Sentimen Analisis Review Film pada Website Rotten Tomatoes Menggunakan Metode SVM Dengan Mengimplementasikan Fitur Extraction Word2Vec

1st Ardhian Fahmi Sabani
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
ardhianfahmi@students.telkomuniversity.ac.id

2nd Adiwijaya
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
adiwijaya@telkomuniversity.ac.id

3rd Widi Astuti
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
widiwdu@telkomuniversity.ac.id

Abstrak

Berkembang pesatnya teknologi jaringan internet menyebabkan masyarakat lebih mudah dalam mengakses berbagai macam film yang ingin mereka tonton. Berkaitan dengan hal tersebut, ulasan film menjadi salah satu yang paling penting untuk diperhatikan dalam industri film. Hal ini didasari karena berdasarkan ulasan penonton dapat diketahui tingkat kepuasan mereka terhadap film yang telah mereka saksikan. Analisis Sentiment adalah salah satu solusi bidang penelitian yang berbasis teks yang cocok untuk membahas masalah kepuasan berdasarkan ulasan penonton film. Analisis yang dilakukan didasarkan pada kelas ulasan positif dan negatif yang diberikan oleh pengguna. Dalam penelitian ini, permasalahan tersebut diselesaikan menggunakan metode Support Vector Machine (SVM) dan Word2Vec. Data yang diambil dari website Rotten Tomatoes menggunakan bahasa inggris yang sudah diolah dan diberi label. Model terbaik dari penelitian ini dengan menggunakan kernel RBF dengan menggunakan K-Fold menghasilkan rata – rata *precission* sebesar 77,2%, *recall* 68,2%, *F1-Score* 70,2% dan *accuracy* 79,0%.

Kata Kunci: analisis sentimen, SVM, word2vec, ulasan, RBF

Abstract
The rapid development of internet network technology makes it easier for people to access various kinds of films they want to watch. In this regard, film reviews are one of the most important things to pay attention to in the film industry. This is based on the fact that based on audience reviews it can be seen their level of satisfaction with the films they have watched. Sentiment analysis is one of the solutions in the field of text-based research that is suitable for discussing the problem of satisfaction based on movie audience reviews. The analysis carried out is based on the class of positive and negative reviews provided by users. In this study, these problems were solved using the Support Vector Machine (SVM) and Word2Vec methods. Data taken from the Rotten Tomatoes website uses English which has been processed and labeled. The best modelfrom this study using the RBF kernel with K-Fold produces an average precision of 77.2%, recall 68.2%, F1-Score 70.2% and accuracy 79.0%.

Keywords: sentiment analysis, SVM, word2vec, review, RBF

I. PENDAHULUAN

Dikarenakan berkembang pesatnya teknologi informasi pada saat ini, menjadikan beberapa media dapat lebih mudah diakses oleh orang – orang dalam mencari suatu informasi dan mengungkapkan opini mengenai film yang telah dilihatnya [1]. Website review film merupakan salah satu yang terkena dampak dari pesatnya perkembangan teknologi informasi yang ada pada saat ini. Salah satu yang terdampak adalah Rotten Tomatoes yang merupakan sebuah situs yang berisi opini yang berupa ulasan dari berbagai individu atau golongan tentang pendapat mereka mengenai berbagai film [2]. Sentiment Analysis juga dikenal sebagai metode untuk menganalisis opini dan mengklasifikasikan teks opini yang didapat berdasarkan tipe nya menjadi beberapa kelas yang berbeda sehingga membuat pemrosesan data menjadi lebih mudah [3]. Data yang telah dikumpulkan kemudian dikategorikan menjadi beberapa kelas seperti positif dan negative [3].

Data yang diambil menggunakan Bahasa Inggris dari website Rotten Tomatoes. Metode klasifikasi yang digunakan adalah Support Vector Machine dengan menggunakan fitur ekstraksi Word2Vec. Penelitian ini [5] menjelaskan bahwa fitur ekstraksi Word2Vec sebagai fitur ekstraksi sangat bagus digunakan karena mewakili setiap kata dengan sebuah vektor. Jadi, pada penggunaan Word2Vec, polaritas skor setiap kata memiliki peran penting terhadap hasil analisis sentimen. Metode Support Vector Machine dipilih karena akurasi yang didapat relatif tinggi dan tingkat generalisasi pada metode ini tidak dipengaruhi banyaknya data latih [6]. Tujuan dari penelitian ini adalah untuk mengetahui model terbaik yang dapat digunakan untuk mengolah dataset menggunakan kernel dalam metode klasifikasi SVM dengan fitur ekstraksi Word2Vec. Karena terdapat beberapa kernel pada metode SVM, maka hasil dari penelitian ini akan disimpulkan berdasarkan model terbaik dari penggunaan kernel dan kombinasi dari penggunaan teknik pengolahan data yang tepat.

II. KAJIAN TEORI

Pada penelitian terkait yang telah dilakukan sebelumnya, dapat diketahui bahwa untuk melakukan penelitian di bidang Sentiment Analysis terdapat beberapa metode yang digunakan. Pada paper [1] penelitian yang dilakukan dengan metode SVM dengan menggunakan fitur ekstraksi TFIDF. Penelitian ini dilakukan dengan membagi data menjadi dua kelas yaitu kelas negative dan kelas positif. Dari pembagian dua kelas tersebut pada penelitian [1]

menggunakan metode SVM dengan fitur ekstraksi TFIDF diperoleh akurasi total 77%.

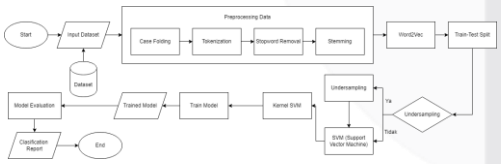
Pada penelitian selanjutnya dapat diketahui bahwa menggunakan metode SVM terbukti dapat meningkatkan hasil akurasi. Pada penelitian ini [4] dilakukan dengan membandingkan metode SVM dan Naïve Bayes pada data movie review untuk memperoleh hasil akurasi yang diinginkan. Hasilnya metode SVM mendapat akurasi sebesar 82.48% sedangkan Naïve Bayes mendapat akurasi sebesar 76.56%. Dapat dilihat dari hasil yang diperoleh dari kedua penelitian tersebut selain fitur ekstraksi, metode yang digunakan sangat mempengaruhi hasil dari akurasi.

Pada penelitian [6] dapat diketahui bahwa terdapat banyak metode yang digunakan dan memiliki tingkat akurasi yang berbeda – beda. Pada penelitian ini, metode Support Vector Machine mendapat akurasi sebesar 87.87%. Pada penelitian selanjutnya [7] dapat diketahui bahwa terdapat penggunaan fitur ekstraksi yang tepat sangat berpengaruh terhadap hasil uji performansi yang akan didapatkan. Penelitian ini [7] juga membagi data menjadi dua kelompok yaitu *balanced* dan *imbalanced*. Akurasi dari data balanced sebesar 87.66% sedangkan data *imbalanced* mendapatkan akurasi sebesar 87.9%.

Pada penelitian [8] dilakukan dataset juga dikelompokan menjadi dua kelas yaitu kelas positif dan kelas negatif dengan menggunakan metode ekstraksi Word2Vec. Akurasi yang diperoleh dengan menggabungkan metode SVM dengan fitur ekstraksi Word2Vec lebih tinggi dari penelitian sebelumnya yang menggunakan metode SVM dengan fitur ekstraksi TFIDF. Akurasi dari kelas positif yang diperoleh sebesar 85.4% dan kelas negative sebesar 83.9%. Preprocessing data dan pemilihan metode ekstraksi merupakan kunci utama untuk memperoleh akurasi yang baik.

III. METODE

Penelitiann ini bertujuan untuk mengetahui dan mengenalisis sentiment mengenai ulasan film pada *website Rotten Tomattoes.* Gambaran sistem yang dibangun dapat dilihat pada gambar 1.



Gambar 1. Flowchart Sistem

A. Dataset

Pada penelitian ini data yang digunakan diambil dari website Rotten Tomatoes yang tersedia pada website Kaggle dalam bentuk file excel. Data berjumlah 1.130.017 baris data. Dikarenakan membutuhkan *running time* yang sangat lama ketika menggunakan dataset secara keseluruhan mulai tahun 2010, dimana dataset tersebut sebanyak 1.130.017 , maka dataset yang digunakan hanya mulai dari tahun 2020 saja. Dataset yang digunakan adalah dataset dengan review_date pada tahun 2020 sebanyak 45.739 baris data.

Ada 2 kelas sentimen dari ulasan penonton yang terdapat pada atribut review_content yang akan memiliki keterkaitan pada atribut review_type yaitu kelas sentimen Rotten untuk sentiment negative dan kelas Fresh untuk sentiment positif. Selain itu, atribut review_date juga memiliki keterkaitan dengan 2 atribut sebelumnya karena digunakan untuk memilih dataset yang akan digunakan berdasarkan tanggal. Berikut adalah contoh dataset yang akan digunakan:

Table 1. Contoh Data

Review_date	Review_content	Review_type
2020-01-13	In the hollywood's godzilla without of personality. we don't care about what happened to that big lizard. unlucky, we don't get a great look at him of the way in the film.	Fresh
2020-03-25	Atomic blonde was falls between james bond and john wick, carrying the violent spy chase of the former while increading the violence level, but not enough reaching that of the latter.	Rotten

Berdasarkan contoh ulasan dari penonton film pada tabel diatas, data ulasan dari penonton film sudah diberi label *Fresh* untuk ulasan yang bersifat positif, sedangkan untuk ulasan yang bersifat negatif diberikan label *Rotten* oleh ahli Bahasa pada *website* Rotten Tomatoes.

B. *Preprocessing*

Pada tahap *preprocessing* ini data dibersihkan dengan menggunakan beberapa tahap diantaranya *case folding*, *tokenisasi*, *stopword removal* dan *stemming* [25]. Proses ini sangat membantu penelitian untuk menghasilkan lebih banyak data yang terstruktur.

3.2.1 Konversi Tipe Data

Dikarenakan program tidak bisa memproses review_type dalam bentuk *string*, maka pada atribut *review_type* akan diubah dalam bentuk data *integer*. Pada tahap ini, label *Fresh* akan diberi label angka 1, dan label *Rotten* akan diberi label angka 0 untuk membedakan tipe label dari dua kelas positif dan negatif.

3.2.2 Case Folding

Case folding merupakan sebuah proses untuk mengubah kata yang memiliki huruf kapital menjadi huruf kecil. Contoh case folding ada pada tabel dibawah :

Table 2. Contoh Proses Case Folding

Sebelum	Sesudah
In the hollywood's godzilla without of personality. we don't care about what happened to that big lizard. unlucky, we don't get a great look at him of the way in the film.	in the hollywood's godzilla without of personality. we don't care about what happened to that big lizard. unlucky, we don't get a great look at him of the way in the film.

3.2.3 Tokenization

Tokenization atau dalam Bahasa Indonesia disebut dengan tokenisasi adalah adalah identifikasi kata dasar dengan proses mengelompokkan teks menjadi kalimat dan kata [11]. Contoh kalimat yang dilakukan proses tokenisasi menjadi kata dasar ada pada tabel berikut.

Table 3. Contoh Hasil Tokenisasi

Sebelum Tokenisasi	Sesudah Tokenisasi
in the hollywood's godzilla without of personality. we don't care about what happened to that big lizard. unlucky, we don't get a great look at him of the way in the film.	['in', 'the', 'hollywood', '’', 's', 'godzilla', 'without', 'of', 'personality', '.', 'we', 'don', '’', 't', 'care', 'about', 'what', 'happened', 'to', 'that', 'big', 'lizzard', '.', 'unlucky', ',', 'we', 'don', '’', 't', 'get', 'a', 'great', 'look', 'at', 'him', 'of', 'the', 'way', 'in', 'the', 'film']

3.2.4 Stopword Removal

Stopword removal adalah salah satu tahap dari preprocessing dimana pada tahap ini berguna untuk menghapus kata – kata tidak penting yang sering muncul karena akan mempengaruhi hasil peneletian [12]. Berikut adalah contoh hasil dari proses stopword removal.

Table 4. Contoh Hasil Stopword Removal

Sebelum Stopword Removal	Sesudah Stopword Removal
['in', 'the', 'hollywood', '’', 's', 'godzilla', 'without', 'of', 'personality', '.', 'we', 'don', '’', 't', 'care', 'about', 'what', 'happened', 'to', 'that', 'big', 'lizzard', '.', 'unlucky', ',', 'we', 'don', '’', 't', 'get', 'a', 'great', 'look', 'at', 'him', 'of', 'the', 'way', 'in', 'the', 'film']	['hollywood', 'godzilla', 'without', 'personality', 'care', 'happened', 'big', 'lizzard', 'unlucky', 'get', 'great', 'look', 'way', 'film']

3.2.5 Stemming

Stemming adalah tahap dimana proses yang dilakukan untuk mengubah kata – kata yang tercampur menjadi kata dasar yang ada [13]. Berikut adalah contoh dari proses stemming.

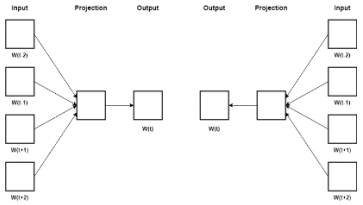
Table 5. Contoh Hasil Stemming

Sebelum Stemming	Sesudah Stemming
['hollywood', 'godzilla', 'without', 'personality', 'care', 'happened', 'big', 'lizzard', 'unlucky', 'get', 'great', 'look', 'way', 'film']	['hollywood', 'godzilla', 'without', 'personal', 'care', 'happens', 'big', 'lizzard', 'unlucky', 'get', 'great', 'look', 'way', 'film']

C. Word2Vec

Word2Vec merupakan sebuah *tool* yang dibuat berdasarkan *deep learning* yang bertujuan untuk merepresentasikan kata dalam sebuah konteks sebagai vektor dengan N dimensi. Word2vec sendiri mengubah kata menjadi vektor dan menghitung kesamaan menurut kosinus antara vektor kata [9].

Terdapat dua jenis model pada fitur ekstraksi Word2Vec diantaranya adalah model Skip-Gram dan Continuous Bag of words atau yang biasa disebut CBOW. Pada penelitian ini menggunakan model *Word2Vec* Skip-Gram karena merupakan sebuah metode yang efisien untuk mempelajari sejumlah besar representasi vektor kata yang ada pada teks yang tidak terstruktur seperti review film.



Gambar 2. Arsitektur Model CBOW dan Skip-Gram [9]

D. Pembagian Data

Data pada penelitian ini dibagi menjadi *data train* dan *data test. Data train* dibagi sebesar 70% dan *data test* dibagi sebesar 30%. Selain membagi data menjadi 30% *data test* dan 70% *data train*, metode pembagian lain yang digunakan dalam skenario pengujian pada penelitian ini adalah *K-Fold*.

E. K-Fold

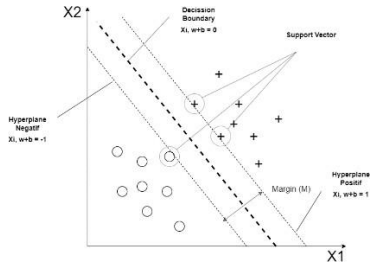
K-Fold Cross Validation merupakan salah satu dari jenis pengujian cross validation yang memiliki tujuan untuk menilai kinerja dari sebuah proses algoritma dalam segi akurasi performansi. K-Fold sendiri bekerja dengan cara membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai K pada k-fold. Pada penilitian ini, K pada K-Fold ditentukan sebanyak K = 5 Fold.

F. Under Sampling

Under sampling merupakan sebuah metode yang digunakan untuk menyesuaikan distribusi kelas dari suatu kumpulan kata agar data tidak mengalami *overfitting*. Menurut Google Developers, kelas data yang tidak seimbang dapat dibagi menjadi 3 kelas ketidakseimbangan yaitu ringan (20–40%), sedang (1–20%), dan ekstrim (<1%) [22]. Pada saat data berjumlah 45.739, terdapat kelas sentiment positif sebanyak 13.097 dan kelas sentiment negatif sebanyak 32.649 baris data dengan perbandingan kelas data sebesar 28:72.

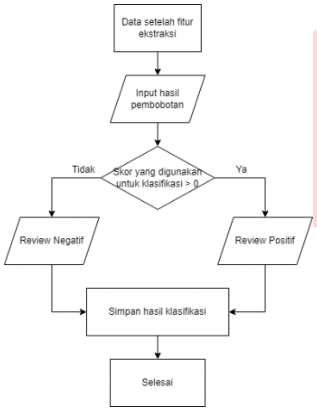
G. Support Vector Machine

Support Vector Machine merupakan salah satu metode dalam *supervised learning* yang biasa digunakan untuk klasifikasi dan regresi. Pada permodelan klasifikasi, SVM memiliki metode yang lebih optimal yang secara sistematis lebih jelas dibandingkan dengan teknik klasifikasi lainnya. SVM mencari hyperplane terbaik dengan cara memaksimalkan jarak antar kelas dan mengoptimalkan margin [14] Berdasarkan hal tersebut, berikut adalah representasi gambar pada metode SVM.



Gambar 3. Representasi Permodelan Metode SVM [20]

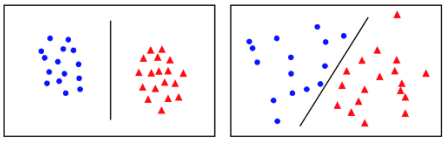
Berdasarkan gambar 3, ada beberapa tahapan dalam melakukan klasifikasi menggunakan metode SVM. Tahapan – tahapan tersebut adalah sebagai berikut:



Gambar 4. Representasi Permodelan Metode SVM

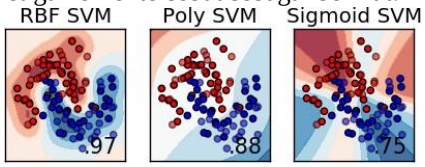
1. Kernel SVM (Support Vector Machine)

Pada metode klasifikasi SVM terdapat 4 kernel yang dapat digunakan diantaranya adalah kernel Linear, Polynomial, RBF dan Sigmoid. Kernel Linear merupakan sebuah kernel yang dapat digunakan ketika dataset memiliki persebaran kelas data yang dapat dipisahkan secara linear. Contoh persebaran kelas data secara linear dapat dilihat pada gambar berikut.



Gambar 5. Contoh Persebaran Data Secara Linear

Jika persebaran data tidak terjadi secara linear, maka dapat menggunakan 3 kernel yang lain yaitu polynomial, RBF dan sigmoid. Ilustrasi hyperplane yang terbentuk dari penggunaan ketiga kernel tersebut sebagai berikut.



Gambar 6. Ilustrasi Hyperplane yang Terbentuk pada Kernel RBF, Polynomial dan Sigmoid

H. Evaluasi

Evaluasi performansi yang akan dilakukan menggunakan metode *Confusion Matrix.* Akurasi setiap kelas dapat diperoleh dengan menghitung jumlah total opini positif dibagi dengan jumlah total data, jumlah total opini netral dibagi dengan jumlah total data dan jumlah total opini negatif dibagi dengan jumlah total data.

Table 6. Tabel Confusion Matrix [1]

	Prediksi Positif	Prediksi Negatif
Prediksi Positif	TP	FP
Prediksi Negatif	FN	TN

Berdasarkan tabel diatas, terdapat beberapa istilah yang digunakan diantaranya sebagai berikut :

True positive (TP) : Data positif yang terdeteksi dengan benar.
True negative (TN) : Data negatif yang terdeteksi dengan benar.
False positif (FP) : Data negatif yang terdeteksi sebagai data positif.
False negative (FN) : Data positif yang terdeteksi sebagai data negatif.

Persamaan yang digunakan untuk mengukur *precision*, *recall* dan *f1-score* adalah [1]

1. Precision

Precission atau dalam Bahasa Indonesia disebut presisi adalah ketepatan antara dua informasi yang diminta terhadap jawaban yang diberikan oleh sistem [1].

$$\text{Precission} = \frac{TP}{TP+FP} \quad (1)$$

2. Recall

Recall merupakan tingkat keberhasilan sistem dalam menemukan sebuah jawaban [1]. Berdasarkan penelitian [1].

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

3. F1-Score

F1-Score merupakan perbandingan dari rata – rata antara presisi dan recall. Berdasarkan penelitian [1].

$$\text{F1} - \text{Score} = 2 \times \frac{Recall \times Precission}{Recall+Precission} \quad (3)$$

1. Accuracy

Accuracy atau dalam Bahasa Indonesia disebut akurasi merupakan prediksi benar dari kelas positif dan negatif secara keseluruhan pada data [1].

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (4)$$

IV. HASIL DAN PEMBAHASAN

A. Skenario Pengujian

Pada evaluasi penelitian ini dilakukan perbandingan data hasil performansi terbaik dari beberapa skenario pengujian yang dibangun dengan menggunakan kernel RBF, Linear, Polynomial dan Sigmoid pada metode klasifikasi SVM yang dibangun. Ekstraksi fitur yang digunakan pada penelitian ini adalah Word2Vec. Penjelasan skenario pengujian secara detail dapat dilihat pada lampiran nomor 2.

B. Analisis Hasil Pengujian

4.2.1 Analisis Hasil Pengujian Skenario 1

Pada skenario pengujian 1 dilakukan pengujian untuk mengetahui performansi terbaik disetiap kernel pada metode klasifikasi SVM dari penggunaan teknik pengolahan data stemming pada dataset. Hasil dari pengujian skenario 1 pada penelitian dapat dilihat pada tabel berikut.

Menggunakan Stemming dengan kernel RBF									
Precission			Recall			F1-Score			Accuracy
Fresh	Rotten	Avg	Fresh	Rotten	Avg	Fresh	Rotten	Avg	
0,760	0,680	0,720	0,950	0,240	0,595	0,850	0,360	0,605	0,750
Menggunakan Stemming dengan kernel Linear									
Precission			Recall			F1-Score			Accuracy
Fresh	Rotten	Avg	Fresh	Rotten	Avg	Fresh	Rotten	Avg	
0,750	0,700	0,725	0,970	0,160	0,565	0,840	0,270	0,555	0,740
Menggunakan Stemming dengan kernel Polynomial									
Precission			Recall			F1-Score			Accuracy
Fresh	Rotten	Avg	Fresh	Rotten	Avg	Fresh	Rotten	Avg	
0,750	0,650	0,700	0,960	0,190	0,575	0,840	0,300	0,570	0,740
Menggunakan Stemming dengan kernel Sigmoid									
Precission			Recall			F1-Score			Accuracy
Fresh	Rotten	Avg	Fresh	Rotten	Avg	Fresh	Rotten	Avg	
0,760	0,380	0,570	0,760	0,380	0,570	0,760	0,380	0,570	0,650

Gambar 7. Hasil dari Skenario 1

Berdasarkan hasil pengujian menggunakan skenario 1, data yang diproses dengan menggunakan proses stemming memiliki akurasi, *precission*, *recall* dan *f1-score* yang beragam pada setiap kernel nya. Kernel RBF dapat memperoleh hasil terbaik karena RBF merupakan fungsi kernel yang digunakan ketika data tidak dapat terpisah secara linear [23]. Pada kernel RBF, diperoleh precision sebesar 72%, recall sebesar 59,5%, F1-Score sebesar 60,5% dan accuracy sebesar 75%. Dari hasil uji performansi yang didapat, running time yang diperlukan kernel RBF untuk memperoleh hasil performansi tersebut selama 9 menit lebih 10 detik.

4.2.2 Analisis Hasil Pengujian Skenario 2

Pada skenario pengujian ke – 2 dilakukan pengujian untuk mengetahui perbandingan penggunaan stemming dan tidak digunakanya stemming pada dataset. Berdasarkan pengujian sebelumnya diperoleh hasil yang lebih baik tanpa menggunakan stemming. Pengujian pada skenario ke – 2 ini bertujuan untuk membandingkan hasil performansi yang lebih baik dari penggunaan stemming pada dataset.

Tanpa Menggunakan Stemming dengan kernel RBF									
Precission			Recall			F1-Score			Accuracy
Fresh	Rotten	Avg	Fresh	Rotten	Avg	Fresh	Rotten	Avg	
0,810	0,710	0,760	0,930	0,430	0,680	0,860	0,540	0,700	0,790
Tanpa Menggunakan Stemming dengan kernel Linear									
Precission			Recall			F1-Score			Accuracy
Fresh	Rotten	Avg	Fresh	Rotten	Avg	Fresh	Rotten	Avg	
0,800	0,700	0,750	0,930	0,410	0,670	0,860	0,520	0,690	0,780
Tanpa Menggunakan Stemming dengan kernel Polynomial									
Precission			Recall			F1-Score			Accuracy
Fresh	Rotten	Avg	Fresh	Rotten	Avg	Fresh	Rotten	Avg	
0,790	0,710	0,750	0,940	0,350	0,875	0,860	0,470	0,665	0,770
Tanpa Menggunakan Stemming dengan kernel Sigmoid									
Precission			Recall			F1-Score			Accuracy
Fresh	Rotten	Avg	Fresh	Rotten	Avg	Fresh	Rotten	Avg	
0,780	0,450	0,615	0,780	0,440	0,610	0,780	0,440	0,610	0,690

Gambar 8. Hasil dari Skenario 2

Berdasarkan hasil percobaan pada skenario 2 yang terdapat pada tabel diatas, dapat disimpulkan bahwa tanpa penggunaan stemming pada preprocessing data dapat meningkatkan performansi yang diperoleh. Dapat dilihat pada tabel 11, pada percobaan skenario 2 ini kernel RBF masih mendapatkan hasil performansi yang paling tinggi dibandingkan dengan ketiga kernel lainnya. Kernel RBF memperoleh precission sebesar 76%, recall sebesar 68%, F1-Score sebesar 70% dan accuracy sebesar 79%. Dibandingkan dengan hasil uji performansi pada skenario 1, hasil tertinggi pada kernel RBF memperoleh precision sebesar 72%, recall sebesar 59,5%, F1-Score sebesar 60,5% dan accuracy sebesar 75%. Hal ini tentu berkaitan dengan tidak digunakannya stemming pada preprocessing data sehingga vektor yang dihasilkan tanpa menggunakan stemming lebih kompleks yang menyebabkan lebih tingginya akurasi pada percobaan skenario 2 ini.

4.2.3 Analisis Hasil Pengujian Skenario 3

Pengujian ini dilakukan berdasarkan hasil tertinggi yang didapat pada skenario pengujian sebelumnya. Pada pengujian ini, dataset tidak dibagi sebanyak 30% *data test* dan 70% *data train*, namun menggunakan K-fold dengan K sebanyak 5.

Menggunakan K-Fold										
Split Data	Precission			Recall			F1-Score			Accuracy
	Fresh	Rotten	Avg	Fresh	Rotten	Avg	Fresh	Rotten	Avg	
1	0,800	0,720	0,760	0,930	0,440	0,685	0,860	0,540	0,700	0,790
2	0,800	0,740	0,770	0,940	0,410	0,675	0,870	0,530	0,700	0,790
3	0,800	0,760	0,780	0,950	0,420	0,685	0,870	0,540	0,705	0,790
4	0,800	0,760	0,780	0,940	0,420	0,680	0,870	0,540	0,705	0,790
5	0,800	0,740	0,770	0,940	0,430	0,685	0,860	0,540	0,700	0,790
Rata - Rata			0,772	Rata - Rata		0,682	Rata - Rata		0,702	0,790

Gambar 9. Hasil dari Skenario 3

Percobaan pada skenario 3 ini, dataset akan langsung dilakukan pengujian menggunakan K-Fold dengan K sebanyak 5. Hasil dari pengujian ini diperoleh rata – rata precission sebesar 77,2%, recall 68,2%, F1-Score sebesar 70,2% dan accuracy sebesar 79%. Jika dibandingkan dengan pengujian pada skenario sebelumnya, dimana pengujian dilakukan tanpa menggunakan K-Fold dan tanpa menggunakan stemming, pada kernel RBF memperoleh precission sebesar 76%, recall sebesar 68%, F1-Score sebesar 70% dan accuracy sebesar 79%. Pada pengujian skenario 3 ini membutuhkan running time yang jauh lebih lama yaitu selama 49 menit 21 detik. Namun, jika dilihat dari hasil performansi tidak dapat dipungkiri bahwa k-fold memiliki hasil performansi yang sedikit lebih baik walaupun tidak signifikan. Hal ini karena K-Fold membagi data menjadi beberapa fold sebanyak K,

sehingga data train dan data test terbagi secara acak dan merata.

4.2.4 Analisis Hasil Pengujian Skenario 4

Pengujian skenario 4 ini bertujuan untuk mencari hasil performansi yang terbaik dari berbagai skenario yang ada dengan menggunakan undersampling. Undersampling sendiri digunakan karena pada dataset yang digunakan mulai tahun 2020 tidak seimbang atau *imbalanced*. Berikut adalah tabel hasil uji skenario 4.

Table 7. Hasil dari Skenario 4

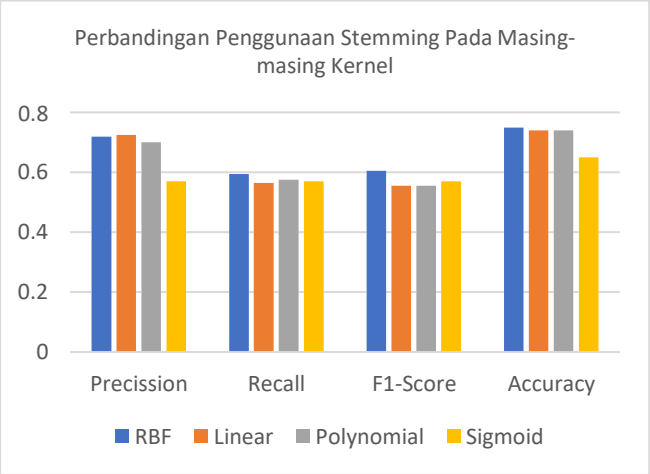
Menggunakan Undersampling									
Precission			Recall			F1-Score			Accuracy
Fresh	Rotten	Avg	Fresh	Rotten	Avg	Fresh	Rotten	Avg	
0,880	0,530	0,800	0,730	0,760	0,745	0,800	0,620	0,710	0,740

Berdasarkan tabel 7, dapat disimpulkan bahwa penggunaan undersampling membagi persebaran data menjadi sama rata disetiap kelasnya. Sehingga pada kelas positif dan negatif pada precission, recall dan F1-Score diperoleh hasil performansi yang tidak terpaut terlalu jauh. Sebelum dilakukan undersampling, perbandingan data set kelas rotten dan fresh sebesar 28:72. Setelah dilakukan undersampling, perbandingan dataset menjadi 50:50. Pada pengujian ini diperoleh precission sebesar 80%, recall sebesar 74,5%, F1-Score sebesar 71% dan accuracy sebesar 74%. Tentu pada pengujian skenario 4 ini mengalami penurunan dibandingkan ketika menggunakan K-Fold dimana diperoleh hasil terbaik yaitu rata – rata precission sebesar 77,2%, recall 68,2%, F1-Score sebesar 70,2% dan accuracy sebesar 79%.

C. Grafik Perbandingan Skenario Pengujian

Pada pengujian skenario program, diperoleh berbagai macam hasil yang berbeda dari masing – masing hasil pengujian skenario. Dari setiap pengujian skenario, dipilih hasil performansi terbaik dari skenario nya. Pada pengujian dari semua skenario program, terdapat nilai precission, recall, F1-Score dan Accuracy. Berikut adalah grafik perbandingan hasil skenario pengujian pada penelitian ini.

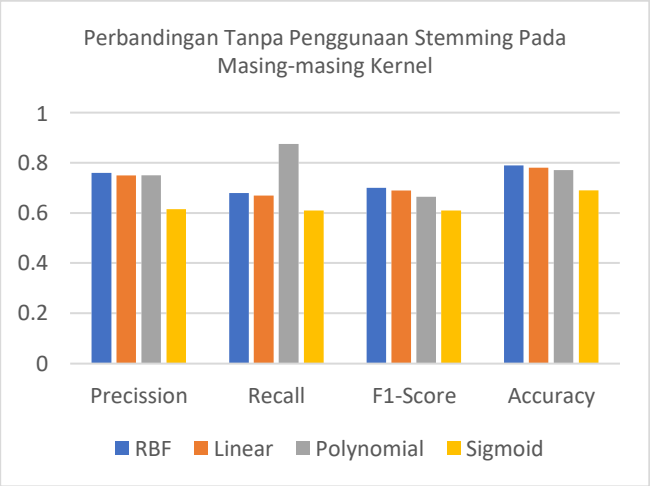
4.3.1 Grafik Perbandingan Skenario 1



Gambar 9. Perbandingan Penggunaan Stemming Pada Masing-masing Kernel

Berdasarkan Gambar 6 diatas, dapat dilihat bahwa pengujian program dengan menggunakan stemming pada penggunaan kernel RBF memperoleh hasil yang konsisten dan selalu memperoleh rata – rata performansi yang paling tinggi dibandingkan dengan kernel linear, polynomial dan sigmoid. Hal ini dikarenakan kernel RBF merupakan fungsi kernel yang digunakan ketika data tidak dapat terpisah secara linear [25].

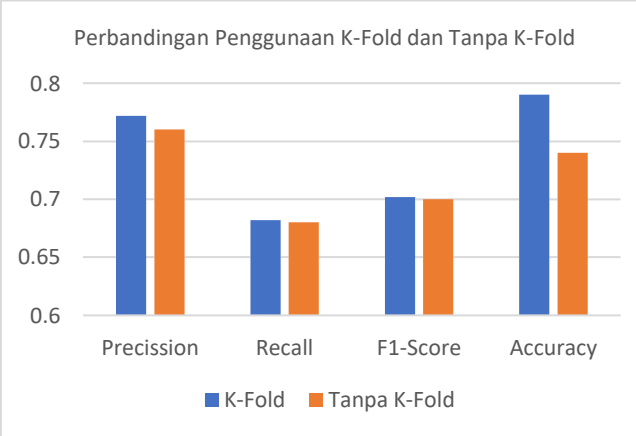
4.3.2 Grafik Perbandingan Skenario 2



Gambar 8. Perbandingan Tanpa Penggunaan Stemming Pada Masing-masing Kernel

Berdasarkan Gambar 8 diatas, dapat dilihat bahwa pengujian program tanpa menggunakan stemming pada penggunaan kernel RBF memperoleh hasil yang konsisten dan selalu memperoleh rata – rata performansi yang paling tinggi dibandingkan dengan kernel linear, polynomial dan sigmoid. Namun pada kernel Polynomial dapat dilihat bahwa recall mendapat performansi yang paling tinggi.

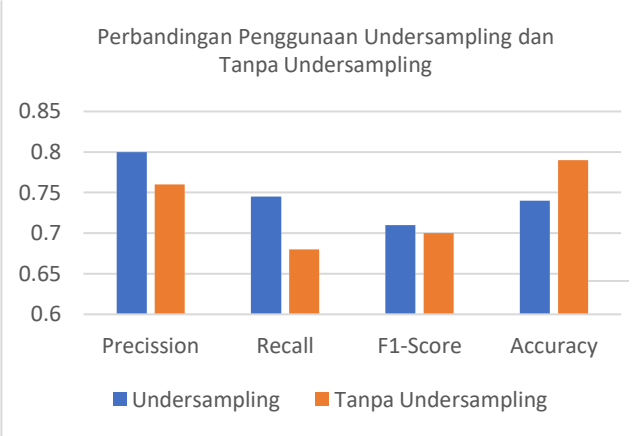
4.3.3 Grafik Perbandingan Menggunakan K-Fold dan Tanpa K-Fold



Gambar 9. Perbandingan Penggunaan K-Fold dan Tanpa Menggunakan K-Fold

Berdasarkan Gambar 8 diatas, dapat dilihat bahwa pengujian program dengan menggunakan *K-Fold* dengan K =5 mendapat hasil performansi yang lebih tinggi dibandingkan dengan pengujian tanpa menggunakan K-Fold. Hal ini terjadi karena setelah digunakan metode K-Fold sebagai metode pembagian data. Sehingga menghasilkan performansi yang lebih tinggi dibandingkan dengan hasil pengujian skenario tanpa menggunakan *K-Fold*.

4.3.4 Grafik Perbandingan Menggunakan Undersampling dan Tanpa Undersampling



Gambar 10. Perbandingan Penggunaan Undersampling dan Tanpa Menggunakan Undersampling

Berdasarkan Gambar 9 diatas, dapat dilihat bahwa pengujian program dengan menggunakan *Undersampling* mendapat hasil precission, recall dan f1-score yang lebih tinggi dibandingkan dengan pengujian tanpa menggunakan Undersampling. Hal ini terjadi karena setelah digunakan metode Undersampling, dataset yang tidak seimbang menjadi seimbang pada setiap kelasnya. Pada pengujian skenario ke – 4 dipeoleh hasil precission, recall dan f1-score yang rata pada kelas rotten dan fresh. Namun, untuk akurasi mengalami penurunan dibandingkan dengan tidak menggunakan undedrsampling dikarenakan dataset menjadi semakin sedikit.

V. KESIMPULAN

Berdasarkan penelitian yang dilakukan, diperoleh kesimpulan bahwa pengolahan data dan pemilihan Teknik pada proses *preprocessing* dan pemilihan kernel pada metode klasifikasi juga mempengaruhi hasil dari pengujian yang dilakukan. Skenario pada preprocessing yaitu dilakukan proses stemming pada data dan tanpa menggunakan proses stemming memberikan hasil yang berbeda. Pengujian skenario ini dilakukan dengan 4 kernel berbeda dan diperoleh hasil bahwa penggunaan kernel RBF menghasilkan hasil performansi yang terbaik pada pengujian skenario pertama dan kedua.

Pada uji performansi menggunakan tipe split data K-Fold dengan menggunakan kernel RBF diperoleh hasil performansi yang lebih besar daripada saat tidak menggunakan K-Fold sebagai proses split data. Hal ini dikarenakan pada saat menggunakan K-Fold data akan dipecah menjadi beberapa bagian sebanyak K yang ditetapkan. Pada skenario berikutnya dimana membandingkan hasil performansi dari sistem yang dibangun menggunakan undersampling. Pada penggunaan undersampling mempengaruhi performansi pada precission, recall dan f1-score karena data menjadi seimbang. Namun untuk accuracy dari performansi masih jauh lebih baik ketika menggunakan K-fold.

Saran untuk penelitian selanjutnya adalah proses labeling yang spesifik dan lebih akurat. Dapat menggabungkan beberapa fitur ekstraksi yang ada untuk menghasilkan performansi yang lebih baik dan juga pemilihan metode klasifikasi yang lebih tepat untuk dataset yang akan digunakan. Atribut *review_date* dinilai terlalu luas karena dimulai dari tahun 2010 hingga tahun 2020 akhir. Selain itu, dalam beberapa ulasan penonton terdapat kata yang membuat model semakin sulit untuk mengklasifikasi sentiment dari ulasan tersebut.

REFERENSI

[1] Suhariyanto, A. Firmanto and R. Sarno, "Prediction of Movie Sentiment Based on Reviews and Score on Rotten Tomatoes Using SentiWordnet," 2018 International Seminar on Application for Technology of Information and Communication, 2018, pp. 202-206, doi: 10.1109/ISEMANTIC.2018.8549704.

[2] Alfianti, Z., Gunawan, D., & Amin, A. (2020). Sentiment Analysis Of Cosmetic Review Using Naive Bayes And Support Vector Machine Method Based On Particle Swarm Optimization. Jurnal Riset Informatika, 2(3), 169-178. https://doi.org/10.34288/jri.v2i3.149.

[3] C. Nanda, M. Dua and G. Nanda, "Sentiment Analysis of Movie Reviews in Hindi Language Using Machine Learning," 2018 International Conference on Communication and Signal Processing (ICCSP), 2018, pp. 1069-1072, doi: 10.1109/ICCSP.2018.8524223.

[4] A. M. Rahat, A. Kahir and A. K. M. Masum, "Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset," 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), 2019, pp. 266-270, doi:

10.1109/SMART46866.2019.9117512.

[5] M. Al-Amin, M. S. Islam and S. Das Uzzal, "Sentiment analysis of Bengali comments with Word2Vec and sentiment information of words," 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE), 2017, pp. 186-190, doi: 10.1109/ECACE.2017.7912903.

[6] A. Alrehili and K. Albalawi, "Sentiment Analysis of Customer Reviews Using Ensemble Method," 2019 International Conference on Computer and Information Sciences (ICCIS), 2019, pp. 1-6, doi: 10.1109/ICCISci.2019.8716454.

[7] J. S. Lee, D. Zuba and Y. Pang, "Sentiment Analysis of Chinese Product Reviews using Gated Recurrent Unit," 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), 2019, pp. 173-181, doi: 10.1109/BigDataService.2019.00030.

[8] E. M. Alshari, A. Azman, S. Doraisamy, N. Mustapha and M. Alkeshr, "Effective Method for Sentiment Lexical Dictionary Enrichment Based on Word2Vec for Sentiment Analysis," *2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, 2018, pp. 1-5, doi: 10.1109/INFRKM.2018.8464775.

[9] Pan, Q., Dong, H., Wang, Y., Cai, Z., & Zhang, L. (2019). *Recommendation of Crowdsourcing Tasks Based on Word2vec Semantic Tags. Wireless Communications and Mobile Computing, 2019, 1–10.* doi:10.1155/2019/2121850

[10] Mikolov, Tomas, Sutskever, Ilya, Chen, Kai, Corrado, Greg S and Dean, Jeff. "Distributed Representations of Words and Phrases and their Compositionality." In *NIPS* , 3111--3119. : Curran Associates, Inc., 2013.

[11] M. Yasen and S. Tedmori, "Movies Reviews Sentiment Analysis and Classification," *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, 2019, pp. 860-865, doi: 10.1109/JEEIT.2019.8717422.

[12] T. Rahman, F. E. M. Agustin and N. F. Rozy, "Normalization of Unstructured Indonesian Tweet Text For Presidential Candidates Sentiment Analysis," 2019 7th International Conference on Cyber and IT Service Management (CITSM), 2019, pp. 1-6, doi: 10.1109/CITSM47753.2019.8965324.

[13] Parwita, W. G. S. (2020). *A document recommendation system of stemming and stopword removal impact: A web-based application. Journal of Physics: Conference Series, 1469, 012050.* doi:10.1088/1742-6596/1469/1/012050

[14] Sari, B., & Haranto, F. (2019). IMPLEMENTASI SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN TELKOM DAN BIZNET. *Jurnal Pilar Nusa Mandiri, 15*(2), 171-176. https://doi.org/10.33480/pilar.v15i2.699

[15] Santoso, Valonia & Virginia, Gloria & Lukito, Yuan. (2017). PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE. Jurnal Transformatika. 14. 72. 10.26623/transformatika.v14i2.439..

[16] Goutte, Cyril & Gaussier, Eric. (2005). A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. Lecture Notes in Computer Science. 3408. 345-359 10.1007/978-3-540-31865-1_25.

[17] A. Rane and A. Kumar, "Sentiment Classification System of Twitter Data for US Airline Service Analysis," *Proc. - Int. Comput. Softw. Appl. Conf.*, vol. 1, pp. 769–773, 2018, doi: 10.1109/COMPSAC.2018.00114.

[18] C. C. P. Hapsari, W. Astuti and M. D. Purbolaksono, "Naive Bayes Classifier and Word2Vec for Sentiment Analysis on Bahasa Indonesia Cosmetic Product Reviews," *2021 International Conference on Data Science and Its Applications (ICoDSA)*, 2021, pp. 22-27, doi: 10.1109/ICoDSA53588.2021.9617544.

[19] Guo, Z., & Jin, P. (2020). *SVM Method and Integrating Pinyin and Dictionary Method for Chinese Spelling Errors Detection. 2020 Prognostics and Health Management Conference (PHM-Besançon).* doi:10.1109/phm-besancon49106.2020.00062

[20] Hutapea, A., Furqon, M., & Indriati, I. Penerapan Algoritme Modified K-Nearest Neighbour Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia. **Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer**, vol. 2, no. 10, p. 3957-3961, peb. 2018. ISSN 2548-964X

[21] Tan, J. (2020, November 12). *How to deal with imbalanced data in python*. Medium. Retrieved January 26, 2022, from https://towardsdatascience.com/how-to-deal-with-imbalanced-data-in-python-f9b71aba53eb

[22] Widayani, W., & Harliana, H. (2021). Analisis support vector machine untuk Pemberian Rekomendasi penundaan Biaya Kuliah Mahasiswa. *Jurnal Sains Dan Informatika*, *7*(1), 20–27. https://doi.org/10.34128/jsi.v7i1.268

[23] Mubarok, M.S., Adiwijaya and Aldhi, M.D., 2017, August. Aspect-based sentiment analysis to review products using Naïve Bayes. In AIP Conference Proceedings (Vol. 1867, No. 1, p. 020060). AIP Publishing LLC.

[24] Pratiwi, A.I., Adiwijaya, 2018. On the feature selection and classification based on information gain for document sentiment analysis. Applied Computational Intelligence and Soft Computing, 2018.

[25] Daeli, N.O.F. and Adiwijaya, A., 2020. Sentiment analysis on movie reviews using Information gain and K-nearest neighbor. Journal of Data Science and Its Applications, 3(1), pp.1-7.

LAMPIRAN

Lampiran dapat dilihat pada link berikut.

- https://docs.google.com/document/d/1jx1069ryY_EJ8by-pY9Ot981FVOLmrUA/edit?usp=sharing&ouid=105001137508719416403&rtpof=true&sd=true