# LAPORAN PROYEK AKHIR PRAKTIKUM DATA SCIENCE

**Analisis Sentiment Dengan Klasifikasi Naïve Bayes Pada *Review* Pesawat Terbang**



ISNAINI KHAIRIAH 123190004

AGUNG RIVALZO 123190017

# PROGRAM STUDI INFORMATIKA JURUSAN TEKNIK INFORMATIKA FAKULTAS TEKNIK INDUSTRI

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” YOGYAKARTA**

**2021**

## PENDAHULUAN

Pada era teknologi yang semakin meningkat dan diikuti oleh perkembangan TI yang semakin modern berpengaruh terhadap kehidupan individu maupun kelompok. Perkembangan TI dapat membantu dalam mengembangkan tugas-tugas baru pada perusahaan berskala pasar global atau pada instansi pemerintah, TI dimanfaatkan untuk mengatasi meningkatkan kemampuan seseorang dalam meraih keunggulan dan kesuksesan yang handal serta canggih.

Pemanfaatan teknologi informasi juga digunakan dalam instansi pemerintahan yang dapat digunakan disegala aspek dengan menggunakan data yang telah tersedia, data ini akan diolah dengan teknologi yang akan mengahasilan informasi transparan, akurat dan mudah dipahami oleh semua kalangan.

Pesawat terbang merupakan salah satu sarana transportasi udara yang banyak digunakan oleh masyarakat. Tingginya minat masyarakat untuk menggunakan pesawat terbang sebagai alat transportasi jarak jauh yang cepat dipengaruhi oleh semakin tingginya tingkat mobilitas masyarakat yang menuntut untuk dapat berada di kota tertentu dalam waktu yang singkat. Hal tersebut sejalan dengan data statistik jumlah penumpang pesawat terbang di Indonesia yang mencapai 94.722.823 penumpang (Oktober 2017) yang meningkat 6,1% dibanding tahun 2016 yang berjumlah 89.281.491 penumpang (Direktorat Jenderal Perhubungan Udara, 2017). Peningkatan jumlah penumpang tersebut tidak sepenuhnya menjamin tingkat keamanan dan keselamatan bagi penumpang.

Hal itu antara lain dapat dibuktikan dengan insiden kecelakaan pesawat AirAsia Indonesia pada 28 Desember 2014 yang mengakibatkan 162 penumpang meninggal dunia (Kompas.com, 2017). Untuk mengetahui kualitas layanan jasa maskapai penerbangan, baik dari segi keselamatan, keamanan, maupun kenyamanannya, para calon penumpang dapat memperhatikan opini penumpang sebelumnya. Saat ini opini dan pengalaman yang ditulis oleh pengguna jasa penerbangan online terus meningkat. Opini tersebut banyak dituangkan pada forum-forum yang ada di internet, antara lain situs web yang bernama Airlinequality.

Airlinequality adalah perusahaan yang melakukan riset mengenai peringkat dan kualitas maskapai penerbangan sipil. Hasil riset tersebut didasarkan pada data situs web yang dikembangkannya. Di situs ini terdapat kolom komentar yang dapat digunakan pengguna jasa maskapai untuk mengomentari kualitas dan layanan maskapai di seluruh dunia, sehingga calon penumpang lainnya dapat menggunakannya sebagai referensi saat mengevaluasi kualitasnya. Dengan membaca ulasan tersebut secara keseluruhan tentu memakan waktu yang lama. Jika hanya sedikit ulasan yang dibaca, evaluasi yang dihasilkan juga akan bias. Usaha mengumpulkan semua ulasan tersebut dan mengolah menjadi informasi yang dapat menjawab keingintahuan para pengguna jasa sangat dibutuhkan untuk mengetahui kualitas maskapai yang nantinya akan dipilih. Usaha tersebut dapat dicapai dengan menggunakan teknik analisis sentimen. Analisis sentimen, sering pula disebut sebagai penambangan opinion mining, opinion mining adalah bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi orang-orang terhadap entitas seperti produk, layanan, organisasi, individu, isu, peristiwa, atau topik tertentu.

Analisis sentimen akan mengelompokan polaritas dari teks yang ada dalam kalimat atau dokumen. Dengan demikian, kecondongan teks untuk bersifat positif atau negatif dapat diketahui dengan mudah. Melakukan analisis sentimen menggunakan metode Naive Bayes yang diimplementasikan pada 10 koleksi data yang diambil dari website Airlinequality. Nive Bayes sendiri merupakan salah satu metode klasifikasi dalam teknik data mining yang dikembangkan oleh ilmuwan Inggris bernama Reverend Thomas Bayes pada abad ke-18 yang secara umum dilakukan dengan pendekatan peluang atau probabilitas dan dapat digunakan untuk melakukan prediksi di masa yang akan datang dengan menggunakan pengalaman di masa sebelumnya (Sipayung et al., 2016).. Penelitian tersebut mampu menghasilkan rata-rata akurasi sebesar 78%. Pada penelitian ini, metode yang digunakan dalam analisis sentimen opini pengguna maskapai penerbangan juga Naive Bayes. Tahapan-tahapan dalam penelitian ini yaitu pemrosesan awal terhadap data, pengekstraksian fitur, pengelompokan data dengan Naaive Bayes dan pengevaluasian hasil.

## METODE

Objek penelitian yang digunakan pada penelitian ini yaitu sebuah data hasil *scraping* darisebuah *website*. Data hasil *scraping* tersebut terdiri dari data *reviews* dan data *reviewers*.Pada proyek akhir ini, yang dilakukan proses *sentiment* hanya pada data *reviews* saja.

Metode pengumpulan data yang digunakan yaitu dengan mengumpulkan data-data yang dapat mendukung dalam proses penelitian. Data-data yang diambil yaitu data relevan dengan penelitian dan sesuai kebutuhan penelitian. Pengumpulan data diambil dari studi literatur dimana dari studi literatur tersebut dapat dijadikan sebagai pertimbangan dalam melakukan penelitian. Sedangkan objek penelitian didapatkan dari hasil scraping sebuah website, yaitu www.airlinequality.com.

Pengolahan data dilakukan setelah data yang dibutuhkan penelitian sudah didapatkan. Tahap-tahap pengolahan data sebagai berikut :

1. *Pre-processing Data* atau *Data Cleaning*

Pada tahap ini, kolom dataset yang sudah dipilih menjadi objek penelitian dilakukan proses *cleaning* atau proses menghapus sesuatu yang tidak diperlukan. *Cleaning* data yang dilakukan pada objek penelitian ini berupa mengubah huruf besar menjadi huruf kecil, menghapus angka, menghapus tanda baca, menghapus *stopwords* dan mengubah *blank space* menjadi *strip*. Dalam melakukan tahapan ini, diimplementasikan dengan bantuan library tm (*corpus*) yang sudah disediakan oleh bahasa pemrograman R.

1. Analisis *sentiment*

*Analisis sentiment* merupakan tahapan utama dari penelitian ini. Setelah data berhasil dibersihkan, kemudian dilanjutkan tahap *sentiment* dimana pada tahap ini dilakukan perhitungan *sentiment* di setiap baris pada kolom *review* atau kolom yang menjadi objek penelitian dengan menggunakan *library* *sentimentr*. Hasil dari *sentiment* ini, digunakan untuk menentukan apakah kalimat yang diidentifikasi termasuk ke dalam kategori kalimat positif atau negatif. Apabila hasil sentiment lebih dari 0 maka dapat dikatakan kalimat tersebut termasuk kalimat positif. Apabila hasil sentiment kurang dari 0 maka dapat dikatakan kalimat tersebut termasuk kalimat negatif.

1. *Naïve Bayes Classifier*

Metode Naïve Bayes digunakan untuk proses klasifikasi dari hasil sentiment yang telah dilakukan pada tahap sebelumnya. Pada tahap ini, dibutuhkannya proses membagi data menjadi 2 bagian yaitu data *training* dan data *testing*. Data *training* pada penelitian ini menggunakan 80% dari dataset yang ada. Sedangkan data *testing* yang digunakan merupakan 20% dari dataset yang ada. Data *training* digunakan untuk proses pelatihan dengan metode *Naïve Bayes*. Kemudian, hasil dari pelatihan data training dan data testing digunakan untuk melakukan proses prediksi. Dari prediksi data testing tersebut, nantinya dapat dilakukan untuk perhitungan *confusion matrix. Confusion matrix* akan menampilkan hasil seperti tingkat akurasi, *sensitivity, specificity*, dan lain sebagainya. Keuntungan menggunakan metode Naïve Bayes Classifier ini yaitu hanya membutuhkan jumlah data pelatihan (*training*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian (Saleh, 2015).

Persamaan dari teorema *Bayes* sebagai berikut :

……..(2.1)

Dimana :

X = Data (*class* belum diketahui)

H = Hipotesis data yang merupakan suatu *class* spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (*posteriori probabilitas*)

P(H) = Probabilitas hipotesis H (*prior probabilitas*)

P(X|H) = Probabilitas X berdasarkan kondisi hipotesis H

P(X) = Probabilitas X

Dalam melakukan proses pengklasifikasian dengan menggunakan metode *Naïve Bayes* dibutuhkannya sebuah persamaan dimana untuk mengetahui kelas apa yang tepat bagi *sample* yang akan dianalisis. Persamaan tersebut sebagai berikut :

…..(2.2)

Atau

…(2.3)

Dimana :

C = Representasi kelas

F1….Fn = Representasi karakteristik petunjuk yang dibutuhkan untuk klasifikasi

P(C|F1…Fn) = Peluang sampel karakteristik tertentu dalam sebuah kelas C (*Posterior*)

P( C ) = Peluang munculnya kelas sebelum masuknya sampel (*Prior*)

P(F1…Fn|C) = Peluang kemunculan karakteristik sampel pada kelas (*Likelihood*)

P(F1…Fn) = Peluang kemunculan karakteristik secara global (*Evidence*)

Nilai *evidence* akan bernilai tetap di setiap kelas pada suatu sampelnya. Lalu, untuk hasil *posterior* akan dibandingkan dengan hasil posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan.

Jika dilakukan sebuah penjabaran pada rumus 2.2, muncul sebuah hasil dimana semakin banyak dan semakin komplek faktor syarat yang mempengaruhi nilai probabilitas, maka hampir mustahil untuk dilakukannya analisa satu persatu dan perhitungan akan semakin sulit. Karena hal tersebut, digunakanlah sebuah asumsi independensi yang sangat tinggi (naif), bahwa semakin banyak faktor atau petunjuk (F1,F2,..Fn) saling bebas satu sama lain, sehingga muncul persamaan yang digunakan untuk proses klasifikasi sebagai berikut :

…(2.4)

Untuk ij, maka sebagai berikut :

…(2.5)

Untuk pengklasifikasian dengan data kontinyu, persamaan sebagai berikut :

…(2.6)

Dimana :

P = Peluang

Xi = Atribut ke-i

xi = Nilai atribut ke-i

Y = Kelas yang dicari

yj = Sub kelas Y yang dicari

= *mean* atau rata-rata dari seluruh atribut

= Deviasi standar dimana menyatakan varian dari seluruh atribut

1. *Wordcloud*

*Wordcloud* merupakan sebuah tahap dimana dari tahap ini dapat diketahui seberapa sering penggunaan kata di dalam kolom reviews pada dataset. Tahap yang perlu dilalui sebelumnya untuk mendapatkan *wordcloud* yaitu tahap cleaning text. Kemudian dari hasil cleaning text tersebut dilakukan proses pengubahan ke dalam bentuk matrix yang nantinya sebuah kata akan terpisah oleh kalimat yang diikuti dan dapat diketahui frekuensi di setiap kata. Kemudian, untuk visualisasi fungsi *wordcloud* yang sudah tersedia di dalam library *wordcloud*.

## HASIL DAN PEMBAHASAN

## KESIMPULAN

Berdasarkan dari hasil dan pembahasan pada penelitian yang dilakukan, didapatkan beberapa kesimpulan sebagai berikut :

1. Analisis *sentiment* dengan menggunakan perhitungan *sentiment* di setiap barisnya menghasilkan kategori kalimat positif kurang lebih sebesar 435 kalimat, kategori kalimat negatif sebesar kurang lebih 225 kalimat, dimana lebih sedikit jika dibandingkan dengan kalimat positif.
2. Analisis *sentiment* dengan menggunakan metode Naïve Bayes jika dilihat dari hasil prediksi maka menghasilkan kategori kalimat positif kurang lebih sebesar 159, kategori kalimat negatif kurang lebih sebesar 166.
3. Tingkat akurasi yang didapatkan pada penggunaan metode Naïve Bayes sebesar 0,8125 sehingga penggunaan metode ini dapat dikatakan berhasil dan optimal untuk proses klasifikasi analisis *sentiment*.

Link Youtube :

Link Github :

**DAFTAR PUSTAKA**

Indrayuni, E. (2016). Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Evolusi Volume 4 Nomor 2 - 2016*, *4*(2), 20–27.

Jumeilah, F. S. (2018). Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE dengan Naïve Bayes. *Jurnal Sistem Informasi Bisnis*, *8*(1), 92. https://doi.org/10.21456/vol8iss1pp92-98

Kristiyanti, D. A. (2015). ANALISIS SENTIMEN REVIEW PRODUK KOSMETIK MELALUI KOMPARASI FEATURE SELECTION . *Konferensi Nasional Ilmu Pengetahuan dan Teknologi (KNIT)* , ISBN 978-602-72850-0-2.

Muhammad Adib Imtiyazi, S. S. (2015). Sentiment Analysis Berbahasa Indonesia Menggunakan Improved Multinomial Naive Bayes . *e-Proceeding of Engineering Vol.2, No. 2*, 6333.

Musyarof, Z. (2019). *Analisis Text Mining terhadap BPS di Twitter Menggunakan R*. *July*.

Pradana, M. G. (2020). *Penggunaan Fitur Wordcloud Dan Document Term Matrix Dalam Text Mining*.

Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter . *JURNAL INOVTEK POLBENG - SERI INFORMATIKA, VOL. 3, NO. 1*, 2527-9866 .

Riany, J., Fajar, M., & Lukman, M. P. (2016). Penerapan Deep Sentiment Analysis pada Angket Penilaian Terbuka Menggunakan K-Nearest Neighbor. *Sisfo*, *06*(01), 147–156. https://doi.org/10.24089/j.sisfo.2016.09.011

Rozi, N. F., Arianto, F., & Hapsari, D. P. (2018) ANALISIS SENTIMEN PADA OPINI PENGGUNA MASKAPAI PENERBANGAN MENGGUNAKAN HYBRID CUCKOO SEARCH. Surabaya: Jurusan Teknik Informatika Institut Teknologi Adhi Tama Surabaya.

Ro’i Fahreza Nur Firmansyah, M. A. (2016). SENTIMENT ANALYSIS PADA REVIEW APLIKASI MOBILE MENGGUNAKAN METODE NAÏVE BAYES DAN QUERY EXPANSION.

Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Creative Information Technology Journal*, *2*(3), 207–217.

Saputra, H. K. (2018). Analisis Data Mining Untuk Pemetaan Mahasiswa Yang Membutuhkan Bimbingan Dan Konseling Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Teknologi Informasi Dan Pendidikan*, *11*(1), 14–26. https://doi.org/10.24036/tip.v11i1.104

Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Informasi*, *8*(1), 958–965.

UPN "Veteran" Yogyakarta. (2020, Januari). Modul Praktikum Data Science. Yogyakarta:UPN "Veteran" Yogyakarta.

Zaky Musyarof, S. (2019). Analisis Text Mining Terhadap BPS (Badan Pusat Statistika) di Twitter Menggunakan R.