



## Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization

Suwanda Aditya Saputra<sup>1</sup>, Didi Rosiyadi<sup>2</sup>, Windu Gata<sup>3</sup>, Syepri Maulana Husain<sup>4</sup>

<sup>1,3</sup> Magister Ilmu Komputer, Fakultas Ilmu Komputer, STMIK Nusa Mandiri Kramat

<sup>2</sup> Pusat Penelitian Informatika LIPI

<sup>4</sup> Universitas Muhammadiyah Tangerang

<sup>1</sup>wanda.adit@gmail.com, <sup>2</sup>didi.rosiyadi@gmail.com, <sup>3</sup>windu@nusamandiri.ac.id, <sup>4</sup>syepri.maulana@umt.ac.id

### Abstract

Increasingly sophisticated technology brings various conveniences both in transportation, information, education to the convenience of transactions in shopping, such as the development of E-wallet can now be easily done using a smartphone. From a number of e-wallet products, researchers took a case study, which is OVO product, which is currently being discussed by many groups, especially in the capital of Jakarta today. Customers or clients who are not satisfied with the services or products offered by a company will usually write their complaints on social media or reviews on Google play. However, monitoring and organizing opinions from the public is also not easy. For this reason, we need a special method or technique that is able to categorize these reviews automatically, whether positive or negative. The algorithm used in this study is Naive Bayes Classifier (NB), with the optimization of the use of Particle Swarm Optimization Feature Selection (FS). The results of cross validation NB without FS are 82.30% for accuracy and 0.780 for AUC. Whereas for NB with FS is 83.60% for accuracy and 0.801 for AUC. Very significant improvement with the use of Feature Selection (FS) Particle Swarm Optimization.

Keywords: particle swarm optimization, naive bayes classifier, e-wallet, OVO

### Abstrak

Teknologi yang semakin canggih menghadirkan berbagai kemudahan baik dalam transportasi, informasi, edukasi hingga kemudahan bertransaksi dalam berbelanja, seperti perkembangan *E-wallet* kini dapat dengan mudah dilakukan dengan menggunakan *smartphone*. Dari beberapa produk *e-wallet* peneliti mengambil studi kasus yaitu produk OVO yang saat ini banyak dibicarakan banyak kalangan khususnya di ibukota Jakarta saat ini. Pelanggan atau klien yang merasa tidak puas dengan layanan atau produk yang ditawarkan oleh sebuah perusahaan biasanya akan menuliskan keluhannya di media sosial atau ulasan di *Google play*. Akan tetapi, memantau dan mengorganisasi opini dari masyarakat juga bukanlah hal yang mudah. Oleh sebab itulah, diperlukan sebuah metode atau teknik khusus yang mampu mengkategorikan *review-review* tersebut secara otomatis, apakah termasuk positif atau negatif. Algoritma yang digunakan dalam penelitian ini adalah *Naive Bayes Classifier* (NB), dengan optimasi penggunaan *Feature Selection* (FS) *Particle Swarm Optimization*. Hasil dari *cross validation* NB tanpa FS adalah 82.30 % untuk *accuracy* dan 0.780 untuk AUC. Sedangkan untuk NB dengan FS adalah 83.60 % untuk *accuracy* dan 0.801 untuk AUC. Peningkatan sangat signifikan dengan penggunaan *Feature Selection* (FS) *Particle Swarm Optimization*.

Kata kunci: *particle swarm optimization, naive bayes classifier, e-wallet, OVO*

© 2019 Jurnal RESTI

### 1. Pendahuluan

Era digital telah membuat manusia memasuki gaya hidup baru yang tidak bisa lepas dari perangkat yang serba elektronik. Teknologi menjadi alat yang membantu kebutuhan manusia, dengan teknologi, apapun dapat dilakukan dengan lebih mudah. Begitu pentingnya peran teknologi inilah yang mulai

membawa peradaban memasuki ke era digital. Meningkatnya kebutuhan akan data dan informasi mendorong manusia untuk mengembangkan teknologi baru agar pengolahan data dan informasi dapat dilakukan dengan mudah dan cepat [1]. Teknologi yang semakin canggih menghadirkan berbagai kemudahan baik dalam transportasi, informasi, edukasi hingga

kemudahan bertransaksi dalam berbelanja, seperti perkembangan *E-wallet* kini dapat dengan mudah dilakukan dengan menggunakan *smartphone*. *E-wallet* sendiri dapat diartikan sebagai dompet digital atau bisa dikatakan sebagai uang elektronik untuk memudahkan bertransaksi dalam pembayaran non tunai. Untuk menggantikan pembayaran tunai seperti *mobile payment* khususnya *e-wallet* dibutuhkan kemauan dari pengguna agar mereka mau menggunakan *mobile payment* sebagai pengganti pembayaran konvensional.

Masuknya teknologi tidak bisa dipungkiri mempengaruhi berbagai segi kehidupan manusia. Termasuk dalam hal jual beli atau penggunaan uang tunai. Di Indonesia bermunculan banyak layanan keuangan berbasis digital dan ini termasuk *e-wallet*. Dengan layanan ini bisa melakukan transaksi untuk berbagai keperluan dengan saldo yang ada pada *e-wallet*. Di kota-kota besar, terutama Jakarta, layanan ini umum digunakan banyak orang terutama mereka yang berada dalam usia produktif. Alasan utama mereka menggunakan *e-wallet* adalah karena praktis. Mereka tidak perlu menyiapkan uang tunai saat bertransaksi. Sedangkan bagi penjual, mereka tidak perlu repot menyediakan kembalian karena pembayaran pasti dilakukan dengan nominal yang sesuai dan beberapa layanan *e-wallet* di Indonesia yang muncul dalam bentuk *mobile wallet* atau aplikasi.

Dari beberapa produk *e-wallet* peneliti mengambil studi kasus yaitu produk OVO yang saat ini banyak dibicarakan banyak kalangan khususnya di ibukota Jakarta saat ini.

*Google Play* adalah layanan konten digital milik *Google* yang terdiri dari toko produk-produk *online* merupakan musik/lagu, buku, aplikasi, permainan, ataupun pemutar media berbasis awan. *Google Play* dapat diakses melalui web, aplikasi android (*Play Store*), dan *Google TV*. Dalam *Google Play* dilengkapi dengan adanya fitur berisi ulasan dari para pengguna yang dapat digunakan untuk melihat ulasan dari pengguna aplikasi. Ulasan dari pengguna sering digunakan sebagai alat yang efektif dan efisien dalam menemukan informasi terhadap suatu produk atau jasa. Bahwa penelitian baru-baru ini menemukan hampir 50% dari pengguna internet bergantung pada rekomendasi *word-of-mouth* (opini) sebelum menggunakan suatu produk, karena *review* dari pengguna lain dapat menyediakan informasi terbaru dari produk tersebut berdasarkan perspektif pengguna-pengguna lain yang sudah menggunakan produk tersebut.

Disadari atau tidak, opini-opini pelanggan yang dituliskan di media sosial atau ulasan di *Google play*, sedikit atau pun banyak, akan memberikan pengaruh pada calon pelanggan. Akan tetapi, memantau dan mengorganisasi opini dari masyarakat juga bukanlah hal yang mudah. Opini yang dimuat di media sosial

jumlahnya terlalu banyak untuk diproses secara manual. Oleh sebab itulah, diperlukan sebuah metode atau teknik khusus yang mampu mengkategorikan *review-review* tersebut secara otomatis, apakah termasuk positif atau negatif.

Analisis *sentiment* atau *opinion mining* adalah studi komputasi mengenai pendapat, perilaku dan emosi seseorang terhadap entitas. Entitas tersebut dapat menggambarkan individu, kejadian atau topik. Analisis sentimen merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat [2].

Pada penelitian yang sudah dilakukan mengenai analisis sentimen, terdapat penelitian tentang analisis sentimen terhadap opini masyarakat mengenai penyedia layanan telekomunikasi seluler. Pada penelitian tersebut, menggunakan metode yang digunakan adalah *Support Vector Machine* dengan menggunakan *Lexicon Based Features* sebagai pembaharuan fiturnya selain memakai fitur *TF-IDF*. Data yang sudah didapatkan akan dilabeli positif, netral dan negatif kemudian dikoreksi oleh ahli bahasa. Data yang digunakan pada penelitian ini sebanyak 300 data yang dibagi menjadi dua jenis data dengan perbandingan 70% untuk data latih dan 30% untuk data uji. Hasil akurasi sistem yang diperoleh dari analisis sentimen dengan metode *Support Vector Machine* dan *Lexicon Based Features* sebesar 79% menggunakan nilai *degree* sebesar 2, nilai konstanta *learning rate* 0,0001, serta jumlah iterasi maksimum sebanyak 50 kali. Sedangkan sistem analisis sentimen tanpa menggunakan *Lexicon Based Features* menghasilkan tingkat akurasi sebesar 84% dengan nilai parameter yang sama [3].

Di dalam penelitian ini, akan dibahas tahapan yang dilalui untuk melakukan proses analisis sentimen terhadap komentar tentang aplikasi OVO pada *Google Play*. Dimulai dari tahap *preprocessing* sampai tahap analisis sentimen dengan *Naive Bayes Classifier* dan *Feature selection Particle Swarm Optimization (PSO)* serta bagaimana mengukur kualitas hasil analisis menggunakan tanpa *feature selection* dan dengan *feature selection*.

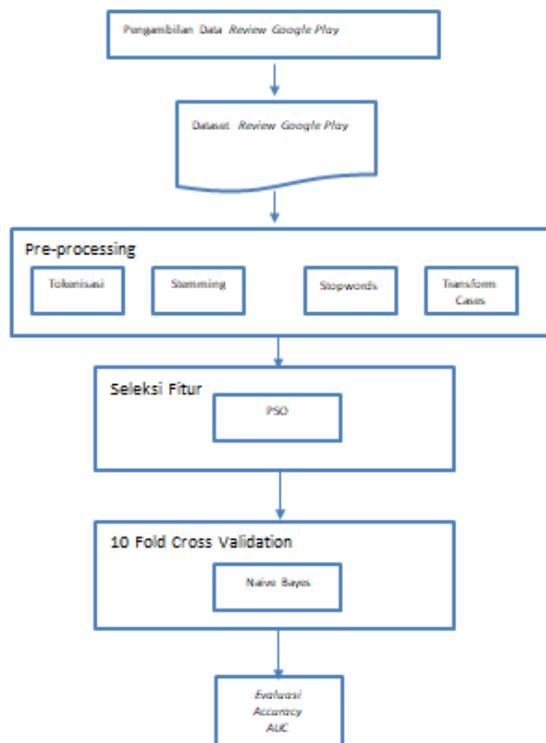
## 2. Metode Penelitian

Pada penelitian ini penulis mencoba menggunakan metode *Naive Bayes* berbasis PSO untuk menghasilkan akurasi terbaik dalam menganalisa komentar pelanggan terhadap aplikasi OVO pada *Google Play*. Kerangka pemikiran pada penelitian ini terdapat pada Gambar 1.

### 2.1. Pengumpulan data

Proses pengumpulan data dikumpulkan dari *review* pengguna aplikasi OVO pada *google play*. Data

tersebut didapatkan dari proses secara manual satu persatu dari komentar pengguna aplikasi, kemudian dikelola dengan bantuan aplikasi *rapidminer* untuk masuk ketahap *preprocessing*.



Gambar 1. Kerangka Berpikir dalam penelitian

## 2.2. Preprocessing data

Tahap selanjutnya adalah melakukan persiapan data yang telah didapat agar dapat terolah pada saat melakukan permodelan. Tahap *preprocessing* mencakup kegiatan membangun data dan juga membersihkan data agar siap untuk dikelola ke tahap selanjutnya. Berikut tahap *preprocessing* data antara lain :

### 1. Tokenization

Proses tokenisasi pada data teks adalah melakukan memecah sekumpulan karakter (kalimat) menjadi potongan karakter atau kata-kata sesuai kebutuhan yang sering disebut token.

### 2. Stemming

Setelah hasil dari *Tokenization (Regex)* akan dilanjutkan dengan proses *stemming* yaitu menghilangkan imbuhan yang terdapat pada masing-masing kata sehingga menjadi kata dasar dengan menggunakan *indonesian stemming* untuk *review* berbahasa Indonesia.

### 3. Indonesian Stopword Removal

Kemudian untuk tahapan selanjutnya adalah *Indonesian stopword removal*, dalam proses ini kata-kata yang tidak relevan akan dihapus, seperti kata tetapi, untuk, dengan, yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang

lain dan tidak terkait dengan kata sifat yang berhubungan dengan *sentiment*. Untuk text contoh tahapan ini diambil berbeda dengan text sebelumnya, dikarena pada text contoh sebelumnya tidak ada perubahan sebelum maupun setelah *Indonesian Stopword Removal*.

### 4. Transform Case

Operator yang digunakan pada tahapan ini adalah untuk mengubah huruf kapital yang masih ada pada *text* akan diubah menjadi huruf kecil semua. Hal ini dilakukan agar ketikan dilakukan proses ke dalam model klasifikasi terdapat keseragaman huruf dan tidak terjadi kesalahan dalam proses *tokenize*.

## 2.3. Feature Selection (Particle Swarm Optimization)

Setelah melakukan tahap *pre-processing* maka tahap selanjutnya akan melakukan seleksi fitur atas data yang telah melalui tahap *pre-processing*. Dalam penggunaan seleksi fitur peneliti menggunakan Algoritma *Particle Swarm Optimization (PSO)*.

*Particle Swarm Optimization (PSO)* ditemukan oleh James Kennedy dan Russ Eberhart pada tahun 1995. *Particle Swarm Optimization iPSO<sub>i</sub>* terinspirasi oleh tingkah laku sosial kawanan burung yang terbang berduyun-duyung (*bird flocking*) atau gerombolan ikan yang berenang berkelompok (*fish schooling*). Ratusan burung mampu terbang tanpa bertabrakan dan ratusan ikan berenang sanggup bergerak cepat tanpa saling bertabrakan. Padahal jarak mereka begitu dekat.

Menurut Basari bahwa *Particle Swarm Optimization (PSO)* merupakan teknik optimasi yang sangat sederhana untuk menerapkan dan memodifikasi beberapa parameter. Dalam *Particle Swarm Optimization (PSO)* terdapat beberapa teknik untuk melakukan pengoptimasian diantaranya meningkatkan bobot atribut (*attribute weight*) terhadap semua atribut atau variabel yang dipakai, menseleksi atribut (*attribute selection*) dan *feature selection* [4].

*Particle Swarm Optimization (PSO)* dimulai dengan suatu populasi yang terdiri atas sejumlah partikel (yang menyatakan calon-calon solusi) yang dibangkitkan secara acak. Selanjutnya dilakukan perbaruan posisi dan kecepatan terbang setiap partikel secara iteratif untuk menghasilkan solusi baru yang lebih baik. *Particle Swarm Optimization (PSO)* akan berhenti ketika solusi optimum telah ditemukan atau kondisi tertentu telah tercapai. Dengan konsep ini, *Particle Swarm Optimization (PSO)* mudah diimplementasikan dengan sedikit pengaturan parameter.

### 2.3. Algoritma Naive Bayes Classifier

Metode ini menggunakan teorema *Bayes*, yang ditemukan oleh Thomas Bayes abad ke-18 [5]. Klasifikasi *Naive Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Menurut Wu dan

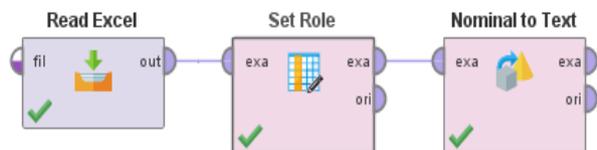
Kumar bahwa *Naive Bayes* merupakan metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik dalam *data mining*. *Naive Bayes* menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada *data training* [6].

### 3. Hasil dan Pembahasan

Pada bab ini akan dijelaskan mengenai proses yang dijalankan dalam penelitian ini pada Gambar 1. Ada beberapa langkah yang diterapkan pada penelitian ini.

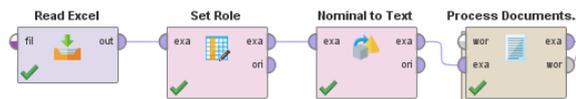
#### 3.1. Pengambilan Dataset

Proses pengambilan data diambil dari *google play* dari ulasan pengguna. Sehingga memudahkan para peneliti untuk mengambil data yang tersedia pada *google play*. Data tersebut dimasukkan dalam *excel* untuk mempermudah proses pengolahan data pada *Tools RapidMiner*.



Gambar 2. Proses pengolahan data pada *Tools RapidMiner*

Setelah *Read Excel* terhubung dengan operator *set role* selanjutnya menetapkan atribut dan label terlebih dahulu pada operator *set role*. Setelah penentuan atribut dan label, langkah selanjutnya melakukan konversi dari nominal *to string*. Setelah terkonversi menjadi binomial, lakukan tahap proses *pre-processing document*.



Gambar 3. Operator *Process Documents from Data*

#### 3.2. Proses Pre-Processing

Setelah dataset terkumpul, maka selanjutnya adalah proses untuk memulai pengolahan data, yaitu proses *pre-processing*. Data tersebut tidak bisa langsung dimasukkan dalam pengolahan untuk sentimen analisis, maka dilanjutkan dengan tahapan *Data Preparation*. Dapat dilihat pada Gambar 4.



Gambar 4. Operator *Process Pre-Processing*

Di dalam proses document terdapat operator *tokenize*, *filter stopwords*, *transform cases* dan *filter tokens*.

##### 1. Tokenize.

proses *Tokenize* yaitu semua kata yang ada di dalam tiap dokumen dikumpulkan dan dihilangkan tanda baca, angka, simbol, karakter khusus atau apapun yang bukan huruf. Tabel 1 memperlihatkan perbedaan *text* sebelum dan sesudah proses ini.

Tabel 1. Hasil dari *Tokenize*

Text	Tokenize
alhamdulillah sangat membantu dalam segala hal banyak promonya jadi lebih hemat dalam pengeluaran terimakasih	alhamdulillah sangat membantu dalam segala hal banyak promonya jadi lebih hemat dalam pengeluaran terimakasih

##### 2. Stopword Removal.

Selanjutnya adalah penggunaan operator *Stopword Removal (by Directory)* yang berfungsi untuk menghilangkan kata-kata yang tidak hubungan dengan isi *text*.

##### 3. Transform Case

Operator yang digunakan pada tahapan ini adalah untuk mengubah huruf kapital yang masih ada pada *text* akan diubah menjadi huruf kecil semua. Hal ini dilakukan agar ketikan dilakukan proses ke dalam model klasifikasi terdapat keseragaman huruf dan tidak terjadi kesalahan dalam proses *tokenize*.

##### 4. Filter Token (by Length)

Ini adalah proses yang ada pada *data preparation* untuk menghilangkan sejumlah kata (setelah proses *tokenize*) dengan panjang karakter tertentu. Pada penelitian ini panjang minimum karakter yang digunakan adalah 4 karakter dan panjang maksimum 25 karakter. Artinya kata yang panjangnya kurang dari 4 karakter dan lebih dari 25 karakter akan dihilangkan. Untuk mendapatkan hasil seperti ini maka dilakukan *setting* pada parameter dari operator ini (Gambar 5).

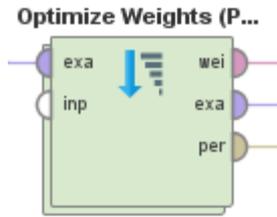


Gambar 5. Parameters dari *Filter Tokens (by Length)*

##### 3.3. Particle Swarm Optimization (PSO)

Setelah proses *pre-processing* atas *dataset* yang ada, maka selanjutnya diteruskan kepada proses melakukan pengoptimasian diantaranya meningkatkan bobot atribut (*attribute weight*) terhadap semua atribut atau variabel yang dipakai, menyeleksi atribut (*attribute*

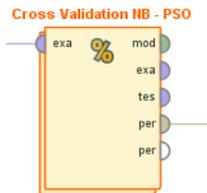
selection) dan feature selection. PSO adalah suatu teknik optimasi yang sangat sederhana untuk menerapkan dan memodifikasi beberapa parameter [7].



Gambar 6. Operator Particle Swarm Optimization (PSO)

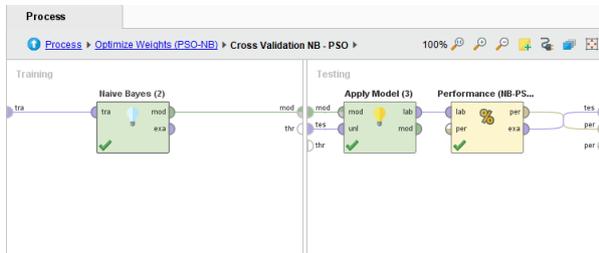
### 3.4. Proses Validasi Algoritma

Pada proses ini menggunakan beberapa operator, sebelumnya menggunakan operator cross validation dengan k-10 fold cross validation. Berikut untuk operator yang digunakan :



Gambar 7. Operator Cross Validation K-Fold

Yang di dalamnya terdapat berbagai operasi di antaranya dapat dilihat pada Gambar 8.



Gambar 8. Kumpulan Operator Cross Validation

### 3.5. Evaluasi

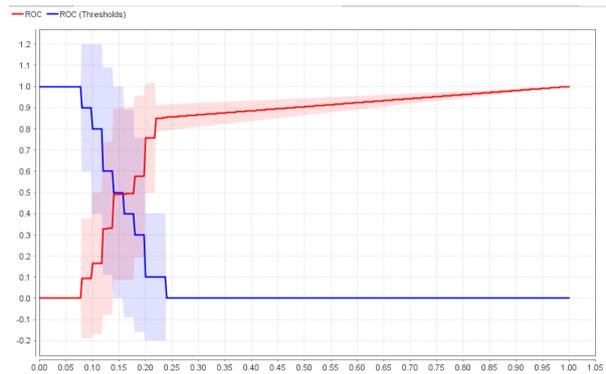
Penelitian ini menggunakan algoritma Naive Bayes Classifier untuk melakukan analisis sentimen, dan untuk melakukan evaluasinya menggunakan Accuracy dan AUC (Area Under Curve). Dari tahapan-tahapan yang dilakukan maka menghasilkan Accuracy dan AUC dari algoritma Naive Bayes Classifier tanpa feature selection dapat dilihat pada Tabel 2.

Tabel 2. Hasil Cross Validation dari Naive Bayes Classifier tanpa feature selection

NBC tanpa feature selection	
Accuracy	82.30 %
AUC	0.780

Diketahui bahwa dari cross validation Naive Bayes Classifier tanpa feature selection hanya mendapatkan

82.30% sedangkan untuk nilai AUC nya sebesar 0.780 yang dapat dilihat pada Gambar 9.



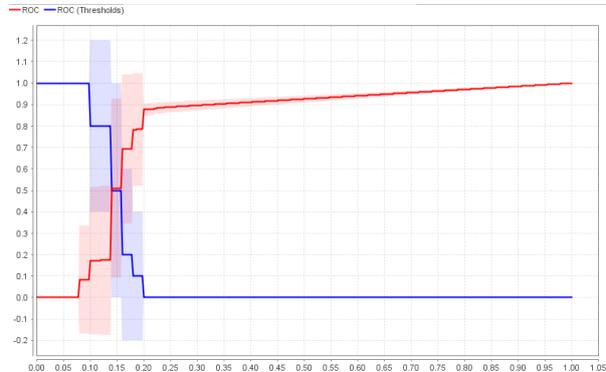
Gambar 9. Bentuk AUC untuk Naive Bayes Classifier tanpa feature selection

Sedangkan untuk nilai accuracy dan AUC untuk cross validation algoritma Naive Bayes Classifier dengan feature selection dapat dilihat di Table 3.

Tabel 3. Hasil Cross Validation dari Naive Bayes Classifier dengan feature selection

NBC dengan feature selection	
Accuracy	83.60 %
AUC	0.801

Dari Tabel 3 bisa terlihat bahwa hasil cross validation dari algoritma Naive Bayes Classifier dengan feature selection naik menjadi 83,60 % dan untuk nilai AUC menjadi 0801. Dapat dilihat pada gambar 10.



Gambar 10. Bentuk AUC untuk Naive Bayes Classifier dengan feature selection

Dari kedua hasil cross validation diatas maka di dapat peningkatan selisih peningkatan dengan penggunaan feature selection untuk nilai accuracy dan AUC. Dengan melihat Tabel 4 bisa dilihat peningkatan nilai accuracy sebesar 1.3% dan untuk AUC sebesar 0.021. Perbedaan sangat terlihat signifikan untuk kenaikan presentasinya.

Tabel 4. Perbandingan Cross Validation dari Naive Bayes Classifier tanpa feature selection dan dengan feature selection

	NBC tanpa FS	NBC + FS	Selisih
Accuracy	82.30 %	83.60 %	1.3%
AUC	0.780	0.801	0.021

#### 4. Kesimpulan

Setelah melakukan penelitian dengan model algoritma *Naive Bayes Classifier* tanpa *feature selection* dan dibandingkan dengan model *Naive Bayes Classifier* dengan *feature selection* dapat disimpulkan bahwa penggunaan *feature selection Particle Swarm Optimization* (PSO) dapat meningkatkan nilai dari *Accuracy* dan AUC. Peningkatan yang didapatkan sangat signifikan yang sebelumnya model *Naive Bayes Classifier* tanpa *feature selection* hanya menghasilkan 82.30% dan nilai AUC 0.780 setelah ditambahkan penggunaan *feature selection* menjadi 83.60% untuk *accuracy* dan nilai AUC 0.801.

*Feature selection Particle Swarm Optimization* (PSO) berpengaruh besar terhadap hasil *accuracy* yang didapatkan, yang memberikan kenaikan sekitar 1.3% untuk hasil yang didapatkan.

#### Daftar Rujukan

- [1] Josi, A., Abdillah, L.A., & Suryayusra. (2014). Penerapan Teknik Web Scraping Pada Mesin Pencari Artikel Ilmiah. Jurnal Sistem Informasi, Volume 5, Nomor 2, September 2014,

hlm. 159-164.

- [2] Buntoro, Asrofi Ghulam. 2017. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter. Jakarta : Integer Journal Vol 1 No 1 Maret 2016:32-41.
- [3] Rofiqoh, U., Perdana, S.P., Fauzi, M.A., 2017. Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. Malang.
- [4] Indrayuni, Elly. 2016. Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization, ISSN: 2338 – 8161. Pontianak : Jurnal Evolusi Volume 4 Nomor 2 – 2016.
- [5] Suyanto. 2017. *Data mining* Untuk Klasifikasi Dan Klusterisasi Data. Bandung: Informatika Bandung.
- [6] Mukminin, Amirul, Dwiza Riana. 2017. Komparasi Algoritma C4.5, *Naive Bayes* Dan *Neural Network* Untuk Klasifikasi Tanah, ISSN: 2355-6579 E-ISSN: 2528-2247. Bandung : JURNAL INFORMATIKA, Vol.4 No.1 April 2017, pp. 21-31.
- [7] Basari, A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J. (2013). Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Procedia Engineering*, 53, 453–462. doi:10.1016/j.proeng.2013.02.059.