Penerapan SVM untuk Klasifikasi Sentimen pada Review Comment Berbahasa Indonesia di Online Shop

Yoshua Refo, Silvia Rostianingsih, Liliana
Program Studi Informatika Fakultas Teknologi Industri Universitas Kristen Petra
Jl. Siwalankerto 121 – 131 Surabaya 60236
Telp. (031) – 2983455, Fax. (031) – 8417658
Email: yoshuarefo19@gmail.com, silvia@petra.ac.id, lilian@petra.ac.id

ABSTRAK

Dengan banyaknya kebutuhan manusia muncul inovasi yang memungkinkan pengguna dapat mengakses toko secara online, dan komentar merupakan aspek penting saat berbelanja di toko online. Pembeli dapat memberikan komentar mengenai barang yang telah dibeli, baik *negative comment* maupun *positive comment*. Dengan mengumpulkan berbagai macam komentar, data tersebut dapat digunakan untuk mengklasifikasikan komentar.

Penelitian ini akan menggunakan algoritma *Support Vector Machine* (SVM) yang dianggap sebagai metode yang tepat untuk *text classification*. Metode tersebut akan diuji kinerjanya, dilihat dari seberapa akurat metode yang digunakan dalam mengklasifikasikan komentar. Selain itu penelitian ini juga menggunakan kernel yaitu kernel Linear, kernel Radial Basis Function (RBF), dan kernel Polynomial sebagai skenario pengujian.

Berdasarkan hasil pengujian menunjukkan, SVM merupakan metode yang baik dalam mengklasifikasikan teks. SVM mengklasifikasikan teks yang telah melalui tahap preprocessing dengan nilai akurasi 88% pada kernel RBF, 87% pada kernel Linear dan 87% pada kernel Polynomial untuk klasifikasi sentimen. Nilai akurasi pada *aspect classification* sendiri sebesar 78% pada kernel RBF, 78% pada kernel Linear dan 74% pada kernel Polynomial.

Kata Kunci: *Text Classification, Positive Comment, Negative Comment, Support Vector Machine, Aspect*

ABSTRACT

With so many users accessing online shops, comments are an important aspect when shopping. Buyers can provide comments about the goods or thing that have been purchased, both negative comments and positive comments. By collecting various kinds of comments, the data can be used to classify comments.

This research will use the Support Vector Machine (SVM) algorithm which is considered the right method for text classification. The method will be tested for its performance, seen from how good and accurate the method used in classifying comments is. In addition, this research also uses kernels, namely Linear kernels, Radial Basis Function (RBF) kernels, and Polynomial kernels as test scenarios.

Based on the test results shown, SVM is a good method in classifying text. SVM classifies text that has gone through the preprocessing stage with an accuracy value of 88% on the RBF kernel, 87% on the linear kernel, and 87% on the polynomial kernel. The accuracy value in the aspect classification itself is 78% on the RBF kernel, 78% on the Linear kernel and 74% on the Polynomial kernel.

Keywords: *Text Classification, Positive Comment, Negative Comment, Support Vector Machine, Aspect*

1. LATAR BELAKANG

Sosial media dan *online shop* bukan lagi hal asing yang tidak pernah didengar, dengan berkembangnya teknologi manusia dapat membuat pekerjaan menjadi begitu efisien. Memanfaatkan sosial media sebagai media promosi merupakan salah satu keuntungan yang bisa digunakan untuk mengembangkan bisnis tanpa perlu mengeluarkan biaya iklan yang begitu besar. Promosi biasanya dilakukan untuk memasarkan suatu produk barang maupun jasa, sehingga memudahkan pelanggan untuk mengetahui ada barang atau jasa yang sedang di jual.

Poin selanjutnya yaitu *online shop*, yang secara definisi singkat dapat diartikan toko berbasis online. Jadi pelanggan dapat melihat-lihat, membeli dan memberikan komentar tanpa harus datang ke toko tersebut. Online shop bekerja sangat efisien dimana semua proses hanya perlu dilakukan didalam satu aplikasi saja seperti (Shopee, Tokopedia, Bukalapak, dsb.) dari pemilihan barang, jumlah barang yang akan dibeli, pembayaran hingga memberikan komentar tentang kualitas barang dapat dilakukan secara *online*.

Berbicara mengenai komentar, komentar dari pelanggan dapat dikategorikan menjadi dua yaitu *positive feedback* dan *negative feedback*. Mungkin sebagai user merasa fitur tersebut tidak terlalu penting karena barang atau produk yang dipesan sudah sesuai dengan yang diinginkan. Tetapi fitur tersebut merupakan fitur yang penting bagi pemilik toko, dimana *owner* dapat melihat apakah pelanggan-pelanggan puas dengan kualitas produk yang ditawarkan, atau pelanggan ternyata kecewa dengan produk yang dibeli. Hal tersebut ditujukan untuk meningkatkan kualitas barang maupun produk yang ditawarkan oleh *owner* kepada planggan, dengan tujuan pelanggan akan setia dan semakin banyak pelanggan yang puas dengan produk yang akan dibeli.

Pada penelitian sebelumnya, Using Naïve Bayes Algorithm in detection of Hate Tweets, menunjukkan cara kerja beberapa metode yang terdapat pada *Machine Learning* seperti Logistic Regression, Naïve Bayes, BernouliNB, dan lain-lain. Penelitian tersebut menyatakan bahwa performa Naïve Bayes cukup baik jika dibandingkan dengan beberapa metode lain [5]. Sedangkan pada penelitian lainnya dituliskan jika menggunakan algoritma SVM (Support Vector Machine) dapat melakukan Analisa pendekatan sentimen dengan hasil yang sangat baik dan kemudian

disimpulkan jika SVM sangat cocok untuk digunakan dalam menganalisa sentiment [1].

Di penelitian kali ini penulis akan mengklasifikasikan komentar user pada suatu produk di dalam *online shop*. Komentar diambil dari salah satu *online shop* yaitu Shopee, komentar yang diambil ada dua jenis *positive feedback* dan *negative feedback*. Pengumpulan dan validasi data dibantu oleh tenaga ahli Ony Jeselin S.Psi agar proses *labelling* data *positive* dan *negative* dapat dipertanggungjawabkan kebenarannya. Pengklasifikasian sentimen dilakukan menggunakan metode SVM, untuk melihat seberapa efektif metode tersebut dalam mengklasifikasikan *positive feedback* dan *negative feedback*. Tidak hanya itu untuk memberikan kontribusi penelitian dilakukan pengklasifikasin berdasarkan aspek kualitas produk dan layanan. Peneliti akan fokus terhadap pengujian terhadap data sebelum dan setelah preprocess yang akan di ujikan kepada tiga kernel (RBF, Linear, dan Polynomial).

2. LANDASAN TEORI

2.1 Online Shop

Secara definisi *online shop* dapat dikatakan toko *online*, atau toko yang dapat diakses menggunakan internet. Indonesia sendiri memiliki berbagai macam toko *online* yang dapat diakses oleh semua orang seperti, (Tokopedia, Shopee, Bukalapak, Blibli dsb). Segala aktivitas yang dilakukan user didalam aplikasi toko *online* bisa beragam jenis seperti membeli maupun menjual produk maupun jasa. User tidak hanya dapat melakukan dua hal tersebut melainkan segala aktivitas yang dilakukan user saat datang ke toko dapat pula dilakukan di *online shop* [4].

2.2 Positive Feedback

Feedback merupakan gabungan dari dua kata dalam bahasa Inggris, *Feed* dan *Back* yang dapat diartikan memberikan kembali. Definisi *feedback* yang sebenarnya adalah memberikan tanggapan atau respon oleh seseorang kepada pemberi berita / pencipta konten. *Positive Feedback* merupakan hal yang sama seperti *feedback* hanya saja isi atau maksud dari respon yang diberikan bersifat *positive* (membangun, menyukai dan lain sebagainya) [10].

2.3 Negative Feedback

Masalah *negative feedback* sangatlah umum dijumpai oleh pemilik toko *online* karena seringkali hal tersebut dapat dimanfaatkan oleh pemilik toko untuk melihat hal-hal apa saja yang perlu ditingkatkan. Apakah hal tersebut berhubungan dengan kualitas produk yang perlu ditingkatkan, atau kesalahan-kesalahan saat melakukan pengiriman atau hal lain yang perlu dibenahi agar pelanggan dapat tetap berbelanja di toko tersebut [10].

2.4 SVM (Support Vector Machine)

Support Vector Machine adalah Teknik *machine learning* yang dapat diterapkan ke dalam klasifikasi dan regresi. SVM bekerja berdasarkan prinsip menemukan hyperlane yang akan memaksimalkan margin antara dua kelas, SVM juga dapat diterapkan pada kumpulan data yang dapat dipisahkan maupun tidak dapat dipisahkan [9]. Hyperplane merupakan garis pemisah antar class, SVM biasa digunakan untuk mengklasifikasikan dua kelas disebut *binary classification* dan lebih dari dua kelas disebut.*multi class classification*.

2.5 Pandas Library Python

Pandas merupakan library python yang bekerja atau biasa digunakan untuk proyek yang berhubungan dengan dataset. Pandas berfungsi untuk melakukan analisis, membersihkan, menjelajahi dan memanipulasi data. Data relevant merupakan data yang penting didalam data science dengan adanya pandas library dapat membersihkan dan memperbaiki dataset yang berantakan menjadi data yang dapat dibaca dan relevant.

Pandas memiliki dua struktur utama yaitu seri dan dataframe yang dapat diatur. Dalam implementasinya pandas dapat digunakan pada data yang berbeda (*string, integer, dan float*) ke dalam satu struktur data dan memiliki kemampuan untuk menerapkan dengan mudah metode atau fungsi ke semua maupun sebagian data. Library pandas merupakan salalh satu library pada python yang berkaitan erat dengan *data analysis context* [2].

2.6 Scikit-Learn Library Python

Scikit-learn merupakan library python yang biasanya digunakan dan dimanfaatkan dalam data science. Scikit-learn memberikan sejumlah fitur untuk data science seperti algoritma regresi, classification, naïve bayes, decision tree, dan berbagai macam algoritma machine learning yang lain. Pemanfaatannya seringkali digunakan untuk penelitian yang memiliki data set dan perlu menerapkan algoritma machine learning ke dalam model [8].

Library ini awalnya dikembangkan oleh Cournapeu pada tahun 2007, tetapi rilis pada tahun 2010. Library ini dibuat untuk keperluan ilmiah, komputasi didalamnya umumnya untuk analisis data. Untuk supervised learning mempelajari kemungkinan pola dari dua atau lebih fitur dari data training yang diberikan juga memanfaatkan library scikit-learn [7].

2.7 Confusion Matrix

Confusion matrix merupakan teknik yang digunakan untuk mengukur performa untuk masalah klasifikasi machine learning. Confusion matrix sendiri adalah tabel yang memiliki 4 kombinasi yaitu True Positive, True Negative, False Positive, False Negative. True Positive (TP) berarti prediksi positive dan benar, True Negative (TN) berarti prediksi negative dan benar, False Positive (FP) berarti prediksi positive dan salah, False Negative (FN) berarti prediksi negative dan salah. Jadi confusion matrix menggunakan nilai TP,TF,FP dan FN untuk menentukan nilai accuracy, precision, dan recall [6].

2.8 K-Fold Cross Validation

K-Fold Cross Validation secara garis besar dapat diartikan metode yang digunakan untuk mengacak dan membagi data kedalam kelompok sesuai k yang ditentukan. Jadi data akan dikelompokkan sebanyak k dan salah satu dari kelompok data tersebut akan dijadikan sebagai data uji dan sisa data yang lain akan dijadikan data training [11].

2.9 NLP (Natural Language Processing)

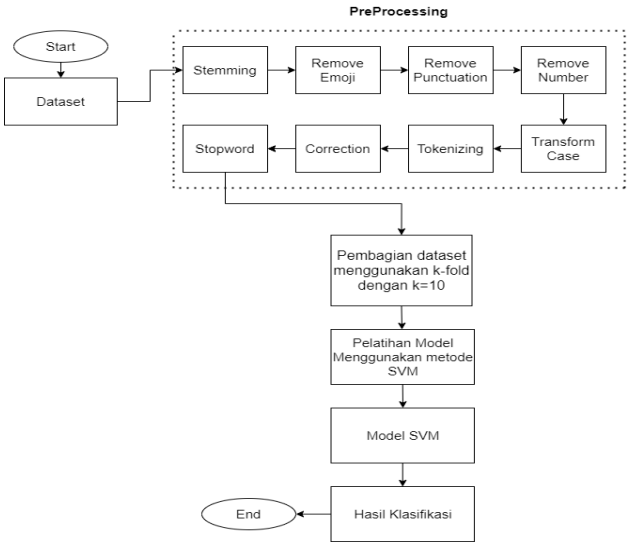
NLP merupakan ilmu komputer yang digunakan untuk mempelajari bahasa manusia atau bahasa sehari-hari. Pemanfaatan NLP sudah cukup banyak beredar seperti, google translate, chatbot, ok google, siri, cortana dan lain-lain. NLP masih memiliki korelasi dengan cabang ilmu dari Artificial Intelligence. Library yang cukup popular untuk digunakan adalah NLTK, merupakan library python yang dapat digunakan untuk memproses kalimat atau bahasa manusia [3].

3. DESAIN SISTEM

3.1 Preprocessing Dataset

Preprocessing data merupakan sebuah tahapan yang dilakukan dengan maksud membuat dataset siap digunakan sebelum dimasukkan kedalam model untuk diklasifikasikan. Jadi data yang

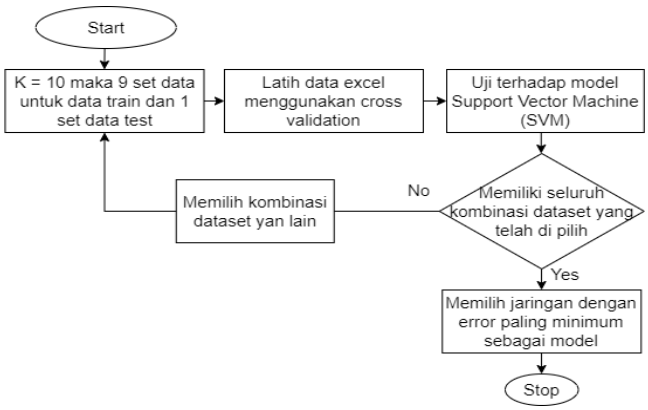
sudah melalui tahapan preprocessing bukan lagi merupakan data mentah melainkan data yang siap untuk diteliti. Manfaat melakukan preprocessing yaitu dapat menghilangkan noise data, mendapatkan kata demi kata yang baku (NLP), sehingga proses klasifikasi dapat berjalan tanpa ada hambatan. Berikut merupakan tahapan preprocessing yang akan dilakukan agar data jadi siap pakai. Proses dapat dilihat pada Gambar 1.



Gambar 1. Proses pada *Preprocessing Data*

3.2 K-Fold Cross Validation

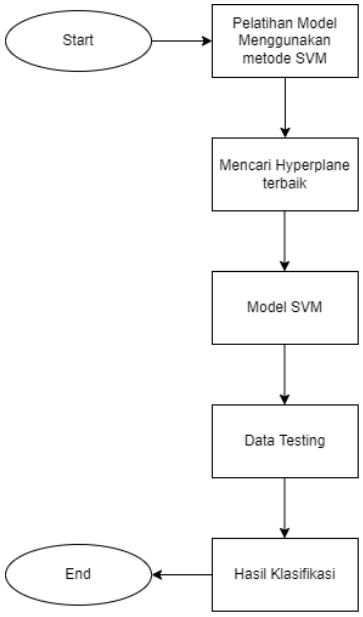
Pada tahap ini data yang sudah dikumpulkan dan telah malalui *preprocessing*, data tersebut sudah dapat di gunakan untuk pengujian. Data tersebut tidak semuanya akan dimasukkan kedalam model, jadi data akan dipisah terlebih dahulu menjadi dua yaitu data training dan data testing. Untuk proses pemisahan data akan menggunakan metode *K-Fold Cross Validation* agar data terbagi secara rata. Data yang telah dipisahkan memiliki tujuan yang berbeda pula, untuk data testing adalah data dengan porsi yang lebih sedikit (data uji). Jadi data testing digunakan untuk menguji model *machine learning* untuk melihat performa nya. Untuk data Training merupakan data dalam porsi yang lebih banyak, data training ditujukan sebagai media pembelajaran untuk model *machine learning* dalam mengklasifikasikan data. Berikut merupakan tahapan dalam data training dan data testing. Proses dapat dilihat pada.Gambar 2.



Gambar 2. Proses K-Fold Cross Validation

3.3 SVM (Support Vector Machine)

Pengklasifikasian menggunakan SVM memanfaatkan hyperplane terbaik, hyperplane merupakan garis yang membagi antar kelas untuk memaksimalkan jarak antar kelas. Jadi dataset yang sudah diberi label akan diproses untuk membuat model machine learning. Pembuatan model tersebut merupakan gabungan antara sistem, dataset dan algoritma machine learning (SVM) yang diproses untuk menghasilkan nilai accuracy, precision dan recall. Di dalam algoritma SVM terdapat juga tiga kernel yang dapat diujikan sebagai skenario percobaan yaitu kernel linear, kernel polynomial dan kernel radial basis function (RBF). Berikut tahapan SVM dalam melakukan klasifikasi. Proses dapat dilihat pada Gambar 3.



Gambar 3. Proses Kerja SVM

4. PENGUJIAN SISTEM

4.1 Hasil Klasifikasi Sentimen pada Kernel (RBF)

Proses yang berlangsung, model svm dijalankan bersamaan dengan kernel RBF untuk menentukan nilai akurasi untuk setiap fold maupun rata-rata nilai akurasi seluruh fold. Jadi nilai akurasi yang tunjukkan berbeda-beda sesuai dengan data yang terdapat dalam fold tersebut. Sebagai skenario uji, model diterapkan pada data yang belum melalui tahap *preprocess* dan data yang telah melalui tahap *preprocess* untuk melihat tingkat akurasi yang dihasilkan.dapat dilihat pada Tabel 1.

Tabel 1. Hasil Akurasi Sentimen Kernel RBF

Data Sebelum Preprocess		Data Setelah Preprocess	
Fold 1	0.915	**Fold 1**	0.910
Fold 2	0.875	**Fold 2**	0.930
Fold 3	0.850	**Fold 3**	0.860

Fold 4	0.885	**Fold 4**	0.855
Fold 5	0.875	**Fold 5**	0.880
Fold 6	0.870	**Fold 6**	0.900
Fold 7	0.845	**Fold 7**	0.885
Fold 8	0.870	**Fold 8**	0.870
Fold 9	0.900	**Fold 9**	0.895
Fold 10	0.860	**Fold 10**	0.870
Average 10 fold = 0.874		**Average 10 fold** = 0.885	

4.2 Hasil Klasifikasi Sentimen pada Kernel (Linear)

Perbedaan dengan pengujian sebelumnya yaitu penggunaan kernel Linear yang digunakan sebeagai skenario uji untuk melihat perbedaan nilai akurasi. Proses yang berjalan dalam pengujian ini sama dengan pengujian sebelumnya dimana data yang diuji berupa data yang sudah melalui *preprocess* dan data yang belum melalui tahap *preprocess*. dapat dilihat pada Tabel 2.

Tabel 2. Hasil Akurasi Sentimen Kernel Linear

Data Sebelum Preprocess		Data Setelah Preprocess	
Fold 1	0.910	**Fold 1**	0.890
Fold 2	0.860	**Fold 2**	0.920
Fold 3	0.865	**Fold 3**	0.850
Fold 4	0.890	**Fold 4**	0.840
Fold 5	0.880	**Fold 5**	0.885
Fold 6	0.840	**Fold 6**	0.885
Fold 7	0.855	**Fold 7**	0.880
Fold 8	0.860	**Fold 8**	0.880
Fold 9	0.880	**Fold 9**	0.885
Fold 10	0.850	**Fold 10**	0.860
Average 10 fold = 0.869		**Average 10 fold** = 0.877	

4.3 Hasil Klasifikasi Sentimen pada Kernel (Polynomial)

Sama halnya dengan penelitian diatas, penelitian kali ini menggunakan kernel polynomial untuk melihat nilai akurasi. Data yang digunakan pada penelitian ini merupakan data yang sudah melalui tahap *preprocess* dan data yang belum melalui tahap *preprocess* sebagai skenario uji. dapat dilihat pada Tabel 3

Tabel 3. Hasil Akurasi Sentimen Kernel Polynomial

Data Sebelum Preprocess		Data Setelah Preprocess	
Fold 1	0.885	**Fold 1**	0.910
Fold 2	0.860	**Fold 2**	0.890
Fold 3	0.815	**Fold 3**	0.870
Fold 4	0.830	**Fold 4**	0.860
Fold 5	0.855	**Fold 5**	0.860
Fold 6	0.840	**Fold 6**	0.895
Fold 7	0.855	**Fold 7**	0.905
Fold 8	0.810	**Fold 8**	0.875
Fold 9	0.840	**Fold 9**	0.900
Fold 10	0.845	**Fold 10**	0.880
Average 10 fold = 0.843		**Average 10 fold** = 0.884	

4.4 Hasil Klasifikasi Aspek pada Kernel (RBF)

Pada pengujian kali ini peneliti akan menggunakan label yang berbeda yaitu label aspek, label aspek terdiri dari kualitas produk dan layanan. Jadi data akan diklasifikasi menjadi dua kelas dan diujikan pada kernel RBF sebagai scenario uji. Data uji berupa data yang sudah melalui tahap preprocess dan data sebelum preprocess. Dapat dilihat pada Tabel 4.

Tabel 4. Hasil Akurasi Aspek Kernel RBF

Data Sebelum Preprocess		Data Setelah Preprocess	
Fold 1	0.720	**Fold 1**	0.800
Fold 2	0.680	**Fold 2**	0.755
Fold 3	0.760	**Fold 3**	0.800
Fold 4	0.735	**Fold 4**	0.810
Fold 5	0.665	**Fold 5**	0.795
Fold 6	0.710	**Fold 6**	0.810
Fold 7	0.745	**Fold 7**	0.785
Fold 8	0.705	**Fold 8**	0.750
Fold 9	0.695	**Fold 9**	0.770
Fold 10	0.735	**Fold 10**	0.790
Average 10 fold = 0.715		**Average 10 fold** = 0.786	

4.5 Hasil Klasifikasi Aspek pada Kernel (Linear)

Pada pengujian kali ini peneliti akan menggunakan label aspek, label aspek terdiri dari kualitas produk dan layanan. Jadi data akan diklasifikasi menjadi dua kelas dan diujikan pada kernel linear sebagai scenario uji. Data uji berupa data yang sudah melalui tahap preprocess dan data sebelum preprocess. Dapat dilihat pada Tabel 5.

Tabel 5. Hasil Akurasi Aspek Kernel Linear

Data Sebelum Preprocess		**Data Setelah Preprocess**	
Fold 1	0.745	**Fold 1**	0.820
Fold 2	0.690	**Fold 2**	0.755
Fold 3	0.780	**Fold 3**	0.815
Fold 4	0.745	**Fold 4**	0.800
Fold 5	0.690	**Fold 5**	0.815
Fold 6	0.740	**Fold 6**	0.795
Fold 7	0.755	**Fold 7**	0.810
Fold 8	0.715	**Fold 8**	0.760
Fold 9	0.710	**Fold 9**	0.745
Fold 10	0.745	**Fold 10**	0.770
Average 10 fold = 0.731		**Average 10 fold** = 0.788	

4.6 Hasil Klasifikasi Aspek pada Kernel (Polynomial)

Pada pengujian kali ini peneliti akan menggunakan label aspek, label aspek terdiri dari kualitas produk dan layanan. Jadi data akan diklasifikasi menjadi dua kelas dan diujikan pada kernel polynomial sebagai scenario uji. Data uji berupa data yang sudah melalui tahap preprocess dan data sebelum preprocess. Dapat dilihat pada Tabel 6.

Tabel 6. Hasil Akurasi Aspek Kernel Linear

Data Sebelum Preprocess		**Data Setelah Preprocess**	
Fold 1	0.670	**Fold 1**	0.735
Fold 2	0.650	**Fold 2**	0.705
Fold 3	0.750	**Fold 3**	0.800
Fold 4	0.710	**Fold 4**	0.815
Fold 5	0.605	**Fold 5**	0.735
Fold 6	0.650	**Fold 6**	0.745
Fold 7	0.665	**Fold 7**	0.745
Fold 8	0.685	**Fold 8**	0.720
Fold 9	0.670	**Fold 9**	0.730
Fold 10	0.675	**Fold 10**	0.720
Average 10 fold = 0.672		**Average 10 fold** = 0.745	

5. KESIMPULAN

Berdasarkan hasil yang telah didapatkan dari berbagai pengujian, maka bisa disimpulkan bahwa perumusan masalah dapat terjawab dengan baik. Berikut kumpulan kesimpulan dari setiap rumusan masalah:

- Algoritma *machine learning* yang digunakan *Support Vector Machine (SVM)* merupakan metode yang tepat untuk melakukan klasifikasi teks. Dalam pengujian algoritma SVM di uji terhadap dua data dengan karakteristik data yang berbeda. Data pertama merupakan data yang belum melalui tahap *preprocess*, sedangkan data kedua merupakan data yang telah melalui tahap *preprocess*. Hasil pengujian menunjukkan data yang telah melalui tahap *preprocess* kemudian diuji kedalam model SVM menunjukkan nilai akurasi yang lebih tinggi dibandingkan data yang tidak melalui tahap *preprocess*.
- Hasil pengujian menunjukkan perbedaan pada data dengan karakteristik berbeda dan kernel yang berbeda. Muncul perbedaan akurasi yang dapat dilihat untuk setiap pengujian terhadap kernel yang berbeda. Pada kernel Radial Basis Function (RBF) menunjukkan nilai rata-rata akurasi sebesar 0.883. Sedangkan pada kernel Linear menunjukkan nilai akurasi sebesar 0.874, dan untuk kernel polynomial menunjukkan nilai akurasi sebesar 0.874. Disetiap kernel yang diuji menunjukkan kesamaan yaitu nilai akurasi yang dihasilkan sangat baik.
- Untuk hasil pengujian pada label aspek menghasilkan nilai akurasi yang baik pula, label aspek juga diujikan pada tiga kernel yang sama. Untuk data setelah preprocess menghasilkan nilai rata-rata akurasi tertinggi, 0.78 pada kernel RBF, 0.78 pada kernel Linear dan 0.76 pada kernel Polynomial. Disetiap kernel yang diuji menunjukkan kesamaan yaitu nilai akurasi yang dihasilkan sangat baik.

6. SARAN

Berdasarkan hasil dan berbagai pengujian yang telah dilakukan, maka saran untuk penelitian selanjutnya yaitu:

- Dapat menggunakan metode atau algoritma lain seperti Naïve Bayes sebagai pembanding dan skenario uji.
- Mempertajam *preprocessing* untuk meningkatkan akurasi menjadi lebih baik. *preprocessing* yang dimaksud yaitu tahapan untuk membuang kata yang tidak bermakna.

7. REFERENSI

[1] Al Amrani, Y., Lazaar, M. and El Kadiri, K., 2018. Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis. *Procedia Computer Science*, 127, pp.511-520.

[2] Betancourt, R. and Chen, S., 2019. pandas Library. *Python for SAS Users*, [online] pp.65-109. URI= https://www.researchgate.net/publication/335669917_pandas_Library [Accessed 7 July 2022].

[3] Bird, S., Klein, E. and Loper, E., 2009. *Natural language processing with Python*. 1st ed. Sebastopol: O'reilly, pp.211-217.

[4] Hermanto, Kuntoro, A., Asra, T., Pratama, E., Effendi, L., & Ocanitra, R. 2020. Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm And Support Vector Machine Based Smote Technique. *Journal Of Physics: Conference Series*, *1641*(1), 012102. DOI= 10.1088/1742-6596/1641/1/012102

[5] Kiilu, K., Okeyo, G., Rimiru, R., & Ogada, K. (2018). Using Naïve Bayes Algorithm in detection of Hate Tweets. *International Journal Of Scientific And Research Publications (IJSRP)*, *8*(3), 99-107. DOI= 10.29322/ijsrp.8.3.2018.p7517

[6] Luque, A., Carrasco, A., Martín, A. and de las Heras, A., 2019. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, pp.216-231.

[7] Nelli, F., n.d. *Python data analytics*. 1st ed. Apress, pp.237-264.

[8] Ofoeda, J., Boateng, R. and Effah, J., 2019. Application Programming Interface (API) Research. *International Journal of Enterprise Information Systems*, 15(3), pp.76-95.

[9] Rohmawati, U., Sihwi, S., & Cahyani, D. (2018). SEMAR: An Interface for Indonesian Hate Speech Detection Using Machine Learning. *2018 International Seminar On Research Of Information Technology And Intelligent Systems (ISRITI)*. DOI= 10.1109/isriti.2018.8864484

[10] Syahputra, H. (2021). Sentiment Analysis of Community Opinion on Online Store in Indonesia on Twitter using Support Vector Machine Algorithm (SVM). *Journal Of Physics: Conference Series*, *1819*(1), 012030. DOI= 10.1088/1742-6596/1819/1/012030

[11] Wong, T., 2015. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, 48(9), pp.2839-2846.