

ANALISA KLASIFIKASI KOMENTAR MOTOGP PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAIVE BAYES

^{1*}Yuda Pratama Wibawa, ²Angga Pramadjaya, ³Andi Romansyah, ⁴Mugni

¹²³⁴Sistem Informasi, Universitas Pamulang

E-mail: ^{1*}dosen03045@unpam.ac.id.

ABSTRAK

Tingginya komentar mengenai *event* suatu perlombaan balap motor motoGP dalam sebuah media cetak maupun media elektronik, menjadikan peristiwa tersebut menjadikan pembicaraan banyak orang di dunia nyata maupun di dunia maya. Terlebih di era *digital* saat ini sangat mempermudah masyarakat ramai untuk mendapatkan suatu informasi yang mereka inginkan, baik melalui *website* atau melalui media *social* yang ada dan terkadang info dimuat secara *real time* pada saat itu juga komentar di tayangkan tentang trending topik yang ada di dunia maya. Akan hal tersebut paper ini akan melakukan klasifikasi seberapa akuratnya komentar mengenai motogp yang ada di dalam media *social* yang ada seperti twitter yang menjadi wadah masyarakat banyak untuk membicarakan tentang perlombaan balap motoGP tersebut. Dalam paper ini akan menerapkan dua algoritma klasifikasi untuk menguji seberapa akuratnya informasi atau komentar yang menjadi banyak pembicaraan orang melalui media *social* twitter. Pada paper ini akan menerapkan algoritma *Support Vector Machine* dan *Naive Bayes* dalam pengolahan *text mining*. Hasil dari nilai akurasi algoritma SVM adalah sebesar 95,50% sedangkan nilai akurasi NB adalah sebesar 90,00%.

Kata kunci : Analisa, Komentar, Twitter, *Support Vector Machine*, Naive Bayes

ABSTRACT

The high comments on the event of a MotoGP motorcycle racing competition in print media and electronic media, make the event a topic of conversation for many people in the real world and in cyberspace. Moreover, in the current digital era, it is very easy for the general public to get the information they want, either through websites or through existing social media and sometimes information is loaded in real time at that time comments are broadcast about trending topics in cyberspace. Regarding this, this paper will classify how accurate the comments about MotoGP are in existing social media such as Twitter, which is a place for many people to discuss the MotoGP race. In this paper, two classification algorithms will be applied to test how accurate the information or comments are that are widely discussed by people through Twitter social media. In this paper, the Support Vector Machine and Naive Bayes algorithms will be applied in text mining processing. The results of the SVM algorithm accuracy value are 95.50% while the NB accuracy value is 90.00%.

Keywords : Analysis, Commentary, Twitter, *Support Vector Machine*, Naive Bayes

PENDAHULUAN

Tingginya komentar mengenai *event* suatu perlombaan balap motor motoGP dalam sebuah media cetak maupun media elektronik, menjadikan peristiwa tersebut menjadikan pembicaraan banyak orang di dunia nyata maupun didunia maya. (Agustina et al, 2012) Terlebih di era *digital* saat ini sangat mempermudah masyarakat ramai untuk mendapatkan suatu informasi yang mereka inginkan, baik melalui *website* atau melalui media *social* yang ada dan terkadang info dimuat secara *real time* pada saat itu juga komentar di tayangkan tentang trending topik yang ada di dunia maya. (Amrullah et al, 2016) Keingin tahuan masyarakat mengenai info-info atau komentar yang beredar mengenai balap motoGP tersebut menjadikan perbincangan dibanyak media *social* yang ada sehingga menjadikan topik tersebut menjadi komentar populer di media *social* yang memposting tentang perlombaan balap motoGP tersebut. Akan hal tersebut paper ini akan melakukan penelitian seberapa akuratkah komentar mengenai motogp yang ada didalam media *social* yang ada seperti twitter yang menjadi wadah masyarakat banyak untuk membicarakan tentang perlombaan balap motoGP tersebut.(Wahono, 2014)

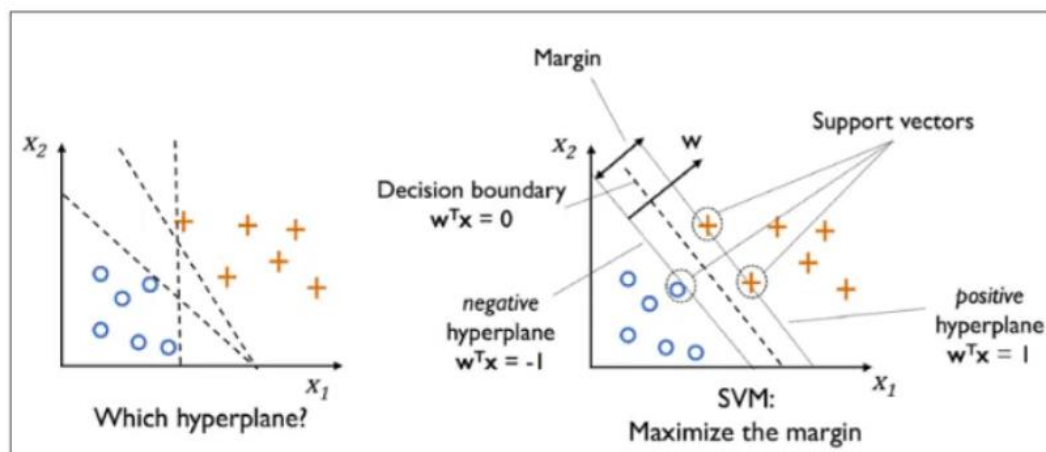
Twitter adalah layanan jejaring sosial dan *mikroblog* daring yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan. Twitter didirikan pada bulan Maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan Juli. Sejak diluncurkannya, twitter telah menjadi salah satu dari sepuluh situs yang paling sering dikunjungi di Internet, dan dijuluki dengan pesan singkat dari Internet. (Ibrahim, Danny 2017). Di twitter, pengguna tak terdaftar hanya bisa membaca kicauan, sedangkan pengguna terdaftar bisa menuliskan kicauan melalui antarmuka sistus web, pesan singkat (SMS), atau melalui berbagai aplikasi perangkat selular. (Aydin et al, 2011).

Bedasarkan permasalahan tersebut, maka penelitian ini akan menerapkan dua algoritma klasifikasi untuk menguji seberapa akuratkah informasi atau komentar yang menjadi banyak pembicaraan orang melalui media *social* twitter (Buntoro, et al, 2017). Pada paper ini akan menerapkan algoritma *Support Vector Machine* SVM dan *Navie Bayes* dalam pengolahan *text mining*, data yang diambil dari cuitan di twitter, yang nantinya akan menghasilkan tingkat akurasi benarkan isi

komentar yang ada di twitter tersebut berkaitan dengan balap motoGP atau tidak yang menjadi pembicaraan di sebuah media *social* dalam kasus ini adalah di twitter. (Hadna et al, 2016).

METODE SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) merupakan salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi (seperti Support Vector Classification) dan regresi (Support Vector Regression). Dalam pemodelan klasifikasi, SVM memiliki konsep yang lebih matang dan lebih jelas secara matematis dibandingkan dengan teknik-teknik klasifikasi lainnya. (Jehan, 2015). SVM juga dapat mengatasi masalah klasifikasi dan regresi dengan linear maupun non linear. SVM digunakan untuk mencari hyperplane terbaik dengan memaksimalkan jarak antar kelas. Hyperplane adalah sebuah fungsi yang dapat digunakan untuk pemisah antar kelas. Dalam 2-D fungsi yang digunakan untuk klasifikasi antar kelas disebut sebagai line whereas, fungsi yang digunakan untuk klasifikasi antar kelas dalam 3-D disebut plane similarly, sedangkan fungsi yang digunakan untuk klasifikasi di dalam ruang kelas dimensi yang lebih tinggi disebut hyperplane.



Gambar 1. Hyperlane memisahkan dua kelas positif dan negatif

Hyperplane yang ditemukan SVM diilustrasikan posisinya berada ditengah-tengah antara dua kelas, artinya jarak antara hyperplane dengan objek-objek data berbeda dengan kelas yang berdekatan (terluar) yang diberi tanda bulat kosong dan positif. Dalam SVM objek data terluar yang paling dekat dengan hyperplane disebut support vector. Objek yang disebut support vector paling sulit diklasifikasikan dikarenakan posisi yang hampir tumpang tindih (overlap) dengan kelas lain. Mengingat sifatnya yang kritis, hanya support vector inilah yang diperhitungkan untuk menemukan hyperplane yang paling optimal oleh SVM.

METODE NAIVE BAYES

Mengklasifikasi Naïve Bayes adalah algoritme machine learning terawasi yang digunakan untuk tugas klasifikasi seperti klasifikasi teks. Mereka menggunakan prinsip probabilitas untuk melakukan tugas klasifikasi. Naive Bayes adalah bagian dari keluarga algoritma pembelajaran generatif, yang berarti bahwa algoritma ini berusaha memodelkan distribusi input dari kelas atau kategori tertentu. Tidak seperti pengklasifikasi diskriminatif, seperti regresi logistik, pengklasifikasi ini tidak mempelajari fitur mana yang paling penting untuk membedakan antar kelas. Pengklasifikasi Naïve Bayes adalah algoritme machine learning terawasi yang digunakan untuk tugas klasifikasi seperti klasifikasi teks. Mereka menggunakan prinsip probabilitas untuk melakukan tugas klasifikasi. Naïve Bayes adalah bagian dari keluarga algoritma pembelajaran generatif, yang berarti bahwa algoritma ini berusaha memodelkan distribusi input dari kelas atau kategori tertentu. (Zubiaga, 2013) Tidak seperti pengklasifikasi diskriminatif, seperti regresi logistik, pengklasifikasi ini tidak mempelajari fitur mana yang paling penting untuk membedakan antar kelas. Adapun rumus teorema bayes yang menjadi dasar naïve bayes:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Keterangan :

X = sampel data yang memiliki class (label) yang tidak diketahui.

C = Hipotesis bahwa X adalah data class (label)

P(C) = Probabilitas hipotesis C.

1. P(X) = Peluang dari data sampel yang diamati (probabilitas C).
Menghitung nilai peluang kasus baru dari setiap hipotesa dengan *class* (label) yang ada di P(C_i).
2. Menghitung nilai akumulasi peluang dari setiap kelas P(X|C_i).
3. Menghitung nilai P(X|C_i) x P(C_i).
4. Menentukan *class* dari kasus baru tersebut.

Namun jika atribut ke-i bersifat kontinu, maka P(X_i|C) diestimasi dengan fungsi densitas Gaussian.

P(X|C) = Probabilitas berdasarkan kondisi pada hipotesis.

Adapun alur dari metode naive bayes sebagai berikut.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ = mean, σ = deviasi standar.

Rumus Teorema Bayes diatas tadi menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam class C (posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, sering kali disebut prior), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada class C (disebut juga likelihood) kemudian dibagi dengan peluang kemunculan karakteristik sampel secara global (disebut juga evidence). Oleh karena itu rumus dapat pula ditulis sebagai berikut.

$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}$$

Nilai evidence selalu tetap untuk setiap class pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai posterior class lainnya untuk menentukan class apa suatu sampel akan diklasifikasikan.

HASIL DAN PEMBAHASAN

1. Tahap Pengolahan Data

Pada tahap ini teks, semua kata yang ada pada tiap komentar dikumpulkan dan dihilangkan tanda baca, serta dihilangkan jika terdapat simbol, karakter khusus atau apapun yang bukan huruf.

Tabel 1.Perbandingan teks sebelum dan sesudah proses *Tokenize*

Teks sebelum dilakukan proses <i>tokenize</i>	Teks sesudah dilakukan proses <i>tokenize</i>
Akankah marcmarquez93 kembali berjaya di SepangGP MalaysianGP Saksikan live race di siang nanti pkl 13.00 Wita hanya di TRANS7	Akankah marcmarquez kembali berjaya di SepangGP MalaysianGP Saksikan live race di siang nanti pkl Wita hanya di TRANS

Sumber: Hasil Penelitian (2020)

Dalam proses ini, kata-kata yang memiliki panjang kurang dari 4 dan lebih dari 25 akan dihapus, seperti kata yang, tidak, j, gak, ane, gan, yang merupakan

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

Tabel 2. Perbandingan teks sebelum dan sesudah di *filter token (by length)*

Teks sebelum dilakukan proses <i>Filter Token (By Length)</i>	Teks sesudah dilakukan proses <i>Filter Token (By Length)</i>
Akankah marcmarquez93 kembali berjaya di SepangGP MalaysianGP Saksikan live race di siang nanti pkl 13.00 Wita hanya di TRANS7	Akankah marcmarquez93 kembali berjaya SepangGP MalaysianGP Saksikan live race siang nanti 13.00 Wita hanya TRANS7

Sumber: Hasil Penelitian (2020)

Dalam proses ini, *stopword removal* yang digunakan adalah operator *Filter Stopwords (Dictionary)* karena *dataset* yang digunakan berbahasa indonesia, yang sebelumnya penulis mengambil daftar kata-kata yang termasuk *stopword* kemudian *filenya* dimasukan dalam operator tersebut. Dalam proses ini, kata-kata yang tidak relevan akan dihapus, seperti kata tetapi, untuk, dengan, yang merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat.

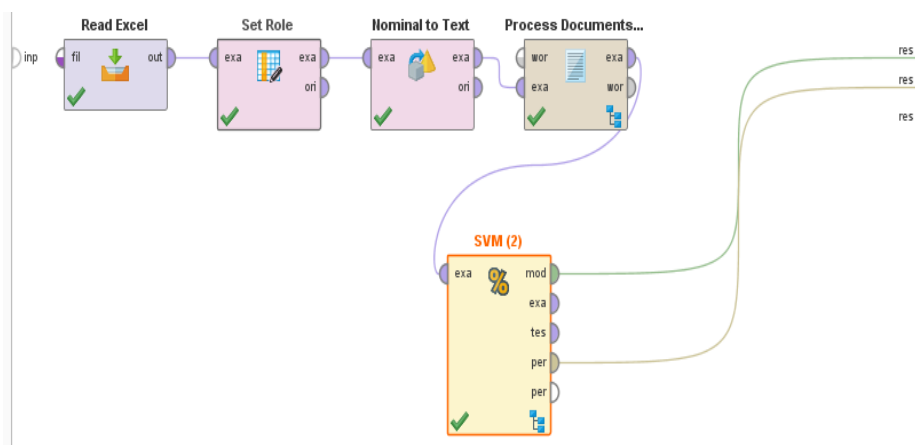
Tabel 3. Perbandingan sebelum dan sesudah di *Filter Stopword*

Teks sebelum dilakukan proses <i>Filter Stopword</i>	Teks sesudah dilakukan proses <i>Filter Stopword</i>
Akankah marcmarquez93 kembali berjaya di SepangGP MalaysianGP Saksikan live race di siang nanti pkl 13.00 Wita hanya di TRANS7	Akankah marcmarquez93 kembali berjaya di SepangGP MalaysianGP Saksikan live race di siang pkl 13.00 Wita hanya di TRANS7

Sumber: Penelitian (2020)

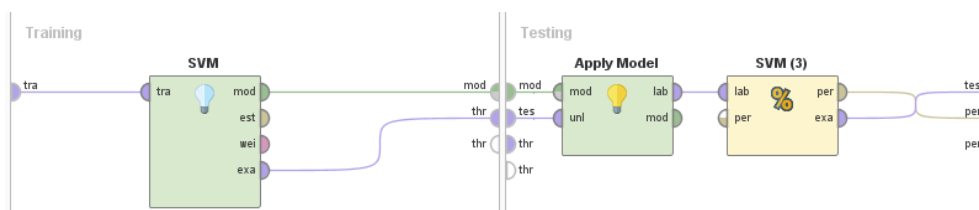
2. Pengujian Model Algoritma SVM

Pengaturan dan penggunaan operator serta parameter dalam *frameworks Rapid Miner* sangat berpengaruh terhadap akurasi dan model yang terbentuk, sebagai contoh dalam penggunaan model SVM dibawah ini:



Gambar 1. Desain Model Algoritma SVM

Model pengujian dari algoritma *support vector machine* (svm) menggunakan *rapidminer*, diawali dari memasukkan data kemudian mengatur set role yang nantinya menentukan label disana dan nominal text lalu keproses dokumen yang berisikan seperti gambar 4.1 setelah itu barulah masuk kemodel perhitungan *support vector machinenya* seperti gambar 4.3.

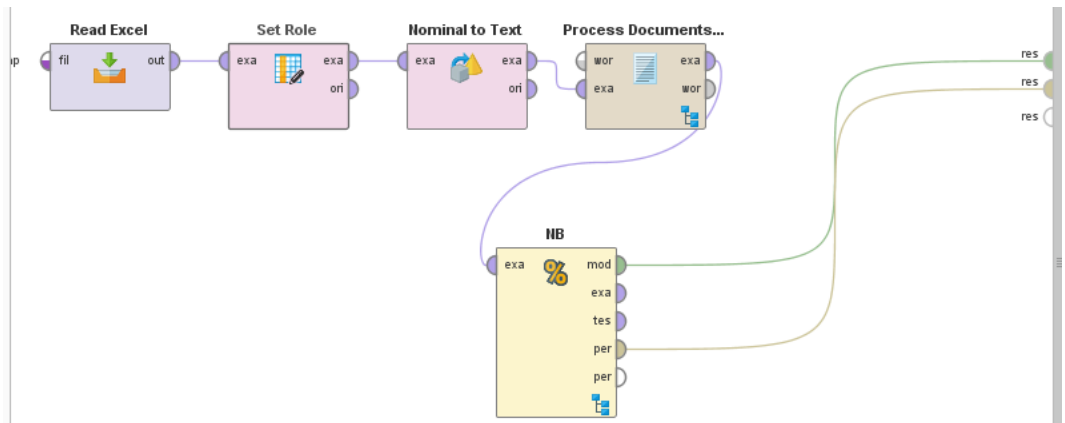


Gambar 2. Desain Proses 10-Fold Cross Validation untuk SVM

Desain proses di dalam operator *cross validation SVM* data digunakan adalah data bersih yang telah melalui *preprocessing*. Data tersebut diambil dari operator *Read Excel*, hal ini dilakukan karena dataset disimpan dalam bentuk Excel (.xlsx). *Process documents from files* untuk mengkonversi *files* menjadi dokumen. Process validasi terdiri dari *data training* dan *data testing*. Kemudian masuk kemodel algoritmanya *support vector machine* didalamnya ada perhitungan algoritmanya kemudian modelnya diapply setelah itu masuk kepenilaian *performancenya* barulah muncul hasil nilai *accuracy* dan *aucnya*.

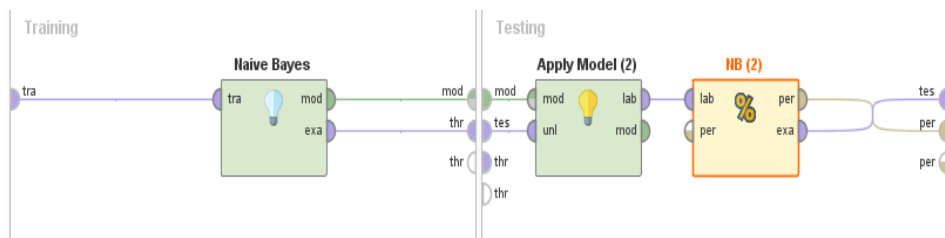
3. Pengujian Model dengan Algoritma Naive Bayes

Pengaturan dan penggunaan operator serta parameter dalam *frameworks Rapid Miner* sangat berpengaruh terhadap akurasi dan model yang terbentuk, sebagai contoh dalam penggunaan model SVM dibawah ini:



Gambar 3. Desain Model Algoritma NB

Model pengujian dari algoritma *naive bayes* (nb) menggunakan *rapidminer*, diawali dari memasukan data kemudian mengatur set role yang nantinya menentukan label disana dan nominal text lalu keproses dokumen yang berisikan seperti gambar 4.1 setelah itu barulah masuk kemodel perhitungan *naive bayes*nya seperti gambar 4.5.



Gambar 4. Proses 10-Fold Cross Validation NB

Data digunakan adalah data bersih yang telah melalui *preprocessing*. Data tersebut diambil dari operator *Read Excel*, hal ini dilakukan karena dataset disimpan dalam bentuk Excel (.xlsx). *Process documents from files* untuk mengkonversi *files* menjadi dokumen. Process validasi terdiri dari *data*

training dan *data testing*. Kemudian masuk kemodel algoritmanya *naive bayes* didalamnya ada perhitungan algoritmanya kemudian modelnya *diapply* setelah itu masuk kepenilaian *performancenya* barulah muncul hasil nilai *accuracy* dan *aucnya*.

4. Nilai Akurasi dari Algoritma SVM

Dari hasil pengujian model diatas dengan menggunakan algoritma SVM maka dapat menghasilkan sebuah nilai *Accuracy (Confusion Matrix)*:

Tabel 4. Nilai Accuracy Algoritma SVM

accuracy: 95.50% +/- 4.15% (mikro: 95.50%)

	<i>true positif</i>	<i>true negatif</i>	<i>class precision</i>
<i>pred. positif</i>	98	7	93.33%
<i>pred. negatif</i>	2	93	97.89%
<i>class recall</i>	98.00%	93.00%	

$$Acc (Accuracy) = \frac{TP + TN}{TP + TN + FP + FN} + \frac{98 + 93}{98 + 7 + 2 + 93} = \frac{191}{200} = 0,955$$

Jumlah *True Positif* (TP) adalah 98 *record* diklasifikasikan sebagai positif dan *Fakse Nigative* (FN) adalah 2 *record* diklasifikasikan sebagai negatif. Berikutnya 7 *False Positif* diklasifikasikan sebagai positif dan 93 *record True Negative* diklasifikasikan sebagai negatif. Berdasarkan tabel 4.1 diatas menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma SVM adalah sebesar 95,00%.

5. Nilai Accuracy dari Algoritma *Naives Bayes*

Dari hasil pengujian model diatas dengan menggunakan algoritma *Naives Bayes* maka dapat menghasilkan sebuah nilai *Accuracy (Confusion Matrix)*

Tabel 5. Nilai Akurasi Algoritma Naives Bayes

accuracy: 93.00% +/- 5.10% (mikro: 93.00%)

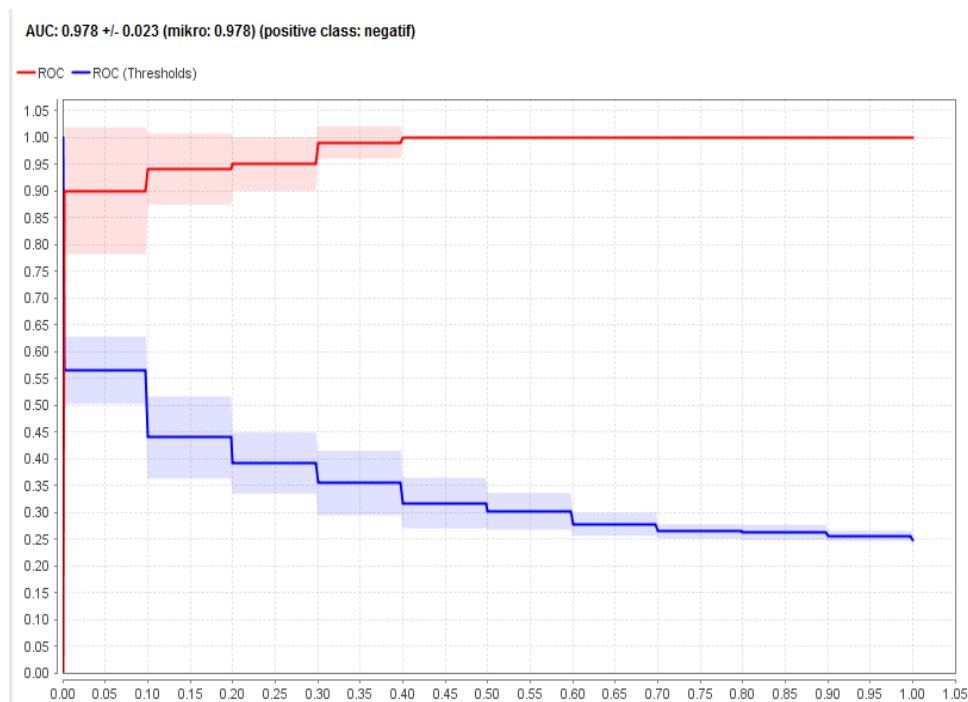
	true positif	true negatif	class precision
pred. positif	95	9	91.35%
pred. negatif	5	91	94.79%
class recall	95.00%	91.00%	

$$\text{Acc (Accuracy)} = \frac{TP + TN}{TP + TN + FP + FN} + \frac{95 + 91}{95 + 9 + 5 + 91} = \frac{186}{200} = 0,93$$

Jumlah *True Positif* (TP) adalah 95 *record* diklasifikasikan sebagai positif dan *Fakse Nigative* (FN) adalah 5 *record* diklasifikasikan sebagai negatif. Berikutnya 9 *False Positif* diklasifikasikan sebagai positif dan 91 *record True Negative* diklasifikasikan sebagai negatif menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma *Naives Bayes* adalah sebesar 93,00%.

6. Nilai AUC dari Algoritma SVM

Berikut ini akan dijelaskan Kurva ROC dan Confusion Matrix dari algoritma *Support Vector Machine*:



Gambar 5. Nilai AUC dalam Algoritma SVM

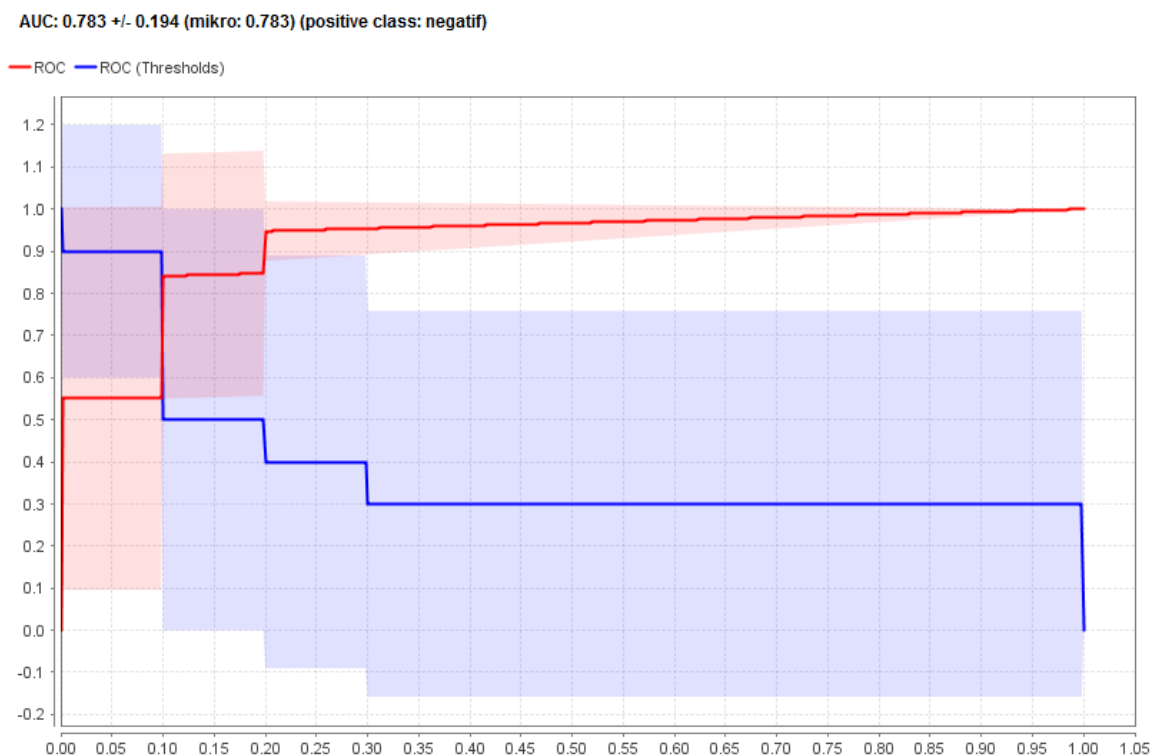
Kategori Klasifikasi AUC:

1. 0.90 - 1.00 = *excellent classification*
2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*

Dari hasil pengujian model yang telah dilakukan adalah untuk mendapatkan hasil akurasi dan *Area Under Curve (AUC)*. Maka mendapatkan hasil grafik ROC dengan nilai *Area Under Curve (AUC)* sebesar 0,978 dengan performance akurasi yaitu excellent.

7. Nilai AUC dari Algoritma *Naives Bayes*

Berikut ini akan dijelaskan Kurva ROC dan Confusion Matrix dari algoritma *Naives Bayes*:



Gambar 6. Nilai AUC dalam Algoritma *Naives Bayes*

Kategori Klasifikasi AUC:

1. 0.90 - 1.00 = *excellent classification*
2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*

Dari hasil pengujian model yang telah dilakukan adalah untuk mendapatkan hasil akurasi dan *Area Under Curve (AUC)*. Maka mendapatkan hasil grafik ROC dengan nilai *Area Under Curve (AUC)* sebesar 0,783 dengan performance akurasi yaitu good.

8. Perbandingan Akurasi

Berdasarkan hasil analisis dari masing-masing algoritma diatas, maka dapat dirangkum hasilnya:

Tabel 4. Perbandingan Performance Algoritma

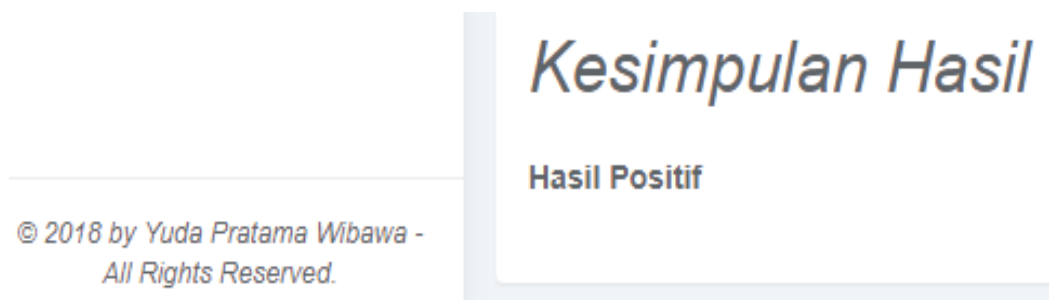
	SVM	<i>Naives Bayes</i>
Akurasi	95,50%	93,00%
AUC	0,978	0,783

Dari hasil perbandingan performance kedua algoritma diatas, maka hasil pengujian *Naives Bayes* nilai akurasinya lebih tinggi dibandingkan dengan algoritma SVM. Nilai akurasi untuk model SVM sebesar 95,50% dan nilai untuk model algoritma *Naives Bayes* sebesar 93,00% dengan selisih sebesar 2,5%.

9. Implementasi Website

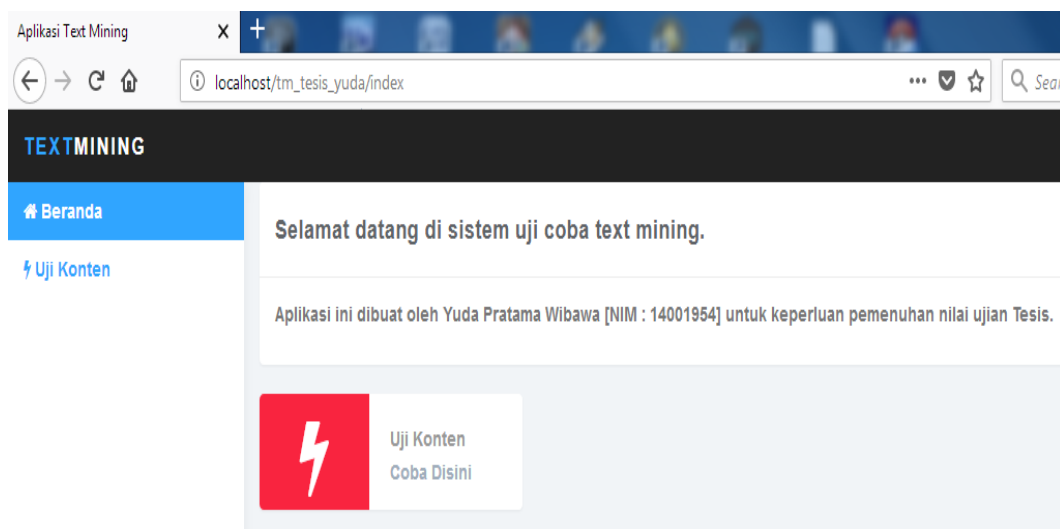
Gambar 7. Tampilan Website Text Mining

Berikut adalah hasil tampilan *text mining transform cases* dari hasil analisa teks sentimen komentar motogp dimalaysia pada tahun 2017 sebagai berikut:



Gambar 8. Tampilan Kesimpulan Hasil Komentar

Menjelaskan tentang kesimpulan dari komentar yang yang diinput. Halaman ini juga menjelaskan bahwa komentar yang di input dapat berupa komentar yang dianggap positif atau negatif



KESIMPULAN

Berdasarkan pembahasan tersebut, Adapun kesimpulan dari hasil penelitian ini adalah, nilai akurasi untuk menentukan bahwa komentar tersebut positif dan negatif, dapat dibuktikan dengan nilai akurasi dan nilai *AUC* dari masing-masing algoritma yaitu untuk SVM nilai akurasi = 95,50% dan nilai *AUC* = 0,978, sedangkan untuk algoritma *Naives Bayes* nilai akurasi = 93,00% dan nilai *AUC* = 0,783.

DAFTAR PUSTAKA

- Aprilla, D, Baskoro, Donny Aji, Ambarwati, Lia, & Wicaksana, IWayan Simri. (2013). *Belajar Data Mining dengan Rapid Miner*. Jakarta.
- Aggrawal, C., & Zhai, C. (2012). *Mining text data*. *Mining Text Data* (Vol. 4) <http://doi.org/10.1007/978-1-4614-3223-4>
- Agustina, P. A., Matulatan, T., Tech, M., Si, M. B. S., Sc, M., Informatika, J., ... Umrah, H. (n.d.). KLASIFIKASI TRENDING TOPIC TWITTER DENGAN PENERAPAN METODE NAÏVE BAYES (The Classification Of The Trending Topic Of Twitter s With Naïve Bayes Method).
- Amrullah, Ahmad Afief, Ahmad Tanton, 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.
- Aydin, I., Karakose, M., & Akin, E. (2011). *A multi-objective artificial immune algoritma for parameter optimization in support vector machine*. *Journal Applied Soft Computing*, 11, 120-129.
- Bellotti, T., & Crook, J. (2007). *Support vector machine for credit scoring and Jiscovery of significant features*. *Expert System with Application: An International Journal*, 36, 3302-3308.
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter, *1*(1), 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.
- Feldman R., Sanger James (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.
- Gerunescu, Florin (2011). *Data Mining: Concept, Model, And Techniques*. Verlag Scoring Berlin Heidelberg: Springer.
- Go, A., Huang, L., & Bhayani, R. (2009). *Twitter Sentiment Analysis. Final Project Report, Stanford University, Department of Computer Science*.
- Hadna, Nurrin Muchammad Shiddieqy, Paulus Insap Santosa dan Wing Wahyu Winarno. 2016. Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter, ISSN :2089-9815. Yogyakarta: Seminar Nasional Teknologi Informasi dan Komunikasi (Sentika 2016).
- Hamzah, Amir. 2012. Klasifikasi Teks Dengan *Naive Bayes Classifier (NBC)* Untuk Pengelompokan Text Berita dan Abstrak Akademis, ISSN:1979-911X. Yogyakarta : Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III.
- Han, Kamber, Pei (2012) *Data Mining Concept and Technique*. Morgan Kaufman Pubisher.
- Hermawati, Fajar Astuti. 2013. *Data mining*. Yogyakarta : Andi.
- Hermawati, Melayu S.P (2002). *Manajemen Sumber Daya Manusia*. Jakarta: Bumi Aksara.

- Ibrahim, Danny. 2017. Analisis Hubungan antar Faktor dan Komparasi Algoritma Klasifikasi pada Penentuan Penundaan Penerbangan, ISSN: 2579-9045 ISBN: 978-602-74355-1-3. Tegal : Seminar Nasional IPTEK Terapan (SENIT) 2017.
- Ipmawati, Joang, Kusrini dan Emha Taufiq Luthfi. 2016. Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen, ISSN :2461-0690. Jakarta : IJSE – Indonesian Journal on Software Engineering Vol. 2 No 2.
- Jehan. S. K. (2015). Perhitungan Analisis Sentimen Berbasis Komparasi Algoritma *Naive Bayes* dan *K-Nearest Neighbour* Berbasis *Particle Swarm Optimization* pada Komentar Insiden Pembalap MotoGP.
- Kusrini, dan Emha Taufiq Luthfi. 2009. Algoritma Data mining. Yogyakarta: Andi Offset.
- Kharde, V. A. (2016). Sentiment Analysis of Twitter Data : A Survey of Techniques, *139*(11), 5–15.
- Lee, K., Palsetia, D., Narayanan, R., Patwary, M. A., Agrawal, A., & Choudhary, A. (2011). *Twitter Trending Topic Classification*, 251–258. <https://doi.org/10.1109/ICDMW.2011.171>
- Manalu, Boy, U. (2014). ANALISIS SENTIMEN PADA TWITTER MENGGUNAKAN TEXT MINING SKRIPSI. *Mashape* (2013). *mashape.com. List of 20+ Sentiment Analysis APIs. Accessed on May 25, 2014 from http://blog.mashape.com/post/48757031167/list-of-20-sentiment-analysis-apis Miner.*
- Mustofa, Mufid. (2016). Pengembangan Sistem Pendukung Keputusan Penjurusan Bagi Siswa Baru Menggunakan Metode *Naive Bayes*. Polteknik Negeri Malang.
- North, Matthew (2012). *Data Mining for The Masses*. A Global Text Project Book.
- Nurhuda, Faishol, Sari Widya Sihwi dan Afrizal Doewes. 2013. Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode *Naive Bayes Classifier*, ISSN :2301-7201. Jakarta : Jurnal ITSMART Vol. 2 No. 2 Desember 2013.
- Pang, B., and Lee, L. (2008). Opinion mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.
- Prasetyo, Eko. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- Safari, Cecep Barkah. 2012. Skripsi Analisis Sentimen Menggunakan Metode Klasifikasi SVM. Jakarta : Universitas Pendidikan Indonesia.
- Sari, Eka Novita. (2013, Agustus). Analisa Algoritma Apriori Untuk Menentukan Merek Pakaian Yang Paling Diminati Pada Mode Fashion Group Medan. *Jurnal Pelita Informatika Budi Darma*, Vol IV, No. 3, pp 35-39.
- Subiyakto, A'ang. 2008. Penggunaan Algoritma Klasifikasi Dalam Data mining. Jakarta : Syarif Hidayatullah State Islamic University.
- Sugiyono. (2013). Metode Penelitian Kuantitatif, Kualitatif dan R&D. Bandung: Alfabeta.
- S, Umajancy & Antony, Selvadoss. T. 2013. *An Analysis On Text Mining-Text Retrieval And Text Extraction*. *International Journal of Advance Research in Computer and Communication Engineering* Vol. 2, Issue 8, August 2013. ISSN (Print) : 2319-5940. ISSN (Online) : 2278-1021.

- Vulandari, Retno Tri.2017. Data mining Teori dan Aplikasi Rapidminer. Yogyakarta : Gava Media.
- Wahono, Romi Satria,Nanna Suryana Herman, dan Sabrina Ahmad. 2014. A Comparison Framework of Classification Models for Software Defect Prediction. America : Adv. Sci. Lett. 20, 1945–1950, 2014.
- Weiss Sholom M., Indurkhya Nitin, Zhang Tong (2010). *Fundamentals of Predictive Text Mining*. Springer-Verlag London Limited.
- Witten, I. H. (2005). The Practical Handbook of Internet Computing. In M. P. Singh, The Practical Handbook of Internet Computing Chapter 14 (pp. 1-23). Danvers, MA: Chapman and Hall/CRC Press. XLSTAT.
- Zubiaga, A., Spina, D., & Mart, R. (2013). Real-Time Classification of Twitter Trends, 2013.