

# KLASIFIKASI KONTEN POST PADA MEDIA SOSIAL INSTAGRAM PERGURUAN TINGGI XYZ MENGGUNAKAN ALGORITMA NAIVE BAYES

Nova Agustina<sup>1</sup>, Andrian Cahyanto<sup>2</sup>, Joko Herwanto<sup>3</sup>, Ratih Ratnasari<sup>4</sup>, Sintia Dewi<sup>5</sup>

Program Studi Teknik Informatika<sup>1, 2, 3, 4, 5</sup>

Sekolah Tinggi Teknologi Bandung

Jl, Soekarno Hatta no. 378, Bandung

[nova@sttbandung.ac.id](mailto:nova@sttbandung.ac.id)<sup>1</sup>, [cahyantoandrian@gmail.com](mailto:cahyantoandrian@gmail.com)<sup>2</sup>, [jokoherwanto68@gmail.com](mailto:jokoherwanto68@gmail.com)<sup>3</sup>,

[sintiad289@gmail.com](mailto:sintiad289@gmail.com)<sup>4</sup>, [ratihratna5@gmail.com](mailto:ratihratna5@gmail.com)<sup>5</sup>

## Abstrak

Media sosial memiliki peran penting sebagai alat komunikasi untuk berbagi informasi dan saling terhubung. media sosial digunakan sebagai ruang untuk berinteraksi dengan orang berbagai kalangan. Salah satu media sosial yang paling banyak digunakan di dunia, termasuk di Indonesia adalah Instagram. Penelitian ini bertujuan untuk mencari tahu konten post yang bagaimana yang paling optimal dalam *Online Engagement*. Penelitian ini menggunakan data post yang diunggah oleh akun Instagram Perguruan Tinggi XYZ pada 22 September 2015 hingga 4 Juni 2019. *Post* yang dianalisis dalam penelitian ini berjumlah 426 konten post. Uji banding dilakukan untuk melihat perbedaan online engagement antara konten *post*. Hasil dari penelitian ini menunjukkan bahwa tipe *post* dan waktu *posting* (bulan dan hari) memiliki pengaruh signifikan terhadap *online engagement* (like dan komentar). Sedangkan bulan yang memiliki pengaruh paling signifikan terhadap like dan komentar adalah bulan Maret, Juni, dan Juli. Hari Kamis, Jumat, dan Sabtu merupakan hari yang memiliki pengaruh signifikan terhadap like dan komentar.

Kata Kunci:

Instagram, Naïve Bayes, Klasifikasi, *Online Engagement*, Media Sosial

## Abstract

*Social media has an important role as a communication tool to share information and connect with each other. social media is used as a space to interact with people from various walks of life. One of the most widely used social media in the world,*

*including in Indonesia is Instagram. This study aims to find out what post content is most optimal in Online Engagement. This study uses post data uploaded by XYZ Higher Education Instagram accounts from September 22 2015 to June 4, 2019. Post analyzed in this study amounted to 426 post content. A comparative test was conducted to see the difference between online engagement between post content. The results of this study indicate that post type and post time (month and day) have a significant effect on online engagement (likes and comments). While the months that have the most significant influence on likes and comments are March, June and July. Thursday, Friday and Saturday are days that have a significant influence on likes and comments.*

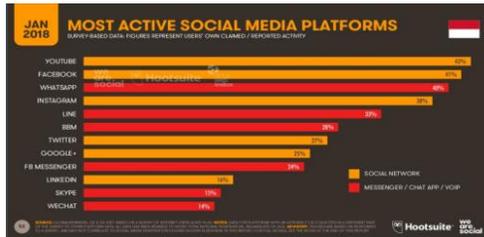
Keywords :

*Instagram, Naïve Bayes, Classification, Online Engagement, Social Media*

## I. PENDAHULUAN

Di era digital saat ini, media sosial memiliki peran penting sebagai alat komunikasi dimana setiap pengguna dapat berbagi informasi, pengetahuan dan saling terhubung. Media sosial merupakan medium di internet yang memungkinkan pengguna merepresentasikan dirinya agar dapat berinteraksi, bekerja sama, berbagi, berkomunikasi dengan pengguna lain, dan membentuk ikatan secara virtual (Nasrullah, 2018). Dari semua jenis media sosial yang ada, pengguna Instagram telah digunakan oleh jutaan user yang aktif.

Bahkan pada awal tahun 2018, Instagram merupakan platform sosial dengan jumlah pengguna terbanyak ke tujuh di dunia (Kemp S. Digital, 2018). Di Indonesia, Instagram merupakan media sosial yang paling sering digunakan keempat setelah Youtube, Facebook, dan Whatsapp (Kemp S. Digital, 2018) seperti yang ditunjukkan pada Gambar 1.1.



Mengingat perkembangannya yang sangat cepat, media sosial dapat menjadi media paling strategis bagi Perguruan Tinggi XYZ untuk berkomunikasi dengan calon mahasiswa dan mahasiswanya. Pernyataan tersebut selaras dengan hasil wawancara. Selain itu informan menuturkan bahwa media sosial yang memiliki interaksi online yang optimal adalah Instagram, hanya saja belum memiliki informasi yang cukup tentang hasil dari konten post yang berada di Instagram.

Dalam penelitian (Santoso, 2017) menunjukkan bahwa konten post bisa berdampak pada interaksi online yang diwakili oleh like dan komentar. Semakin banyak jumlah like dan komentar berbanding lurus dengan banyaknya follower (Huang Jianghua dkk, 2018). Dengan adanya respon yang banyak, hal ini menunjukkan bahwa suatu akun Instagram dapat menjalin hubungan dengan follower secara intens dan baik. Selain itu, jumlah like dan komentar yang banyak juga akan membuka peluang untuk sebuah akun agar lebih populer dan dikenal.

Berdasarkan fakta-fakta tersebut, konten post yang diteliti dalam penelitian ini menggunakan kriteria yang menjadi acuan terdiri dari like dan komentar yang didapatkan dari media sosial Instagram Perguruan Tinggi XYZ. Data-data yang telah didapatkan kemudian diolah untuk dapat mengambil informasi-informasi yang tersembunyi dari data-data tersebut. Metode pengolahan data seperti ini sering disebut sebagai data mining.

Metode data mining yang digunakan untuk mengelola datanya menggunakan metode Naïve Bayes. Metode Naïve Bayes juga digunakan pada penelitian ini untuk menampilkan konten post yang

diteliti berdasarkan jumlah like, jumlah komentar serta jumlah like dan komentar. Dari data tersebut akan menampilkan karakteristik konten post yang terdiri dari video, gambar, caption, jam posting, hari posting dan bulan posting yang menimbulkan interaksi online secara optimal. Alasan menggunakan metode naïve bayes diantaranya ialah karena metode ini dinilai berpotensi baik dalam mengklasifikasi dokumen dibandingkan dengan metode klasifikasi yang lain dalam hal akurasi dan efisiensi komputasi (Simamora, 2018)

Berdasarkan dari uraian diatas penulis mengambil topik yaitu “Klasifikasi Konten Post Pada Media Sosial Instagram Perguruan Tinggi XYZ Menggunakan Naïve Bayes” yang bertujuan sebagai bahan pertimbangan untuk pengelola media sosial Perguruan Tinggi XYZ untuk membuat konten post yang akan diposting untuk mendapatkan interaksi online secara optimal.

## II. KAJIAN LITERATUR

### II.1 Instagram

Instagram adalah sebuah aplikasi sosial yang populer dalam kalangan pengguna telepon pintar (Smartphone). Nama Instagram diambil dari kata “Insta” yang asalnya “Instan” dan “gram” dari kata “telegram” (Prayoga, 2019). Jadi Instagram merupakan gabungan dari kata Instan-Telegram. Dari penggunaan kata tersebut dapat diartikan sebagai aplikasi untuk mengirimkan informasi dengan cepat, yakni dalam bentuk foto yang berupa mengelola foto, mengedit foto, dan berbagi ke jejaring sosial yang lain

### II.2 Konten

Konten bisa diartikan sebagai teks yang ditulis, foto, video, suara, dan sebagainya yang disebarluaskan melalui media sosial (Nasrullah, 2018). Konten merupakan bahan dasar yang membentuk budaya yang ada di media sosial, karena pada dasarnya komoditas yang diproduksi sekaligus di konsumsi oleh khalayak di media sosial adalah konten. Konten menjadi istilah yang digunakan untuk mengidentifikasi dan mengkuantifikasi berbagai macam format dan genre informasi sebagai komponen nilai tambah media (Prayoga, 2019).

### II.3. Online engagement

Online engagement bisa didefinisikan sebagai kondisi psikologis dari user yang dikategorikan oleh keinteraktifan, pengalaman kokreatif user dengan seorang agen dan objek (Santoso, 2017). Untuk mengoperasikan online engagement di Instagram, like dan komentar yang digunakan untuk mewakili online engagement (Huang Jianghua dkk, 2018). Suatu post dengan banyak like bisa mengindikasikan bahwa sebuah konten dikatakan menarik, meningkatkan kemungkinan untuk juga disukai oleh seseorang, dan mengarah pada penyebaran informasi dari suatu merek kepada pelanggan potensial (Huang Jianghua dkk, 2018). Di sisi lain, jumlah komentar yang banyak pada suatu post merepresentasikan tingkat kesuksesan atau dampak karena komentar membuat user meluangkan waktunya untuk menyampaikan pendapatnya. Like dan komentar telah banyak digunakan sebagai pengukuran untuk dampak suatu publikasi (Simamora, 2005). Online engagement menjadi penting karena semakin banyak online engagement yang diraih, maka semakin besar kemungkinan sebuah merek bisa lebih dikenal. Selain itu, online engagement juga menjadi penting karena dapat menunjukkan hubungan baik dengan konsumen dan berdampak pada loyalitas konsumen [7].

### II.4. Naive Bayes

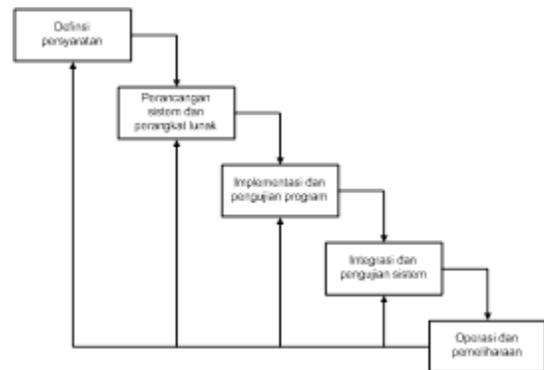
Naive Bayes Bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Karakteristik Naive Bayes yaitu :

1. Bekerja teguh terhadap data-data yang terisolasi, yang biasanya merupakan data karakteristik berbeda (*outliner*). Naive Bayes dapat menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model prediksi.
2. Tangguh menghadapi nilai atribut yang tidak relevan.
3. Apabila atribut memiliki korelasi dapat mendegrasi kerja klasifikasi Naive Bayes, karena asumsi independensi atribut tersebut sudah tidak ada.

## III. ANALISIS DAN PERANCANGAN

### III. 1. Metodologi Penelitian

Metodologi pengembangan sistem yang digunakan adalah metode yang dikenal dengan nama SDLC atau Software Development Life Cycle atau sering disebut juga Systems Development Life Cycle, metode ini merupakan proses mengembangkan atau mengubah suatu sistem perangkat lunak dengan menggunakan model-model dan metodologi yang digunakan orang untuk mengembangkan sistem-sistem perangkat lunak sebelumnya. Adapun model yang digunakan adalah Model Waterfall.



Gambar 3.1. Model Waterfall

#### 1. Requirement

Pada tahap ini peneliti membuat kriteria awal penilaian yang akan diukur berdasarkan hasil wawancara dan observasi dengan bagian Media Perguruan Tinggi XYZ.

#### 2. Design

Peneliti melakukan tahap design sistem, yakni dimulai dari perancangan Unified Modelling Language (UML) hingga ke perancangan basis data.

#### 3. Implementation

Setelah dilakukan perancangan, implementasi dilakukan setelah melakukan pengkodean yang bertujuan untuk menerapkan aplikasi pada studi kasus.

#### 4. Verification

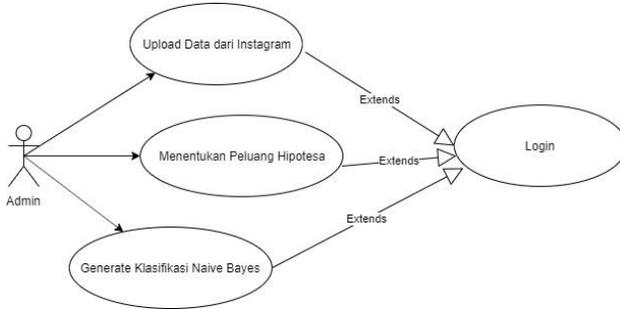
Tahap ini dilakukan untuk melakukan pengujian sistem pada studi kasus, memastikan bahwa sistem dapat berjalan dengan baik dan semua fungsi berjalan sesuai kebutuhan.

5. Maintenance

Tahap ini

**III. 2. Perancangan Sistem**

Perancangan system digunakan menggunakan Unified Modelling Language (UML). Sistem dirancang untuk memenuhi kebutuhan penggunaan yang Naïve Bayes, yakni Upload Data dari Instagram, Menentukan Probabilitas hingga *Generate* Klasifikasi Naïve Bayes.



**Gambar 3.2. Use Case Diagram Sistem**

**III. 3. Implementasi Sistem**

| No | Like Count | Comment Count | Optimal |
|----|------------|---------------|---------|
| 1  | 7652       | 19            | YES     |
| 2  | 5056       | 2             | YES     |
| 3  | 7935       | 6             | YES     |
| 4  | 1595       | 115           | YES     |
| 5  | 6116       | 0             | NO      |
| 6  | 6362       | 0             | NO      |
| 7  | 170        | 0             | NO      |
| 8  | 159        | 0             | NO      |
| 9  | 121        | 6             | YES     |
| 10 | 155        | 21            | YES     |

**Gambar 3.3. Data Upload dari Instagram**

Data ini merupakan data yang sudah digenerate dan diambil menggunakan API Instagram. Data yang diambil adalah jumlah Like dan Komentar di Instagram Perguruan Tinggi XYZ.

| No | Like Count | Comment Count | Optimal |
|----|------------|---------------|---------|
| 1  | 7652       | 19            | YES     |
| 2  | 5056       | 2             | YES     |
| 3  | 7935       | 6             | YES     |
| 4  | 1595       | 115           | YES     |
| 5  | 6116       | 0             | NO      |
| 6  | 6362       | 0             | NO      |
| 7  | 170        | 0             | NO      |
| 8  | 159        | 0             | NO      |
| 9  | 121        | 6             | YES     |
| 10 | 155        | 21            | YES     |

$P(\text{Like Count} + \leq 7370 \mid \text{optimal}=\text{YES}) = 4/6 = 0,66$   
 $P(\text{Like Count} + \leq 7370 \mid \text{optimal}=\text{NO}) = 4/4 = 1$

**Gambar 3.4. Data Like Count <= 7370 Hipotesa Awal**

Hipotesa Awal menghitung jumlah Like dimana jumlahnya adalah <=7370 dan status optimal adalah YES untuk melakukan hipotesa pertama.

| No | Like Count | Comment Count | Optimal |
|----|------------|---------------|---------|
| 1  | 7652       | 19            | YES     |
| 2  | 5056       | 2             | YES     |
| 3  | 7935       | 6             | YES     |
| 4  | 1595       | 115           | YES     |
| 5  | 6116       | 0             | NO      |
| 6  | 6362       | 0             | NO      |
| 7  | 170        | 0             | NO      |
| 8  | 159        | 0             | NO      |
| 9  | 121        | 6             | YES     |
| 10 | 155        | 21            | YES     |

$P(\text{Comment Count} \leq 22 \mid \text{Optimal}=\text{YES}) = 5/6 = 0,83$   
 $P(\text{Comment Count} \leq 22 \mid \text{Optimal}=\text{NO}) = 4/4 = 1$

**Gambar 3.5. Data Comment Count <= 22 Hipotesa Awal**

Hipotesa selanjutnya adalah menghitung jumlah komentar yang berjumlah <= 22 komentar sebagai hipotesa awal dengan status optimal yaitu YES.

| No | Like Count | Comment Count | Optimal |
|----|------------|---------------|---------|
| 1  | 7652       | 19            | YES     |
| 2  | 5056       | 2             | YES     |
| 3  | 7935       | 6             | YES     |
| 4  | 1595       | 115           | YES     |
| 5  | 6116       | 0             | NO      |
| 6  | 6362       | 0             | NO      |
| 7  | 170        | 0             | NO      |
| 8  | 159        | 0             | NO      |
| 9  | 121        | 6             | YES     |
| 10 | 155        | 21            | YES     |

$P(\text{Like Count} \leq 256 \mid \text{Optimal}=\text{YES}) = 2/6 = 0,33$   
 $P(\text{Like Count} \leq 256 \mid \text{Optimal}=\text{NO}) = 2/4 = 0,5$

**Gambar 3.6. Data Like Count <= 256**

| No | Like Count | Comment Count | Optimal |
|----|------------|---------------|---------|
| 1  | 7652       | 19            | YES     |
| 2  | 5056       | 2             | YES     |
| 3  | 7935       | 6             | YES     |
| 4  | 1595       | 115           | YES     |
| 5  | 6116       | 0             | NO      |
| 6  | 6362       | 0             | NO      |
| 7  | 170        | 0             | NO      |
| 8  | 159        | 0             | NO      |
| 9  | 121        | 6             | YES     |
| 10 | 155        | 21            | YES     |

$P(\text{Comment Count} \leq 5 \mid \text{Optimal}=\text{YES}) = 1/6 = 0,16$   
 $P(\text{Comment Count} \leq 5 \mid \text{Optimal}=\text{NO}) = 4/4 = 1$

**Gambar 3.7. Data Like Count <= 5**

Hipotesa berikutnya adalah menghitung data Like dengan jumlah <=5 dengan status NO sebagai bahan awal hipotesa.

**Tabel 1. Hitung P(Xk | Ci) utk setiap class I**

| Hitung P(X   Ci) utk setiap class I       |      |
|---|------|
| P(like count <=7370   optimal = YES = 4/6 | 0.66 |
| P(like count <=7370   optimal = NO = 4/4  | 1    |
| P(comment count <=22   optimal YES = 5/6  | 0.83 |
| P(comment count <=22   optimal NO = 4/4   | 1    |
| P(like count <=256   optimal = YES = 2/6  | 0.33 |
| P(like count <=256   optimal = NO = 2/4   | 1    |
| P(comment count <=5   optimal YES = 1/6   | 0.16 |
| P(comment count <=5   optimal YES = 4/4   | 1    |

Keterangan :

P(X) = Peluang dari X yang diamati

Ci = Kelas

Hitung P (Xk | Ci) untuk setiap class i

- Hitung P(X | Ci) untuk setiap class :
- P ( X | Optimal = “YES”)
- $0.66 \times 0.83 \times 0.33 \times 0.16 = 0.289$
- P ( X | Optimal = “NO”)
- $1 \times 1 \times 0.5 \times 1 = 0.5$
- P ( X | Ci ) \* P ( Ci )
- P ( X | Optimal = “YES” ) \* P ( Optimal= “YES” )
- $0.289 * (6/10) = 0.173$
- P ( X | Optimal = “NO” ) \* P ( Optimal=“NO” )

$$0.5 * (4/10) = 0.2$$

X Memiliki Class “Optimal=NO”

Karena

P(X | Optimal=“NO”) Memiliki nilai maksimum pada perhitungan diatas.

#### IV. KESIMPULAN DAN SARAN

Dari hasil penelitian ini ditemukan bahwa Tipe post memiliki pengaruh signifikan terhadap online engagement (like dan komentar). Menghasilkan kesimpulan bahwa perangkat lunak *clustering* konten post pada media sosial Instagram Perguruan Tinggi XYZ menggunakan Algoritma Naïve Bayes yang telah berhasil dibangun dan dapat digunakan untuk mengelompokan data konten *post* Instagram Perguruan Tinggi XYZ.

Untuk pengembangan lebih lanjut disarankan melakukan perbandingan dengan Naive Bayes untuk mencari Algoritma paling optimal dengan permasalahan yang serupa.

#### REFERENSI

Nasrullah D. Media Sosial Perspektif Komunikasi, Budaya dan Siosioteknologi. Edisi Kelima. Bandung: Simbioasa Rekatama Media. 2018: 13

Kemp S. Digital In 2018. Singapore: We Are Social. 2018

Santoso Amanda P., Baihaqi Imam, dan Persada Satria F. Pengaruh Konten Post Instagram terhadap Online Engagement: Studi Kasus pada Lima Merek Pakaian Wanita. Jurnal Sains dan Seni ITS. 2017; Vol. 6, No. 1: D-50 - D-54.

Huang Jianghua dkk. Inspecting Influences on Likes and Comments of Photos in Instagram. IEEE SmartWorld. 2018: 938-945

Simamora B. Analisis Multivariat Pemasaran. Jakarta: PT. Gramedia Pustaka Utama. 2005

Primartha R. Belajar Machine Learning Teori dan Praktek. Bandung :Informatika. 2018

Prayoga Dewa E. Konten dan Konten Media di Era Teknologi Digital.<https://billionairecoach.co.id/marketings/konten-dan-konten-media-di-era->

---

teknologi-digital. Tanggal Akses Terakhir:  
13 Februari 2019