

# Peningkatan Performa *Naive Bayes* Dengan Seleksi Atribut Menggunakan *Chi Square* Untuk Klasifikasi Loyalitas Pelanggan GRAB

Rianti Yunita Kisworini<sup>#1</sup>, Muhammad Akbar Setiawan<sup>\*2</sup>

<sup>#</sup> Prodi Teknik Informatika STMIK Widya Utama Purwokerto

Jl. Sunan Kalijaga, Dusun III, Karangnanas, Kec. Sokaraja, Kabupaten Banyumas, Jawa Tengah 53146

<sup>1</sup> riantiyunita32@gmail.com

<sup>2</sup> akbar@swu.ac.id

Accepted on 19-05-2020

## Abstrak

Transportasi menjadi salah satu dari beberapa area yang dipengaruhi oleh kemajuan perkembangan teknologi informasi. Dimana transportasi merupakan sarana yang diperlukan untuk mendukung aktivitas manusia dan mobilitas setiap saat, oleh karena itu transportasi harus dipersiapkan dengan baik dan aman. Layanan Grab *online* menjadi salah satu alat transportasi yang banyak digunakan oleh masyarakat Indonesia. *Naive Bayes* adalah salah satu algoritma *classifying* data yang biasa digunakan untuk mengklasifikasikan data. Tetapi kinerja yang baik dari *Naive Bayes* sangat dipengaruhi oleh penggunaan atribut pada *dataset*. Untuk mengatasi ini, penelitian ini mengusulkan peningkatan kinerja *Naive Bayes* dengan pemilihan atribut menggunakan *Chi Square*. Pada tahap pra pengolahan dilakukan pembersihan data untuk menghapus data yang kosong. Dari percobaan yang dilakukan, menunjukkan hasil pengujian dengan nilai akurasi yang diperoleh *Naive Bayes* dan *Chi Square* sebesar 99,51%., Tetapi standar *Naive Bayes* sebesar 92,73%.

**Kata kunci:** Penambahan data, klasifikasi, *Naive Bayes*, pra pemrosesan, pemilihan atribut, *Chi Square*, klasifikasi loyalitas pelanggan Grab

## I. PENDAHULUAN

Transportasi menjadi salah satu dari beberapa bidang yang ikut terpengaruh dari kemajuan perkembangan teknologi informasi. Dimana transportasi adalah sarana yang dibutuhkan dalam mendukung kegiatan dan mobilitas manusia setiap waktunya, maka dari itu transportasi harus disiapkan dengan baik dan aman. Jasa Grab *online* menjadi salah satu bagian sarana transportasi yang banyak dipergunakan masyarakat Indonesia, yang menawarkan berbagai alternatif penyediaan jasa yang tersedia diberbagai kota besar di wilayah Indonesia pada khususnya, untuk membantu dan memudahkan masyarakat dalam melakukan berbagai kegiatan dalam kesehariannya, seperti kegiatan pekerjaan, perkuliahan, perjalanan, penelitian, perniagaan ataupun pekerjaan penunjang lainnya[1].

Grab memiliki keunggulan tersendiri yang mampu dinikmati dan memuaskan kebutuhannya sekaligus mempunyai kekurangan yang membuat masyarakat mengeluh dan bisa berpindah ke penyedia

layanan ojek *online* lain. Sehingga Grab perlu memperhatikan apa yang menjadi keunggulan dan kelemahan produk layanannya agar tidak terjadi penurunan layanan dan berpindahkannya konsumen ke penyedia jasa lain.

Agar dapat mengetahui tingkat loyalitas dari pengguna jasa terhadap suatu penyedia layanan yang diberikan Grab, maka diperlukan pengklasifikasian loyalitas dari pelanggan Grab dengan menggunakan data primer atau menyebar kuesioner yang dilakukan di wilayah Kabupaten Banyumas Provinsi Jawa Tengah dengan mengadopsi dari jurnal yang ditulis oleh Ida Farida, dkk terkait kepuasan pelanggan pengguna Grab *online*.

Metode yang digunakan pada penelitian ini adalah *Naive Bayes*, pemilihan metode ini karena merupakan salah satu algoritma terbaik yang sering digunakan untuk menangani data mining dalam menyelesaikan masalah klasifikasi. Dimana *Naive Bayes* memanfaatkan teori probabilitas, yaitu memprediksi peluang atau probabilitas dimasa depan berdasarkan pengalaman dimasa sebelumnya. Algoritma *Naive Bayes Classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat dan relevan [2]. *Chi-Square* yang merupakan algoritma pendamping digunakan untuk meningkatkan performa *Naive Bayes* dan sebagai pembanding. Dalam penelitian ini yang diuji adalah mengklasifikasikan dataset yang diperoleh dari data primer berbentuk kuesioner untuk menghasilkan review Grab yang memiliki nilai akhir loyal atau tidak loyal.

## II. TINJAUAN PUSTAKA

Dalam penelitian ini, digunakan beberapa referensi sumber pustaka yang berasal dari jurnal dan penelitian yang sudah dilakukan sebelumnya, yang pertama Harun Al Rasyid, 2017 dalam penelitian yang berjudul “Pengaruh Kualitas Layanan dan Pemanfaatan Teknologi Informasi Terhadap Kepuasan dan Loyalitas Pelanggan Grab”. Sampel yang digunakan sebanyak 140 sampel, melalui penggunaan AMOS versi 19.00 dengan pemodelan SEM Penelitian ini memperlihatkan bahwa variabel yang signifikan saling berhubungan, ini menjadikan hasil uji dalam penelitian merupakan model akhir.

Penelitian lainnya dilakukan oleh Teguh Budi Santoso, 2017. Dengan judul “Analisa loyalitas pelanggan dengan variabel berdasarkan harga, sikap dalam melayani konsumen”. Pendekatan Penelitian ini adalah pembentukan model prediksi memakai metode C4.5 pada algoritma C4.5 dilakukan perhitungan entropy dan information gain dimana atribut harga, pelayanan, promosi, citra perusahaan dan kepercayaan sebagai atribut sumber untuk memperoleh node akar sedangkan atribut loyalitas pelanggan sebagai atribut tujuan.

Rujukan penelitian yang lain dilakukan oleh Jitendra Rout Kumar, Kim-Kwang Raymand Choo, Amiya Kumar Dash, Sambit Bakshi, Sanjay Kumar Jena, Karen L, Williams, 2018. Dengan judul “Identifikasi emosi dan sentimen data yang tidak terstruktur”. Dengan pendekatan *Maximum entropy dan Support Vector Machines* digunakan untuk mengidentifikasi tweets yang lebih baik. *Machine Learning Algorithms* dipergunakan untuk identifikasi sentimen

Penelitian lain dilakukan oleh Wa’el Hadi, Qasem Al-Radies, Samer Alhawari 2018. Dengan permasalahan nilai atribut yang mungkin tidak memiliki aturan dalam model AC untuk menebak data pengujian baru dengan benar. Untuk mengatasi masalah ini, sebuah algoritma baru yang disebut Hybrid Associative Classification (HAC) diusulkan dalam makalah ini.

Perbedaan dari penelitian sebelumnya, pada penelitian ini menggunakan algoritma *Naive Bayes* dengan fitur seleksi untuk mengklasifikasikan loyal dan tidak loyal dari data primer kuesioner pelanggan Grab. Algoritma *Naive Bayes* sebagai salah satu algoritma yang bisa digunakan untuk klasifikasi suatu produk. Proses klasifikasi dari komentar kadang kurang maksimal karena penggunaan fitur yang kurang relevan. Fitur yang tidak relevan adalah fitur yang memberikan sedikit informasi atau ketidakterkaitan dengan data. Sehingga perlu dilakukan penambahan seleksi fitur. Untuk itu perlu dilakukan seleksi fitur, dengan algoritma seleksi fitur agar didapatkan fitur-fitur yang relevan, sehingga memotong aturan-aturan (rule) klasifikasi lebih pendek. Salah satu metode seleksi fitur adalah dengan *Chi Square*. Dataset pada umumnya perlu diklasifikasikan antara kelas positif dan negatif, dalam penelitian ini mengusulkan metode *chi square* untuk pengklasifikasian loyal dan tidak loyal dari data primer kuesioner pelanggan Grab.

## II. METODE PENELITIAN

Penelitian ini merupakan penelitian eksperimen dengan data primer berupa kuesioner bagi pelanggan Grab, penelitian ini bertujuan untuk menguji model klasifikasi yang terbaik untuk klasifikasi dengan label akhir loyal dan tidak loyal dari pelanggan Grab.

### 3.1 Pengumpulan Data

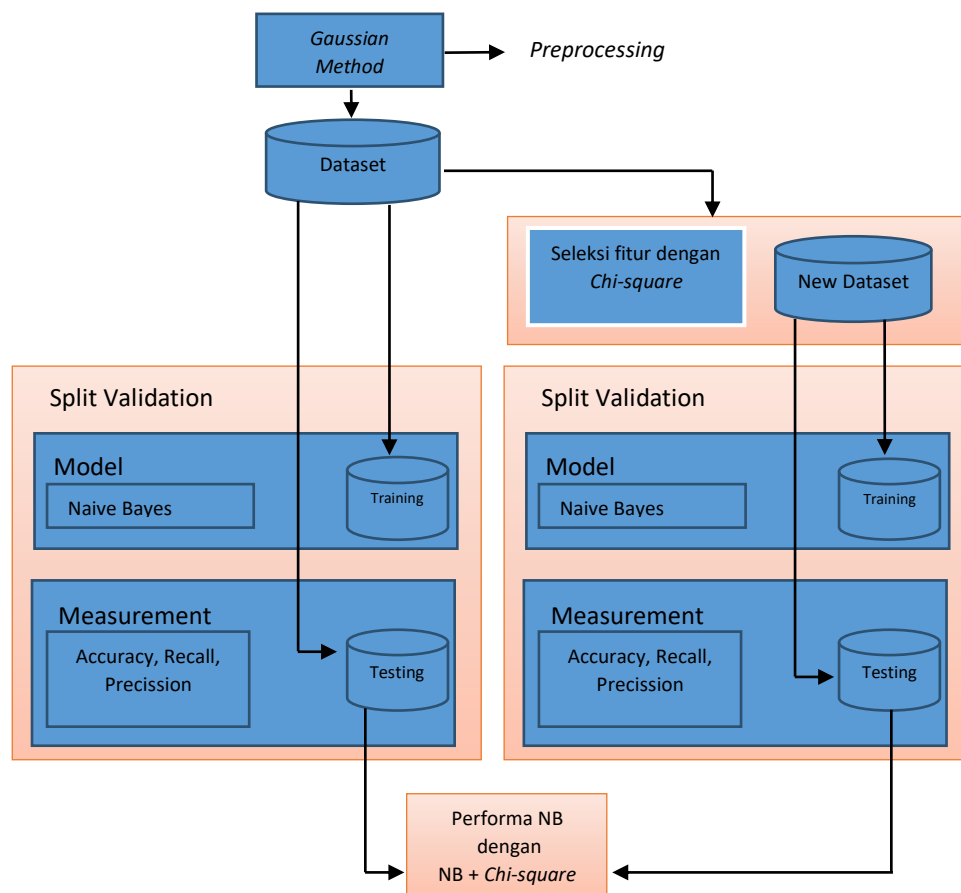
Data yang digunakan pada penelitian ini dikumpulkan dari pelanggan Grab berupa kuesioner dari 331 pelanggan Grab. Data tersebut diambil dari pengguna konsumen Grab di sekolah, kampus, pusat perbelanjaan serta dinas instansi yang berada di wilayah kota Purwokerto yaitu Purwokerto Utara, Barat, Timur dan Selatan serta beberapa daerah di Kabupaten Banyumas..

### 3.2 Pengolahan Data Awal

Pada tahap ini dilakukan analisa terkait atribut sebelum digunakan untuk proses selanjutnya, atribut yang digunakan dalam penelitian ini sebanyak 4 (empat) atribut yaitu kualitas pelayanan, harga, kepuasan pelanggan dan loyalitas pelanggan. Atribut penilaian kepuasan pelanggan digunakan sebagai label[5][8]. Kemudian dari hasil tersebut langkah berikutnya adalah melakukan *replace missing value* berdasarkan model *average* yaitu menggantikan nilai yang kosong dengan nilai rata-rata yang sering muncul pada data kuesioner pelanggan Grab.

### 3.3 Metode yang Diusulkan

Menggambarkan metode yang digunakan dan menerangkan bagaimana metode yang diusulkan untuk menyelesaikan permasalahan pada penelitian diperlihatkan pada gambar 1.



Gambar 1. Metode Yang Diusulkan

### 3.4 Eksperimen dan Pengujian Metode

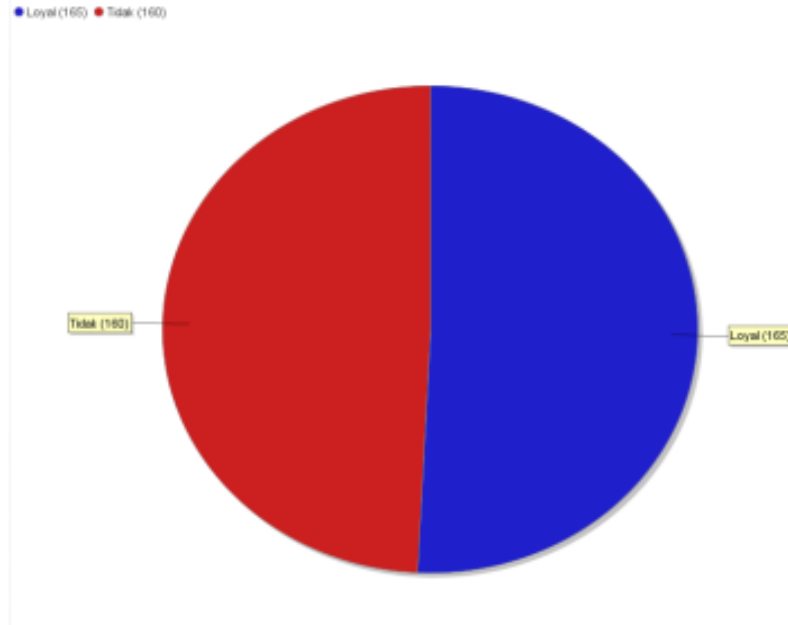
Pada tahap ini data dianalisis, dikelompokkan variabel mana yang berhubungan dengan satu sama lainnya. Setelah data dianalisis lalu diterapkan model-model yang sesuai dengan jenis data. Pembagian data kedalam data latihan (training data) dan data uji (testing data) juga diperlukan untuk pembuatan model.

### 3.5 Evaluasi dan Hasil Validasi

Metode yang dihasilkan akan diuji dengan menggunakan *split validation* untuk mengetahui tingkat akurasi, *split validation* digunakan untuk menghindari *overlapping* pada data testing. Tahapan *split validation* terdiri dari proses pembagian data menjadi subset yang berukuran sama kemudian menggunakan setiap subset untuk data testing dan sisanya untuk data training.

## IV. HASIL DAN PEMBAHASAN

Gambar dibawah ini menjelaskan bahwa hasil pengklasifikasian menggunakan *rapid miner* sesuai dengan perhitungan manual, yaitu loyal sebanyak 168 dan tidak loyal sebanyak 163.



Gambar 2. Hasil Grafik hasil pengklasifikasian menggunakan *rapid miner*

Berdasarkan hasil eksperimen menggunakan Rapid Miner, *Split validation* dilakukan untuk mengevaluasi dan membandingkan *learning algorithms* dengan membagi data menjadi dua. Percobaan menggunakan *Split Validation* dilakukan sebanyak 5 kali percobaan dengan perbandingan pada tahap 1 (dengan perbandingan data training sebanyak 90% dan data testing sebanyak 10%), tahap 2 (dengan perbandingan data training sebanyak 80% dan data testing sebanyak 20%), tahap 3 (dengan perbandingan data training sebanyak 70% dan data testing sebanyak 30%), tahap 4 (dengan perbandingan data training sebanyak 60% dan data testing sebanyak 40%), dan tahap 5 (dengan perbandingan data training sebanyak 50% dan data testing sebanyak 50%). Setelah dilakukan percobaan beberapa kali, maka rangkuman hasil uji terdapat pada tabel berikut ini.

Tabel 1  
 Rangkuman Hasil Percobaan

Nomor	PROPORSI DOKUMEN		JUMLAH DOKUMEN		HASIL
	TRAINING	TESTING	TRAINING	TESTING	ACCURACY
1	83%	23%	68	263	98,76%
2	75%	35%	102	233	98,01%
3	63%	43%	133	198	97,44%

Pada hasil tersebut dapat kita ketahui bahwa pengujian dengan tingkat akurasi tertinggi didapat dengan hasil 98,76%. Selanjutnya pada pengujian model dan *wordlist* akan diimplementasikan untuk melakukan prediksi terhadap data yang ada.

Dari eksperimen yang telah dilakukan, selanjutnya akan dibandingkan hasil penelitian yang telah diperoleh untuk kedua model yang telah digunakan. Tabel dibawah ini menunjukkan hasil yang diperoleh dari klasifikasi loyalitas pelanggan Grab.

Tabel 2  
 Hasil Penelitian

Hasil Pengujian	Naive Bayes	Naive Bayes + seleksi atribut <i>Chi Square</i>
Accuracy	92,73%	99,51%
Precision	98,81	98,32%
Recall	86,30%	100,00%

## V. KESIMPULAN

Performa *Naive Bayes* sangat sensitif dengan penggunaan atribut dari *dataset* yang digunakan, sehingga akan mempengaruhi hasil dari performa *Naive Bayes* khususnya pada nilai *accuracy*. Dengan menerapkan *attribute selection* menggunakan *chi square*, masih dapat meningkatkan performa *Naive Bayes* untuk loyalitas pelanggan Grab. Dengan peningkatan nilai *accuracy* sebesar 99,51%, sedangkan nilai *accuracy Naive Bayes* standar sebesar 92,73%. Hasil dari penelitian ini menunjukkan bahwa penggunaan *attribute selection* menggunakan *chi square* dapat mempengaruhi performa *Naive Bayes*. Tetapi pada penelitian ini belum sepenuhnya menggunakan atribut yang sama dengan penelitian dari Ida Farida, sehingga disarankan untuk penelitian berikutnya dapat menggunakan seluruh atribut yang digunakan oleh Ida Farida. Atribut tersebut diantaranya adalah produk, promosi dan proses.

## UCAPAN TERIMAKASIH

Penulis menyadari bahwa tanpa bantuan dan dukungan dari berbagai pihak penyusunan jurnal dan penelitian ini tidak dapat berjalan dengan baik. Sehubungan dengan hal tersebut, maka pada kesempatan ini, perkenankanlah penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Allah S.W.T karena telah memberikan kesehatan dan kekuatan untuk menyelesaikan penelitian ini.
2. Kedua orang tua dan mertua tercinta untuk dukungan moril maupun spiritual.
3. Suami tercinta Nuryanto dan juga anak-anak tersayang Arshaka Nur Azzam dan Hamas Nur Hafiz yang selalu memberikan semangat dan dukungan disetiap keadaan.
4. Muh Sofi'i, S.E., M.Si. selaku Ketua STMIK Widya Utama Purwokerto.
5. Seluruh dosen dan staf tata usaha pada lingkungan STMIK Widya Utama Purwokerto

Akhir kata semoga Allah SWT memberikan balasan yang sepadan dengan kebaikan-kebaikan yang telah diberikan dan semoga karya ini dapat bermanfaat bagi semua pihak serta memberikan sumbangan kepada Ilmu Pengetahuan dan Teknologi khususnya di bidang Informatikan, *Amin Yaa Robbal Alamin.*

## DAFTAR PUSTAKA

- [1] Ida Farida, dkk, "Analisis Pengaruh Bauran Pemasaran 7P Terhadap Kepuasan Pelanggan Pengguna GOJEK Online" *Jurnal Riset Manajemen dan Bisnis* Vol.1, Juni 2016
- [2] K. Widhiyantil, "Dengan HMM dan Rule Based," *Jurnal Informatika UKDW*, vol. 8, no. 2, 2012.
- [3] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Third Edition*. Elsevier Inc., 2011.
- [4] Kusriani and L. E. Taufiq, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.
- [5] N. Zaidi and J. Cerquides, "Alleviating Naive Bayes attribute independence assumption by attribute weighting," *J. Mach. ...*, vol. 14, pp. 1947–1988, 2013.
- [6] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [7] J. Ling and T. B. Oka, "Analisis Sentimen Menggunakan Metode Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *E Jurnal Matematika Universitas Udayana*, vol. 3, no. 3, pp. 92–99, 2014.
- [8] A. Mudry and A. Tjellström, "Historical background of bone conduction hearing devices and bone conduction hearing aids," *Adv. Otorhinolaryngol.*, vol. 71, pp. 1–9, 2011.
- [9] U. Topics, *Principles of Data Mining*, Second edi. London: Springer London.