View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Asosiasi Profesi Multimedia Indonesia (APMI): E-Journals

Komparasi Metode K-Nearest Neighbors dan Support Vector Machine Pada Sentiment Analysis Review Kamera

Rizki Aulianita
Program Studi Sistem Informasi, STMIK Nusa Mandiri
rizki.rzk@nusamandiri.ac.id

Abstract *- Sentiment analysis is becoming one of the research growing trend, especially in text classification. In this study, the authors use as a camera review dataset by comparing two methods of KNN and SVM. Each method trials conducted so as to produce Accuracy KNN = 79.00% and the AUC of 0.929. While the data processing method SVM its accuracy is 72.00% and the AUC of 0845. Based on these results, proving that the rate of KNN text classification using more accurate than the method of SVM*
Keywords: Sentiment Analysis, Review, KNN, SVM, Text Classification

Abstrak – Sentiment Analisis menjadi salah satu trend riset yang semakin berkembang, khususnya dalam klasifikasi teks. Pada penelitian ini, penulis menggunakan review kamera sebagai dataset dengan membandingkan dua metode yaitu KNN dan SVM. Masing-masing metode dilakukan uji coba sehingga menghasilkan Akurasi KNN= 79.00% dan AUC sebesar 0.929. Sedangkan hasil pengolahan data metode SVM akurasi-nya adalah 72.00% dan AUC sebesar 0.845. Berdasarkan hasil penelitian tersebut, membuktikan bahwa tingkat klasifikasi teks menggunakan KNN lebih akurat dibandingkan dengan metode SVM.
Kata Kunci: Analisis Sentimen, Ulasan, KNN, SVM, Klasifikasi Teks

A. PENDAHULUAN

Perkembangan era digital saat ini, mempengaruhi gaya hidup seseorang sehingga dapat mengubah pola pikir, tingkah laku, trend dan *life style*. Termasuk fenomena selfie. Dengan melihat pangsa pasar tersebut, produk kamera berbondong-bondong merilis atau mengeluarkan kamera dengan berbagai spesifikasi dan merk. Mulai dari kamera Digital, DSLR, Semi Pro DSLR dan sebagainya. Untuk itulah diperlukan review yang baik apabila ingin membeli sebuah produk kamera. Data yang bertebaran di internet mengenai review kamera tidak dapat dijadikan keputusan untuk menentukan Customer dalam memilih kamera. Menurut Bagheri, Saraee, de Jong (2013) *Review Customers* sangatlah penting sebagai sumber informasi bisnis marketing dan intelligent dalam upaya memahami berbagai opini dari *Customers* dan membantu mereka dalam mengambil keputusan yang terbaik.

Analisis sentimen telah menjadi bidang penelitian utama sejak awal 2000-an. dampaknya dapat dilihat di banyak aplikasi praktis, mulai dari menganalisis ulasan produk (Stepanov & Riccardi, 2011) dalam Habernal et al untuk memprediksi penjualan dan stok market menggunakan pemantauan media sosial (Yu, Wu, Chang, & Chu, 2013). Analisis sentimen yang banyak diteliti yaitu analisis sentimen terhadap produk tertentu, *twitter brand* sentimen analisis (Ghiassi, Skinner, Zimbra, 2013), sentimen analisis terhadap *social media* (Habernal, Ptáček, Steinberger, 2015), sentimen analisis terhadap *mobile users* (Zhang et al, 2014), sentimen analisis terhadap tweet World Cup 2014 (Yu, Wang, 2015), sedangkan penulis akan melakukan penelitian terhadap sentimen analisis review kamera.

Machine Learning yang memperkenalkan klasifikasi teks seperti KNN, Centroid-Based Classifier, Naive Bayes, Decision Tree dan SVM (Tan, 2005). Naive Bayes. Eksperimental serta evaluasi menunjukan bahwa SVM, KNN dan NB merupakan tradisional teks klasifikasi. Eksperimen dan evaluasi menunjukan teks klafikasi yang valid (Yao, Min, 2012). Untuk itulah penelitian ini menggunakan metode K-Nearest Neighbor dan Support Vector Machine untuk klasifikasi teks.

SVM merupakan metode *supervised learning* yang menganalisa data dan mengenali pola-pola yang digunakan untuk klasifikasi (Basari et al., 2013). Support Vector Machines (SVM) adalah kasus khusus dari keluarga algoritma yang disebut sebagai *regularized* metode klasifikasi linier dan metode yang kuat untuk meminimalisasi resiko (Weiss, Indurkhya, & Zhang, 2010). SVM memiliki kelebihan yaitu mampu mengidentifikasi *hyperplane* terpisah yang memaksimalkan margin antara dua kelas yang berbeda (Chou et al., 2014). Namun *Support Vector Machine* memiliki kekurangan terhadap masalah pemilihan parameter atau fitur yang sesuai (Basari et al., 2013). Pemilihan fitur sekaligus penyetingan parameter di SVM secara signifikan mempengaruh hasil akurasi klasifikasi (Zhao, et al., 2011). Dalam masalah aplikasi tertentu, tidak semua fitur ini sama

pentingnya. Kinerja yang lebih baik dapat dicapai dengan membuang beberapa fitur. Dengan demikian, dapat dihilangkannya data yang *noise*, data yang tidak relevan dan berlebihan (Zhao et al., 2011).

K-Nearest Neighbor adalah metode sederhana namun efektif untuk teks kategorisasi tetapi memiliki 3 kelemahan yaitu kompleksitas pada sample yang komputasi kesamaan besar, *performance* KNN mudah dipengaruhi oleh sample tunggal, seperti *noisy sample* dan KNN tidak membangun model klasifikasi karena termasuk ke dalam *lazy learning method* (Jiang et al, 2012).

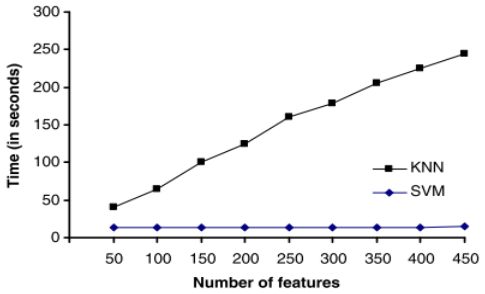
Berdasarkan penjelasan di atas, penelitian ini menggunakan metode k-NN dan SVM dalam mengklasifikasikan teks negatif dan teks positif terhadap review kamera.

B. TINJAUAN PUSTAKA

1. Penelitian Terkait

Beberapa penelitian terdahulu yang dijadikan rujukan terkait dengan topik penelitian.

Pada penelitian (Hmeidi et al., 2008) difokuskan untuk membandingkan dua metode *machine learning* pada kategori teks arab. Dataset yang digunakan dalam training dan testing terdiri dari review olahraga dan review ekonomi. Metode *machine learning* yang digunakan yaitu SVM dan k-NN dengan menggunakan TF.IDF sebagai metode optimasi untuk seleksi fitur. Berdasarkan kedua metode di atas, diperoleh hasil SVM memiliki performa yang lebih baik dibandingkan k-NN. Hal ini terjadi karena k-NN termasuk ke dalam kategori lazy algoritm yang tergantung pada statistik dann perbandingan dan memilih sejumlah besar fitur. SVM menggunakan *offline learning* untuk menemukan hyperplane yang optimal.



Gambar 1. Hasil Penelitian Hmeidi et al

(Basari et al, 2013) dalam penelitiannya dengan mengggunakan objek twitter seperti opini mengenai buku, movie, produk, politik dan sejenisnya. Metode yang digunakan adalah SVM yang merupakan metode supervised learning untuk menghasilkan klasifikasi. Klasifikasi disini terfokus pada dua class yaitu negatif klasifikasi dan positif klasifikasi. SVM ditingkatkan akurasinya dengan hybrid PSO dan menggunakan validasi 10 Fold-Cross Validation.

Tabel 1. Hasil pengujian Basari et al

	SVM	SVM-PSO
Accuracy	71.87	**77.00**
Precision	68.81	77.56
Recall	**81.87**	76.13

Peneliti (Jiang, 2012) dalam penelitiannya yang berjudul An improved K-nearest-neighbor algorithm for text categorization. Membahas mengenai teks kategori, dimana metode K-Nearest Neighbor merupakan metode klasifikasi efektif tetapi kurang efisien. Penulis mengusulkan algoritma KNN ditingkatkan untuk teks kategorisasi yang membangun klasifikasi model dengan menggabungkan algortima clustering dengan KNN teks kategori. Hasilnya diperoleh bahwa algoritma yang diusulkan mengungguli performa dari Naive Bayes, KNN dan Support Vector Machine.

Tabel 2. Hasil Pengujian Jiang

Categoryname	Train	Test	INNTC (K=45)			KNN (K=10) F1 value	NB F1 value	SVM F1 value
			Cluster vectors	Compression ratio	F1 value			
ACQ	1650	719	686	58.42%	0.9360	0.8998	0.9674	0.9608
Corn	181	56	41	77.35%	0.9126	0.8870	0.9444	0.8807
Crude	389	189	119	69.41%	0.7821	0.8235	0.8429	0.7930
Earn	2877	1087	915	68.20%	0.9643	0.9500	0.9843	0.9799
Interest	347	131	70	79.83%	0.9143	0.8945	0.9231	0.9243
Ship	197	89	98	50.25%	0.7364	0.6503	0.6154	0.6115
Trade	369	117	120	67.48%	0.8872	0.8943	0.8561	0.8739
	6010	2388	2049	65.91%	0.8850	0.8597	0.8843	0.8628

Berikut merupakan Tabel hasil penelitian terkait:

Tabel 3. Hasil Penelitian Terkait

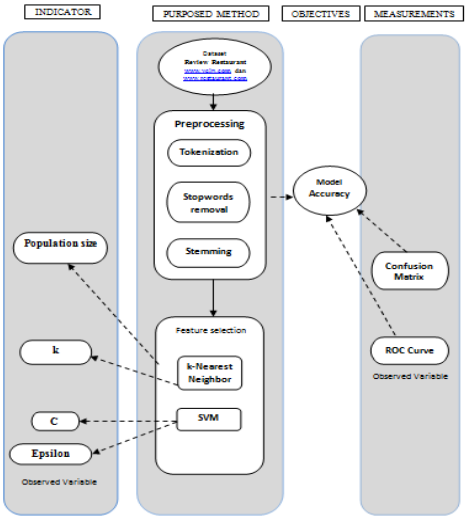
Judul	Peneliti	Classifier and Feature Selection	Akurasi	Hasil
Performance of KNN and SVM classifiers on full word Arabic articles	Hmeidi et al	SVM, KNN, TF+IDF	Recall KNN=0.96, SVM=1	Peneliti menghasilkan pengujian data dengan SVM +TFIDF memiliki akurasi tertinggi
An improved K-nearest-neighbor algorithm for text categorization	Jiang	KNN	83%	Peneliti mengusulkan algoritma INNTC yang memiliki akurasi tertinggi
Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization	Basari et al	Hybrid SVM PSO	77%	Peneliti menggunakan PSO sebagai optimasi dengan 10 Fold-Cross Validation

2. Sentiment Analysis

Menurut Tang dalam Haddi (Haddi, Liu, & Shi, 2013), analisa sentimen pada *review* adalah proses menyelidiki *review* produk di internet untuk menentukan opini atau perasaan terhadap suatu produk secara keseluruhan. Menurut Thelwall dalam Haddi (Haddi, Liu, & Shi, 2013), analisa

sentimen diperlakukan sebagai suatu tugas klasifikasi yang mengklasifikasikan orientasi suatu teks ke dalam positif atau negatif.

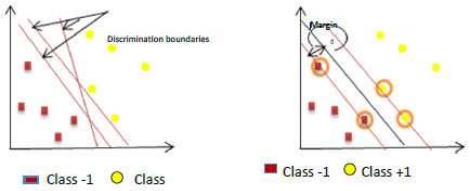
3. Kerangka Pemikiran



Sumber: Peneliti
Gambar 2. Kerangka Pemikiran

4. K-Nearest Neighbors
   Algoritma Nearest Neighbor melakukan klasifikasi berdasarkan kemiripan suatu data dengan data lain (Tan et al, 2005). Prinsip sederhana yang diadopsi oleh algortima NN adalah "Jika suatu hewan berjalan seperti bebek, maka hewan itu mungkin bebek". Semakin dekat lokasi data latih terhadap data uji, maka bisa dikatakan bahwa data latih tersebut yang lebih dipandang mirip oleh data uji. Semakin dekat maka akan semakin mirip, yang berarti juga semakin kecil jarak maka akan semakin mirip. Dengan kata lain, semakin kecil nilai ketidakmiripan (jarak) maka semakin miriplah data uji terhadap sejumlah k tetangga data latih terdekat. (Prasetyo, 2014).

5. Support Vector Machine
   Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah class pada input space. Untuk dimensional space, input data x (i=1. . .k), dimana milik kelas 1 atau kelas 2 dan label yang terkait menjadi -1 untuk kelas 1 dan +1 untuk kelas 2. Gambar di bawah ini memperlihatkan beberapa pattern yang merupakan anggota dari dua buah class: positif (dinotasikan dengan +1) dan negatif (dinotasikan dengan –1). *Pattern* yang tergabung pada class negatif disimbolkan dengan kotak, sedangkan *pattern* pada *class* positif, disimbolkan dengan lingkaran. Jika data input dapat dipisahkan secara linear, pemisahan *hyperplane* dapat diberikan dalam:



Sumber: (Nugroho, 2008)

Gambar 3. SVM berusaha menemukan *Hyperplane* terbaik yang memisahkan kedua class negatif dan positif

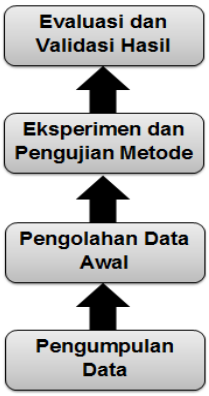
6. Evaluasi dan Validasi
   Ada banyak metode yang digunakan untuk memvalidasi suatu model berdasarkan data yang ada, seperti holdout, random sub-sampling, cross- validation, stratified sampling, bootstrap dan lain sebagainya. Menurut Han (Han & Kamber, 2007) confusion matrix adalah alat yang sangat berguna untuk menganalisa seberapa baik pengklasifikasi bias mengenali tuple dari class yang berbeda. Dalam confusion matrix dikenal beberapa istilah seperti True positive yang merujuk pada tuple positif yang secara benar dilabeli oleh pengklasifikasi, sementara True negative adalah tuple negatif yang secara benar dilabeli oleh pengklasifikasi. Adapula False positive yang merupakan tuple negatif yang secara tidak benar dilabeli oleh pengklasifikasi, dan False negative yang merupakan tuple positif yang secara tidak benar dilabeli oleh pengklasifikasi. K-fold cross validation yaitu memecah set data menjadi k bagian set data dengn ukuran yang sama.(Prasetyo, 2014). Kurva ROC akan digunakan untuk mengukur AUC (Area Under Curve). Kurva ROC membagi hasil positif dalam sumbu y dan hasil negative dalam sumbu x (Witten, Frank, & Hall, 2011).

   Berikut merupakan hasil klasifikasi:

   a. Akurasi 0.90 – 1.00 = *Excellent*
   b. Akurasi 0.80 – 0.90 = *Good*
   c. Akurasi 0.70 – 0.80 = *Fair*
   d. Akurasi 0.60 – 0.70 = *Poor*
   e. Akurasi 0.50 – 0.60 = *Failure*

C. METODE PENELITIAN

Metode eksperimen merupakan suatu penelitian yang kondisi-kondisi tertentu dikendalikan sehingga satu atau beberapa variabel dapat dikontrol untuk menguji hipotesis (Zikmund dalam Hermawan, 2005). Adapun metode penelitian yang penulis gunakan melalui beberapa tahapan, yaitu:



Sumber: Peneliti
Gambar 4. Metode Penelitian

1. Pengumpulan Data
   Data yang digunakan untuk melakukan eksperimen dikumpulkan melalui website *amazon.com* data kamera tersebut diseleksi dan dikumpulkan ke dalam notepad untuk diolah dalam pengujian data.
2. Pengolahan Data awal
   Teks yang belum diolah biasanya memiliki karakteristik dimensi yang tinggi, terdapat *noise* pada data dan terdapat struktur teks yang tidak baik. Untuk itu, dalam pengolahan data awal, teks mining harus melalui beberapa tahapan yang disebut dengan *preprocessing*
3. Eksperimen dan Pengujian Metode
   Memilih metode yang akan digunakan pada saat pengujian data. Metode yang dipilih, berdasarkan penelitian yang terdahulu. Penulis menggunakan Metode Algoritma k-Nearest Neighbor dan SVM. Eksperimen yang dilakukan peneliti, menggunakan framework RapidMiner 5.3 untuk mengolah data sehingga menghasilkan nilai akurasi yang akurat dan untuk pengujian metode.
4. Evaluasi dan Validasi Hasil Evaluasi
   Evaluasi berfungsi untuk mengetahui akurasi dari model algoritma yang diusulkan. Validasi digunakan untuk melihat perbandingan hasil akurasi dari model yang digunakan dengan hasil yang telah ada sebelumnya. Teknik validasi yang digunakan adalah *Cross Validation*. Akurasi algoritma akan diukur menggunakan *Confusion Matrix* dan hasil perhitungan akan ditampilkan dalam bentuk *Curve ROC (Receiver Operating Characteristic).*

D. HASIL DAN PEMBAHASAN

1. Pengumpulan Data
   Berdasarkan data yang peneliti kumpulkan melalui *www.amazon.com*. Dataset review kamera terdiri dari 100 review positif dan 100 review negatif
2. Pengolahan Data Awal
   a. *Tokenization*
      Proses memootong setiap kata dalam teks dan mengubah huruf dalam dokumen menjadi huruf kecil. Hanya huruf yang diterima, sedangkan karakter khusus atau tanda baca akan dihilangkan. Jadi hasil dari proses tokenization adalah kata-kata yang merupakan penyusun kalimat atau string yang dimasukan tanpa ada tanda baca.
   b. *Stopwords Removal*
      Penghapusan kata-kata yang tidak relevan seperti *the, on, of, with, and* dan sebagainya.
   c. *Stemming*
      Proses untuk menggabungkan atau memecahkan setiap varian-varian suatu kata menjadi kata dasar. Stem (akar kata) adalah bagian dari akar yang tersisa setelah dihilangkan imbuhannya (awalan dan akhiran).

Berikut akan disajikan Tabel Hasil Pengolahan Data Awal:
Tabel 4. Hasil Pengolahan Data Awal

Review Kamera	Hasil Tokenization	Stopwords Removal	Stemming
I was very disappointed with this camera. I bought this camera because my son was graduating from Marine Corp boot camp. Every picture taken has a pink strip at top of every one.....even with cropping or enlarging. Those once in a lifetime picures.....ruined!!! I even researched this camera extensively before buying, with no mentiin of pink strip in pictures.	I was very disappointed with this camera I bought this camera because my son was graduating from Marine Corp boot camp Every picture taken has a pink strip at top of every one even with cropping or enlarging Those once in a lifetime picures ruined I even researched this camera extensively before buying with no mentiin of pink strip in pictures	I disappointed camera I bought camera son graduating Marine Corp boot camp picture taken pink strip top cropping enlarging lifetime picures ruined I researched camera extensively buying mentiin pink strip pictures	I disappoint camera i bought camera son graduat marin corp boot camp pictur taken pink strip top crop enlarg lifetim picur ruin i research camera extens bui mentiin pink strip pictur
Continues to freeze up until the battery dies. I contacted customer service with a video of it not working and have not heard anything. When it works the product is great, but it does not work, it'll take 2 or 3 pictures then freeze up and is unable to use or even turn off. See attached video.	Continues to freeze up until the battery dies I contacted customer service with a video of it not working and have not heard anything When it works the product is great but it does not work it ll take or pictures then freeze up and is unable to use or even turn off See attached video	Continues freeze battery dies I contacted customer service video working heard works product great work ll take pictures freeze unable use turn See attached video	continu freez batteri di i contact custom servic video work heard work product great work ll take pictur freez unabl us turn see attach video

Sumber: Peneliti

Proses menentukan bagian kalimat sebagai anggota *class* positif atau *class* negatif berdasarkan nilai perhitungan naive bayes, jika hasilnya kalimat tersebut untuk *class* positif lebih besar dari pada *class* negatif, maka kalimat tersebut termasuk ke dalam class positif, jika probabilitas untuk *class* positif lebih kecil dari pada *class* negatif, maka kalimat tersebut termasuk ke dalam *class* negatif. Peneliti hanya menampilkan sample 2 dokumen dari keseluruhan 200 data training dan 3 kata yang berhubungan dengan sentimen yaitu:

Tabel 5. Hasil Klasifikasi dengan label class
Sumber: Peneliti

Dokumen	Accept	Disappoint	Class
KNN			
Negatif24.txt	0	1	Negatif
Positif97.txt	0	0	Positif
SVM			
Positif83.txt	0	0	Positif
Negatif26.txt	0	1	Negatif

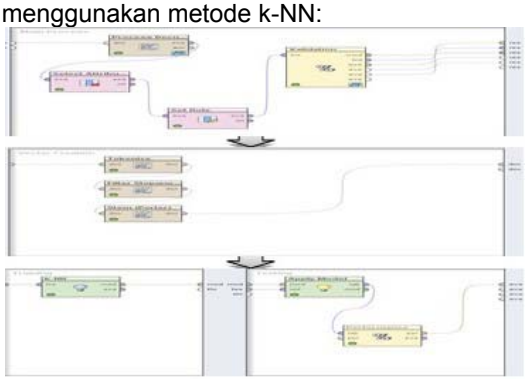
Sumber: Peneliti

3. Pengujian Data k-Nearest Neighbors

a. Pengujian Data metode k-NN

Metode k-Nearest Neighbor merupakan salah satu metode berbasis NN yang paling populer. Nilai k yang digunakan menyatakan jumlah tetangga terdekat yang dilibatkan dalam penentuan prediksi label kelas pada data uji. Untuk memperkirakan nilai k yang terbaik, bisa dilakukan dengan menggunakan teknik validasi silang *(Cross Validation)*. Jika nilai k terlalu kecil, maka berakibat hasil prediksi yang didapat bisa sensitif terhadap keberadaan *noise,* namun jika k terlalu besar maka tetangga terdekat yang terpilih mungkin terlalu banyak dari kelas lain yang sebenarnya tidak relevan karena jarak yang terlalu jauh. Berikut adalah hasil pengujian data dengan cara melakukan uji coba memasukan nilai k (jumlah tetangga terdekat) pada RapidMiner.

Berikut merupakan hasil pengolahan data menggunakan metode k-NN:



Gambar 5. Hasil Pengujian Data Menggunakan k-NN - Sumber: Peneliti

Berdasarkan pengujian yang terdiri dari 200 *sample* menghasilkan 4 spesial attribute, 1253 reguler attribute. Adapun hasil akurasinya akan ditampilkan pada tabel di bawah ini:

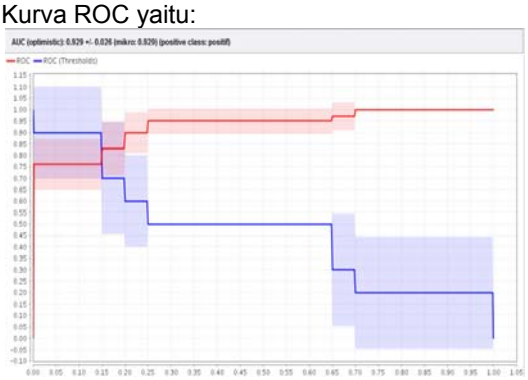
Tabel 6. Hasil Pengujian k-NN
Sumber: Peneliti

Nilai k	Hasil Pengujian Data KNN (Accuracy)
1	76.00%
2	**79.00%**
3	70.50%
4	69.00%
5	66.00%
6	71.50%
7	69.50%
8	71.50%
9	69.50%
10	70.50%

Hasil pengujian data di atas, menunjukan bahwa dengan memasukan uji coba nilai k=10, menghasilkan nilai **k=2** merupakan nilai dengan akurasi tertinggi yaitu **Accuracy= 79.00%**

Dalam penelitian ini, pengujian nilai k hanya sampai k=10, hal ini dikarenakan pengambilan nilai k yang tertinggi. Untuk k>10 membuat hasil akurasi menjadi turun. Untuk itulah peneliti hanya menguji data sampai k=10. Menurut Wu dan Kumar dalam Prasetyo (2014:151) "Pemilihan nilai k, jika K terlalu kecil maka berakibat hasil prediksi yang didapat bisa sensitif terhadap keberadaan noise dan sebaliknya jika nilai K terlalu besar maka tetangga terdekat yang terpilih mungkin terlalu banyak dari kelas lain yang sebenarnya tidak relevan karena jarak yang terlalu jauh". Hal ini telah dibuktikan dalam penelitian nilai k>10 maka hasil akurasinya ada di rentang 63.00% - 69.00%. Untuk itulah pada pengujian ini digunakan uji coba nilai k=10

Evaluasi dan Validasi hasil pengujian metode k-NN akan ditampilkan dalam bentuk Kurva ROC yaitu:



Sumber: Peneliti

Gambar 6. Kurva ROC Metode k-NN

Kurva ROC k-NN memiliki **AUC= 0.929** yang termasuk ke dalam kategori *Excellent Classification* dengan akurasi 79.00%.

Tabel 7. *Confusion Matrix* k-NN

C	Epsilon	SVM		*Population Size*
		Accuracy	AUC	
0.0	0.0	58.00%	0.826%	5
0.1	**0.1**	**72.00%**	**0.845%**	5
0.2	0.2	70.50%	0.847%	5
0.3	0.3	71.00%	0.835%	5
0.4	0.4	69.50%	0.832%	5
0.5	0.5	68.50%	0.830%	5
0.6	0.6	68.50%	0.834%	5
0.7	0.7	68.00%	0.829%	5
0.8	0.8	68.50%	0.828%	5
0.9	0.9	69.00%	0.835%	5
1.0	1.0	50.00%	0.500%	5

Sumber: Peneliti

Confusion Matrix yang dihasilkan dari pengujian data di atas yaitu:

$$\text{Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{85+73}{85+27+15+73} = \frac{158}{200} = 0.79$$

Data review positif yang termasuk ke dalam prediksi negatif terdapat 15 data kamera. Sedangkan data review negatif yang sesuai prediksi negatif terdapat 73 data.

b. Pengujian Data dengan metode SVM

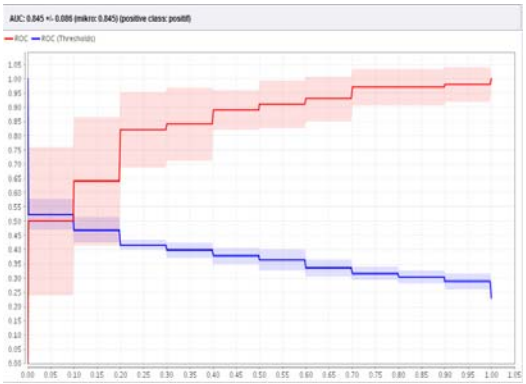
Pada SVM, dilakukan uji coba dengan memasukan nilai c dan epsilon. Berikut merupakan tabel hasil *training* SVM.

Tabel 8. Hasil Pengujian Data Metode SVM
Sumber: Peneliti

Accuracy :72.00% +/- 7.14% (mikro: 72.00%)			
	True Positif	**True Negatif**	**Class Precission**
Prediksi Positif	96	52	64.86%
Prediksi Negatif	4	48	92.31%
Class Recall	96.00%	48.00%	

Berdasarkan pengujian data yang telah dilakukan, maka diperoleh hasil SVM dengan **Accuracy=72.00%** dan **AUC=0.845** dengan cara memasukan nilai C=0.1 dan epsilon=0.1 dengan population size=5 yang merupakan hasil akurasi tertinggi.

Adapun Kurva ROC pada penelitian dengan metode SVM akan ditampilkan di bawah ini:



Sumber: Peneliti

Gambar 7. Kurva ROC Metode SVM

Kurva ROC pada metode SVM di atas memiliki **AUC=0.845** dan **Accuracy=72.00%** termasuk ke dalam klasifikasi *Good.* Untuk *Confusion Matrix* pada SVM dilakukan perhitungan dengan hasil sebagai berikut:

Tabel 9. *Confusion Matrix* SVM
Sumber: Peneliti

$$\text{Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{96+48}{96+52+4+48} = \frac{144}{200} = 0.72$$

Berdasarkan perhitungan pada Confusion Matrix, maka disimpulkan: data review positif yang termasuk prediksi negatif yaitu 4 data. Sedangkan data review negatif yang sesuai dengan prediksi negatif ada 48 data.

E. KESIMPULAN

1. Penelitian ini membandingkan dua metode yaitu k-NN dan SVM yang diimplementasikan pada sentiment *analysis review* kamera untuk mendapatkan hasil klasifikasi teks terbaik.
2. Kedua metode di atas dipilih berdasrkan tinjauan terdahulu dan ingin membuktikan bahwa metode k-NN merupakan metode dengan akurasi terbaik dibandingkan dengan SVM dengan berbagai kelebihan dan kekurangannya masing-masing.
3. Uji coba pada pengujian data training yang telah dilakukan mendapatkan hasil berupa ***Accuracy*** KNN= **79.00%** dan **AUC= 0.929,** sedangkan hasil Accuracy SVM= **72.00%** dan **AUC=0.845** pada review kamera.

DAFTAR PUSTAKA

[1] Bagheri, Saraee, de Jong - 2013 – Elsevier, *Care more about customers Unsupervised domain-independent aspect detection for sentiment analysis of customers review*

[2] Habernal, Ptáček, Steinberger - 2015 – Elsevier, *Reprint of "Supervised sentiment analysis in Czech social media"*

[3] Ghiassi, Skinner, Zimbra - 2013 – Elsevier, *Twitter brand sentiment analysis A hybrid system using n-gram analysis and dynamic artificial neural network*

[4] Zhang et al. - 2014 – Elsevier, *Sentiment Analysis on Reviews of Mobile Users*

[5] Yu, Wang - 2015 – Elsevier, *World Cup 2014 in the Twitter World A big data analysis of sentiments in U.S. sports fans' tweets*

[6] Tan. (2005), *Neighbor-weighted K-nearest neighbor for unbalanced text corpus*. *Expert Systems with Applications*, 28, 667-671.

[7] Yao, Zhi-Min. (2012), *An Optimized NBC Approach in Text Classification*. *Physics Procedia*, 24, 1910-1914

[8] *Basari et al. - 2013* – Elsevier, *Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization*

[9] *Zhang, Yoshida, Tang - 2011 - A comparative study of TFIDF, LSI and multi-words for text classification*

[10] Chou, J.-S., Cheng, M.-Y., Wu, Y.-W., & Pham, A.-D. (2014). Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification. *Expert Systems with Applications*, *41*(8), 3955–3964. doi:10.1016/j.eswa.2013.12.035

[11] Zhao, M., Fu, C., Ji, L., Tang, K., & Zhou, M. (2011). Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. *Expert Systems with Applications*, *38*(5), 5197–5204. doi:10.1016/j.eswa.2010.10.041

[12] Jiang et al. (2012), *An improved K-nearest-neighbor algorithm for text categorization*. *Expert Systems with Applications*, 39, 1503-1509.

[13] Hmeidi, Hawashin, El-Qawasmeh. (2008), *Performance of KNN and SVM classifiers on full word Arabic articles*. *Advanced Engineering Informatics*, 22, 106-111

[14] *Haddi, Liu, Shi - 2013 - The Role of Text Pre-processing in Sentiment Analysis*

[15] Prasetyo, Heri. (2014). Data Mining Mengolah Data Menjadi Informasi. Yogyakarta: Andi Offset.

[16] Han, J., & Kamber, M. (2007). *Data Mining Concepts and Techniques*. San Francisco: Diane Cerra