

Analisis Sentimen Dengan Metode Naïve Bayes Terhadap Ulasan Hotel New Saphir Yogyakarta Pada Platform Tripadvisor

Apfia Normalita¹; Dede Brahma Arianto^{2*}

Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana¹

Informatika, Fakultas Sains dan Teknik, Universitas Faletehan^{2*}

apfia.normalita@si.ukdw.ac.id¹; dedebrahma@uf.ac.id^{2*}

* dedebrahma@uf.ac.id

Abstrak

Salah satu pondasi yang mendukung pembangunan sektor pariwisata adalah industri perhotelan. Saat ini beberapa hotel telah mengikuti perkembangan teknologi dengan menggunakan platform digital online. Tripadvisor menjadi platform yang banyak digunakan sebagai pilihan pelanggan dalam mencari, memesan hingga memberikan ulasan pada hotel. Hotel New Saphir Yogyakarta merupakan salah satu hotel terbaik dan memiliki rating kinerja hotel yang cukup tinggi di situs Tripadvisor. Akan tetapi, pihak hotel membutuhkan data yang kuat yang menunjukkan data valid mengenai komentar hotel secara ringkas dan cepat. Maka dari itu, dilakukan analisis sentimen komentar pelanggan untuk mengetahui seberapa banyak data yang positif dan negatif. Analisis ini mengumpulkan dataset yang berjumlah 118 baris data bersumber. Dataset yang terkumpul kemudian dilakukan pemodelan analisis sentimen dengan metode Naïve Bayes. Hasil penelitian ini menunjukkan bahwa kinerja model Naïve Bayes memiliki nilai akurasi sebesar 92%. Maka, dapat disimpulkan bahwa model dapat memprediksi sentimen ulasa dengan baik.

Kata Kunci: Analisis Sentimen; Naïve Bayes; Scraping; Ulasan

Abstract

One of the foundations that supports the development of the tourism sector is the hotel industry. Currently, several hotels have followed technological developments by using online digital platforms. Tripadvisor is a platform that is widely used as a choice for customers in searching, booking and providing reviews of hotels. Hotel New Saphir Yogyakarta is one of the best hotels and has a fairly high hotel performance rating on the Tripadvisor site. However, the hotel needs strong data that shows valid data regarding hotel comments in a concise and fast manner. Therefore, a sentiment analysis of customer comments was conducted to determine how much data was positive and negative. This analysis collected a dataset of 118 rows of sourced data. The collected dataset was then subjected to sentiment analysis modeling using the Naïve Bayes method. The results of this study indicate that the performance of the Naïve Bayes model has an accuracy value of 92%. So, it can be concluded that the model can predict review sentiment well.

Keywords: Sentiment Analysis; Naïve Bayes; Scraping; Reviews.

PENDAHULUAN

Hotel menjadi salah satu usaha yang bergerak dibidang akomodasi dalam objek material ilmu pariwisata. Hotel menyediakan jasa penginapan, makanan, minuman dan fasilitas jasa lainnya. Jasa penginapan sangat dibutuhkan bagi para wisatawan saat berlibur dan menjadi salah satu hal yang penting dalam suatu perjalanan. Hotel yang memberikan jasa mereka yang memuaskan akan memberi kesan baik dan memungkinkan tamu untuk kembali lagi. Standar kamar dan fasilitas pada hotel harus dipertimbangkan untuk menarik perhatian tamu. Kepuasan dapat dinilai dari pengalaman tamu yang dirasakan saat merasakan jasa hotel yang disajikan. Penilaian tamu akan sangat berpengaruh besar bagi usaha hotel kedepannya, maka pelayanan yang diberikan harus memuaskan agar tamu memberikan nilai terbaik bagi hotel tersebut.

Untuk memahami bagaimana jasa pihak hotel berikan kepada pelanggan sudah sesuai diperlukan sebuah

analisis. Analisis sentimen menjadi salah satu pilihan yang tepat untuk melihat hal tersebut. Analisis sentimen dapat melihat gambaran tentang jasa hotel dari pendapat dan perasaan pelanggan yang ditulis melalui kolom komentar. Sehingga, perusahaan dapat mengidentifikasi jasa yang paling disukai oleh pelanggan dan meningkatkan pemasaran. Analisis Sentimen merupakan proses *analytics text* yang merupakan teknik Natural Language Processing (NLP) untuk menganalisis emosi dari suatu text yang sumber data berasal dari internet dan berbagai macam platform media sosial [6]. Melakukan analisis sentimen bertujuan untuk mendapatkan feedback dari pengguna berdasarkan komentar yang diberikan pada media platform tersebut.

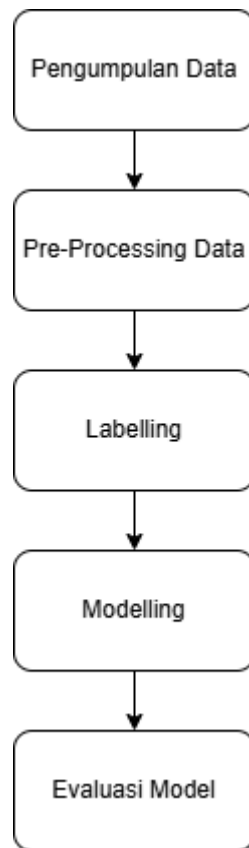
Penerapan analisis sentimen dapat didukung dengan adanya penggunaan machine learning. Machine learning merupakan salah satu kecerdasan butanan atau Artificial Intelligence (AI) yang mempelajari dari data dan meningkatkan kinerja tanpa perlu diprogram ulang secara manual [7]. Cara kerja machine learning dalam analisis ini adalah dengan mempelajari pola dalam data teks. Salah satu metode dalam machine learning yang dapat digunakan adalah metode Naïve Bayes. Naïve Bayes Classifier merupakan salah satu algoritma machine learning yang paling populer yang berfungsi untuk mengklasifikasikan data. Data yang telah diklasifikasikan berfungsi untuk memprediksi kelas dalam suatu data. Naïve Bayes Classifier sendiri berasal dari teorema Bayes untuk menghitung probabilitas suatu kelas dengan melihat fitur satu persatu dan mengumpulkan statistik sederhana setiap kelas dari fitur yang ada [8]. Metode ini bekerja dengan mengasumsikan bahwa setiap kata dalam teks memiliki probabilitas independen untuk mewakili golongan sentimen positif maupun negatif.

Penelitian mengenai analisis sentimen sudah banyak dilakukan oleh para ahli diluar sana. Penelitian analisis setimen yang dilakukan oleh Alqis Rausanfitra dan kedua penulis lainnya [3] dengan menggunakan data set sebanyak 72 *tweet* dan menerapkan metode *Extreme Learning Machine* (ELM) menghasilkan tingkat akurasi sebesar 42,857%. Redahnya tingkat akurasi pada penelitian ini dikatakan karena adanya kemunculan *sparse* dan *matrix* sehingga terjadinya *overfitting*. Penelitian analisis sentimen selanjutnya dilakukan oleh Aditiya Rahman dan keempat penulis lainnya [5] menggunakan dataset sebanyak 1500 data komentar pada platform youtube dengan menerapkan metode naïve bayes menghasilkan akurasi sebesar 78,17% hal ini dapat dikatakan bahwa metode naïve bayes memiliki akurasi yang cukup tinggi. Penelitian lain yang dilakukan oleh Vincentius dan Fitriah [10] yang melakukan analisis ulasan hotel bahasa indonesia dengan melakukan pembobotan menggunakan *Term Frequency - Inverse Document Frequency* (TF-IDF) kemudian menerapkan metode *Support Vector Machine* (SVM) diperoleh hasil akurasi dari model tersebut cukup sebesar 85%.

Berdasarkan permasalahan diatas, penelitian ini bertujuan untuk melakukan analisis sentiment pada ulasan hotel New Saphir Yogyakarta menggunakan algoritma Naïve Bayes untuk mengetahui klasifikasi sentimen. Diharapkan kedepannya industri perhotelan dapat memanfaatkan penelitian ini untuk meningkatkan dan mengembangkan kualitas.

METODE PENELITIAN

Metode penelitian diawali dengan tahapan pengumpulan data, kemudian dilanjutkan pada tahap pembersihan data (*pre-processing data*) dan proses pelabelan data. Selanjutnya masuk kedalam tahap pemodelan dengan naïve bayes dan evaluasi model. Diagram alir metode penelitian ditunjukkan pada gambar 1 berikut.



Gambar 1. Diagram Alir Metode Penelitian

1. Pengumpulan Data

Pengambilan data komentar berasal dari website Tripadvisor. Tripadvisor merupakan sebuah platform perjalanan online terbesar di dunia yang sudah banyak sekali user yang menggunakan platform tersebut untuk perencanaan hingga pemesanan berbagai macam tiket. Tripadvisor merupakan platform yang cukup besar, maka data yang akan diambil juga banyak sehingga diperlukan bantuan dalam pengambilan data. Salah satu teknik yang digunakan untuk mengambil data dalam jumlah banyak adalah teknik *scraping*. Teknik *scraping* adalah proses pengambilan data dari situs web secara otomatis.

2. Data Pre-Processing

Tahap Pre-Processing Data merupakan tahapan dimana merubah data mentah menjadi data bersih, distandarisasi, dan merubah format sesuai dengan kebutuhan model sehingga data akan menjadi lebih mudah dipahami dan dianalisis. Tahap ini juga memiliki beberapa bagian yang harus diterapkan, antara lain:

- a) **Cleaning Text**, pada tahap ini dilakukan pembersihan setiap baris value data dengan membersihkan dari karakter simbol (@#\$\$%^*, dll), angka, link, merubah enter menjadi spasi dan menghapus karakter spasi di sisi kiri dan kanan pada text.
- b) **Case Folding**, tahap merubah semua huruf pada seluruh baris data menjadi huruf kecil.

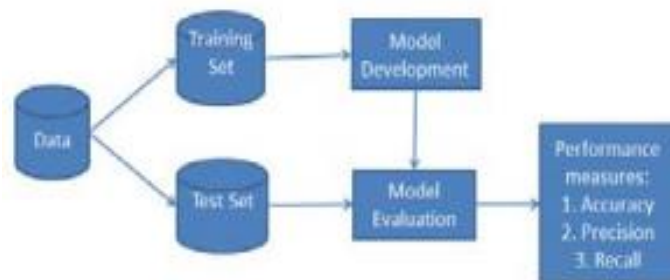
- c) Tokenizing, pada tahap ini dilakukan pemecahan data text menjadi sebuah list data kata.
- d) Stopword, tahap dimana dilakukan penghapusan kata yang tidak memiliki makna dalam teks. Sehingga data yang dihasilkan akan menjadi lebih ringkas.
- e) Stemming, tahap terakhir dengan mengubah kata yang ada menjadi kata dasar.

3. Labelling

Tahap labelling merupakan tahap pemberian label pada pada baris data. Labelling bertujuan untuk menghitung frekuensi kemunculan kata-kata dalam dokumen yang kemudian dihitung bobot berdasarkan frekuensi kemunculan tersebut. Proses labelling dapat menggunakan TF-IDF dan dapat dilakukan sebelum dilakukan proses modeling Naïve Bayes. Dari frekuensi data yang sering muncul dalam baris data dokumen positif dan negatif dapat digunakan untuk menghitung probabilitas dalam penggolongan sentimen positif dan negatif.

4. Modelling Naïve Bayes

Dalam menggunakan model Naïve Bayes terdapat dua pengelompokan data, yaitu data train dan data test. Data training merupakan data yang akan digunakan untuk melatih model, sedangkan data test adalah data yang digunakan untuk mengevaluasi kinerja model. Kemudian pada data training dilakukan pemodelan yang nantinya hasil dari model akan diterapkan pada data test. Dari hasil hasil data test model didapatkan hasil performance berupa akurasi, presisi, dan recall. Alur diagram pembuatan model ditunjukkan pada gambar 2 berikut.



Gambar 2. Diagram Alir Pemodelan Data

5. Evaluasi Model

Dari hasil pemodelan yang berupa perfoma akurasi, presisi dan recall menjelaskan bahwa kinerja model dalam memprediksi kelas data baru dapat diukur dari ketiga metrik tersebut dengan penjelasan masing-masing sebagai berikut:

- a) Akurasi, digunakan untuk mengukur kinerja model dengan melihatkan presentase data testing yang diprediksi.
- b) Presisi, melihatkan presentase data testing yang diprediksi sebagai positif dan benar-benar positif.
- c) Recall, melihatkan presentase data positif yang diprediksi dengan benar

HASIL PENELITIAN DAN PEMBAHASAN

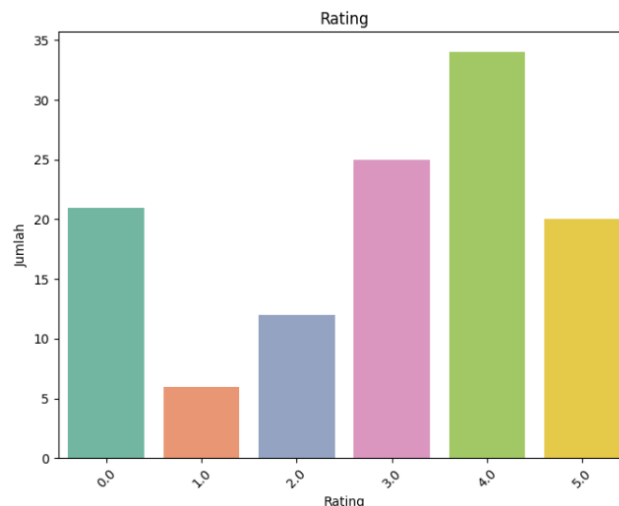
1. Eksplorasi Data

Pengumpulan data menggunakan teknik scraping pada website Tripadvisor. Dengan mengumpulkan data komentar pelanggan Hotel New Sahpir Yogyakarta. Dataset yang terkumpul sebanyak 1118 komentar. Selanjutnya dataset yang telah terkumpul dilakukan pre-processing data. Tahap preprocessing data yang digunakan untuk mengubah data mentah menjadi data yang memiliki kualitas baik dan sesuai untuk dilakukan analisis menggunakan metode Naïve Bayes. Hasil dari tahap ini adalah jumlah kata dalam komentar (word_count), jumlah karakter dalam text (char_count), dan hasil dari tahap Preprocessing data (Cleaning Text, Case Folding, Tokenizing, Stopword, dan Stemming). Gambar 3 menunjukkan hasil pre-processing data.

index	rating	komen	word_count	char_count	text_preprocessed
0	1.0	4 stars hotel but feels like not a 4 stars hotel. Bad service felt directly when I arrived at lobby. No security or staff help us to bring our luggage. Receptionist named Anjas was unable to respond well when I gave a feedback for improvement. Safety box was not ready to be used, their staff had to change the battery and other stuff. Room cleanliness was bad, I found used medicine on the floor. Overall experience was bad and will never come back again to stay there.	88	471	stars hotel but feels like not a 4 stars hotel bad service felt directly when i arrived at lobby no security or staff help us to bring our luggage receptionist named anjas was unable to respond well when i gave a feedback for improvement safety box was not ready to be used their staff had to change the battery and other stuff room cleanliness was bad i found used medicine on the floor overall experience was bad and will never come back again to stay ther
1	0.0	Second stay, Cek in room 103, room humid, ac trouble then found worm in the shower room so disgusting. No apology from the staff and asking me to take the new room key to front desk insted of deliver the new room key very uncomfortable. Food is ok, staff in swimming pool is friendly	54	283	second stay cek in room room humid ac trouble then found worm in the shower room so disgusting no apology from the staff and asking me to take the new room key to front desk insted of deliver the new room key very uncomfortable food is ok staff in swimming pool is friendli
2	1.0	THIEF!!! The valet parking staff swapped the full load E-Money card balance Rp650.000, with E-Money card balance Rp.4.500. . I put the card in the dashboard pocket, and didnt realize until we use it at the Toll gate with balance Rp4.500 only. . The toll officer helped us the verify the history of the card. (We stay in the hotel 25-27 Dec, and only valet once in 26 night due to no lot available). . History from the swapped card : last card used was in 21 Dec 2019, while our stolen card last used is 25 Dec 2019 (arrival to hotel) . It is clearly the card was swapped in the night of 26.	119	594	thief the valet parking staff swapped the full load e money card balance rp with e money card balance rp i put the card in the dashboard pocket and didnt realize until we use it at the toll gate with balance rp only the toll officer helped us the verify the history of the card we stay in the hotel dec and only valet once in night due to no lot available history from the swapped card last card used was in dec while our stolen card last used is dec arrival to hotel it is clearly the card was swapped in the night of

Gambar 3. Hasil Pre-Processing Data

Data yang telah dilakukan tahap pre-processing selanjutnya perlu dilakukan tahap preparation dalam bentuk pemberian label pengelompokan sentimen pada data. Pengelompokan dilakukan dengan menggunakan Vender_Lexicon. Dari hasil tahap labelling ini, dapat dilakukan visualisasi untuk melihat seberapa jumlah maupun presentasi dari penggolongan setiap segmen sentimentasi dan rating penilaian data pelanggan. Visualisasi antara lain, yaitu pada grafik pertama melihatkan seberapa banyak data rating yang diberikan oleh konsumen dari rentang 0-5. Dengan hasil tertinggi adalah pada rating 4. Grafik rating ditunjukkan pada gambar 4.



Gambar 4. Grafik Rating Komentar

2. Eksplorasi Model

Sebelum dilakukan pemodelan Naïve Bayes, dilakukan penentuan nilai frekuensi sebuah kata dalam sebuah dokumen dengan menggunakan Term Frequency–Inverse Document Frequency (TF-IDF). Selanjutnya, dilakukan klasifikasi Naïve Bayes menggunakan machine learning dengan membagi data dalam dua bagian yaitu data training dan testing. Perbandingan data training dengan testing yang digunakan adalah 80:20 dari 1118 Data. Berikut adalah potongan script pembentukan TF-IDF dan pemodelan naïve-bayes.

```
tfidf = TfidfVectorizer(lowercase=True)

text_tf =
tfidf.fit_transform(df_clean['text_preprocessed'].dropna().values.astype('U'))
text_tf

X_train_tf, X_test_tf, y_train, y_test =
train_test_split(df_clean['text_preprocessed'], df_clean['polarity'],
test_size = 0.2, random_state = 42)

cv = CountVectorizer(max_features=10000,ngram_range=(1, 3))
pipeline = Pipeline([
    ('vectorizer', cv),
    ('classifier', nb)
])
sentiment_fit = pipeline.fit(X_train_le, y_train)
y_pred = sentiment_fit.predict(X_test_le)
print(classification_report(y_test, y_pred,
target_names=['Negative','Positive']))
print("accuracy score :")
print(accuracy_score(y_test, y_pred))
```

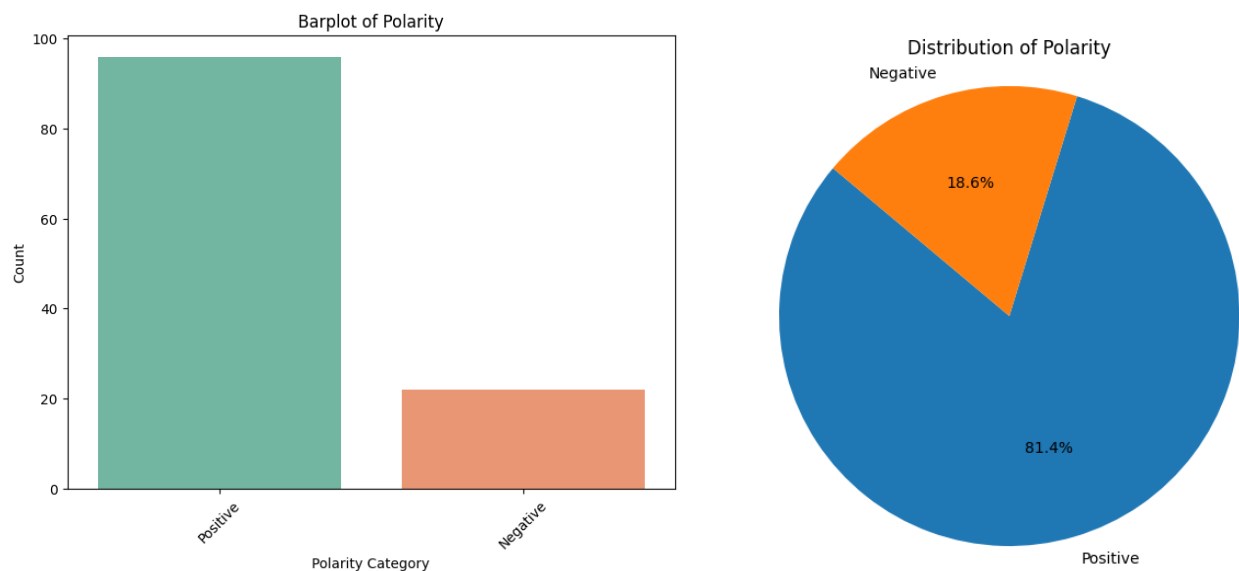
Tahap selanjutnya adalah mengukur akurasi dari model naïve bayes yang telah dibuat dengan evaluasi model. Evaluasi model dilakukan untuk menguji hasil klasifikasi dengan mengukur nilai kebenaran dari sistem yang telah buat. Hasil evaluasi model naïve bayes yang tampilkan dalam bentuk classification report yang menunjukkan precision, recall, f1-score, support dan accuracy. Dihasilkan tingkat akurasi model Naïve Bayes sebesar 92%. Gambar 5 menunjukkan hasil evaluasi model naïve bayes.

	precision	recall	f1-score	support
Negative	0.50	0.50	0.50	2
Positive	0.95	0.95	0.95	22
accuracy			0.92	24
macro avg	0.73	0.73	0.73	24
weighted avg	0.92	0.92	0.92	24
accuracy score				
0.9166666666666666				

Gambar 5. Hasil Evaluasi Model

3. Eksplorasi Hasil

Hasil dari pengukuran evaluasi model dibuktikan dengan hasil klasifikasi sentiment. Gambar 6 menunjukkan bahwa model naïve bayes berhasil melakukan klasifikasi pada sentiment positif dan sentiment negatif dengan proporsi sentiment positif lebih dari 80%.



Gambar 6. Presentase Klasifikasi Sentimen

SIMPULAN

Berdasarkan hasil pnegujian pada penelitian dapat disimpulkan bahwa naive bayes berhasil melakukan klasifikasi sentiment dengan sangat baik. Model yang dirancang menggunakan metode Naïve Bayes

mendapatkan tingkat akurasi sebesar 92%. Dari hasil analisis yang diberikan, terlihat bahwa hasil penilaian rating hotel dan komentar yang diberikan oleh konsumen memiliki presentase nilai positif lebih besar dari pada komentar bernilai negatif. Hal tersebut membuktikan bahwa Hotel New Saphir Yogyakarta mendapatkan penilaian dari pelanggan dengan komentar positif dikarenakan pelayanan yang diberikan memuaskan bagi para pelanggan. Saran untuk penelitian selanjutnya, adalah untuk dapat menambahkan model model machine learning lainnya seperti random forest dalam proses klasifikasi sentiment selanjutnya.

DAFTAR PUSTAKA

- [1] Kurniawan, S., Kusumaningrum, R., & Timu, M. E. (2019). *Hierarchical Sentence Sentiment Analysis of Hotel Reviews Using The Naïve Bayes Classifier*. IEEE.org.
- [2] Morama, H. C., Ratnawati, D. E., & Arwani, I. (2022). Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1702-1708.
- [3] Rahman, A., Rahmat, F., Fariqi, M. Y., & Adi, S. (2020). Metode Naive Bayes untuk Menganalisis Akurasi Sentimen Komentar di Youtube. *Jurnal EECCIS*, 31-34.
- [4] Santosa, B. (2021). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- [5] Saputro, I. W., & Sari, B. W. (2019). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Citec Journal*.
- [6] Singgalen, Y. A. (2022). Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naïve Bayes Classifier. *Journal Building of Informatics, Technology and Science (BITS)*, 1343–1352.
- [7] Thomas, V. W., & Rumaisa, F. (2022). Analisis Sentimen Ulasan Hotel Bahasa Indonesia Menggunakan Support Vector Machine dan TF-IDF. *Jurnal Media Informatika Budidarma*, 1767-1774.