

29

**ANALISIS SENTIMEN ARTIKEL BERITA TOKOH SEPAK BOLA DUNIA
MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN
NAIVE BAYES BERBASIS PARTICLE SWARM OPTIMIZATION**

Fuad Nur Hasan, Mochamad Wahyudi

**Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK
Nusa Mandiri)**

(Naskah diterima: 12 Agustus 2018, disetujui: 12 Oktober 2018)

Abstract

Information about the actual news that occurs every day, or what happens every minute that can now be easily obtained such as general online news sites containing various actual information, as well as news sites that have special rubrics, for example news about politics, economics, education , entertainment, sports and so on. by using sentiment analysis by classifying documents with text mining. The algorithm used in this study is Naive Bayes and Support Vector Machine based on Particle Swarm Optimization. The results obtained from testing NB, NB (PSO), SVM and SVM (PSO) data will be compared. SVM (PSO) accuracy has a higher accuracy compared to SVM, NB and NB (PSO). So it can be concluded that the best optimization application in this model is that Support Vector Machine based on Particle Swarm Optimization (PSO) can provide a solution to classification problems in the case of sentiment analysis of world football figures Lionel Messi.

Keywords: *Sentiment Analysis, Mining Text, Classification, Naive Bayes, Support Vector Machine, Particle Swarm Optimization*

Abstrak

Informasi tentang berita aktual yang terjadi setiap hari, atau apa yang terjadi setiap menit yang kini dapat dengan mudah diperoleh seperti situs berita online yang bersifat umum berisi berbagai informasi aktual, serta situs berita yang memiliki rubrik khusus, untuk misal berita tentang politik, ekonomi, pendidikan, hiburan, olahraga dan sebagainya. dengan menggunakan analisis sentimen dengan mengklasifikasikan dokumen dengan penambahan teks. Algoritma yang digunakan dalam penelitian ini adalah Naive Bayes dan Support Vector Machine berdasarkan Particle Swarm Optimization. Hasil yang diperoleh dari pengujian NB, NB (PSO), SVM dan SVM (PSO) data akan dibandingkan. Akurasi SVM (PSO) mendapat akurasi lebih tinggi dibandingkan dengan SVM, NB dan NB (PSO). Jadi dapat disimpulkan bahwa aplikasi optimasi terbaik dalam model ini adalah Support Vector Machine berdasarkan Particle Swarm Optimization (PSO) dapat memberikan solusi untuk masalah klasifikasi dalam kasus analisis sentimen dari tokoh sepakbola dunia Lionel Messi.

Kata kunci: Analisis Sentimen, Pertambangan Teks, Klasifikasi, Naive Bayes, Mesin Support Vector, Particle Swarm Optimization

I. PENDAHULUAN

Informasi mengenai berita-berita aktual yang terjadi setiap hari, atau yang terjadi setiap menit yang saat ini bisa dengan mudah didapatkan seperti situs berita online yang sifatnya umum memuat berbagai informasi teraktual, maupun situs berita yang menampilkan rubrik secara khusus, misal berita tentang politik, ekonomi, pendidikan, hiburan, olahraga dan lain sebagainya. Hal tersebut bisa didapatkan dengan membuka berbagai media online yang saat ini sangat beragam jenisnya (Zain, 2015). Berita yang disajikan biasanya dalam bentuk teks pada media digital yang dikelompokkan berdasarkan isi pembahasan dari masing-masing kategori berita (Widodo et.al, 2016).

Dewasa ini, di Indonesia terdapat banyak media online yang menyediakan berbagai macam berita bermunculan seperti www.detik.com, www.kompas.com, www.antaranews.com, www.liputan6.com, www.cnnindonesia.com dan sebagainya. Penggunaan situs berita online sekarang ini tidak hanya terbatas sebagai media untuk membaca

artikel berita saja, tetapi juga dapat digunakan untuk melihat isu-isu yang sedang terjadi.

Hal tersebut dimungkinkan dengan menggunakan analisis sentimen. Analisis sentimen merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat (Buntoro, 2017). Pengukuran sentimen pada informasi open source saat ini merupakan area penelitian yang aktif (Dhande et.al, 2014).

Penerapan metode machine learning digunakan untuk mengklasifikasi polaritas suatu berita dari sumber data yang sangat banyak. Untuk melakukan hal itu, bisa menggunakan salah satu fungsi dari text mining, dalam hal ini adalah klasifikasi dokumen (Nurhuda et al, 2011). Text mining merupakan penerapan konsep dari teknik data mining untuk mencari pola dalam teks, bertujuan untuk mencari informasi yang bermanfaat dengan tujuan tertentu (Anjani,

2015). Text mining dapat diolah untuk berbagai macam keperluan diantaranya adalah untuk summarization, pencarian dokumen teks dan sentimen analisis (Maulana et.al, 2016).

Pada penelitian yang sudah dilakukan mengenai analisis sentimen, terdapat penelitian tentang analisis sentimen terhadap Presiden Jokowi. Pada penelitian tersebut, menggunakan search techniques dalam pengambilan data yaitu Boolean searching dengan operator “AND”. Data yang sudah didapat dilabeli positif, netral dan negatif kemudian dikoreksi oleh ahli bahasa. Setelah itu dilakukan preprocessing baik itu mengubah kata tidak baku menjadi baku atau biasa disebut normalisasi menggunakan kamus dan mencari akar kata yaitu stemming dengan bantuan aplikasi Sastrawi Master. Selanjutnya dilakukan juga tokenisasi N-Gram, Unigram, Bigram, dan Trigram terhadap kalimat, kemudian menghilangkan kata-kata yang umum digunakan dan tidak mempunyai Informasi yang berharga pada suatu konteks atau biasa disebut stopword removal, dan mempertahankan emoticon karena emoticon merupakan simbol yang menunjukkan ekspresi seseorang ke dalam tulisan. Algoritma klasifikasi yang digunakan yaitu Naive Bayes Classifier dan Support Vector

Machine. Akurasi yang terbaik dalam penelitian ini adalah dengan dilakukan normalisasi dan stemming pada data sebesar 89,2655% menggunakan metode SVM, dan kemudian data yang dinormalisasi saja sebesar 88,7006% menggunakan metode SVM (Saputra et al, 2015).

Di dalam penelitian ini, akan dibahas tahapan yang dilalui untuk melakukan proses analisis sentimen terhadap artikel berita sepak bola dengan tokoh dunia Lionel Messi. Dimulai dari tahap preprocessing sampai tahap analisis sentimen dengan Naive Bayes Classifier dan Support Vector Machine berbasis Particle Swarm Optimization serta bagaimana mengukur kualitas hasil analisis menggunakan dari masing-masing algoritma klasifikasi. Particle Swarm Optimization (PSO) merupakan teknik optimasi yang digunakan untuk menerapkan dan memodifikasi beberapa parameter dan meningkatkan bobot atribut.

II. KAJIAN TEORI

2.1 Text Mining

Text mining merupakan area penelitian baru dan menarik dimana mencoba memecahkan permasalahan informasi yang berlebih dengan menggunakan teknik data mining, machine learning, Natural Language

Processing (NLP), Information Retrieval (IR), dan manajemen pengetahuan. Text mining melibatkan tahapan preprocessing koleksi dokumen seperti kategorisasi teks, ekstraksi informasi, ekstraksi istilah (Feldman dan Sanger, 2007).

Menurut Feldman dan Sanger bahwa tahap-tahap text mining secara umum adalah text preprocessing dan feature selection. Tahap text preprocessing adalah tahap awal dari text mining. Tahap ini mencakup semua rutinitas, dan proses untuk mempersiapkan data yang akan digunakan pada operasi knowledge *discovery* sistem *text mining*.

2.2 Analisis Sentimen

Analisis sentimen adalah sebuah proses untuk menentukan sentimen atau opini dari seseorang yang diwujudkan dalam bentuk teks dan bisa dikategorikan sebagai sentimen positif atau negatif (Hadna, 2016). Analisis sentimen mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan text mining yang bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu (Manalu, 2014).

Analisis sistem mempunyai tugas untuk mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen kemudian menentukan pendapat yang dikemukakan dalam dokumen tersebut apakah bersifat positif, negatif atau netral (Manalu, 2014).

2.3 Teknik Klasifikasi

Teknik klasifikasi adalah sebuah model dalam data mining dimana classifier dikonstruksi untuk memprediksi categorical label seperti “aman” atau “beresiko” untuk data aplikasi peminjaman uang, “ ya” atau “tidak untuk data marketing atau “*treatment A*”, “*treatment B*”, “*treatment C*” untuk data medis. Kategori tersebut dapat direpresentasikan dengan nilai yang sesuai dengan kebutuhannya (Vulandari, 2017). Klasifikasi merupakan tugas yang sama dengan *data mining*, dimana tujuan utama dari klasifikasi adalah prediksi label kelas.

2.4 Naïve Bayes Classifier

Klasifikasi *Naïve Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Menurut Wu dan Kumar bahwa *Naïve Bayes* merupakan metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik dalam data mining. *Naïve Bayes* menggunakan cabang matematika yang

dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada *data training* (Mukminin & Riana, 2017).

Metode NBC menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses analisis terhadap sampel dokumen berupa pemilihan *vocabulary* dan selanjutnya melakukan penentuan probabilitas bagi tiap kategori berdasarkan sampel dokumen. Pada tahap klasifikasi ditentukan nilai kategori dari suatu dokumen berdasarkan term yang muncul dalam dokumen yang diklasifikasi (Hamzah, 2012).

2.5 Support Vector Machine (SVM)

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep-konsep unggulan dalam bidang *pattern recognition*. SVM adalah metode learning machine yang bekerja atas prinsip *Structural Risk Minimization (SRM)* dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space* (Nugroho et.al, 2003).

Berikut ini merupakan kekuatan dari *Support Vector Machine (SVM)* antara lain (Suyanto, 2017):

- a. Mempunyai kemampuan generalisasi yang tinggi.
- b. Mampu menghasilkan model klasifikasi yang baik meskipun dilatih dengan himpunan data yang relatif sedikit hanya dengan pengaturan parameter yang sederhana. SVM memiliki konsep dan formulasi yang jelas dengan sedikit parameter yang harus diatur.
- c. Relatif mudah diimplementasikan karena penentuan SVM dapat dirumuskan dalam masalah QP (*Quadratic Programming*).

2.6 Particle Swarm Optimization

Menurut Basari bahwa *Particle Swarm Optimization (PSO)* merupakan teknik optimasi yang sangat sederhana untuk menerapkan dan memodifikasi beberapa parameter. Dalam PSO terdapat beberapa teknik untuk melakukan pengoptimasian diantaranya meningkatkan bobot atribut (*attribute weight*) terhadap semua atribut atau variabel yang dipakai, menseleksi atribut (*attribute selection*) dan *feature selection*. (Indrayuni, 2016).

2.7 Model CRISP-DM

CRISP-DM (*CRoss-Industry Standard Process for Data mining*) merupakan suatu konsorsium perusahaan yang didirikan oleh Komisi Eropa pada tahun 1996. CRISP-DM adalah standarisasi *data mining* yang disusun oleh tiga pengagas *data mining market* yaitu Daimler Chrysler, SPSS, NCR (Budiman et al, 2012). CRISP-DM tidak menentukan standar atau karakteristik tertentu karena setiap data yang akan dianalisis akan diproses kembali pada fase-fase di dalamnya (Imtiyaz et al, 2015).

Proses *data mining* berdasarkan CRISP-DM terdiri dari enam fase sebagai berikut :

1. *Business Understanding*

Pada tahapan pertama ini harus didefinisikan apa pengetahuan yang ingin didapatkan dalam bentuk pertanyaan-pertanyaan yang sifatnya umum, misalnya bagaimana cara meningkatkan keuntungan, bagaimana cara mengantisipasi kesalahan cacat produk, dan sebagainya.

2. *Data Understanding*

Tahapan kedua ini bertujuan untuk mengumpulkan, mengidentifikasi, dan memahami aset data yang kita miliki. Data tersebut juga harus dapat diverifikasi kebenaran dan realibilitasnya.

3. *Data Preparation*

Tahapan ini meliputi banyak kegiatan, seperti membersihkan data, memformat ulang data, mengurangi jumlah data, dan sebagainya yang bertujuan untuk menyiapkan data agar konsisten sesuai format yang dibutuhkan.

4. *Modelling*

Model adalah representasi komputasi dari hasil pengamatan yang merupakan hasil dari pencarian dan identifikasi pola-pola yang terkandung pada data.

5. *Evaluation*

Evaluasi bertujuan untuk menentukan nilai kegunaan dari model yang telah berhasil kita buat pada langkah sebelumnya.

6. *Deployment*

Pada tahap ini, hasil yang diperoleh dari seluruh tahapan sebelumnya digunakan secara nyata.






2.8 Evaluasi dan Model Validasi

Confusion matrix ini berisi jumlah elemen yang telah dikelompokkan dengan benar atau tidak benar untuk setiap kelas. Salah satu manfaat dari *confusion matrix* adalah mudah untuk melihat sistem *confusion* dua kelas. Untuk setiap contoh di *test set*, akan membandingkan kelas yang sebenarnya dengan kelas *classifier*. Contoh positif (negatif) yang diklasifikasikan dengan benar

oleh *classifier* disebut *True Positive (true negative)*, contoh positif (negatif) yang salah diklasifikasikan adalah disebut *False Negative (false positive)* (Rokach & Maimon, 2015).

Area under curve (AUC) sebagai salah satu indikator untuk mengevaluasi kinerja dari *classifier*. AUC adalah area di bawah kurva ROC. AUC memiliki potensi secara signifikan untuk meningkatkan konvergensi secara empiris dari percobaan, karena itu memisahkan kinerja prediktif dari pengoperasian kondisi, dan merupakan ukuran umum prediktif. Selanjutnya, AUC memiliki statistik yang jelas interpretasinya.

Tabel 2.2 Tabel Nilai AUC

<i>AUC</i>	<i>Meaning</i>	<i>Symbol</i>
0.90 – 1.00	<i>excellent classification</i>	
0.80 – 0.90	<i>good classification</i>	
0.70 – 0.80	<i>fair classification</i>	
0.60 – 0.70	<i>poor classification</i>	
< 0.60	<i>Failure</i>	

Sumber : Gorunescu (Wahono,2014)

Kurva ROC (*Receiver Operation Characteristic*) Menurut Provost and Fawcett menjelaskan pengukuran lain menggunakan

kurva ROC yang menggambarkan *trade off* antara *true positive* terhadap *false positive*.

Kurva ROC di mana sumbu X mewakili tingkat *false positive* dan Y -axis merupakan tingkat *true positive*. Titik ideal pada ROC Kurva akan menjadi (0,100), Artinya, semua contoh positif diklasifikasikan dengan benar dan tidak ada contoh negatif yang salah klasifikasi sebagai positif.

III.METODE PENELITIAN

3.1 Pengumpulan Data

Data yang akan digunakan adalah data dari artikel berita olahraga mengenai bintang sepak bola dunia yaitu Lionel Messi. Data tersebut di peroleh dari tiga situs online berita www.news.detik.com, www.bolalob.com, dan www.bola.com dengan beberapa kata kunci pencarian (keyword) yaitu Messi, Lionel Messi, La Pulga, Barcelona, Argentina serta kata kunci lain yang berhubungan dengan Lionel Messi. Pada situs berita tersebut banyak terdapat artikel atau berita mengenai Lionel Messi sehingga penulis gunakan untuk mengklasifikasikan data berita positif dan data berita negatif. Data berita yang digunakan dalam penelitian ini sebanyak 200 data.

3.2 Pengolahan Data Awal

Pada tahap ini dilakukan klasifikasi teks atau sentimen dengan tahapan preprocessing

agar teks yang memiliki isi yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Data tersebut lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi. Untuk itu, dalam pengolahan data awal, text mining harus melalui beberapa tahapan yang disebut dengan preprocessing. Tahapan preprocessing yang dapat dilakukan yaitu: *Tokenize, Filter Tokens (By Length), Stopwords Removal, Transform Cases* (Utami,2017).

3.3 Metode Yang Diusulkan

Metode atau model yang diusulkan yaitu menggunakan dua algoritma *Naive Bayes, Support Vector Machine* dengan masing-masing menggunakan seleksi fitur Particle Swarm Optimization (PSO). Penulis melakukan perbandingan terhadap kedua metode tersebut NB dan SVM berbasis PSO. Selain itu metode yang diusulkan

IV.HASIL DAN PEMBAHASAN

4.1 Business Understanding

Pada tahapan business understanding, dilakukan pemahaman terhadap objek penelitian. Pemahaman mengenai objek penelitian dilakukan dengan menggali informasi melalui beberapa situs berita online terhadap tokoh sepak bola dunia Lionel Messi. Motivasi pada

fase ini yaitu berita yang disajikan biasanya dalam bentuk teks pada media digital yang dikelompokkan berdasarkan isi pembahasan dari masing-masing kategori berita. Rivalitas antar pemain sepak bola juga sering diamati melalui konten di media sosial. Analisis sentimen ini dilakukan untuk mencari metode klasifikasi yang dapat membantu dalam menentukan artikel berita positif dan negatif.

Pada tahap ini juga dilakukan pemahaman untuk mencari metode klasifikasi yang terbaik agar dapat membantu pada saat proses pengolahan data yang akan dilakukan dengan cara membandingkan hasil dari algoritma yang digunakan dan untuk meningkatkan performa dari metode klasifikasi dapat dilakukan dengan menggunakan feature selection.

4.2 Data Understanding

Pada tahap data understanding, dilakukan proses pengambilan data mentah sesuai dengan atribut yang dibutuhkan. Data diperoleh dari situs berita online yaitu www.detik.com, www.bola.com dan www.bolalob.com. Data yang diambil adalah data terbaru mengenai bintang sepak bola dunia Lionel Messi. Data primer yang diperoleh sebanyak 200 data artikel yaitu Detik Sport dengan 86 data, Bola dengan 58 data dan

Bolalob dengan 56 data. Dengan menggunakan sumber data yang diperoleh, dibuat dataset dengan atribut yaitu data berita yang berisi semua data-data berita tentang tokoh sepak bola dunia Lionel Messi, data dikelompokkan menjadi data positif sebanyak 100 data dan data negatif 100 data. Semua data berita tersebut dikelompokkan menjadi satu baik itu berita positif atau berita negatif dan disimpan dalam bentuk ekstensi .xlsx

4.3 Data Preparation

Tahap data preparation merupakan tahap dengan proses penyiapan data yang bertujuan untuk mendapatkan data yang bersih dan siap untuk digunakan dalam penelitian. Dalam *text mining* tahapan awal yang akan dilakukan adalah tahap *preprocessing*. Berikut merupakan tahapan yang dilakukan dalam *preprocessing*:

1. Tokenize

Dalam proses tokenisasi ini, semua kata yang ada di dalam tiap dokumen akan dikumpulkan dan dihilangkan tanda baca, serta dihilangkan juga simbol, karakter khusus ataupun bukan huruf.

2. Filter Token (By Length)

Dalam proses ini, kata-kata yang memiliki panjang kurang dari 4 atau lebih dari 25 akan dihapus, seperti kata yg, tdk, jd, ga, ane, gan

yang merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentimen.

3 Stopwords Removal

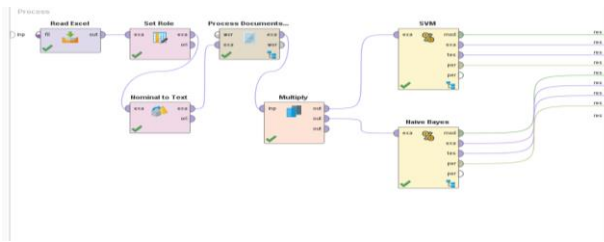
Pada tahap ini, operator yang digunakan adalah *filter stopword (dictionary)* karena dataset yang digunakan berbahasa Indonesia. Pada proses ini terlebih dahulu dibuat daftar kata-kata yang termasuk stopwords kemudian *file* nya akan diupload ke dalam operator *filter stopword (dictionary)*. Dalam tahap ini, kata-kata yang tidak relevan akan dihapus seperti kata tetapi, untuk, dengan, yang merupakan kata-kata yang tidak memiliki makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan sentimen.

4. Transform Cases

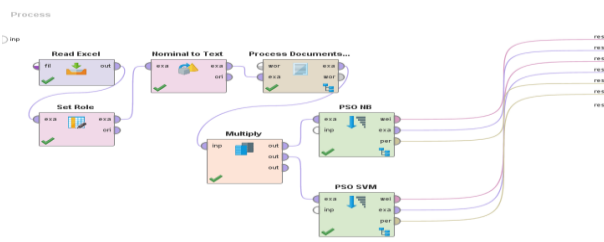
Dalam proses ini, kata-kata yang tidak relevan akan diubah, seperti kata yang mengandung huruf besar yang diubah menjadi huruf kecil sehingga dapat berhubungan dengan sentimen.

4.4 Modelling

Merupakan fase pemilihan teknik mining dengan menentukan algoritma yang akan digunakan. *Tool* yang digunakan adalah RapidMiner versi 7.3. Berikut adalah desain model Rapidminer yang digunakan yaitu :



Gambar 4.1 Desain Model Perbandingan Algoritma (SVM dan NB)

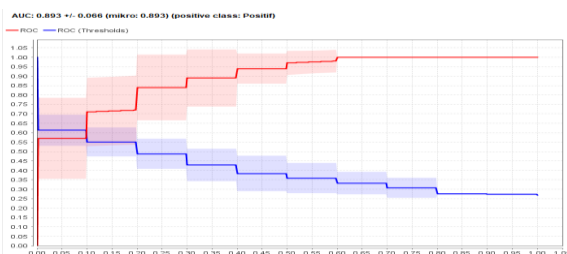


Gambar 4.2 Desain Model Perbandingan Algoritma (SVM PSO dan NB PSO)

4.5 Evaluation

Tahapan evaluasi bertujuan untuk menentukan nilai kegunaan dari model yang telah berhasil dibuat pada langkah sebelumnya. Untuk evaluasi digunakan *10-fold cross validation*. Dari hasil pemodelan yang telah dilakukan sebelumnya. Berikut ini akan dijelaskan Kurva ROC dan Confusion Matrix dari masing-masing algoritma :

1. Kurva ROC Support Vector Machine



Gambar 4.3 Kurva ROC SVM

Kurva ROC SVM dengan nilai AUC (Area Under Curve) yang dihasilkan dari gambar 4.3 diatas sebesar 0.893 dimana diagnosa hasilnya adalah *Excellent Classification*.

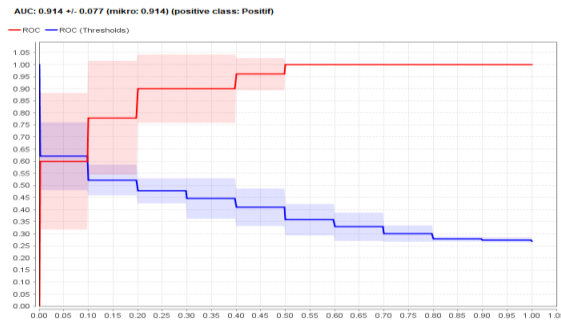
2. Confusion Matrix (Accuracy) SVM

accuracy: 78.50% +/- 10.97% (mikro: 78.50%)			
	true Positif	true Negatif	class precision
pred. Positif	80	23	77.67%
pred. Negatif	20	77	79.38%
class recall	80.00%	77.00%	

Tabel 4.1 Confusion Matrix SVM

Akurasi yang diperoleh yaitu 78.50% dari 100 data artikel berita positif dan 100 artikel berita negatif tentang tokoh sepak bola dunia Lionel Messi. Data berita positif yang sesuai prediksi yaitu 80 data. Data berita negatif yang termasuk ke dalam prediksi positif yaitu 23 data. Data berita positif yang termasuk kedalam prediksi negatif yaitu 20 data dan data berita negatif yang sesuai prediksi yaitu 77 data.

3. Kurva ROC Support Vector Machine (PSO)



Gambar 4.4 Kurva ROC SVM PSO

Kurva ROC *Support Vector Machine* berbasis *Particle Swarm Optimization (PSO)* dengan nilai AUC (Area Under Curve) yang dihasilkan dari gambar 4.4 diatas sebesar 0.914 dimana diagnosa hasilnya adalah *excellent classification*.

4. Confusion Matrix (Accuracy) SVM PSO

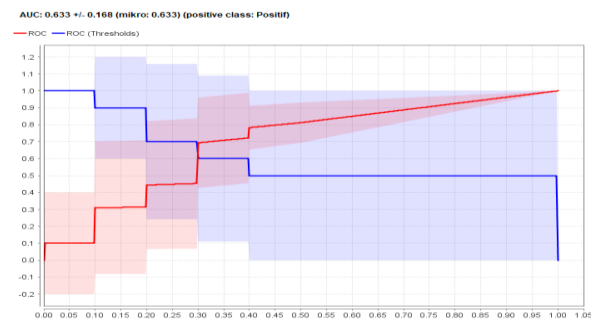
Tabel 4.2 Confusion Matrix SVM PSO

accuracy: 84.00% +/- 9.17% (mikro: 84.00%)			
	true Positif	true Negatif	class precision
pred. Positif	86	18	82.69%
pred. Negatif	14	82	85.42%
class recall	86.00%	82.00%	

Akurasi yang diperoleh yaitu 84.00% dari 100 data artikel berita positif dan 100 artikel berita negatif tentang tokoh sepak bola dunia Lionel Messi. Data berita positif yang sesuai prediksi yaitu 86 data. Data berita negatif yang termasuk ke dalam prediksi positif yaitu 18 data. Data berita positif yang termasuk kedalam prediksi negatif yaitu 14 data dan

data berita negatif yang sesuai prediksi yaitu 82 data.

5. Kurva ROC Naive Bayes



Gambar 4.5 Kurva ROC Naive Bayes

Kurva ROC NB dengan nilai AUC (Area Under Curve) yang dihasilkan dari gambar 4.5 diatas sebesar 0.633 dimana diagnosa hasilnya adalah *Failure*.

6. Confusion Matrix (Accuracy) NB

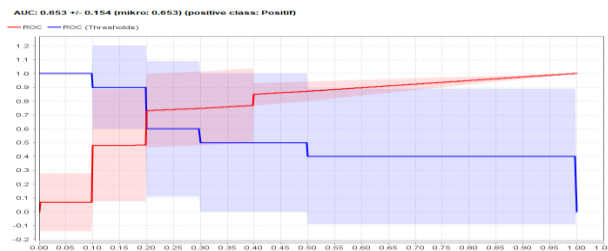
Tabel 4.3 Confusion Matrix NB

accuracy: 76.50% +/- 11.41% (mikro: 76.50%)			
	true Positif	true Negatif	class precision
pred. Positif	80	27	74.77%
pred. Negatif	20	73	78.49%
class recall	80.00%	73.00%	

Akurasi yang diperoleh yaitu 76.50 % dari 100 data artikel berita positif dan 100 artikel berita negatif tentang tokoh sepak bola dunia Lionel Messi. Data berita positif yang sesuai prediksi yaitu 80 data. Data berita

negatif yang termasuk ke dalam prediksi positif yaitu 27 data. Data berita positif yang termasuk kedalam prediksi negatif yaitu 20 data dan data berita negatif yang sesuai prediksi yaitu 73 data.

7. Kurva ROC *Naive Bayes* PSO



Gambar 4.6 Kurva ROC NB PSO

Kurva ROC *Naive Bayes Classifier* berbasis *Particle Swarm Optimization (PSO)* dengan nilai AUC (Area Under Curve) yang dihasilkan dari gambar 4.6 diatas sebesar 0.653 dimana diagnosa hasilnya adalah *Failure*.

8. Confusion Matrix (Accuracy) NB PSO

Tabel 4.4 Confusion Matrix NB PSO

accuracy: 83.00% +/- 6.78% (mikro: 83.00%)			
	true Positif	true Negatif	class precision
pred. Positif	85	19	81.73%
pred. Negatif	15	81	84.38%
class recall	85.00 %	81.00%	

Akurasi yang diperoleh yaitu 83.00% dari 100 data artikel berita positif dan 100 artikel berita negatif tentang tokoh sepak bola dunia

Lionel Messi. Data berita positif yang sesuai prediksi yaitu 85 data. Data berita negatif yang termasuk ke dalam prediksi positif yaitu 19 data. Data berita positif yang masuk ke dalam prediksi negatif yaitu 15 data dan data berita negatif yang sesuai prediksi yaitu 81 data.

Adapun perbandingan hasil komparasi akurasi dan AUC Algoritma telah digunakan sebagai berikut:

Tabel 4.5 Perbandingan Akurasi dan AUC

Algoritma	Accuracy	AUC
SVM	78.50%	0.893
SVM + PSO	84.00%	0.914
NB	76.50%	0.633
NB +PSO	83.00%	0.653

Dalam penelitian ini, hasil perhitungan metode SVM mendapatkan nilai akurasi 78.50% sedangkan hasil perhitungan SVM (PSO) mendapatkan nilai akurasi 84.00% artinya ada kenaikan tingkat akurasi sebesar 5.50% setelah menggunakan fitur seleksi PSO. Hasil perhitungan metode NB menghasilkan nilai akurasi 76.50%, sedangkan hasil perhitungan NB (PSO) mendapatkan nilai akurasi 83.00% artinya ada kenaikan tingkat akurasi sebesar 6.50% setelah menggunakan fitur seleksi PSO. Akurasi SVM dan NB memiliki perbedaan nilai sekitar 2.00%, sedangkan SVM (PSO) dan NB (PSO)

memiliki perbedaan nilai sekitar 1.00 %. Berdasarkan Tabel 4.8, dapat disimpulkan bahwa akurasi SVM berbasis PSO mendapatkan akurasi yang lebih tinggi dibandingkan dengan SVM, NB, NB (PSO) dan fitur seleksi PSO dapat meningkatkan nilai akurasi untuk metode SVM dan NB.

4.6 Deployment

Pada fase deployment akan dilakukan perancangan aplikasi dengan menggunakan dataset berita tokoh sepak bola dunia Lionel Messi.

V.KESIMPULAN

Hasil perhitungan metode SVM mendapatkan nilai akurasi 78.50% dan nilai AUC 0.893 sedangkan hasil perhitungan SVM (PSO) mendapatkan nilai akurasi 84.00% dan nilai AUC 0.914 artinya ada kenaikan tingkat akurasi sebesar 5.50% dan kenaikan nilai AUC sebesar 0.021 setelah menggunakan fitur seleksi PSO. Hasil perhitungan metode NB menghasilkan nilai akurasi 76.50% dan nilai AUC 0.633, sedangkan hasil perhitungan NB (PSO) mendapatkan nilai akurasi 83.00% dan nilai AUC 0.653 artinya ada kenaikan tingkat akurasi sebesar 6.50% dan kenaikan nilai AUC sebesar 0.020 setelah menggunakan fitur seleksi PSO. Akurasi SVM dan NB memiliki perbedaan nilai sekitar 2.00%, sedangkan

SVM (PSO) dan NB (PSO) memiliki perbedaan nilai sekitar 1.00 %. Hasil yang diperoleh dari pengujian data NB, NB (PSO), SVM dan SVM (PSO) akan dibandingkan. Akurasi SVM (PSO) mendapatkan akurasi yang lebih tinggi dibandingkan dengan SVM, NB dan NB (PSO). Sehingga dapat disimpulkan bahwa penerapan optimasi yang terbaik ada pada model ini adalah Support Vector Machine berbasis Particle Swarm Optimization (PSO) dapat memberikan solusi terhadap permasalahan klasifikasi pada kasus analisis sentimen tokoh sepak bola dunia Lionel Messi.

DAFTAR PUSTAKA

- Alindu, Paulina. 2013. *Twitter Used by Indonesian President: An Sentiment Analysis of Timeline*. Kupang : Information Systems International Conference (ISICO) 2-4 Desember 2013.
- Amrullah, Ahmad Afief, Ahmad Tantoni, Nahrowi Hamdani, Rahmat Taufik R.L.Bau, Muhammad Rafiqudin Ahsan dan Ema Utami. 2016. Review Atas Analisis Sentimen Pada Twitter Sebagai Representasi Opini Publik Terhadap Bakal Calon Pemimpin.
- Anjani, D. 2015. Bab II Landasan Teori Text Mining. <http://repository.widyatama.ac.id/xmlui/bitstream/handle/123456789/5867/Bab%202.pdf?sequence=9>.
- Yogyakarta : PROSIDING SEMINAR

NASIONAL MULTI DISIPLIN ILMU & CALL FOR PAPERS UNISBANK (SENDI_U) KE-2 Tahun 2016 Kajian Multi Disiplin Ilmu dalam Pengembangan IPTEKS untuk Mewujudkan Pembangunan Nasional Semesta Berencana (PNSB) sebagai Upaya Meningkatkan Daya Saing Global ISBN: 978-979-3649-96-2.

Basaria , Abd. Samad Hasan, Burairah Hussina, dan I. Gede Pramudya Anantaa,Junta Zeniarja. 2012. Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. Malaysia : Procedia Engineering 53 (2013) 453 – 462 Malaysian Technical Universities Conference on Engineering & Technology 2012, MUCET 2012 Part 4 Information And Communication Technology.

Budiman, Irwan, Toni Prahasto dan Yuni Christyono. 2012.Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma,ISSN : 1907-5022. Yogyakarta : Seminar Nasional Aplikasi dan Teknologi Informasi 2012 (SNATI 2012).

Buntoro, Asrofi Ghulam. 2017. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter. Jakarta : Integer Journal Vol 1 No 1 Maret 2016:32-41.

Chandani, Vinita, Romi Satria Wahono,dan Purwanto.2015. Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film, ISSN 2356-

3982.___. Journal of Intelligent Systems, Vol. 1, No. 1, February 2015.

Dhande, Lina L, dan Dr. Girish Patnaik. 2014. Review of Sentiment Analysis using Naive Bayes and Neural Network Classifier, ISSN :2319-8885. India : International Journal of Scientific Engineering and Technology Research Vol. 03 Issue.07 Mei 2014 Page 1110-1113.