



**ANALISIS SENTIMEN TERHADAP PRESEPSI KONSUMEN KOPI  
DI MEDIA SOSIAL TWITTER MENGGUNAKAN *NAÏVE BAYES*  
*MULTINOMIAL***

**JURNAL ILMIAH**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika**

**Oleh**

**Nurul Qomariyah  
B2A016021**

**PROGRAM STUDI STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS MUHAMMADIYAH SEMARANG**

2020



# ANALISIS SENTIMEN TERHADAP PRESEPSI KONSUMEN KOPI DI MEDIA SOSIAL TWITTER MENGGUNAKAN NAÏVE BAYES MULTINOMIAL

Nurul Qomariyah

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang

e-mail : [nurulqomariyah1809@gmail.com](mailto:nurulqomariyah1809@gmail.com)

<b>Article history</b>	<b>ABSTRACT</b>
<b>Submission</b> : <b>Revised</b> : <b>Accepted</b> :	<i>Coffee is the result of Indonesian plantations that have high competitiveness in the international market. Coffee demand can be influenced by people's perception of coffee in Indonesia. Society is easy to give opinions and opinions through social media. One of the most popular social media in all circles, namely Twitter. Types of coffee that are trending topics on Twitter today are civet coffee, black coffee, sick coffee and bitter coffee. The purpose of this study is to provide an overview of the positive and negative sentiments of the community towards coffee in Indonesia. To determine the public's good or bad perception that can influence coffee demand, sentiment analysis using Naïve Bayes Multinomial is used. Naïve Bayes Multinomial is a method used in text processing, is easy to implement and has a high accuracy value and handles a large number of vocabulary sizes and reduces high error rates. The data used is based on data scrapping from Twitter. The positive sentiment obtained is that Indonesian people like civet coffee, black coffee, and coffee connoisseurs namely men, and the negative sentiment obtained is an upset stomach after drinking coffee. The test results obtained an accuracy value of 94%, precision on positive sentiment of 97% and negative 80%, and recall on positive sentiment of 96% and negative 86%.</i>
<b>Keyword :</b> <i>Coffee, Twitter, Sentiment Analysis, Naïve Bayes Multinomial</i>	

## PENDAHULUAN

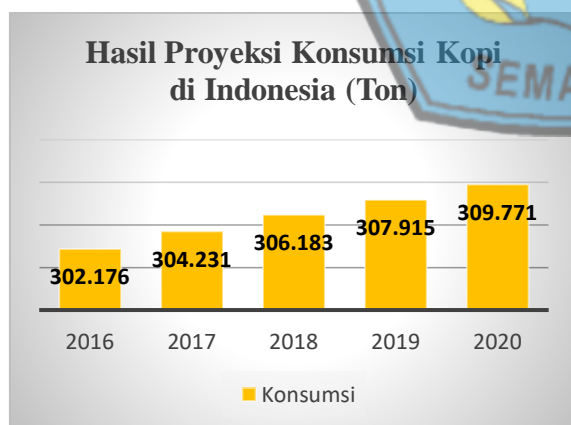
Dalam era globalisasi, perkembangan suatu negara dapat dilihat dari kualitas pembangunan negara tersebut. Salah satunya Negara Indonesia yang memiliki program pembangunan yakni konsumsi dan pola produksi. Sektor yang mendapat tujuan dalam keberlangsungan konsumsi dan pola

produksi yang menjamin yaitu pada sektor pertanian. Salah satu sub sektor pertanian yang potensial adalah perkebunan kopi (Syaad dkk, 2013). Perkebunan kopi merupakan sub sektor yang memberikan kontribusi tertinggi dibandingkan dengan sub sektor lainnya di sektor pertanian.

Kopi adalah salah satu hasil perkebunan yang sampai saat ini menjadi

komoditas unggulan nasional bahkan mempunyai daya saing di pasar Internasional. Bahkan perkebunan kopi mempunyai peran penting dalam perekonomian Indonesia, antara lain yaitu sumber pendapatan devisa, penyedia lapangan kerja, dan sumber pendapatan bagi pekebun kopi bahkan pelaku ekonomi lainnya (Kustiari, 2016). Pada faktanya, menunjukkan bahwa kopi tersebut menjadi komoditas nomor dua yang paling banyak diperdagangkan setelah minyak bumi.

Menurut, data Kementerian Pertanian tahun 2018, produktivitas kopi Indonesia berada pada angka 731kg/ha. Berdasarkan data Organisasi kopi Indonesia (ICO) diperkirakan bahwa kopi global tahun 2018-2019 diperkirakan meningkat 7,1% menjadi 171 juta kantong. Begitupula, konsumsi kopi yang berada pada kisaran 163.200 juta kantong. Selain itu, proyeksi konsumsi terhadap kopi oleh Pusat Data dan Sistem Informasi Pertanian Sekretariat Jenderal Kementerian Pertanian pada tahun 2016 hingga 2020 terus mengalami peningkatan sebesar 7,1% per tahunnya. Berikut grafik hasil proyeksi konsumsi kopi di Indonesia :



Sumber data : SUSENAS BPS

**Gambar 1. Hasil Proyeksi Konsumsi Kopi di Indonesia**

Dilihat dari grafik proyeksi konsumsi terhadap kopi di Indonesia terus meningkat setiap tahunnya. Dari 302.176 Ton hingga mencapai 309.771

Ton pertahunnya. Hal ini dikarenakan adanya tren permintaan kopi yang meningkat tajam dan secara tidak sadar kopi juga sebagai bagian dari gaya hidup masyarakat Indonesia (Suisa et al., 2017).

Penetrasi internet Indonesia yang menunjukkan peningkatan disetiap tahunnya. Berdasarkan data survei yang diselenggarakan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menunjukkan jumlah pengguna internet di Indonesia tahun 2018 naik 10,12% dari tahun sebelumnya. Kenaikan mencapai 27 juta pengguna, artinya ada 171,17 juta pengguna internet dari total 246,15 juta jiwa penduduk Indonesia yang hampir semua pengguna internet memiliki media sosial. Hal ini dapat memunculkan perilaku untuk mengunggah status bahkan kegiatan lainnya yang sedang dilakukan ke media sosial, termasuk mengunggah opini-opini dan ulasan terhadap produk atau jasa yang pernah dicoba.

Pada saat ini permintaan kopi meningkat dengan diikuti oleh beberapa macam persepsi masyarakat. Menurut penelitian Tambunan, dkk (2018) melalui judul jurnal yang diteliti yakni Analisis Sikap Konsumen Terhadap Kopi Bubuk Lokal dan Kopi Instan di Kota Pamarangsiantar yang bertujuan untuk menjelaskan dan menganalisis sikap konsumen terhadap konsumsi kopi tersebut. Penelitian diperoleh hasil yang menunjukkan bahwa alasan utama dari konsumen dalam konsumsi kopi adalah kesegaran dengan pertimbangan awal berupa rasa dan kualitas kopinya. Konsumen puas terhadap kopi yang dibeli dan dikonsumsi serta bersedia membeli kembali meskipun harga produk kopi mengalami kenaikan. Selain itu, konsumen juga cenderung dalam mengonsumsi kopi instan dibandingkan bubuk lokal (Tambunan, 2018).

Seiring dengan permintaan dan produksi kopi terus meningkat dan

diikuti oleh berbagai macam persepsi yang ada di masyarakat. Oleh karena itu, dapat dibuat suatu pendekatan untuk mengetahui persepsi masyarakat terhadap kopi dengan menggunakan data analisis sentimen di media sosial. Masyarakat dapat mudah menyalurkan pendapat dan opininya melalui berbagai media, salah satunya media sosial *Twitter*. Semakin banyak masyarakat membeli produk kopi maka semakin meningkat juga permintaan terhadap kopi. Jika sebagian besar masyarakat suatu daerah memiliki persepsi baik mengenai produk kopi, maka permintaan pada suatu daerah tersebut juga baik. Sedangkan pada suatu daerah memiliki persepsi buruk terhadap kopi maka permintaan kopi juga rendah. Berdasarkan hal tersebut, bila permintaan produk kopi meningkat maka produsen akan lebih giat untuk memproduksi kopi tersebut (Oktavian and Maulana, 2019)

Berbagai penelitian yang berkaitan dengan penilaian persepsi terhadap konsumsi kopi telah dilakukan penilitan yang dilakukan (Dewi, Anwar and Amalia, 2009) yaitu mengenai persepsi terhadap kopi dan teh serta dampaknya dikalangan mahasiswa. Hasil yang diperoleh yaitu dampak positif yang paling banyak dirasakan oleh kalangan mahasiswa setelah mengkonsumsi kopi dan teh yaitu rasa kantuk berkurang pada kopi (92,3%) dan Lelah berkurang pada teh (52,1%). Sedangkan untuk dampak negatif yang paling banyak dirasakan oleh kalangan mahasiswa setelah mengkonsumsi kopi dan teh yaitu merasa ketagihan kopi (42,9%) dan sering buang air kecil pada konsumsi teh (33,3%). Terdapat hubungan yang signifikan antara pengetahuan gizi dan kafein dengan konsumsi kopi yaitu semakin tinggi pengetahuan tentang gizi dan kafein, maka jumlah masyarakat yang mengkonsumsi kopi semakin rendah.

*Twitter* yang merupakan salah satu *microblog* yang memiliki banyak

pengguna didunia. Penggunaan *twitter* di Indonesia menempati peringkat ke 5. Dimana posisi Indonesia berada di antara USA, Brazil, Jepang dan Inggris. Pengguna *twitter* di Indonesia mencapai 19,5 juta pengguna di Indonesia dari total 500 juta pengguna global yakni *instagram*, *facebook*, dan lainnya (Kominfo, 2018). Akhi-akhir ini, berdasarkan berita kompas tahun 2019 timbul beberapa persepsi masyarakat yang sempat menjadi *trending topic* di *twitter* mengenai produk kopi di Indonesia yang meningkat tajam. Pemantauan dilakukan menggunakan platform *Evello Inteligent Tagging System*. Terdapat 43.965 *tweet* membicarakan tentang produk kopi. Dari seluruh masyarakat dapat di manfaatkan dengan menganalisis sentimen terhadap produktivitas kopi. Hal ini, dilakukan sesuai dengan persepsi serta tanggapan masyarakat disemua kalangan berdasarkan produk kopi menggunakan big data *twitter*.

Analisis sentimen juga disebut dengan *opining mining* yang tujuannya untuk menganalisis, memahami, mengolah dan mengekstrak data tekstual yang berupa opini terhadap entitas seperti halnya dengan organisasi dan topik tertentu untuk memperoleh suatu informasi. Analisis sentimen juga bertujuan pada pengolahan opini yang mengandung polaritas yakni memiliki nilai sentimen positif maupun negatif. Biasanya, masalah yang ada didalam analisis sentimen sulit untuk didefinisikan, ditentukan oleh konsep masalah, sub masalah, dan tujuannya yang berfungsi sebagai kerangka kinerja dalam berbagai penelitian lain (Lin, Wang and Zhou, 2016).

Penelitian yang dilakukan oleh (Ma'arif, 2016) mengklasifikasikan judul artikel dengan menggunakan metode *Naïve Bayes Classifier* dengan *Support Vector Machine*. Dari hasil penelitiannya menunjukkan bahwa NBC memiliki nilai akurasi performa akurasi yang lebih baik.

Penelitian analisis sentimen juga telah dilakukan Yanuar, Saiful, dan Rahmat (2018) menggunakan *Naïve Bayes Classifier* pada klasifikasi sentimen, pernah dilakukan penelitian dalam prosiding IOP Conference dengan judul Sistem Analisis Sentimen Ulasan Film dalam Bahasa Indonesia diperoleh dari data situs Movienthusiast berupa 1201 *dataset* yang terdiri dari 783 sentimen positif dan 483 sentimen negatif. Diperoleh hasil klasifikasi dengan akurasi nilai 88,37% dalam lima kali pengujian menggunakan *dataset* tersebut (Yanuar, dkk 2018).

Selanjutnya penelitian lain oleh Artanti dkk (2018), yakni penggunaan *Naïve Bayes Classifier* dalam melakukan klasifikasi sentimen yang berjudul “Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma *Naïve Bayes*”. Tujuan dari penelitian ini untuk mengklasifikasikan sentimen ke dalam kelas negatif, positif dan netral. Data yang digunakan sebanyak 1200 data. Hasil yang diperoleh dalam penelitian tersebut menggunakan metode *Naïve Bayes Classifier* memiliki akurasi nilai yang baik sebesar 93,3% (Artanti *et al.*, 2018).

Demikian juga penelitian lain oleh (Saputra and Ayuningtias, 2016) melakukan penelitian mengenai Penerapan Algoritma *Naïve Bayes* untuk Penentuan Calon Penerima Beasiswa Pada SMK Pasim Plus Sukabumi dengan data testing diperoleh hasil dari evaluasi dan validasi diketahui bahwa memiliki nilai akurasi dan AUC yang tinggi sebesar 96,67%. Dengan demikian metode *Naïve Bayes* layak digunakan untuk menentukan calon penerima beasiswa lebih efektif dan efisien. Untuk penggabungan metode dilakukan oleh (Devita, Herwanto and Wibawa, 2018) tentang penggunaan data *Text Mining* dalam Seminar Nasional Aplikasi Teknologi Informasi Tahun 2016 yang membahas tentang penerapan *Text*

*Mining* menggunakan 2 metode yakni *Naïve Bayes Classifier* dan *Support Vector Machine (SVM)*. Secara keseluruhan diperoleh hasil *F-measure* pada metode *Naïve Bayes Classifier* yakni nilai akurasi yang tinggi.

Serta penelitian berikutnya dilakukan oleh Devita dkk (2018) yaitu perbandingan metode *naïve bayes* dan *K-Nearest Neighbor* untuk klasifikasi artikel berbahasa Indonesia. Metode *naïve bayes* digunakan karena menghasilkan nilai akurasi yang maksimal dengan data latih yang sedikit dengan tingkat akurasi sebesar 70% (Devita, Herwanto and Wibawa, 2018).

Penelitian lain oleh Sondakh (2016) membandingkan kinerja lima algoritma klasifikasi teks berbasis pembelajaran mesin yakni *Decision Rules*, *Decision Tree*, *K-Nearest Neighbor*, *Naïve Bayes* Dan *Support Vector Machine*, dengan ukuran efektifitas yang digunakan adalah *precision*, *recall*, *F-measure* dan akurasi. Hasil yang didapatkan dari kelima algoritma tersebut adalah nilai akurasi tertinggi adalah pada metode *Naïve Bayes* persentase 90/10 (Sondakh, 2016).

Setianingrum dkk (2018) melakukan penelitian yakni implementasi algoritma multinomial *Naïve Bayes Classifier* untuk mengklasifikasi surat keluar sehingga dapat menentukan nomor surat secara otomatis. Diperoleh hasil pengujian yang menunjukkan implementasi pada sistem klasifikasi surat memiliki tingkat *accuracy*, *precision*, *recall* dan *F-measure* berturut-turut sebesar 89,58%; 79,1%; 78,72% dan 77,05% (Setianingrum, Kalokasari and Shofi, 2018)

Penelitian terdahulu diatas, dapat diterapkan teknik klasifikasi teks menggunakan *Multinomial Naïve Bayes*. Klasifikasi teks adalah proses pengelompokan dokumen ke dalam kategori-kategori yang berbeda. Teknik klasifikasi yang baik digunakan adalah

*Multinomial Naïve Bayes*. Algoritma ini pengembangan dari algoritma *Naïve Bayes* yang mempunyai keunggulan dalam memproses teks, mudah diimplementasikan serta memiliki nilai akurasi tinggi. *Multinomial Naïve Bayes* juga dapat menangkap ukuran kosakata dalam jumlah yang besar serta mereduksi tingkat eror yang tinggi (Suryanendra, 2018). Oleh karena itu, pada penelitian ini penulis akan melakukan Klasifikasi Analisis Sentimen pada data *twitter* dengan metode *Multinomial Naïve Bayes* dengan studi kasus konsumen kopi di Indonesia.

## TINJAUAN PUSTAKA

### 1. Produksi Kopi

Produksi kopi Indonesia tidak respon terhadap perubahan harga kopi dan komoditas substitusi dipasar domestik, tingkat upah dan luas area. Penawaran kopi di Indonesia dipengaruhi oleh tingkat teknologi dan jumlah penawaran setahun sedangkan pengaruh harga kopi dan teh secara statistik tidak berpengaruh nyata. Tanda koefisien peubah teh yang negatif menunjukkan bahwa kopi dan teh di Indonesia merupakan *competing produc*. Untuk jenis kopi yang sering menjadi perbincangan di sosial media terutama pada *twitter* meliputi kopi pahit, kopi sakit, kopi luwak dan kopi hitam. Dari topik kopi memiliki arti masing-masing. Kopi pahit adalah produk kopi yang rasanya pahit atau tanpa gula. Kemudian, kopi sakit merupakan hastag yang menjadi trending topik di media sosial khususnya *twitter*, kopi sakit artinya seseorang akan sakit setelah mengkonsumsi kopi. Untuk kopi luwak adalah produk kopi yang berasal dari biji luwak yang sampai saat ini banyak digemari masyarakat di semua kalangan. Dan kopi hitam adalah produk kopi yang berasal dari biji kopi asli yang berwarna hitam yang biasanya sering diseduh oleh orang dewasa (Pradipta *et al.*, 2017).

### 2. Konsumen Kopi

Menurut Zulfi, dkk (2018), perilaku konsumen yaitu sebagian dari pemahaman terhadap tindakan secara langsung oleh konsumen dalam memperoleh, mengkonsumsi, dan juga menjadikan produk dan jasa yang diperoleh habis. Hasil dari pengolahan bahan minuman dari sector pertanian sangat diminati masyarakat untuk membantu proses metabolisme tubuh, menghilangkan dahaga serta sebagai bentuk kebiasaan masyarakat tersebut. Salah satu jenis minuman yang populer di kalangan masyarakat saat ini adalah kopi (Zulfi *et al.*, 2018).

Mengkonsumsi kopi tidak hanya untuk menghilangkan dahaga, tetapi juga untuk menikmati aktivitas kehidupan masyarakat. Misalnya dalam kebiasaan sehari-hari, serta dalam acara kantor. Hal ini menjadi tren untuk mengkonsumsi kopi yang terus meningkat. (Suisa *et al.*, 2016). Tetapi sebagian masyarakat suka terhadap kopi karena beberapa alasan. Ada beberapa faktor yang menjadikan masyarakat enggan untuk minum kopi. Faktor tersebut yaitu rasa pahit terhadap kopi, penyebab *maag*, perut kembung dan jantung yang berdebar-debar, bahkan menjadikan susah tidur dan sebagainya. Hal ini tentu membuat persepsi masyarakat terhadap kopi berbeda-beda (Pradipta *et al.*, 2017).

### 3. Text Mining

*Text mining* adalah proses pengetahuan intensif yang memungkinkan pengguna berinteraksi dan bekerja dari waktu ke waktu pada sekumpulan dokumen menggunakan berbagai macam analisis. *Text mining* atau *text analytics* adalah istilah dari cara mengekstraksi atau mendeskripsikan sebuah teknologi yang mampu dalam menganalisis sebuah teks semi-terstruktur maupun tidak terstruktur, hal inilah yang membedakannya dengan *data mining*

dimana data mining mengolah data yang sifatnya terstruktur. *Text mining* juga sebagai proses untuk menemukan suatu informasi atau tren terbaru yang sebelumnya tidak diketahui dengan memproses dan menganalisis data yang sejumlahnya besar atau *big data* (Feldman dan Sanger, 2007).

#### 4. Analisis Sentimen

Analisis sentimen adalah metode untuk menganalisis sebagian data untuk mengetahui emosi manusia. Analisis sentimen dapat dikategorikan ke dalam tiga *task*, yaitu *informative text, detection, information extraction dan sentimen interestiness classification (emotional, polarity identification)*. *Sentimen classification* (negatif dan positif) digunakan untuk memprediksi sentimen *polarity* berdasarkan data sentimen dari pengguna (Liu, 2012).

*Sentimen Analysis* atau *opinion mining* mengacu dalam bidang yang meluas mulai pengolahan Bahasa alami, komputasi *linguistic* dan *text mining* yang tujuannya untuk menganalisa sentimen, pendapat, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara tersebut atau penulis dapat memberikan suatu topik, produk, layanan, organisasi, individu ataupun kegiatan tertentu (Liu, 2012).

#### 5. Klasifikasi

Teknik klasifikasi adalah salah satu teknik data mining yang termasuk *supervised learning*. Dimana, *supervised learning* artinya proses pembentukan sebuah korespondensi menggunakan sebuah *training* dataset. Tujuannya adalah untuk memprediksi target dari beberapa atribut (Zaki & Meira, 2014). Terdapat pada dua pekerjaan utama pada klasifikasi yaitu melakukan *training* untuk disimpan sebagai prediksi dan melakukan *testing* untuk proses klasifikasi agar diketahui di label mana objek data tersebut (Liu, Loh, & Sun, 2009).

#### 6. K-fold Cross Validation

*K-fold cross validation* adalah salah satu metode statistik yang digunakan untuk mempartisi data menjadi data *training* dan data *testing*. Pada metode ini sering digunakan peneliti karena dapat mengurangi bias yang terjadi dalam pengambilan sampel. Metode ini digunakan secara berulang-ulang membagi data menjadi dua yakni data *training* dan data *testing*, setiap data memperoleh kesempatan menjadi data *testing* (Wibowo, 2017).

*10-fold cross validation* adalah salah satu *K-fold cross validation* yang direkomendasikan untuk pemilihan model terbaik karena cenderung memberikan estimasi akurasi yang kurang bias dibandingkan dengan *cross validation* biasa, *leave-one-out cross validation* dan *bootstrap*. Dalam *10-fold cross validation*, data dibagi menjadi *10-fold* berukuran kira-kira sama, sehingga kita memiliki *10 subset* data untuk mengevaluasi kinerja model atau algoritma. Untuk masing-masing dari *10 subset* data tersebut, *cross validation* akan menggunakan *9-fold* untuk pelatihan (data *training*) dan *1-fold* untuk pengujian (data *testing*) (Wibowo, 2017).

#### 7. Naïve Bayes Classifier

*Naïve bayes classifier* merupakan Teknik memprediksi dalam mengklasifikasi berbasis probabilistik sederhana berdasarkan pada penerapan teorema *Bayes* (aturan *Bayes*) dengan asumsinya yaitu independensi (ketidaktergantungan) yang kuat (naif). *Naïve Bayes Classifier* mengasumsikan bahwa kehadiran (atau ketiadaan) fitur tertentu dari suatu kelas yang tidak saling berhubungan dengan kehadiran fitur lainnya (Kaku, Mulyanto, & Rohandi, 2014).



Berikut prediksi *Naïve Bayes* didasarkan pada teorema *Bayes* dengan rumus untuk klasifikasi sebagai berikut :

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)}$$

Sedangkan *Naïve Bayes* dengan fitur kontinu memiliki formula sebagai

berikut :

$$P(X|Y) = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-(x-\mu)^2}{2\sigma^2}$$

Dengan keterangan

- P : Parameter  
P (Y|X) : Probabilitas data dengan atribut X pada kelas Y (*posterior probability*)  
P (Y) : Probabilitas awal kelas Y (*prior probability*)  
σ : Standar deviasi  
μ : *mean* atau nilai rata-rata dari atribut dengan fitur kontinu  
 $\prod_{i=1}^q P(X_i|Y)$  : Probabilitas independent kelas Y dari semua fitur dalam vektor X (*Likelihood*)

## 8. Multinomial Naïve Bayes

*Multinomial naïve bayes* merupakan proses pengambilan jumlah kata yang muncul dalam setiap dokumen, dimana metode ini mengansumsikan dokumen yang memiliki beberapa kejadian dalam kata dengan panjang tidak bergantung dari kelasnya dalam dokumen tersebut.

Berikut formula dalam *multinomial naïve bayes*, dimana untuk memperkirakan *prior* probabilitas P(c) dapat dilihat formula berikut

$$P(c) = \frac{Nc}{N}$$

Keterangan :

- Nc : jumlah dokumen *training* dalam kelas c

N : jumlah keseluruhan dokumen *training* dari seluruh kelas

Untuk memperhatikan *conditional probability* P(t|c) dinyatakan dengan rumus berikut :

$$P(t|c) = \frac{Tct}{\sum_{t' \in V} Tct}$$

Keterangan

Tct : jumlah kemunculan kata t dalam sebuah dokumen *training* pada kelas c

$\frac{Tct}{\sum_{t' \in V} Tct}$  : jumlah total keseluruhan kata dalam dokumen *training* pada kelas c

t' : jumlah total kata dalam dokumen *training*

Untuk menghilangkan nilai nol pada sebuah dokumen. Digunakan *laplace smoothing* sebagai proses penambahan nilai 1 pada setiap nilai Tct pada perhitungan *conditional probability* yang dinyatakan dengan formula berikut

$$P(tPtk|c) = \frac{Tct+1}{\sum_{t' \in V} Tct+B'}$$

Keterangan

B' : total kata unik pada keseluruhan kelas dalam dokumen *training*

Untuk mendapatkan nilai probabilitas yang tinggi pada setiap kata maka digunakan *laplace smoothing* atau *add-one, laplace smoothing* ini digunakan agar nilai dari probabilitas masing-masing kata dapat memenuhi syarat yaitu tidak sama dengan nol. Jadi, jika suatu nilai pada probabilitas kata adalah nol, maka data tidak baik pada data *training* maupun *testing* juga tidak akan pernah cukup untuk mewakili frekuensi saat terdapat kejadian langka.

## 9. Ukuran Hasil Evaluasi

Evaluasi pada suatu klasifikasi pada umumnya dilakukan dengan menggunakan sebuah himpunan data yang diuji, tidak digunakan dalam pelatihan klasifikasi tersebut, pada suatu ukuran tertentu. Pada tahap ini terdapat sejumlah ukuran yang dapat digunakan untuk menilai kembali atau mengevaluasi model klasifikasi, yakni *accuracy* atau tingkat pengenalan, *error rate* atau tingkat kesalahan atau kekeliruan klasifikasi, *recall* atau *sensitivity* atau *true positif*, *specificity* atau *true negative* dan *precision* (Lim, dkk 2006).

Ada beberapa formula dalam perhitungan performa klasifikasi yaitu nilai akurasi, presisi, dan *recall* biasa ditampilkan dalam presentase :

a. *Accuracy*

Akurasi adalah nilai ketepatan dimana, pengguna memprediksi suatu kata sesuai dengan jawaban suatu system (Lim, dkk 2006). Berikut formula nilai akurasi

$$Accuracy = \frac{TP+FN}{TP+FP+TN+FN}$$

b. *Precision*

*Precision* adalah proposi jumlah dokumen teks yang relevan yang dikenali diantara semua dokumen teks yang dipilih oleh sistem atau kesesuaian terhadap system (Lim, dkk 2006).

$$Precision = \frac{TP}{TP+FP}$$

c. *Recall*

*Recall* adalah proporsi jumlah dokumen teks yang relevan dikenal diantara semua dokumen teks relevan yang ada pada koleksi (Lim, dkk 2006).

$$Recall = \frac{TP}{TP+FN}$$

## METODE PENELITIAN

### 1. Sumber Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder. Data diambil dari *database Tweet* menggunakan *Twitter API (Application Programming Interface)* pada *tweet* mengenai kata kunci produk kopi yakni ‘Kopi Luwak’, ‘Kopi Sakit’, ‘Kopi Hitam’ dan ‘Kopi Pahit’. Sampel yang digunakan adalah data yang diambil sebanyak satu bulan November 2019.

### 2. Variabel Penelitian

Variabel yang digunakan pada penelitian ini sebanyak dua atribut dan satu label. Atribut yang digunakan antara lain tanggal komentar, komentar. Label yang digunakan adalah data hasil pelabelan dari analisis sentimen. Jenis data label adalah kategorik yakni positif dan negatif.

### 3. Struktur Data

Berikut adalah struktur data penelitian sebelum *preprocessing* :

**Tabel 1. Struktur Data Penelitian**

No	Tweet	Sentimen
1	Siaap dinikmati ya.. kopi luwak yang cocok menemaniku mengerjakan skripsi...	Positif
2	Sob, sudah pernah cobain kopi hitam? Kopi hitam warna hitam, tapi rasanya maknyuss poll	Positif
3	Tiap kali minum kopi pasti mules teruss	Negatif
4	Aku fikir minum kopi arabika lebih lumayan dari hitam ternyata sama aja, asem juga	Negatif

#### 4. Langkah-langkah Penelitian

Langkah analisis data yang digunakan dalam penelitian ini sebagai berikut :

1. Mengambil data *tweet* dengan menggunakan *Twitter API*
2. Menyiapkan data *tweet*, daftar *stopword* dan kata dasar. Kemudian diperoleh *stopword* dengan kata dasar Bahasa Indonesia baku
3. Praproses Teks
4. Lalu dilakukan seleksi dengan TF-IDF yaitu merubah teks menjadi vektor dan pembobotan kata
5. Pembagian data *training* dan data *testing* menggunakan *10-Fold Cross Validation* dengan pembagian data 90% : 10%
6. Klasifikasi data menggunakan *Naïve Bayes Multinomial Classifier* untuk masing-masing *keyword*
7. Melakukan evaluasi hasil klasifikasi
8. Melakukan visualisasi *tweet* dengan *wordcloud*
9. Interpretasi dan mengambil kesimpulan

#### HASIL DAN PEMBAHASAN

##### 1. *Scrapping data twitter*

*Scrapping data twitter* menggunakan *Twitter API (Application Programming Interface)* yang diakses dengan program R *consule* sebagai data teks yang akan diolah. Data yang digunakan adalah *tweet* trending topik tentang kopi selama bulan November 2019 yakni kopi sakit, kopi pahit, kopi hitam dan kopi luwak diperoleh sebanyak 3275 *tweet*.

##### 2. *Preprocessing data*

Berikut adalah hasil *preprocessing data twitter* :

**Tabel 2. Contoh Hasil *Preprocessing Data Tweet***

Hasil Case Folding dan Tokenisasi	
Abis minum kopi hitam	selesai minum
aku malah mules banget	kopi hitam aku

rasa mulesnya kayak	jadi mules
nanggung mules 5	seperti
minggu	menanggung
HASHAJXSJCJDJ	mules 5 minggu

##### Hasil Stemming dan Filtering

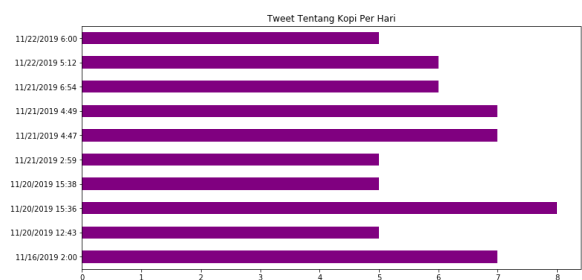
selesai minum kopi	selesai minum
hitam aku jadi mules	kopi hitam aku
seperti menanggung	jadi mules
mules 5 minggu	seperti
	nanggung
	mules 5 minggu

##### Hasil Data *Preprocessing*

selesai minum kopi	selesai minum
hitam aku jadi mules	kopi hitam aku
seperti nanggung mules	jadi mules
5 minggu	seperti
	nanggung
	mules 5 minggu

#### 3. Analisis Deskriptif

Setelah melakukan tahap *preprocessing* maka dilanjutkan dengan analisis deskriptif pada *tweet* pada kata yang memiliki frekuensi kemunculan paling banyak. Analisis deskriptif dilakukan untuk mengetahui gambaran umum mengenai *tweets* tentang kopi di media sosial *twitter*. Informasi dapat diambil adalah jumlah *tweet* yang diberikan konsumen perharinya dengan frekuensi 10 *tweet*. Analisis deskriptif dapat dilihat pada grafik berikut :



**Gambar 2. Grafik *Tweet* tentang kopi**

Berdasarkan hasil grafik diatas dapat dilihat *tweet* perhari dengan *tweet* terbanyak pada tanggal 20 bulan



## KESIMPULAN

Berdasarkan hasil analisis dan pembahasan, dapat diambil kesimpulan sebagai berikut :

1. Gambaran umum yang didapatkan dari hasil analisis deskriptif pada *tweet* per hari tentang kopi paling banyak pada bulan November tanggal 22 dengan waktu yang berbeda-beda. Sedangkan untuk *tweet* yang sering muncul pada sentimen negatif adalah sakit perut, sedangkan sentimen positif adalah kopi luwak dan kopi hitam.
2. Hasil klasifikasi sentimen pada *tweet* tentang kopi di Indonesia menggunakan Algoritma *Multinomial Naive Bayes* dengan pembagian data menggunakan *10-fold Cross Validation* diperoleh hasil tingkat akurasi sebesar 94%, *precision* pada sentimen positif sebesar 97% dan sentimen negatif sebesar 86%, sedangkan pada nilai *recall* setiap kelas, kelas positif sebesar 96% dan kelas negatif sebesar 86%, yang berarti bahwa tingkat keberhasilan sistem data opini positif dalam menentukan kembali hasil sebuah informasi yaitu hanya memiliki sedikit kesalahan saat proses klasifikasi. Hal ini menunjukkan bahwa kemampuan prediksi dengan menggunakan Algoritma *Multinomial Naive Bayes* layak digunakan untuk dijadikan metode rujukan dalam mengolah data *text*.
3. Dari hasil mengklasifikasi ke dalam sentimen positif dan negatif pada *tweet* tentang kopi di Media Sosial *Twitter* diperoleh ulasan negatif yang sering muncul tentang kopi diantaranya, setelah minum kopi susah tidur, kopi dapat menyebabkan sakit perut dan sakit kepala dan terutama pada kopi hitam.

## DAFTAR PUSTAKA

- APJII. (2018). Penetrasi & Profil Perilaku Pengguna Internet Indonesia. *Apjii*. Retrieved from [www.apjii.or.id](http://www.apjii.or.id)
- Arkhamsiagustinah. 2015. *Wordcloud of Corruption Eradication Commission*. Makalah Pengantar Data Mining Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.
- Anonim. Kompas. Kamis, 31 Oktober 2019, *Pengguna Twitter Indonesia Diklaim Terbanyak* (diakses Oktober-November 2019) <https://tekno.kompas.com/read/2019/10/30/16062477/pengguna-aktif-harian-twitter-indonesia-diklaim-terbanyak>
- Artanti, Dian Putri, Abdul Syukur, Adi Prihandono, and De Rosal I Moses Setiadi. 2018. "Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes". *Jurnal Konferensi Nasional Sistem Informasi*. Universitas Dian Nuswantoro (Hal. 8–9).
- Dewi, Febriana Ira, Faisal Anwar, dan Leily Amalia. 2009. "Presepsi Terhadap Konsumsi Kopi Dan The Mahasiswa Tpb-Ipb Tahun Ajaran 2007-2008". *Jurnal Gizi dan Pangan*, Institut Pertanian Bogor.
- Devita, Riri Nada, Heru Wahyu Herwanto, and Aji Prasetya Wibawa. 2018. "Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia." *Jurnal Teknologi Informasi Dan Ilmu Komputer* 5 (4): 427.

- Feldman, R., & Sanger, J. 2007. *The Text Mining Handbook Advanced Approaches In Analyzing Unstructured Data*. New York : Cambridge University Press.
- Han, J., & Kamber, M. 2006. *Data Mining : Concepts and Techniques Second Edition*. San Francisco : Morgan Kauffman.
- International Coffe Organization. 2018. <http://www.ico.org/> (diakses Oktober-November 2019).
- Kustiari, Reni. 2016. "Perkembangan Pasar Kopi Dunia Dan Implikasinya Bagi Indonesia." *Forum Penelitian Agro Ekonomi* 25 (1): 43. <https://doi.org/10.21082/fae.v25n1.2007.43-55>.
- Kementerian Pertanian. 2018. Outlook Kopi Komoditas Pertanian Subsektor Perkebunan. hal 54-55. (diakses Oktober-November 2019)
- Lin, Yuming, Xiaoling Wang, and Aoying Zhou. 2016. "Opinion Spam Detection." *Opinion Analysis for Online Reviews*.
- Liu, B. *Sentimen Analysis and Opinion Mining*. Diakses bulan Oktober 2019, Tersedia di <https://www.cs.uic.edu/~liub/FB/S/SentimentAnalysis-and-OpinionMining.pdf>.2012
- Lim, S.Y., Song, M.H., & Lee, S.J. 2006. *Ontology-based automatic classification of web document*. Springer-Verlag, 690-700.
- Liu, Y., Loh, H. T., & Sun, A. 2009. Imbalanced text classification: A term weighting approach. *Expert Systems with Applications*, 36(1), 690-701.
- Ma'arif, Muhammad Rifqi. 2016. Perbandingan Naïve Bayes Classifier dan Support Vector Machine Untuk Klasifikasi Judul Artikel. Program Studi Manajemen Informatika STMIK Jenderal Achmad Yani Yogyakarta.
- Nurdiansyah, Yanuar., Bukhori, Saiful., & Hidayat, R. 2018. Sentiment Analysis System for Movie Review In Bahasa Indonesia using Naïve Bayes Classsifier Method. *IOP Conf.Seies: Journal of Physics* (hal. 1008). United Kingdom : IOPscience.
- Nuritha, I. 2017. *Pemanfaatan Text Mining dalam Analisis Sentimen Presepsi Masyarakat Terhadap Kopi Organik dengan Naïve Bayes Classifier*. Jember : Universitas Jember.
- Pradipta, Kristianto. 2017. "JURNAL REVIEW PERBEDAAN AIR SEDUH TERHADAP PERSEPSI MULTISENSORIS KOPI". *Jurnal Pangan dan Agroindustri Vol 5 No1* : 85-91.
- Syaad., Jujur. 2013. Pengembangan Potensi Kopi Sebagai Komoditas Unggulan Kawasan Agropolitan Kabupaten Dairi, *Jurnal Ekonomi dan Keuangan*.
- Saraswati, N.W.S. 2011. Text Mining Dengan Metode Naïve Bayes Classifier Dan Support Vector Machines Untuk Sentimen Analysis, *Tesis*, Universitas Udayana, Denpasar.
- Suryanendra, Adjie. 2018. *Identifikasi Presepsi Konsumen Kopi Di Media Sosial Twitter Menggunakan Analisis Sentimen Dengan Metode Naïve Bayes Multinomial*. Skripsi.

- Program Studi Sistem Informasi  
Fakultas Ilmu Komputer  
Universitas Negeri Jember.
- Setianingrum, Anif Hanifa, Dea Herwinda Kalokasari, and Imam Marzuki Shofi. 2018. "Implementasi Algoritma Multinomial Naive Bayes Classifier." *Jurnal Teknik Informatika* 10 (2): 109–18.
- Sondakh, Debby Erce. 2016. "Comparative Study of Classification Algorithms: Holdouts As Accuracy Estimation." *CogITO Smart Journal* Vol 1, No. 1 : 13.
- Saputra, Rizal Amegia, and Shinta Ayuningtias. 2016. "Penerapan Algoritma Naive Bayes Untuk Penentuan Calon Penerima Beasiswa Pada Smk Pasim Plus Sukabumi". *Swabumi* IV No. 2 (Hal 114–20).
- Suisa, Kelvianto, Veronica Febrilia. 2018. "Gaya Hidup Minum Kopi Konsumen Di," *Manajemen Perhotelan*, and Universitas Kristen Petra. 326–43.
- Tambunan, Salmiah. 2018. "Analisis Sikap Konsumen Terhadap Kopi Bubuk Lokal Dan Kopi Instan Di Kota Pematangsiantar". *Jurnal Agribisnis Fakultas Pertanian Universitas Sumatera*.
- United States Department of Agriculture. 2018. <http://www.usda.gov/> (diakses Oktober-November 2019)
- Zaki, M. J., & Meira, W. J. 2014. *Data Mining and Analysis: Fundamental Concepts and Algorithms*, 562 Retrieved from <http://book.google.com/book?hl=en&lr=&id=Gh9GAWAAQBAJ&pgis=1>
- Zafikri, A. 2008. *Implementasi Metode Term Frequency Inverse Document Frequency (TF-IDF) Pada Sistem Temu Kembali Informasi*. Skripsi. Program Studi S-1 Ilmu Komputer FMIPA USU.
- Zulfi, Jivana., Kusnandar. 2018. "Analisis Preferensi Konsumen Terhadap Pembelian Kopi Instan White Coffee Di Kecamatan Kebumen Kabupaten Kebumen" 14 (2): 159–66. Program Studi Agribisnis Fakultas Pertanian Universitas Sebelas Maret.