**MAKALAH INFORMATION RETRIEVAL**

**Dr. Drs. Eri Zuliarso,M.KOM**

****

**Disusun Oleh :**

**Susatya Raka Winarsa**

**(15.01.55.0028)**

**FAKULTAS TEKNOLOGI INFORMASI**

**UNIVERSITAS STIKUBANK SEMARANG**

**2018**

**KATA PENGANTAR**

Puji syukur kehadirat Tuhan Yang Maha Esa atas segala rahmat-Nya sehinggadapat menyusun makalah dengan sebaik-baiknya dan tersusun hingga selesai.Saya ucapkan terima kasih kepada seluruh pihak yang telah membantu, memfasilitasi, memberi masukan dan mendukung saya hngga makalah ini selesai.

Harapan saya semoga makalah ini dapat menambah ilmu pengetahuan dan pengalaman bagi para pembaca. Untuk ke depannya dapat memperbaiki dari bentuk maupun bagian isi agar makalah menjadi lebih baik lagi.

Meski telah menyusun makalah ini dengan maksimal, karena keterbasan pengalaman maupun pengetahuan saya tidak menutup kemungkinan masih banyak kekurangan. Oleh karena itu sangat diharapkan kritik dan saran yang membangun dari pembaca demi makalah yang lebih baik dan sempurna.

Semarang, 14 Juli 2018

Penyusun

**BAB I PENDAHULUAN**

1. **Latar Belakang**

Analisis sentimen adalah bidang ilmu yang mempelajari bagaimana menganalisa opini, sentimen, evaluasi, penilaian, sikap dan emosi dari sebuah entitas yang dapat berupa produk, pelayanan, organisasi, individu, isu-isu, peristiwa, dan topik (Liu, 2012). Analisis sentimen biasa digunakan untuk menilai kesukaan atau ketidaksukaan publik terhadap suatu barang atau jasa. Sentimen tersebut merupakan informasi tekstual yang bersifat subjektif dan memiliki polaritas positif dan negatif. Nilai polaritas ini dapat digunakan sebagai parameter dalam menentukan suatu keputusan.Tugas dasar dalam analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen kemudian menentukanpendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral (Dehaff, M., 2010).

Analisa sentimen atau biasa disebut opinion mining merupakan salah satu cabang penelitian Text Mining.Text mining merupakan analisis teks dimana sumber data  biasanya di dapatkan dari dokumen. Definisi lain, text mining melingkupi sebuah  proses ekstrasi informasi yang terpola yang berasal dari sejumlah besar sumber data teks, seperti dokumen Word, PDF, kutipan teks, atau bahkan sms (tweet). Teks mining dibagi menjadi 2 tahap, pertama diawali degan merubah data teks unstrukstur ke data semi atau terstruktur dilanjutkan dengan melakukan ekstraksi informasi yang diteliti dari data teks terstruktur.

Salah satu metode yang sedang dikembangkan saat ini untuk masalah di atas adalah klasifikasi sentimen. Metode ini merupakan perkembangan dari text mining yang sedang popular dan masih terus berkembang. Dapat dilihat dari nama metode di atas, metode ini bertujuan untuk memisahkan suatu opini dalam jumlah yang sangat besar ke dalam dua kelas, dalam hal ini kelas positif dan negatif.

1. **Manfaat**
2. Mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan, keterkaitan dan kelas antar dokumen.
3. Identifikasi masalah menentukan topik pilihan menentukan sumber data untuk analisasentimen.
4. Untuk mencari analisa keterhubungan, keterkaitan dan kelas antar dokumen.

**BAB II PEMBAHASAN**

Di dalam bab ini saya akan menjelaskan mengenai macam-macam dari analisa sentimen dan pengolahan data menggunakan analisa sentimen. Penggunaan analisa sentimen dilihat dari beberapa aspek. Di bawah ini saya akan membahas tentang materi pembahasan analisa sentimen.

**Analisa Sentimen Dengan Metode Machine Learning Approach**

Machine Learning memerlukan dataset untuk digunakan sebagai data training. Dibutuhkan effort untuk mengumpulkan dan melakukan class tag pada sampel dataset tersebut, selain itu proses training juga membutuhkan waktu. Akurasi dari pendekatan klasifikasi machine learning sangat baik, akan tetapi performa klasifikasinya domain dependent terhadap dataset yang digunakan pada saat training.Metode-metode yang masuk ke dalam kategori ini adalah sebagai berikut:

* Naïve Bayes
* Maximum Entropy
* SVM
* Neural Network

Semua metode pada pendekatan machine learning membutuhkan proses training, beberapa metode machine learning, yaitu Naïve Bayes, Maximum Entropy dan SVM. Sentiment analysis dilakukan pada data review film yang didapat dari IMD. Pelabelan data dilakukan secara otomatis menggunakan star rating yang diberikan oleh penulis review. Kemudian star rating tersebut dibagi menjadi 3 kategori, yaitu positive, negative dan netral. Hanya data yang berlabel positive dan negative yang digunakan sebagai data training. Untuk menghindari bias dari penulis review, jumlah review dari seorang reviewer dibatasi hanya 20 review per kategori label. Hasil penelitian menunjukan bahwa SVM memiliki persentase akurasi tertinggi dari ketiga metode yang digunakan. Selain itu, penggunaan unigram dengan negation-tag pada proses training memiliki persentase akurasi lebih baik dibandingkan jika menggunakan bigram. Penggunaan word presence dan word count juga mempengaruhi akurasi sentiment analysis. Penggunaan word presence memberikan akurasi lebih baik dibandingkan menggunakan word count pada proses training. Selain itu dicoba juga proses part-of-speech atau POS tagging pada preprocessing. Tujuannya adalah untuk menghindari penggunaan kata yang tidak tepat, seperti misalnya “I love this movie” yang memiliki sentiment positif dengan “This is a love story” yang bersentimen netral. Akan tetapi hasil percobaan tidak lebih baik dibandingkan dengan tanpa menggunakan POS tagging.

**Analisa Sentimen Dengan Metode Knowledge-Based Method Approach**

Knowledge-Based adalah pendekatan analisa sentimen pada word level, dimana entitas yang diproses adalah kata. Metode-metode yang masuk di dalam pendekatan ini adalah sebagai berikut:

* Lexicon-Based
* PMI (Pointwise Mutual Information)

Knowledge-Based bergantung pada dictionary atau kamus lexicon yang digunakan untuk melakukan penilaian terhadap fitur yang didapat.

* **Lexicon-Based**

Lexicon-Based menggunakan dictionary atau kamus lexicon untuk melakukan penilaian terhadap kata. Pada dictionary, kata-kata dipasangkan dengan nilai polaritasnya. Yang harus dilakukan sebelum melakukan analisis menggunakan lexicon adalah menentukan kata yang akan dianalis dari corpus. Pemilihan kata tersebut dapat dilakukan dengan melakukan Part-Of-Speech Tagging dan lalu mencari kata-kata dengan tipe yang ingin diambil, seperti misalnya adjective dan adverb.

Makna kata dapat berubah bergantung pada konteks dari kalimat, oleh karena itu, pendekatan lexicon-based terkadang tidak bisa menangkap makna sebenarnya dari kata yang diprosesnya. Akan tetapi, lexicon-based memiliki performa klasifikasi yang baik pada kasus lintas domain, dan knowledge dapat ditambahkan kapan saja kedalam dictionary.

* **PMI (Pointwise Mutual Information)**

Dapat digunakan sebagai pendekatan semi-unsupervised method, dimana seed words yang dipilih adalah ‘excellent’ dan ‘poor’. Dengan PMI, kalimat tersebut dihitung semnatic orientation-nya terhadap seed words yang dipilih.

Penggabungan dilakukan karena metode lexicon-based seringkali tidak tepat dalam mendeteksi polaritas karena kata-kata dapat memiliki polaritas yang berbeda bergantung pada konteks kalimatnya (Low Recall). Metode lexicon-based digunakan untuk mengkategorikan tweet kedalam kelas polaritas. Setelah proses tersebut, dilakukan proses aggregating opinion pada entity (NN & NNP) pada tweet tersebut. Hal ini dilakukan untuk menghindari low recall dari metode lexicon-based. Setelah proses tersebut, dilakukan klasifikasi polaritas menggunakan opinion rules, dimana kalimat dipecah-pecah menggunakan POS tagger dan dikenali pattern-nya. Pengenalan pattern dilakukan untuk mengetahui polaritas dari kalimat tersebut. Polaritas dari tweet tersebut kemudian digunakan sebagai label untuk proses training SVM classifier.

**Analisa Sentimen berdasarkan level sumber datanya**

Berdasarkan level sumber datanya Analisa Sentimen terbagi menjadi 2 kelompok besar yaitu :

1. **Coarse-grained Sentiment Analysis**

Pada analisa sentimen jenis ini, analisa sentimen yang dilakukan adalah pada level dokumen. Secara garis besar fokus utama dari analisa sentimen jenis ini adalah menganggap seluruh isi dokumen sebagai sebuah sentimen positif atau sentimen negatif.

1. **Fined-grained Sentiment Analysis**

Fined-grained Sentiment Analysis adalah analisa sentimen pada level kalimat. Fokus utama fined-greined Sentiment Analysis adalah menentukan sentimen pada setiap kalimat.

**Analisa Sentimen terdiri dari 3 subproses besar**

Masing-masing subproses ini bisa kita jadikan bahan/topik riset secara terpisah karena masing-masing subproses ini membutuhkan teknik yang tidak mudah yaitu dengan :

1. **Subjectivity Classification** menentukan kalimat yang merupakan opini.
2. **Orientation Detection** setelah berhasil diklasifikasi untuk kategori Opini, sekarang kita tentukan apakah dia positif, negatif, netral.
3. **Opinion Holder and Target Detection** menentukan bagian yang merupakan Opinion Holder dan bagian yang merupakan Target.

**Analisa Sentimen dengan Pengumpulan Data& Pengolahan Data**

1. **Membaca Data**

Sebelum melakukan pengolahan data, terlebih dahulu kita baca data yang akan kita olah. Semua informasi ditabulasikan dalam dataset sentimen, dan tidytext menyediakan fungsi get\_sentiments () untuk mendapatkan sentimen spesifik tanpa kolom yang tidak digunakan.

1. **Sentiment Analysis with Inner Join**

Kita perlu mengambil teks dan mengubah teks menjadi format rapi menggunakan unnest\_tokens. Kita juga membuat beberapa kolom lain untuk melacak garis dan bagian dari masing-masing kata, menggunakan group\_by dan bermutasi untuk membangun kolom tersebut. Apabila sudah di uji maka plot berubah menjadi sentimen positif atau negatif lebih jauh dari lintasan cerita.

1. **Membandingkan Tiga Sentimen**

Gunakan ketiga leksikon sentimen tersebut dan periksa bagaimana sentimen berubah melintasi busur naratif Pride and Prejudice. Pertama, gunakan filter () untuk memilih hanya kata-kata dari satu teks yang diminati.

1. **Kata Positif dan Negatif yang Sering Muncul**

Salah satu keuntungan memiliki bingkai data dengan sentimen dan kata adalah bahwa kita dapat menganalisis jumlah kata yang berkontribusi pada setiap sentimen. Dengan menerapkan hitungan () di sini dengan argumen dari kedua kata dan sentimen, kita mengetahui berapa banyak setiap kata berkontribusi pada setiap sentiment.

1. **Wordclouds**

Kami telah melihat bahwa pendekatan penambangan teks yang rapi ini bekerja dengan baik dengan ggplot2, namun data dalam format yang rapi juga berguna untuk plot lainnya.Kita dapat menggunakan visualisasi ini untuk melihat kata-kata positif dan negatif yang paling penting, namun ukuran kata-katanya harus sebanding dengan sentimen.

**Contoh Program analisa sentimen twitter:**

library(e1071)

library(twitteR)

library(ROAuth)

library(tm)

library(ggplot2)

library(wordcloud)

library(sentimentr)

library(plyr)

library(RTextTools)

library(sentiment)

library(Rstem)

setup\_twitter\_oauth("xQNIFX6fY0C5epCtZUEZKyFr9", "z4CpFNRGfSV8y2o1t4wiocgcZZaqN2eOZvAW4RDILXNMlsyDZp", "1632612013-IpZOn5NATxdIXRQRJ4pdg4wiVv3VtxpG9U0M113", "33VIKDivUSf690PQcdiKihD5ahTxnL8lDg5vtCfAyZsu7")

# harvest some tweets

some\_tweets = searchTwitter("Real Madrid C.F", n=500, lang="en")

# get the text

some\_txt = sapply(some\_tweets, function(x) x$getText())

# remove retweet entities

some\_txt = gsub("(RT|via)((?:\\b\\W\*@\\w+)+)", "", some\_txt)

# remove at people

some\_txt = gsub("@\\w+", "", some\_txt)

# remove punctuation

some\_txt = gsub("[[:punct:]]", "", some\_txt)

# remove numbers

some\_txt = gsub("[[:digit:]]", "", some\_txt)

# remove html links

some\_txt = gsub("http\\w+", "", some\_txt)

# remove unnecessary spaces

some\_txt = gsub("[ \t]{2,}", "", some\_txt)

some\_txt = gsub("^\\s+|\\s+$", "", some\_txt)

# define "tolower error handling" function

try.error = function(x)

{

# create missing value

y = NA

# tryCatch error

try\_error = tryCatch(tolower(x), error=function(e) e)

# if not an error

if (!inherits(try\_error, "error"))

y = tolower(x)

# result

return(y)

}

# lower case using try.error with sapply

some\_txt = sapply(some\_txt, try.error)

# remove NAs in some\_txt

some\_txt = some\_txt[!is.na(some\_txt)]

names(some\_txt) = NULL

# classify emotion

#library(sentiment)

class\_emo = classify\_emotion(some\_txt, algorithm="bayes", prior=1.0)

# get emotion best fit

emotion = class\_emo[,7]

# substitute NAâs by "unknown"

emotion[is.na(emotion)] = "unknown"

# classify polarity

class\_pol = classify\_polarity(some\_txt, algorithm="bayes")

# get polarity best fit

polarity = class\_pol[,4]

# data frame with results

sent\_df = data.frame(text=some\_txt, emotion=emotion,

polarity=polarity, stringsAsFactors=FALSE)

# sort data frame

sent\_df = within(sent\_df,

emotion <- factor(emotion, levels=names(sort(table(emotion), decreasing=TRUE))))

# plot distribution of emotions

ggplot(sent\_df, aes(x=emotion)) +

geom\_bar(aes(y=..count.., fill=emotion)) +

scale\_fill\_brewer(palette="Dark2") +

labs(x="emotion categories", y="number of tweets")

######

ggplot(sent\_df, aes(x=polarity)) +

geom\_bar(aes(y=..count.., fill=polarity)) +

scale\_fill\_brewer(palette="RdGy") +

labs(x="polarity categories", y="number of tweets")

######

emos = levels(factor(sent\_df$emotion))

nemo = length(emos)

emo.docs = rep("", nemo)

for (i in 1:nemo)

{

tmp = some\_txt[emotion == emos[i]]

emo.docs[i] = paste(tmp, collapse=" ")

}

####

# remove stopwords

emo.docs = removeWords(emo.docs, stopwords("english"))

# create corpus

corpus = Corpus(VectorSource(emo.docs))

tdm = TermDocumentMatrix(corpus)

tdm = as.matrix(tdm)

colnames(tdm) = emos

# comparison word cloud

comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"),

scale = c(3,.5), random.order = FALSE, title.size = 1.5)

**BAB III PENUTUP**

Demikian yang dapat saya paparkan mengenai materi yang menjadi pokok pembahasan dalam makalah ini. Tentunya masih banyak kekurangan dan kelemahan dalam penulisan makalah ini, karena terbatasnya pengetahuan dan pengalaman serta refrensi yang berhubungan dengan judul makalah ini.

Saya berharap kepada para pembaca yang bijaksana untuk memberikan kritik dan saran yang membangun untuk saya demi kesempurnaan makalah ini dan untuk penulisan makalah di kesempatan yang lainnya. Semoga makalah ini berguna bagi diri saya sendiri khususnya dan juga para pembaca yang bijaksana pada umumnya.

**REFRENSI**

* www.Academia.edu/33480625/makalah\_analisa\_sentimen\_menggunakan\_bahasa\_R.pdf.
* www.Researchgate.net/publication/303942719\_ANALISIS\_SENTIMEN.
* Kundi, F. M., & Asghar, M. Z. (2014). Lexicon-based sentiment analysis in the social web. Journal of Basic and Applied Scientific Research, 4(6).
* Asghar, M. Z. (2014). Lexicon based Approach for Sentiment Classification of User Reviews. Life Science Journal, 11(10).