

# Klasifikasi Sentimen Terhadap Pinjaman *Online* (Pinjol) Menggunakan Algoritma *Naive Bayes*

Topan Dedi Ramadhan  
Universitas Buana Perjuangan  
Karawang, Indonesia  
if18.topanramadhan@mhs.ubpkarawang.ac.id

Deden Wahiddin  
Universitas Buana Perjuangan  
Karawang, Indonesia  
deden.wahiddin@ubpkarawang.ac.id

Elsa Elvira Awal  
Universitas Buana Perjuangan  
Karawang, Indonesia  
elsaelvira@ubpkarawang.ac.id

**Abstract**— Program pinjaman *online* merupakan salah satu contoh platform jasa keuangan yang ditawarkan langsung oleh pelaku usaha *fintech*. Seiring banyaknya penyedia pinjaman *online*, masyarakat akan lebih banyak lagi membicarakan aplikasi pinjaman *online*. Pendapat yang diberikan masyarakat terhadap pinjaman *online* (pinjol) berupa positif dan negatifnya dari penyedia jasa pinjaman *online* tersebut. Analisis sentimen merupakan sebuah studi dari komputasi yang mengekspresikan sebuah sentimen, opini, evaluasi, sikap, emosi, penilaian maupun pandangan terhadap suatu teks. Pada suatu teks tersebut dapat dikategorikan kedalam nilai positif, negatif maupun netral. Klasifikasi naive bayes adalah perhitungan untuk pengelompokan berdasarkan hipotesis bayes dan digunakan untuk mencari kemungkinan atau kemungkinan suatu kelas dengan kecurigaan bahwa satu kelas dengan kelas lain biasanya tidak bergantung atau bebas. Berdasarkan hasil penelitian menggunakan algoritma *naive bayes* terhadap 650 data dengan 80% digunakan untuk data *training* serta 20% digunakan untuk data *testing* diperoleh hasil akurasi sebesar 75%.

**Kata kunci** — Analisis sentimen, Klasifikasi *Naive Bayes*, Pinjaman *Online*

## I. PENDAHULUAN

Program pinjaman online adalah contoh *platform* layanan keuangan yang ditawarkan langsung oleh pemain *fintech* komersial. Munculnya *fintech* dalam program pinjaman online sangat memudahkan siapa saja untuk mendapatkan uang dengan proses cepat hanya dalam beberapa jam sesuai keinginan [1]. Seiring banyaknya penyedia pinjaman *online*, masyarakat akan lebih banyak lagi membicarakan aplikasi pinjaman *online*. Banyak berbagai tempat masyarakat bisa menyuarakan opini atau pendapatnya, mulai dari mulut ke mulut sampai melalui sosial media, termasuk media sosial twitter. Pendapat yang disampaikan dari masyarakat terhadap pinjaman *online* (pinjol) berupa positif dan negatifnya dari penyedia jasa pinjaman *online* tersebut. Dengan banyaknya pendapat yang diberikan oleh masyarakat terhadap layanan jasa pinjaman *online*, maka masyarakat yang lain akan lebih berhati-hati lagi dalam memilih jasa pinjaman *online*. Guna mengantisipasi permasalahan tersebut, dari pihak penyedia jasa pinjaman *online* harus meningkatkan lagi dari kualitas layanan yang diberikan, agar masyarakat semakin nyaman dalam menggunakan jasa pinjaman *online* tersebut.

Menurut data statistika per juli 2021 Indonesia berada di urutan ke enam dalam penggunaan media sosial twitter dengan jumlah pengguna mencapai 15.7 juta dibawah negara Amerika Serikat, Jepang, India, Britania Raya, dan Brazil. Twitter dapat dijadikan sebagai sumber data opini dan sentimen dari masyarakat. Dari data yang diperoleh dapat digunakan secara efisien untuk pemasaran maupun studi sosial [2]. Analisis sentimen merupakan sebuah studi dari komputasi yang mengekspresikan sebuah sentimen, opini, evaluasi, sikap, emosi, penilaian maupun pandangan terhadap suatu teks. Pada suatu teks tersebut dapat dikategorikan kedalam nilai positif, negatif maupun netral [3].

Telah banyak dilakukan penelitian sebelumnya mengenai analisis sentimen. Salah satu penelitian mengenai analisis sentimen yaitu menggunakan algoritma *Naive Bayes* terhadap pelanggan shopee. Jumlah data yang dipakai pada penelitian tersebut sebanyak 600 data yang terdiri dari 300 data latih yaitu 150 untuk komentar positif dan 150 untuk komentar negatif serta 300 data digunakan sebagai data uji. Hasil akurasi yang didapatkan dari pengujian dengan algoritma *Confusion Matrix* mencapai 97% [4]. Penelitian mengenai analisis sentimen selanjutnya menggunakan algoritma *K Nearest Neighbor* terhadap pengguna twitter. Data yang digunakan pada penelitian tersebut berjumlah 2000 data *tweet* dengan 1500 digunakan sebagai data latih dan 500 sebagai data uji. Hasil akurasi yang didapatkan dari pengujian dengan algoritma *Confusion Matrix* mencapai 67,2% [5]. Penelitian mengenai analisis sentimen berikutnya yaitu menggunakan algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization* terhadap *review* hotel. Pada penelitian tersebut data yang digunakan diambil dari situs [www.tripadvisor.com](http://www.tripadvisor.com) dengan mendapatkan 300 data *review* hotel yang terdiri dari 150 *review* untuk opini positif dan 150 *review* untuk opini negatif. Hasil akurasi yang didapatkan dari pengujian dengan *Cross Validation*, *Confusion Matrix* dan *ROC Curve* mendapatkan nilai akurasi mencapai 91.33% [6].

Berdasarkan pemaparan dari permasalahan dan penelitian sebelumnya mengenai analisis sentimen, maka pada penelitian ini melakukan klasifikasi sentimen terhadap pinjaman *online* (pinjol) dengan menggunakan algoritma *naive bayes*, karena algoritma tersebut memiliki tingkat akurasi yang lebih besar dibandingkan algoritma lainnya.

## II. TINJAUAN PUSTAKA

### A. Analisis Sentimen

Analisis sentimen merupakan salah satu metode yang digunakan untuk ekstraksi dari sebuah opini, yang kemudian diolah menjadi data yang berupa sentimen dari sebuah opini [7]. Cara kerja dari analisis sentimen adalah mengelompokkan teks-teks ke dalam kalimat atau dokumen yang kemudian ditentukan pendapat yang dituangkan pada dokumen untuk dianalisis, dicari tahu apakah bersifat positif, negatif maupun netral [8].

### B. Pinjaman Online

Pinjaman *online* merupakan salah satu kredit yang ditawarkan langsung secara *online* oleh pihak *fintech*. Ada banyak aplikasi-aplikasi penyedia pinjaman *online* yang bisa langsung diunduh melalui *smartphone* antara lain seperti kredivo, kredit pintar, ada kami dan masih banyak lainnya. Seiring banyaknya penyedia pinjaman *online*, menarik antusias masyarakat pada program yang diberikan pinjaman *online* walaupun bunganya lebih besar dari pada meminjam langsung dibank, karena syarat dan ketentuan lebih cepat dan mudah dibandingkan dengan harus meminjam di bank langsung [9].

### C. Klasifikasi Naive Bayes

Klasifikasi Naive Bayes merupakan algoritma untuk klasifikasi berdasarkan teorema bayes dan digunakan untuk menghitung sebuah probabilitas atau peluang dari suatu kelas dengan anggapan bahwa salah satu kelas dengan kelas lainnya tidak saling tergantung atau berdiri sendiri (Putra, 2018).

Tahapan-tahapan penggunaan algoritma Naive Bayes antara lain sebagai berikut :

1. Menghitung jumlah kelas/label
2. Menghitung jumlah kasus per kelas
3. Kalikan semua variabel kelas
4. Bandingkan hasil per kelas

Menurut (Ruger dkk, 2018) persamaan algoritma *Naive Bayes* dapat dilihat pada persamaan (1)

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Dimana:

$P(X|H)$  : Probabilitas X berdasarkan kondisi pada hipotesis H

X : Merupakan sebuah *class* data kosong / belum diketahui

H : Hipotesis atau spekulasi data

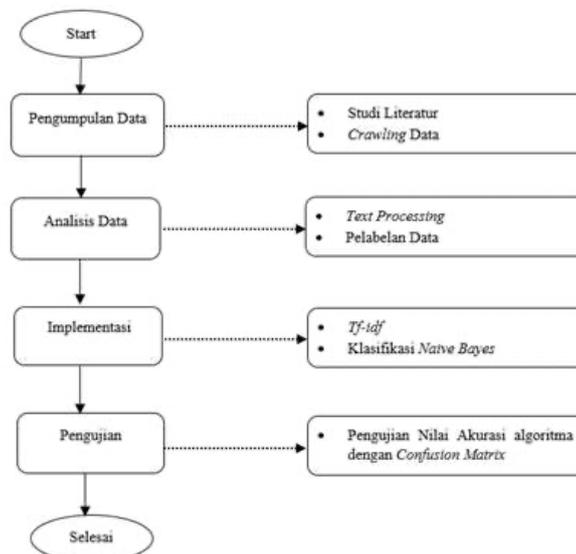
$P(H|X)$  : Merupakan hipotesis kemungkinan spekulasi H berdasarkan kondisi X (posterior probabilitas)

$P(H)$  : Merupakan hipotesis kemungkinan spekulasi H (prior probabilitas)

$P(X)$  : Probabilitas / kemungkinan X

### D. Metode Penelitian

Penelitian ini dimulai dengan pengumpulan data lalu dilakukan analisis data, kemudian hasil analisis data diimplementasi dengan *Term Frequency - Inverse Document Frequency* atau (*TF - IDF*) dan klasifikasi menggunakan algoritma *Naive Bayes*, kemudian data yang sudah diimplementasikan dilakukan pengujian dengan menggunakan *confusion matrix*.



Gambar 1 Metode Penelitian

- Pengumpulan Data
 

Pada penelitian proses pengumpulan data dibagi menjadi dua, antara lain sebagai berikut:

  1. Studi Literatur
 

Studi Literatur yaitu menemukan landasan teori dan memperoleh data dari berbagai jurnal, buku dan internet, seperti definisi, contoh penelitian yang relevan, metode, data mining, analisis sentimen, dan data yang terkait dengan penelitian ini.
  2. *Crawling* Data
 

Data yang akan dikumpulkan yaitu diambil dari sebuah *tweet* melalui media sosial twitter. Proses pengumpulan data dilakukan dengan menggunakan aplikasi RapidMiner dengan cara *crawling* data dengan sebuah *query* “pinjol”.
- Analisis Data
 

Data hasil *crawling* merupakan data yang masih mentah dengan masih banyak kalimat-kalimat yang noise atau tidak terstruktur, untuk itu perlu dilakukan analisis data, adapun analisis data mempunyai tahapan sebagai berikut :

  1. *Text Processing*

Sebelum memasuki tahap klasifikasi, diperlukan sebuah *text* dokumen untuk menyiapkan data terlebih dahulu, proses tersebut dikenal dengan istilah *text processing*. *Text Processing* sendiri mempunyai manfaat untuk digunakan agar *text* dalam sebuah data yang masih memiliki *noise* menjadi lebih terstruktur.

Ada beberapa tahapan dari *Text Processing* untuk digunakan pada penelitian ini, antara lain sebagai berikut :

    - 1) *Case Folding*

*Case Folding* yaitu merubah semua huruf yang awalnya huruf besar menjadi *lowercase* atau kecil semua. Proses tersebut dilakukan agar memudahkan pada saat pencarian, karena tidak semuanya dari dokumen teks cocok dengan huruf besar.
    - 2) *Cleaning*

*Cleaning* adalah sebuah proses untuk membersihkan data yang tidak diperlukan dari hasil *crawling*.
    - 3) *Stopwords Removal*

*Stopwords Removal* yaitu menghilangkan kata-kata yang tidak penting dalam suatu data. *Stopwords* adalah kata-kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak berarti. Contoh *stopword* bahasa Indonesia antara lain "yang", "dan", "di", "dari", dll. Maksud penggunaan *stopwords* adalah dengan menghilangkan kata-kata informasi rendah dari teks, kita dapat fokus pada kata-kata penting.
    - 4) *Stemming*

*Stemming* adalah menghilangkan gabungan kata, dan menyebabkan kata yang telah ditambahkan menjadi kata dasar.
    - 5) Normalisasi
 

Normalisasi yaitu bertujuan untuk memeriksa kata pada sebuah teks dengan memperbaikinya menjadi kata yang normal, contoh pada singkatan kata.
  2. Pelabelan Data
 

Data yang digunakan pada penelitian ini sebanyak 650 data, kemudian data tersebut diberikan suatu label secara manual untuk membentuk suatu sentimen berupa positif, negatif dan netral.
- Implementasi
 

Setelah dilakukan analisis data dengan melewati proses *text processing* dan pelabelan data, kemudian data diimplementasikan dengan tahapan sebagai berikut:

  1. *TF – IDF*

*Term Frequency - Inverse Document Frequency* atau *TF-IDF* adalah teknik kalkulasi yang sering digunakan untuk menghitung bobot setiap kata. Teknik ini juga dikenal produktif, mendasar dan tepat. Strategi ini akan memastikan nilai *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)* dari setiap tag (kata) di setiap laporan di korpus. Secara sederhana, teknik *TF-IDF* digunakan untuk mengetahui seberapa sering sebuah kata muncul dalam sebuah record.
  2. Klasifikasi *Naive Bayes*

Setelah dilakukan pembobotan kata atau proses *TF-IDF*, kemudian data dilakukan klasifikasi dengan menggunakan algoritma *Naive Bayes*. Klasifikasi adalah metode untuk menemukan itemset dan mengelompokkannya sesuai dengan nilai bobot dari setiap item set. Data yang dilakukan pada proses klasifikasi yaitu 700 data dengan 80% sebagai data latih dan 20% sebagai data uji. Pada penelitian ini, data akan dikelompokkan serta dibagi menjadi tiga kelas, yaitu kelas positif, kelas netral dan kelas negatif.
- Pengujian
 

Pengujian dilakukan pada penelitian ini bertujuan untuk menemukan/mengevaluasi dan mengetahui performa terhadap algoritma (yang digunakan). Proses evaluasi dilakukan dengan menggunakan metode confusion matrix

untuk menilai tingkat akurasi dari algoritma tersebut. Metode tersebut sangat membantu untuk menilai kinerja dari analisis klasifikasi [4]. Pengujian tersebut dilakukan dengan cara menghitung tingkat accuracy, recall, dan precision yang dapat ditampilkan dalam bentuk persentase.

1. *Accuracy*

*Accuracy* atau akurasi adalah hasil dari skala prediksi yang akurat.

2. *Precision*

*Precision* merupakan suatu jumlah teks yang signifikan dalam kalangan dokumen yang sebelumnya disortir menggunakan sebuah sistem.

3. *Recall*

*Recall* merupakan suatu jumlah teks relevan dalam kalangan dokumen teks dari koleksi.

Adapun yang digunakan untuk mengukur *accuracy*, *precision*, dan *recall* seperti tabel 1 dan persamaan berikut ini:

Tabel 1 *Confusion Matrix*

Manual	Sistem		
	Positif	Netral	Negatif
Positif	True Positive (TP)	False Netral (FNt)	False Negative (FN)
Netral	False Positive (FP)	True Netral (TNt)	False Negative (FN)
Negatif	False Positive (FP)	False Netral (FNt)	True Negative (TN)

Sumber : Eza Ananda Putra (2021)

$$Accuracy = \frac{TP+TNt+TN}{TP+TN+TNt+FP+FNt+FN} * 100\% \quad (2)$$

$$Precision = \frac{TP}{FP+TP} * 100 \quad (3)$$

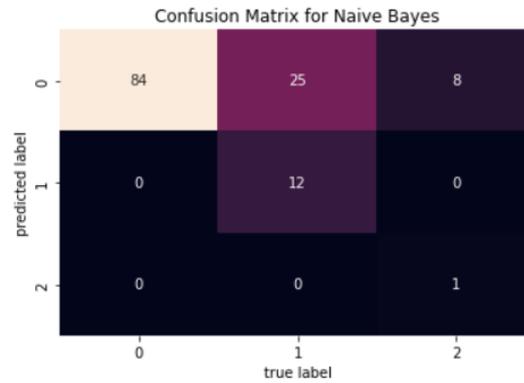
$$Recall = \frac{TP}{FN+FNt+TP} * 100\% \quad (4)$$

Dari persamaan 2, 3, dan 4 dijelaskan bahwa:

- TP adalah *True* Positif, yaitu semua data benar pada kelas positif yang diklasifikasi oleh sistem.
- TNt adalah *True* Netral, yaitu semua data benar pada kelas netral yang diklasifikasi oleh sistem.
- TN adalah *True* Negatif, yaitu semua data benar pada kelas negatif yang diklasifikasi oleh sistem
- FP adalah *False* Positif, yaitu semua data pada kelas positif namun terklasifikasi salah oleh sistem.
- FNt adalah *False* Netral, yaitu semua data pada kelas netral namun terklasifikasi salah oleh sistem.
- FN adalah *False* Negatif, yaitu semua data pada kelas negatif namun terklasifikasi salah oleh sistem..

### III. HASIL DAN PEMBAHASAN

Data hasil *crawling* yang sudah dipersiapkan, kemudian dilakukan *text processing* terhadap data dan diberikan lebel / kelas pada data secara manual dengan sentimen positif, negatif dan netral pada setiap dokumennya. kemudian data yang sudah diberikan label sebanyak 650 data diklasifikasikan menggunakan algoritma *naive bayes* dengan 80 % digunakan sebagai data *training* dan 20 % digunakan sebagai data *testing*, selanjutnya dilakukan pengujian menggunakan *confusion matrix*, proses pengujian bertujuan untuk mengetahui performa dari algoritma yang digunakan. Hasil dari pengujian menggunakan *confusion matrix* dengan 650 data, dimana 130 data digunakan sebagai data *testing* menghasilkan *confusion matrix* seperti gambar 2.



Gambar 2. Hasil *Confusion Matrix*

Dari Gambar 2 dijelaskan bahwa 0 merupakan kelas dari negatif, 1 merupakan kelas dari netral dan 2 merupakan kelas dari positif. Setelah mengetahui model dari *confusion matrix* diatas, kemudian dibuat tabel *confusion matrix* untuk memudahkan mengetahui kinerja algoritma *naive bayes* terhadap data. Adapun tabel *confusion matrix* seperti tabel 2.

Tabel 2 *Confusion Matrix* Dari *Naive Bayes*

Manual	Sistem		
	Positif	Netral	Negatif
Positif	1	0	8
Netral	0	12	25
Negatif	0	0	84

Dari Tabel 2 dapat dijelaskan bahwa model mengklasifikasi data benar apabila kelas pada sistem sesuai dengan data *training*, gambar diatas memiliki kesamaan antara lain 1 data sebagai positif, 3 data sebagai netral dan 82 sebagai data negatif. Pada perhitungan manual akurasi dari matriks diatas adalah sebagai berikut :

$$\begin{aligned}
 Accuracy &= \frac{TP + TNt + TN}{Total\ Data\ Yang\ diuji} * 100\% \\
 &= \frac{97}{130} * 100\% \\
 &= 75\%
 \end{aligned}$$

Dari penjumlahan diatas 97 adalah jumlah dari data benar / sesuai pada setiap kelas sedangkan 130 merupakan jumlah data *training* atau data yang digunakan sebagai data uji. Akurasi menggambarkan sejauh mana model yang dihasilkan dapat mengklasifikasikan data dengan benar. Akurasi diperoleh dari menghitung rasio antara klasifikasi yang benar dari kumpulan data. Dengan mengetahui nilai *precision* untuk kinerja sistem secara keseluruhan, seseorang dapat menyatakan sejauh mana sistem mampu menemukan akurasi antara informasi yang diinginkan oleh pengguna dan jawaban yang diinginkan pengguna, kata yang ditamirkan dari sistem. Sistem dalam mencari informasi tingkat keberhasilan pada penelitian ini yaitu 75%.

#### IV. KESIMPULAN DAN SARAN

Berdasarkan hasil dari penelitian terhadap pinjol menggunakan algoritma *Naive Bayes* maka diperoleh kesimpulan yaitu, Penerapan algoritma *Naive Bayes* pada klasifikasi sentimen terhadap pinjaman *online* (pinjol) memiliki beberapa tahapan antara lain seperti, pengumpulan data, analisis data, implementasi serta pengujian. Kemudian dapat disimpulkan bahwa pinjaman *online* mendapat sentimen negatif berdasarkan penelitian ini. Pengujian akurasi menggunakan *confusion matrix* pada penelitian dengan algoritma *Naive Bayes* menghasilkan tingkat akurasi sebesar 75%. Hasil dari akurasi tersebut dihasilkan dari data sebesar 650 data berupa sentimen positif, negatif dan netral. Pada pengujiannya data pisahkan menjadi 2 yaitu data 80% digunakan sebagai data *training* dan 20% data digunakan sebagai data *testing*.

Saran untuk penelitian selanjutnya yaitu Menerapkan algoritma klasifikasi lain untuk mengetahui tingkat akurasi yang berbeda seperti algoritma *Support Vector Machine*, *K-Nearest Neighbor*, dan *Decision Tree* kedalam kedalam sentimen terhadap pinjaman *online* (pinjol). Menambahkan data *training* kedalam sentimen terhadap pinjaman *online* (pinjol) untuk menyeimbangkan jumlah data pada setiap kelasnya.

## PENGAKUAN

Naskah ilmiah ini adalah Sebagian dari penelitian Tugas Akhir milik Topan Dedi Ramadhan dengan judul Klasifikasi Sentimen Terhadap Pinjaman *Online* (Pinjol) Berdasarkan Opini Masyarakat Pada Media Sosial Twitter Menggunakan Algoritma *Naive Bayes* yang dibimbing oleh Bapak Deden Wahiddin, M.Kom dan Ibu Elsa Elvira Awal, M.Kom.

## DAFTAR PUSTAKA

- [1] R. A. E. Wahyuni and B. E. Turisno, "Praktik Finansial Teknologi Ilegal Dalam Bentuk Pinjaman Online Ditinjau Dari Etika Bisnis," *J. Pembang. Huk. Indones.*, vol. 1, no. 3, pp. 379–391, 2019, doi: 10.14710/jphi.v1i3.379-391.
- [2] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," *Proc. 7th Int. Conf. Lang. Resour. Eval. Lr. 2010*, pp. 1320–1326, 2010, doi: 10.17148/ijarce.2016.51274.
- [3] B. Liu and B. Liu, "Sentiment Analysis : A Multi-Faceted Problem Sentiment Analysis : A Multi-Faceted Problem," no. March, pp. 2–7, 2010.
- [4] A. H. Ruger, M. Suyanto, and M. P. Kurniawan, "Sentimen Analisis Pelanggan Shopee di Twitter dengan Algoritma Naive Bayes," *J. Inf. Technol.*, vol. 1, no. 2, pp. 26–29, 2021, doi: 10.46229/jifotech.v1i2.282.
- [5] A. Deviyanto and M. D. R. Wahyudi, "Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018, doi: 10.14421/jiska.2018.31-01.
- [6] E. Indrayuni, "Komparasi Algoritma Naive Bayes Dan Support Vector Machine Untuk Analisa Sentimen Review Film," *J. Pilar Nusa Mandiri*, vol. 14, no. 2, p. 175, 2018, doi: 10.33480/pilar.v14i2.918.
- [7] A. Sari, F. V., & Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 2, no. 2, pp. 681–686, 2019.
- [8] T. A. Aliah, "PENERAPAN METODE NAIVE BAYES DALAM ANALISIS SENTIMEN PADA DATA TWITTER (STUDI KASUS: HASIL DEBAT CALON PRESIDEN 2019)," *Digit. Repos. Univ. Jember*, no. September 2019, pp. 2019–2022, 2019.