

Aplikasi Web Analisis Sentimen Dengan Algoritma Multinomial Naïve Bayes

Joko Winahyu¹, Imam Suharjo²
Program Studi Pendidikan Teknik Informatika
Universitas Mercu Buana Yogyakarta
Yogyakarta, Indonesia

jwinahyu6@gmail.com¹, imam@mercubuana-yogya.ac.id²

Abstract— Teknologi adalah keseluruhan sarana untuk menyediakan sebuah pelayanan yang diperlukan bagi kelangsungan dan kenyamanan hidup manusia salah satunya adalah aplikasi. Pada jejaring media sosial Twitter telah banyak pengguna yang menulis tweet tentang kaitannya dengan web aplikasi. Tahap dilakukan penambangan opini pengguna terkait kata kunci Aplikasi Web mengacu pada penggunaan pemrosesan bahasa alami, analisis teks untuk mengidentifikasi, mengekstrak, mengukur, dan mempelajari keadaan efektif dan informasi subjektif secara sistematis pada data opini di twitter. Dengan menggunakan metode text classification yang telah banyak secara umum digunakan adalah algoritma Multinomial Naive Bayes. Metode ini memanfaatkan teorema probabilitas yang mana teorema bayes dan fungsionalitas data mining yaitu klasifikasi naïve bayes. Penelitian ini berfokus pada opini pengguna jejaring sosial twitter dengan kata kunci aplikasi web data ini di ambil dari API twitter dengan jumlah item data 1000 selanjutnya klasifikasi opini pengguna dan preproses data. Dari hasil pengujian Dalam mengevaluasi performance algoritma kita menggunakan acuan Confusion Matrix dan merepresentasikan prediksi dan kondisi sebenarnya(aktual) dari data yang dihasilkan oleh algoritma kita bisa menentukan Accuracy 69%, Precision merupakan rasio jumlah yang benar diprediksi pada label kelas positif 0,67, netral 0, negatif 1 dan recall merupakan rasio jumlah yang diprediksi benar terhadap keseluruhan data yang benar yaitu kelas positif 1, netral 0, negatif 0,14 dan f1-score merupakan rata-rata dari precision dan recall yaitu kelas positif 0,81, netral 0, negatif 0,25.

Keywords: *sentimen analisis, naïve bayes multinomial, confusion matrix.*

I. PENDAHULUAN

Twitter adalah salah satu media sosial yang sering digunakan orang untuk mengekspresikan pendapatnya. Pendapat itu sendiri adalah salah satu sumber informasi yang

berupa teks yang dikategorikan menjadi opini. Opini adalah kalimat subjektif yang berisi persepsi seseorang mengenai suatu objek atau peristiwa. Teknologi adalah keseluruhan sarana untuk menyediakan sebuah pelayanan atau barang-barang yang diperlukan bagi kelangsungan dan kenyamanan hidup manusia salah satunya adalah aplikasi. Teknologi telah memengaruhi masyarakat dan sekelilingnya dalam banyak cara. Salah satu

penulis menunjukkan pada jejaring media sosial Twitter yang banyak sekali pengguna menulis tweet tentang kaitannya dengan web aplikasi. Twitter merupakan media sosial yang cukup populer media sosial tersebut yang di bangun Jack Dorsey ini cukup banyak penggunaannya, terdapat lebih dari 500 juta pengguna terdaftar di Twitter, 200 juta di antaranya adalah pengguna aktif. (Kazeniak, 2009) Di sini penulis melakukan penambangan opini pengguna terkait kata kunci Aplikasi Web dengan batas crawling 1000 item data, mengacu pada penggunaan pemrosesan bahasa alami, analisis teks untuk mengidentifikasi, mengekstrak, mengukur, dan mempelajari keadaan efektif dan informasi subjektif secara sistematis pada data opini di twitter. Penulis menggunakan metode text classification yang telah banyak secara umum digunakan adalah algoritma Multinomial Naive Bayes. Metode ini memanfaatkan teorema probabilitas yang mana teorema bayes dan fungsionalitas data mining yaitu klasifikasi naïve bayesian. Multinomial Naive bayes adalah salah satu metode bayes yang dipakai dengan memperhitungkan frekuensi masing-masing kemunculan token atau kata dalam sebuah dokumen dan probabilitas.

Media sosial ini semakin memperluas jangkauan sosial setiap

penggunanya. Di media sosial, kita dapat menemukan, berkenalan dan berinteraksi dengan orang asing secara mudah. Kemudahan ini membuat setiap orang bebas (hampir tanpa batas) berbagi informasi dan saling berpendapat (Fajar Agus Maulana, 2020). Teks mining merupakan proses ekstraksi pola data dari sejumlah data tidak terstruktur yang akan diperoleh pola data, tren dan ekstraksi pengetahuan yang potensial dari data teks (Turban & et.al, 2011). Pada sosial media Twitter yang dilakukan oleh Muhamad Syarifudin universitas brawijaya. penelitian ini tentang Pendapat Seseorang terkadang sulit untuk disampaikan kepada pihak penanggung jawab secara langsung, hal itu mendorong masyarakat untuk menyampaikan pendapat, kritik dan sejenisnya melalui media sosial (Syarifuddin, 2020). Twitter dalam penelitian digunakan sebagai sumber data untuk menganalisis tweet berbahasa Indonesia (Diana Ikasari, 2020). Teks mining bertujuan menghasilkan informasi dari satu set dokumen. Text Mining mampu menghasilkan informasi melalui pemrosesan, pengelompokan, dan analisis data-data tidak terstruktur dalam jumlah besar (A. Sholihin, 2019). Analisis sentimen adalah proses menentukan sentimen dan mengelompokkan polaritas teks dalam dokumen atau kalimat sehingga kategori dapat ditentukan sebagai sentimen positif, negatif, atau netral (Setiawan, 2019). Metode Naïve Bayes adalah metode klasifikasi dalam penambangan teks yang digunakan dalam analisis sentimen. Metode ini berpotensi baik dalam klasifikasi dalam hal presisi dan komputasi data (H, 2014)

3. METODE

Penelitian tentang analisis sentimen dari tweet twitter tentang keyword Aplikasi web. Secara umum suatu analisis sentimen digunakan untuk memprediksi kecenderungan suatu opini atau pendapat masyarakat terhadap sesuatu hal, apakah cenderung ke opini positif, netral atau negatif. Pada penelitian ini di buat sebuah web aplikasi dimana fitur dari web aplikasi tersebut berfungsi untuk memprediksi opini atau pendapat user dengan cara live sehingga pengguna yang awam dengan lebih mudah mengetahui hasil analisis sentimen dengan keyword yang di inginkan. Selain dengan keyword pengguna juga bisa melakukan input dengan hastag untuk mengetahui hasil sentimen analisis. Penambangan data crawling dapat digunakan untuk tingkat analisis sentimen, Dalam menganalisis sentimen, ada beberapa tahapan yang perlu dilakukan untuk mendapatkan hasil tes. Berikut menunjukkan

langkah-langkah yang penulis gunakan :

3.1 Bahan Penelitian.

3.1.1 Pengumpulan Data

Tahap pertama dalam melakukan proses analisis sentimen adalah pengumpulan data. Data diambil dari twitter dengan permintaan pencarian tentang “Web Aplikasi” sebanyak 1000 menggunakan aplikasi Rapidminer. Search twitter digunakan untuk pengambilan data dari sosial media twitter yang dilanjut dan disimpan dalam bentuk excel.csv dengan write excel.csv. Dilanjut dengan remove duplicates yang berfungsi untuk menghapus konten yang berganda atau duplikat. Dilanjut dengan select atributte, Replace Missing Value, SubProses dan Sentimen Analisis kemudian hasil ahir akan menjadi suatu data yang akan di gunakan sebagai data latih pada algoritma Multinomial Naïve Bayes.

3.2 Alat Penelitian

Alat penelitian yang diperlukan berupa komputer yang akan diimplementasikan prototype Dalam penelitian dengan minimum required perangkat lunak sebagai berikut :

Operating System	: Microsoft Windows 7/8/10
Memory	: 1 GB of RAM required
Hard Disk Space	: 200 MB of free space required
Procesor	: Intel Dual Core or later

3.2.1 Perangkat Keras / Hardware

Dalam penelitian dibutuhkan perangkat keras dengan spesifikasi minimum seperti yang dapat dilihat pada Tabel 3.1.5

TABEL 3.1 PERANGKAT KERAS

Hardware	Spesifikasi
Jenis Komputer	Asus K42F
RAM	8.00 GB
Procesor	Intel core i5 2,53GHz
Hardisk	500 GB

3.2.2 Perangkat Lunak

Perangkat lunak yang digunakan dalam penelitian ini dapat dilihat pada Tabel 3.2

TABEL 3.2 PERANGKAT LUNAK

Software	Spesifikasi
Sistem Operasi	Microsoft Windows 10 pro 64 bit
Aplikasi Database	Rapidminer
Web Browser	Google Chrome
Medsos	Twitter
Tool	Rapidminer, <i>Python language Program</i> ,
Teks editor	Vs Code

3.3 Jalan Penelitian

3.3.1 Crawling Data Pada Twitter Dengan Rapidminer

Hal yang harus di lakukan adalah Mendownload Tool rapid miner dan melakukan instalasi pada komputer, selanjutnya membuat akun Twitter yang nantinya akan di hubungkan atau koneksi dengan rapid miner guna untuk crawling data yang akan di teliti. Dari data yang telah di crawling akan di simpan di local database dan di simpan dalam extensi excel.csv.

3.3.2 Tahap Preprocessing Data dengan Rapid Miner

RapidMiner adalah platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan bernama sama dengan yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam, penambangan teks, dan analisis prediktif. Hal ini digunakan untuk bisnis dan komersial, juga untuk penelitian, pendidikan, pelatihan, rapid prototyping, dan pengembangan aplikasi serta mendukung semua langkah dalam proses pembelajaran mesin termasuk persiapan data, hasil visualisasi, validasi model, dan optimasi (Markus Hofmann, 2013).Selanjutnya dilakukannya *preprocessing* dokumen seperti di atas untuk menghilangkan *noise*, menyeragamkan bentuk kata dan mengurangi volume kata. Tahapan yang dilakukan dari dokumen preprocessing dapat dilihat seperti di bawah ini :

1. Pembersihan Data

Pembersihan data atau pembersihan data adalah proses

mendeteksi dan mengoreksi (atau menghapus) catatan yang rusak atau tidak akurat

dari kumpulan catatan, tabel , atau basis data dan mengacu pada pengidentifikasian bagian data yang tidak lengkap, tidak benar, tidak akurat, atau tidak relevan dan kemudian mengganti, memodifikasi, atau menghapus data kotor (Wu, 2013). pembersihan data dilakukan untuk menghilangkan beberapa data tidak memiliki value. Data yang harus dihilangkan ialah seperti username yang di-mention, menghapus hashtag, RT, dan menghapus @.

2. Remove Duplicates

Operator ini menghapus contoh duplikat dari *ExampleSet* dengan membandingkan semua contoh satu sama lain berdasarkan atribut yang ditentukan. Dua contoh dianggap duplikat jika atribut yang dipilih memiliki nilai yang sama.

3. Select Attribute

Operator ini memilih subset dari *Attributes* dari sebuah *ExampleSet* dan menghapus *Attributes* lainnya.

4. SubProcessing

Pada tahap ini Operator ini memperkenalkan proses di dalam suatu proses. Kapanpun operator subproses dicapai selama eksekusi proses, pertama-tama seluruh subproses dijalankan. Setelah eksekusi subproses selesai, aliran dikembalikan ke proses (proses induk). Sebuah subproses dapat dianggap sebagai unit kecil dari suatu proses, seperti dalam proses, semua operator dan kombinasi operator dapat diterapkan dalam sebuah subproses. Itulah sebabnya subproses juga dapat didefinisikan sebagai rantai operator yang diterapkan.

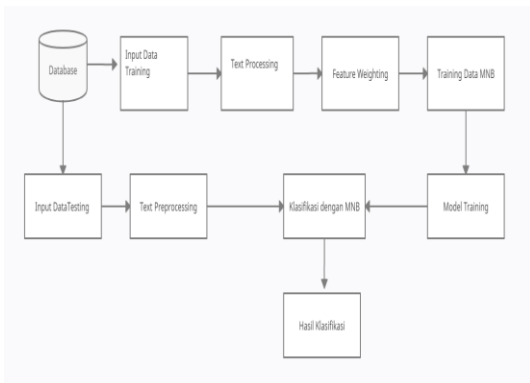
5. Sentimen Analisis

Dikutip dari MonkeyLearn, sentiment analysis adalah proses memahami dan mengelompokkan emosi (positif, negatif, dan netral) yang terdapat dalam tulisan dengan teknik analisis teks.Sentiment analysis ini juga menganalisis teks online untuk menentukan emosional dari penulis. Analisis sentimen adalah riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual (Liu. B, 2010)

3.3.3 Deskripsi Flow Program

Input Data Training dan Testing di tahap ini, data yang telah di crawling dibagi menjadi 2 yaitu artikel training dan artikel

testing. Artikel training digunakan untuk membuat model klasifikasi sedangkan artikel testing digunakan untuk menguji model.



Gambar 3. 1 Flow Program

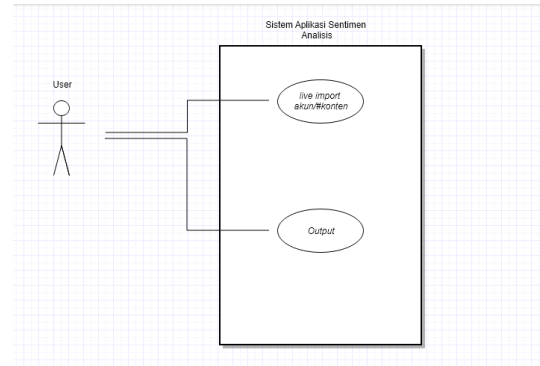
3.3.4 Data training

Data training yang sebelumnya telah diberi label dimasukkan ke dalam sistem untuk diproses. Data yang didapatkan dari Twitter diberi kategori 1, 0, -1 (positive, netral, negative). Setelah Data diberi label, dilakukan *preprocessing* dan pembobotan pada data, yang kemudian di-training menggunakan *Multinomial Naïve Bayes*.

3.3.5 Data Testing

Data testing merupakan data yang diambil dari dataset tetapi tidak diberi label seperti artikel training. Artikel testing dimasukkan ke sistem untuk di prediksi target atau label. Sebelum diklasifikasikan, artikel testing juga terlebih dahulu melewati tahap *preprocessing*.

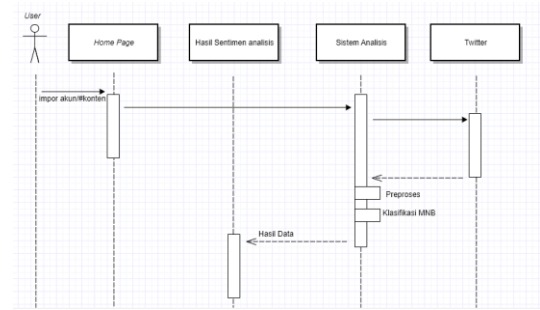
3.3.6. Usecase Diagram



Gambar 3.2 Usecase Diagram

Berdasarkan gambar di atas user yang interaksi dengan sistem dapat melakukan action pada aplikasi yaitu import live dengan akun atau hastag lalu akan di tampilkan halaman hasil analisis dari aplikasi tersebut.

3.3.7. Sequence Diagram



Gambar 3.3 Sequence Diagram

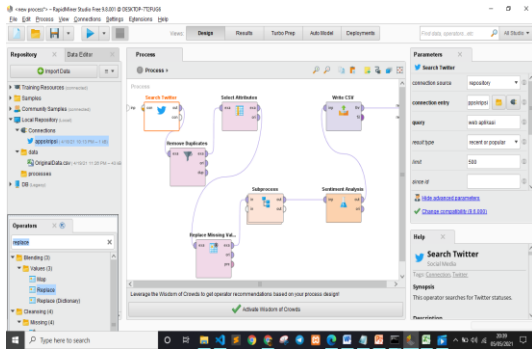
Hasil Dan Pembahasan

4.1 Hasil Penelitian

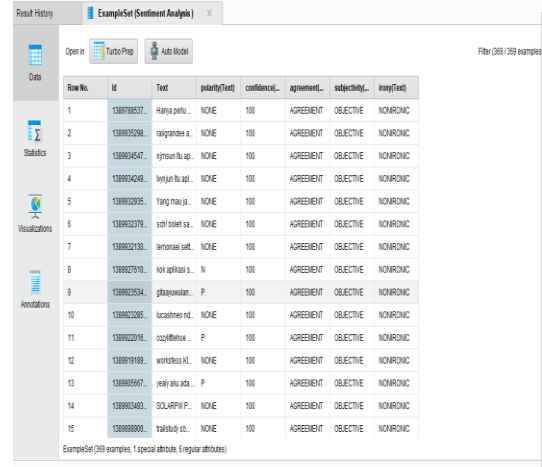
Sebagai pembahasan hasil penelitian disini penulis memberikan proses dari paling awal yaitu :

4.1.1.Melakukan Crawling Data Twitter Dengan Tools Rapid Miner

Melakukan crawling data twitter dengan tools rapid miner berikut design seperti gambar di bawah ini :



Gambar 4.1 proses crawling data di twitter



Row No.	Id	Text	polarity(Text)	confidence...	agreement...	subjectivity...	lang(Text)
1	1386788537	Harjo paku...	NEG	100	AGREEMENT	OBJECTIVE	NONRNC
2	1386925298	raigranda s...	NEG	100	AGREEMENT	OBJECTIVE	NONRNC
3	1386934547	ngman bu ap...	NEG	100	AGREEMENT	OBJECTIVE	NONRNC
4	1386934249	hngun bu apl...	NEG	100	AGREEMENT	OBJECTIVE	NONRNC
5	1386922035	Yang mau ja...	NEG	100	AGREEMENT	OBJECTIVE	NONRNC
6	1386922379	scf' tohai sa...	NEG	100	AGREEMENT	OBJECTIVE	NONRNC
7	1386921330	hormasi saff...	NEG	100	AGREEMENT	OBJECTIVE	NONRNC
8	1386927110	kat apkalid s...	NEG	100	AGREEMENT	OBJECTIVE	NONRNC
9	1386922534	gtaqwan...	P	100	AGREEMENT	OBJECTIVE	NONRNC
10	1386922295	lucathes ind...	NEG	100	AGREEMENT	OBJECTIVE	NONRNC
11	1386922016	o2qBlah...	P	100	AGREEMENT	OBJECTIVE	NONRNC
12	1386919189	workshes KL...	NEG	100	AGREEMENT	OBJECTIVE	NONRNC
13	1386905567	hsjg ato ada...	P	100	AGREEMENT	OBJECTIVE	NONRNC
14	1386934483	SOLARPH P...	NEG	100	AGREEMENT	OBJECTIVE	NONRNC
15	1386888000	trahduy ch...	NEG	100	AGREEMENT	OBJECTIVE	NONRNC

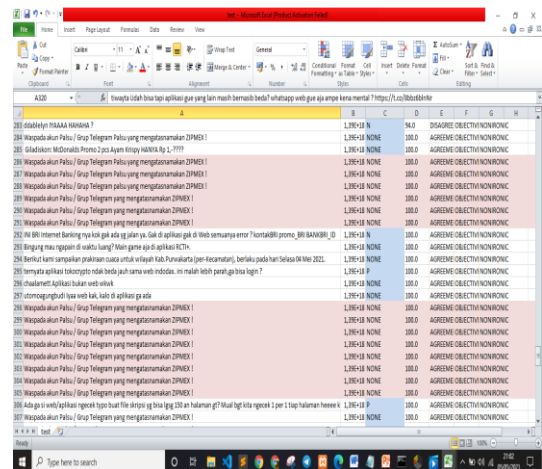
Gambar 4.2 hasil data crawling

Pada gambar 4.1 diatas adalah mulai pengumpulan data menggunakan tools Rapid Miner. Data yang digunakan adalah data yang di *crawling* dari sosial media twitter kemudian pada tools menggunakan berbagai operator.

Berikut operator yang di pakai pada gambar 4.1 ialah :

1. *Search twitter*
2. *Remove duplicate*
3. *Select attribute*
4. *Replace missing value*
5. *Sub process*
6. *Sentimen analisis*
7. *Write.csv*

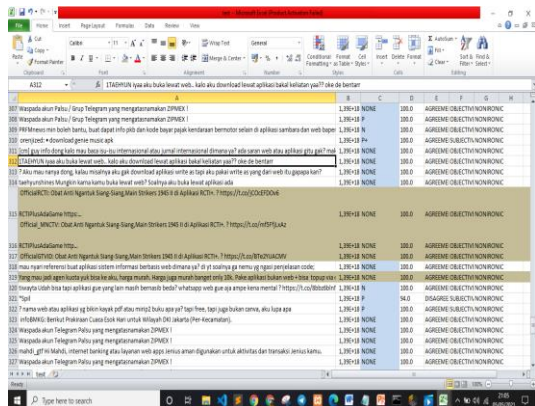
Sebagai hasil crawling di dapatkan 369 data crawling berdasarkan sistem di rapid miner. Kemudian hasil sebagai berikut:



Gambar 4.3 dataset.csv

Kemudian setelah melakukan filter hasil data duplikat peneliti juga menghapus data yang bermuatan kalimat iklan.

Total	180	45	225
-------	-----	----	-----



Gambar 4.4 contoh data iklan dan duplikat data yang di hapus

Setelah semua data di filter kemudian dilakukan labeling secara manual, Karena dalam sentimen yang di dihasilkan tools rapid miner belum akurat dan masih terdapat hasil sentimen yang tidak benar misalnya hasil sentimen pada atribut polarity adalah none, P (positif) , P+ , Neu (netral) , N (negative).

4.1.2 Hasil Test Data I

Di bawah ini merupakan hasil penelitian test data dengan jumlah item data 57 item dan data training 168 item data. Sebagai perbandingan 75% data training dan 25 % data testing.

4.1.3. Keterangan Data Set

Keterangan data set di jelaskan sebagai berikut :

- Jumlah dataset : 225
- Jumlah atribut : 7
- Jumlah tiga kelas yaitu : positif (1), negatif (-1), netral (0)

Dari semua data (225) training set 80% = $225 \times 80 / 100 = 180$ data latih dan data tes set 20% = $225 \times 20 / 100 = 45$ data tes

TABEL 4.1 JUMLAH CLASS / TARGET DATASET

Sample	Training 80%	Testing 20%	Total
Positif	100	25	125
Netral	8	2	10
Negatif	72	18	90

4.1.3. Confusion Matrix multiclass

TABEL 4.1 PERHITUNGAN KELAS POSITIF

		Actual Value		
		1	0	-1
Predicted Value	1	TP (1)	FP	FP
	0	FN	TN	TN
	-1	FN	TN	TN

TABEL 4.2 PERHITUNGAN KELAS NETRAL

		Actual Value		
		1	0	-1
Predicted Value	1	TN	FN	TN
	0	FP	TP (0)	FP
	-1	TN	FN	TN

TABEL 4.4 PERHITUNGAN KELAS NEGATIF

		Actual Value		
		1	0	-1
Predicted Value	1	TN	TN	FN
	0	TN	TN	FN
	-1	FP	FP	TP (-1)

```
(Env) C:\Dev\Skripsi>naive_bayes_training_unitest.py
Skipped: Creating Pre-processing data (already exist)
Creating Training Model
Training data: (180,) Test data: (45,)
[[29 0 0]
 [ 2 0 0]
 [12 0 2]]
[[[ 2 14]
 [ 0 29]]

 [[43 0]
 [ 2 0]]

 [[31 0]
 [12 2]]]
      precision    recall  f1-score   support

 Negative         1.00     0.14     0.25     14
  Neutral         0.00     0.00     0.00     2
  Positive         0.67     1.00     0.81     29

 accuracy          0.69         45
 macro avg         0.56     0.38     0.35     45
 weighted avg      0.75     0.69     0.60     45
```

Gambar 4.5 Hasil preproses data test I

Dalam mengevaluasi *performance* algoritma dari *Machine Learning* (ML) kita menggunakan acuan *Confusion Matrix*. *Confusion Matrix* merepresentasikan prediksi dan kondisi sebenarnya (aktual) dari data yang dihasilkan oleh algoritma. Berdasarkan *Confusion Matrix*, kita bisa menentukan *Accuracy*. Gambar diatas menunjukkan Akurasi prediksi sebesar 69% yang perhitungannya sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{f1-score} = 2 * \frac{\text{recall} * \text{precision}}{\text{recall} + \text{precision}} \quad (4)$$

$$= \frac{29+2}{29+2+14+0} = \frac{31}{45} = 0,6888888888888889 = 0,69$$

$$\text{Rate error} = 1 - \text{akurasi} = 1 - 0,69 = 0,31$$

Precision merupakan rasio jumlah yang benar diprediksi pada label perhitungan sebagai berikut:

$$\text{Kelas positif} = \frac{TP}{TP+FP} = \frac{29}{29+14} = \frac{29}{43} = 0,6744186046511628 = 0,67$$

$$\text{Kelas netral} = \frac{TP}{TP+FP} = \frac{0}{0+0} = 0$$

$$\text{Kelas negatif} = \frac{TP}{TP+FP} = \frac{2}{2+0} = \frac{2}{2} = 1$$

recall merupakan rasio jumlah yang diprediksi benar terhadap keseluruhan data yang benar sebagai perhitungan *recall* sebagai berikut :

$$\text{kelas positif} = \frac{TP}{TP+FN} = \frac{29}{29+0} = \frac{29}{29} = 1$$

$$\text{kelas netral} = \frac{TP}{TP+FN} = \frac{0}{0+2} = \frac{0}{2} = 0$$

Confusion Matrix multiclass

TABEL 4.3 CONFUSION MATRIX SEMUA KELAS

		Prediksi			Jml_T
		1	0	-1	
Target	1	29	0	0	29
	0	2	0	0	2
	-1	12	0	2	14
	Jml_P	43	0	2	

TABEL 4.4 EVALUASI HASIL CONFUSION MATRIX

Kelas	1	0	-1
TP	29	0	2
FP	14	0	0
TN	2	43	31
FN	0	2	12

Berikut keterangan cara perhitungan tabel *confusion matrix*:

Nilai TP diambil dari diagonal tabel *confusion matrix*

Nilai FP diambil dari total prediksi – hasil diagonal TP

Nilai TN diambil dari penjumlahan nilai dari selain baris kelas dan kolom kelas

Nilai FN diambil dari total target – hasil diagonal TP

$$\text{kelas negatif} = \frac{TP}{TP+FN} = \frac{2}{2+12} = \frac{2}{14} = 0,14$$

f1-score merupakan rata – rata dari precision dan recall, sedangkan

akurasi merupakan nilai keseluruhan prediksi yang benar sebagai perhitungan sebagai berikut :

$$\text{F1 Score} = \frac{2*(\text{Recall}*Precision)}{(\text{Recall}+\text{Precision})}$$

$$\text{Kelas positif} = 2 * \frac{\text{recall} * \text{precision}}{\text{recall} + \text{precision}} = 2 * \frac{1 * 0,67}{1 + 0,67} = 2 * \frac{0,67}{1,67} = 0,81$$

$$\text{Kelas netral} = 2 * \frac{\text{recall} * \text{precision}}{\text{recall} + \text{precision}} = 2 * \frac{0 * 0}{0 + 0} = 0$$

$$\text{Kelas negatif} = 2 * \frac{\text{recall} * \text{precision}}{\text{recall} + \text{precision}} = 2 * \frac{0,14 * 1}{0,14 + 1} = 0,25$$

4.2 Analisis dan Pembahasan

yaitu Penelitian tentang mengetahui sentimen dari tahap pelabelan sampai pemodelan memiliki beberapa permasalahan:

4.2.1 Memahami Makna Kalimat

Kendala tersebut terjadi dikarenakan di dalam sebuah kalimat mengandung beberapa makna. Sebagai contoh : keluhan, sindiran, serta apresiasi.

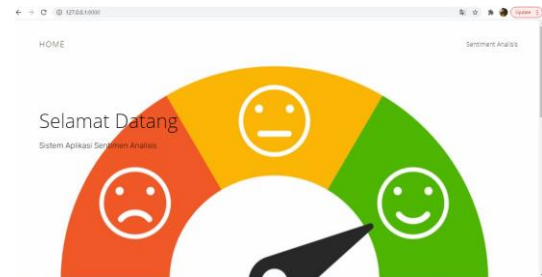
4.2.2 Kendala Pelabelan Secara Manual

Ada beberapa dataset yang di awal, tengah dan akhir kalimat mengandung kata yang berkategori positif tetapi di tengah kalimat mengandung kata berkategori negatif. Hal ini menyebabkan sebuah kalimat bersifat ambigu yang menyebabkan kesalahan pada proses pelabelan.

4.2.3 Data Tidak Seimbang

Data yang tidak seimbang merupakan penyebab dari kendala memahami opini dan kendala dalam proses pelabelan, walaupun selisih data yang berlabel positif dan berlabel negatif berbeda jumlah datanya tetapi akan berdampak pada hasil.

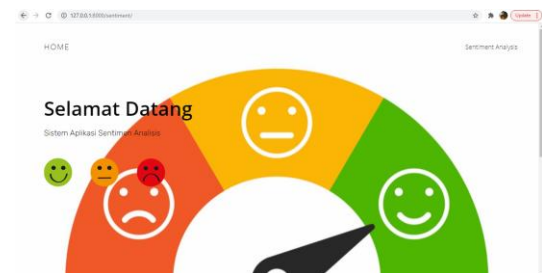
4.2.4. Aplikasi Sentimen Analisis



Gambar 4.10 Home aplikasi

Gambar 4.10 diatas adalah bentuk tampilan utama atau Home yang

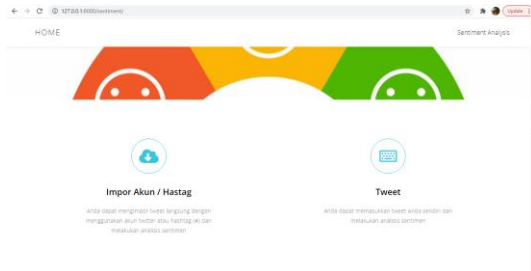
akan pertama kali di akses oleh user atau pengguna pada saat akan melakukan sentimen analisis dengan aplikasi tersebut. Di dalam sumber daya Home ter dapat fitur Sentimen analisis yang akan di gunakan untuk memprediksi tweet berdasarkan keyword yang di input oleh pengguna.



Gambar 4.11 Fitur Sentimen Analisis

Selanjutnya gambar 4.10 adalah gambaran halaman sentimen analisis. Pada halaman ini pengguna dapat melakukan input keyword sesuai keinginan, Tapi jika keyword yang telah di inputkan ternyata pada API twitter tidak di temukan maka akan

mengembalikan pesan error dan solusinya pengguna harus input dengan keyword yang lain.



Gambar 4.12 Live Sentimen Analisis

Selanjutnya yaitu gambar 4.11 merupakan halaman yang sama dengan gambar 4.10.

6. PENUTUP

5.1 Kesimpulan

Sebagai kesimpulan Analisis sentimen tweet dengan keyword aplikasi web berbagai opini yang ada di bagi 3 klasifikasi yaitu opini yang bermakna positif, opini bermakna netral dan opini bermakna negatif. Dari semua item data menghasilkan output dengan metode algoritma multinomial naive bayes karna klasifikasi lebih dari dua kelas dan proses di lakukan dengan 80% data untuk training dan 20 % untuk testing. Pada sistem melakukan uji evaluasi dengan confusion matrix. Berdasarkan data yang di uji dan pembagian data train dan tes yang berbeda akan mendapatkan performa akurasi yang berbeda. Kemudian hasil dari preproses dan uji algoritma di import ke *django framework* untuk dilakukan *input live* sehingga pengguna awam dengan sentimen analisis bisa mengoperasikan dan dengan mudah mengetahui sentimen analisis dengan input keyword atau hastag yang di cari atau di inginkan. Berdasarkan penelitian yang sudah dilakukan dapat dihasilkan output data memberikan pengaruh terhadap precision maka hal ini dikarenakan Naïve Bayes Classifier dengan tipe Multinomial Naïve Bayes menghitung jumlah istilah dan kata yang ada pada sebuah kelas. Penelitian dengan 225 data yang telah di crawling, terdiri dari 90 data berlabel negatif, 125 data positif, 10 data netral, lebih banyak mengandung sentimen positive yang berisi makna informasi dan diskusi dan negatif yang berisi makna ketakutan, bullying, sedih, kecewa, unsur protes lebih sedikit. Pengujian analisis sentimen pada data crawling dengan keyword Web Aplikasi pada sosial media twitter dengan metode Multinomial Naive

Bayes dengan data latih 80% dan data uji 20% menghasilkan akurasi sebesar 69%. Kemudian saran yang dapat penulis berikan adalah Melakukan pengembangan untuk penelitian analisis sentimen menggunakan algoritma klasifikasi yang lain atau menggunakan dua algoritma atau lebih yang berbeda supaya di hasilkan output klasifikasi analisis

sentimen yang lebih baik dan bermanfaat bagi pengguna. Menambah konversi gambar ke dalam teks pada sistem ini, karena selain teks banyak sekali user yang membuat capture untuk memberikan opini atau pendapat mereka pada hasil crawling dataset. Selanjutnya perbaikan menggunakan data latih yang lebih banyak agar mendapat hasil performa yang lebih akurat.

DAFTAR PUSTAKA

- A. Sholihin, H. N. (2019). "Analisis Penyakit Difteri Berbasis Twitter Menggunakan Algoritma Naïve Bayes". *SAKTI – Sains, 1*.
- Diana Iksari, Y. F. (2020, Agustus 1). ANALISIS SENTIMEN DAN KLASIFIKASI TWEETS BERBAHASA. *Jurnal Ilmiah Informatika Komputer, 25*.
- Fajar Agus Maulana, I. E. (2020, Agustus 14). ANALISA SENTIMEN CYBERBULLYING DI JEJARING SOSIAL. *SENAMIKA*.
- H, J. M. (2014). "Opinion Mining For Sentiment Data Classification". *Int. J. Res. Inf. Technol, 3*.
- Liu, B, S. A. (2010). Sentiment Analysis and Subjectivity.
- Markus Hofmann, R. K. (2013, Oktober 25). "RapidMiner: Data Mining Use Cases and Business Analytics Applications (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series). *CRC Press*.
- Setiawan, D. A. (2019). "ANALISIS SENTIMEN PROGRAM ACARA DI SCTV PADA TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE,". *e-Proceeding of Engineering*.
- Syarifuddin, M. (2020, Agustus 1). ANALISIS SENTIMEN OPINI PUBLIK TERHADAP EFEK PSBB PADA. *Inti Nusa Mandiri*.
- Turban, E., & et.al. (2011). Decision Support and Business Intelligence Systems. *New Jersey: Pearson Education, Inc*.
- Wu, S. (2013). "A review on rough warranty data and analysis". *Reliability Engineering and System*.