budi, "Twitter sentiment analysis of online transportation service providers," *2016 Int. Conf. Adv. Comput. Sci. Inf. Syst.*, pp. 359–365, 2016.
- [6] B. Liu, "Sentiment Analysis and Subjectivity," pp. 1–38, 2010.
- [7] C. R. Fink, D. S. Chou, J. J. Kopecky, and A. J. Llorens, "Coarse- and fine-grained sentiment analysis of social media text," *Johns Hopkins APL Tech. Dig. (Applied Phys. Lab.)*, vol. 30, no. 1, pp. 22–30, 2011.
- [8] O. Uryupina, B. Plank, A. Severyn, A. Rotondi, and A. Moschitti, "SenTube: A corpus for sentiment analysis on YouTube social media," *Proc. Lang. Resour. Eval. Conf.*, vol. 2, pp. 4244–4249, 2014.
- [9] A. F. Hidayatullah, "The Influence of Stemming on Indonesian Tweet Sentiment Analysis," no. August, pp. 19–20, 2015.
- [10] T. Singh and M. Kumari, "Role of Text Pre-Processing in Twitter Sentiment Analysis," *Procedia - Procedia Comput. Sci.*, vol. 89, pp. 549–554, 2016.
- [11] S. Gharatkar, A. Ingle, T. Naik, and A. Save, "Review Preprocessing Using Data Cleaning And Stemming Technique," *Int. Conf. Innov. Inf. Embed. Commun. Syst. Rev.*, 2017.
- [12] L. Kumar and P. K. Bhatia, "Text mining: concepts, process and applications," *J. Glob. Res. Comput. Sci.*, vol. 4, no. 3, pp. 36–39, 2013.



[Halaman ini sengaja dikosongkan]