

## **KLASIFIKASI METODE NAIVE BAYES UNTUK KELANCARAN PEMBAYARAN KREDIT LEASING SEPEDA MOTOR**

Fakultas Teknologi Informasi  
Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin

**Budi Ramadhani**

Budiramadhani99@gmail.com

### **ABSTRACT**

Permasalahan yang sering timbul pada perusahaan leasing adalah banyaknya pelanggan yang mengalami kesulitan dalam membayar cicilannya, maka diperlukan suatu sistem yang dapat mengklasifikasikan konsumen yang masuk ke grup saat ini, kelompok kurang lancar dan konsumen yang masuk ke dalam kelompok tidak lancar dalam membayar cicilan sepeda motor. Sehingga sewa bisa mengatasi masalah awal. Sebuah perusahaan leasing harus memiliki data yang sangat besar. Banyak yang tidak menyadari bahwa pengolahan data tersebut bisa memberikan informasi seperti klasifikasi data konsumen yang akan bergabung dengan perusahaan itu sendiri. Penerapan teknik data mining diharapkan dapat memberikan informasi yang berguna mengenai teknik klasifikasi data konsumen yang akan bergabung dengan grup saat ini, kelompok kurang lancar atau tidak lancar dalam membayar premi.

Langkah penelitian meliputi pengumpulan dan pengujian data algoritma Naive Bayes. Dalam penelitian ini, kumpulan data yang digunakan adalah Customer, Employment, Number of Children, Status Houses, region, angsuran.

Penelitian ini bertujuan untuk mengetahui Klasifikasi Metode Naive Bayes Berbasis Metode PSO Untuk Smooth Credit Leasing Motorcycles

Hasil percobaan menggunakan metode Naive Bayes untuk mengukur pengukuran lancar dan tidak lancar yang diperoleh pengukuran memiliki Naive Bayes tertinggi adalah 96,43% namun sekarang metode algoritma Naive Bayes Particle Swarm Optimization sebesar 96,88%, adalah akurasi namun baik

**Keywords:** *Current and Non Current, Naive Bayes Method Based PSO*

### **PENDAHULUAN**

Data yang ada dianalisis menggunakan algoritma *Naive Bayes*. *Naive Bayes* merupakan salah satu metode pada probabilistic reasoning. Algoritma *Naive Bayes* bertujuan untuk melakukan klasifikasi data pada kelas

tertentu, kemudian pola tersebut dapat digunakan untuk memperkirakan nasabah yang bergabung, sehingga perusahaan bisa mengambil keputusan menerima atau menolak calon nasabah tersebut. Guna memperoleh nilai prediksi yang akurat diperlukan

metode yang tepat agar dapat menghasilkan nilai prediksi dengan tingkat akurasi tinggi. Saat ini banyak permasalahan nyata yang diselesaikan dengan metode *softcomputing* dari pada *hardcomputing*.

### **TUJUAN PENELITIAN**

Berdasarkan latar belakang maka penelitian ini bertujuan “Untuk mengetahui Klasifikasi Metode Naive Bayes Untuk Kelancaran Pembayaran Kredit Leasing Sepeda Motor.

### **MANFAAT PENELITIAN**

Hasil penelitian ini diharapkan dapat membantu pihak terkait, yang dalam hal ini adalah pihak perusahaan dalam membuat perencanaan Metode Naive Bayes Untuk Kelancaran Pembayaran Kredit Leasing Sepeda Motor dengan hasil yang lebih baik.

### **KLASIFIKASI NAIVE BAYES**

Model dalam klasifikasi mempunyai arti yang sama dengan kotak hitam, dimana ada suatu model yang menerima masukan, kemudian mampu melakukan pemikiran terhadap masukan tersebut dan memberikan jawaban sebagai keluaran dari hasil pemikirannya.

1. Model yang sudah dibangun pada saat pelatihan kemudian dapat digunakan untuk memprediksi label kelas baru yang belum diketahui. Dalam pembangunan model selama proses pelatihan tersebut diperlukan suatu algoritma

untuk membangunnya, yang disebut algoritma pelatihan (*learning algorithm*).

2. Prediksi Klasifikasi adalah pemrosesan untuk menemukan sebuah model (atau fungsi) yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu, yang bisa menggunakan pemodelan untuk memprediksi kelas objek yang label nya tidak diketahui.

Tidak seperti klasifikasi dan prediksi, yang menganalisa pelabelan objek data, clustering menganalisis objek data tanpa mengkonsultasikan label kelas yang dikenal. Secara umum label kelas bukan didapat dalam pengolahan data sederhana karena mereka tidak tahu bagaimana memulainya. Clustering bisa digunakan untuk megenerate label.

### **PENELITIAN**

#### **Metode Pengolahan Data Awal**

Data yang didapatkan dari Perusahaan leasing di Wilayah Banjarmasin, Banjarbaru dan Martapura mengambil data pembayaran kredit sepeda motor dengan atribut No LOAN Nasabah, nama nasabah, merk kendaraan.

#### **Metode Yang Diusulkan**

Metode yang diusulkan untuk menentukan kelancaran pembeyaran kredit sepeda motor Perusahaan leasing di Wilayah Banjarmasin, Banjarbaru dan Martapura adalah algoritma algoritma naive bayes.

**Eksperimen dan Pengujian Model/Metode**

Metode yang telah dikembangkan dalam penelitian ini akan diterapkan pada data nasabah Perusahaan leasing di Wilayah Banjarmasin, Banjarbaru dan Martapura tahun 2014 melalui suatu model simulasi menggunakan R Software atau RapidMiner. Sebanyak 90 % dengan metode algoritma naive bayes.

**Tabel. 3.1 Data Nasabah**

No	Nama	Pekerjaan	Jumlah Anak	Status Rumah	Wilayah	Angsuran		Jumlah
						Tipe	Tingkat	
1	Bunai	1	4	4	1	1	1	Lancar
2	Widada	2	1	3	4	1	1	Tidak Lancar
3	Yuda	1	1	4	1	1	1	Lancar
4	Eda	1	1	4	1	1	1	Lancar
5	Eda	1	1	4	1	1	1	Lancar
6	Eda	1	1	4	1	1	1	Lancar
7	Eda	1	1	4	1	1	1	Lancar
8	Eda	1	1	4	1	1	1	Lancar
9	Eda	1	1	4	1	1	1	Lancar
10	Eda	1	1	4	1	1	1	Lancar
11	Eda	1	1	4	1	1	1	Lancar
12	Eda	1	1	4	1	1	1	Lancar
13	Eda	1	1	4	1	1	1	Lancar
14	Eda	1	1	4	1	1	1	Lancar
15	Eda	1	1	4	1	1	1	Lancar
16	Eda	1	1	4	1	1	1	Lancar
17	Eda	1	1	4	1	1	1	Lancar
18	Eda	1	1	4	1	1	1	Lancar
19	Eda	1	1	4	1	1	1	Lancar
20	Eda	1	1	4	1	1	1	Lancar
21	Eda	1	1	4	1	1	1	Lancar
22	Eda	1	1	4	1	1	1	Lancar
23	Eda	1	1	4	1	1	1	Lancar
24	Eda	1	1	4	1	1	1	Lancar
25	Eda	1	1	4	1	1	1	Lancar
26	Eda	1	1	4	1	1	1	Lancar
27	Eda	1	1	4	1	1	1	Lancar
28	Eda	1	1	4	1	1	1	Lancar
29	Eda	1	1	4	1	1	1	Lancar
30	Eda	1	1	4	1	1	1	Lancar
31	Eda	1	1	4	1	1	1	Lancar
32	Eda	1	1	4	1	1	1	Lancar
33	Eda	1	1	4	1	1	1	Lancar
34	Eda	1	1	4	1	1	1	Lancar
35	Eda	1	1	4	1	1	1	Lancar
36	Eda	1	1	4	1	1	1	Lancar
37	Eda	1	1	4	1	1	1	Lancar
38	Eda	1	1	4	1	1	1	Lancar
39	Eda	1	1	4	1	1	1	Lancar
40	Eda	1	1	4	1	1	1	Lancar
41	Eda	1	1	4	1	1	1	Lancar
42	Eda	1	1	4	1	1	1	Lancar
43	Eda	1	1	4	1	1	1	Lancar
44	Eda	1	1	4	1	1	1	Lancar
45	Eda	1	1	4	1	1	1	Lancar
46	Eda	1	1	4	1	1	1	Lancar
47	Eda	1	1	4	1	1	1	Lancar
48	Eda	1	1	4	1	1	1	Lancar
49	Eda	1	1	4	1	1	1	Lancar
50	Eda	1	1	4	1	1	1	Lancar

lancar. Jumlah sampel data keseluruhan ada 221 Nasabah. Dari semua data ditentukan atribut proses pengolahan data awal dimulai dari kolom Nasabah, Pekerjaan, Jumlah Anak, Status Rumah, Wilayah, Angsuran, untuk lebih jelasnya dapat dilihat tabel atribut dibawah ini :

**Tabel 3.2. Atribut**

No	Nama Atribut	Tipe	Keterangan
1	Pekerjaan	Nominal	PNS, TNI/POLRI, Swasta, Wiraswasta
2	Jumlah Anak	Numerik	1, 2, 3, 4, 5

3	Status Rumah	Nominal	Milik Sendiri, Sewa, Pinjam, Ikut Orang tua
4	Angsuran	Numerik	Tepat, Tidak tepat

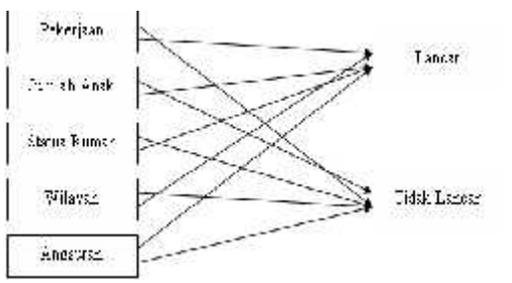
**Metode yang Diusulkan (Proposed Method)**

Pada penelitian ini penulis mengusulkan metode *Naive Bayes* untuk membantu menentukan lancar dan tidak lancar pembayaran angsuran kendaraan bermotor. Metode ini dipilih karena kemampuannya untuk melakukan pengelompokkan suatu objek atau data yang belum memiliki klasifikasi, ke dalam kelas tertentu menurut kesamaan yang dimilikinya berdasarkan derajat keanggotaan dengan cara minimalisasi nilai fungsi objektifnya. Metode naive bayes akan diimplementasikan pada penelitian ini.

**Eksperimen dan Pengujian Model**

Algoritma yang telah dikembangkan dalam penelitian ini akan diterapkan pada Nasabah, Pekerjaan, Jumlah Anak, Status Rumah, Wilayah, Angsuran pada saat pembayaran angsuran oleh nasabah.

Pada tahap awal dilakukan pemetaan korelasi antara nilai, data lancar dengan hasil akhir penempatan peminatan, hasilnya ditunjukkan pada gambar berikut:



Gambar 3.1 Pemetaan antara lancar dan tidak lancar

Menentukan hasil akhir merupakan hasil yang diperoleh dari nasabah dalam membayarn angsuran yaitu lancar dan tidak lancar, ditambahkan Pekerjaan, Jumlah Anak, Status Rumah, Wilayah, Angsuran. Kemudian Data ini akan digunakan sebagai data variabel ujicoba peminatan menggunakan *Metode Naïve Bayes* (gambar 3.1).

**Evaluasi**

Evaluasi dilakukan dengan mengambil hasil prediksi dibandingkan dengan data awal yang diolah.

**HASIL EKSPERIMEN**

**Hasil Eksperimen dan Pengujian Model**

Data yang digunakan merupakan data sekunder hasil penerimaan angsuran dari nasabah tahun 2016. Data yang terkumpul kemudian dilakukan *pre-processing* dengan cara melakukan identifikasi dan inspeksi sejumlah atribut-atribut yang tidak relevan. Kemudian setelah diketahui sejumlah atribut yang tidak relevan maka tahap selanjutnya adalah mengeliminasi sejumlah atribut tersebut dan melakukan substitusi data sehingga data dapat diproses dalam *data mining*.



**Gambar 4. 1** Proses *Pre-processing*

*Dataset* tersebut memiliki label Lancar dan Tidak Lancar 5 variabel sebagai atribut. Ada dua tipe data yaitu data nominal dan data numeric. Data nominal tersebut yang diinginkan adalah pekerjaan, status rumah dan wilayah sedangkan data numeriknya yaitu jumlah anak dan angsuran.

Hasil perhitungan data angsuran nasabah sebagai berikut :

1. Perhitungan Data Nominal

**Diketahui :**

- Tipe Data = Nominal
- Pekerjaan
- Status Rumah
- Wilayah

**Misal :**

- $E_1$  = Pekerjaan
- $E_2$  = Status Rumah
- $E_3$  = Wilayah
- Yes = Lancar
- No = Tidak Lancar

**Tabel 4. 1** Perhitungan Data Nominal

Tipe Data	Perhitungan	Hasil	Label
Tipe Data	17%	24%	1.00%
Pekerjaan	4%	22%	0.80%
Status Rumah	13%	17%	0.57%
Wilayah	10%	10%	0.19%
Angsuran	10%	10%	0.22%
Angsuran	0.22%	0.22%	0.19%
Angsuran	0.22%	0.22%	0.19%

perhitungan fakta umum yaitu PB (Yes = Lancar) adalah 0,809 sedangkan PB (No = Tidak lancar) adalah 0,190. Selanjutnya

menghitung berdasarkan rumus *Naive Bayes* atau HMAP data *training* dengan hipotesa

**Tabel 4. 2 Nilai rata-ratadan standar deviasi Perhitungan Data Numerik**

Nilai perhitungannya menggunakan metode	
1. Rata-rata	0,75
2. Standar deviasi	0,25
3. Jumlah data	100
4. Jumlah data yang akurat	84,96%

Oleh karena itu, the naïve bayes classifier predicts lancer dan tidak lancer = yes for the new example.

Berdasarkan sampel data pada tabel 4.3 diatas dapat diketahui perbandingan hasil keputusan *Naive Bayes* dengan data original yaitu 221 data teridentifikasi akurat dan 42 yang tidak akurat.

Pada tahap pertama eksperimen dan pengujian model dilakukan dengan data original yang sebagian masih memiliki data kosong. Kemudian pengujian dilakukan menggunakan algoritma *Naive Bayes* dengan validasi model klasifikasi dilakukan terhadap data *testing* dengan teknik *10-folds cross validation*.

Ketika diimplementasi menghasilkan data sebagai berikut :

**Tabel 4.4 Akurasi Naive Bayes Menggunakan Data Original**

Kelas	Data Asli		Data Hasil	
	Jumlah	Persentase	Jumlah	Persentase
Lancer	100	100%	85	85%
Tidak Lancer	100	100%	15	15%

menggunakan metode *Naive Bayes* dengan data original tersebut

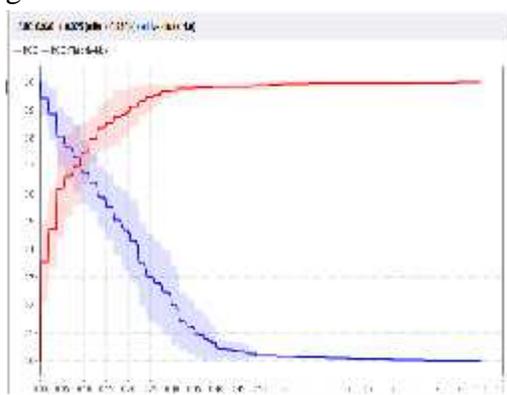
didapatkan akurasi yang sudah cukup baik sekitar 84,96%. Oleh sebab itu maka untuk membuat model keputusan yang baik, harus menggunakan data yang baik pula (lengkap, benar, konsisten, terintegrasi).

Dari beberapa percobaan menggunakan teknik pengisian yang berbeda, didapatkan hasil bahwa tidak semua teknik tersebut sesuai atau cocok digunakan pada data yang diuji. Hal tersebut terjadi pada percobaan menggunakan teknik *minimum*, *maximum* dan *zero* karena ketiga teknik tersebut tidak cocok dengan karakter nilai yang terdapat pada data yang diuji yang sebagian besar bersifat nominal sehingga ketika diimplementasikan meskipun telah menggunakan metode *replace missing values* terutama yang bertipe *minimum*, *maximum* dan *zero*, hasil akurasi yang didapatkan tidak terjadi perubahan yang berarti hasil akurasi sama dengan hasil akurasi tanpa dilakukan pengisian data kosong pada percobaan sebelumnya atau penggunaan ketiga teknik tersebut tidak berpengaruh terhadap perubahan akurasi.

Perbedaan terjadi ketika dilakukan percobaan menggunakan metode *replace missing values* bertipe *average*, berhasil didapatkan hasil akurasi yang berbeda sehingga teknik tersebut cocok dan sesuai untuk digunakan pada data yang akan diuji yang sebagian besar berkarakter nominal, oleh sebab itu selanjutnya metode *replace missing values* bertipe

average akan digunakan pada percobaan selanjutnya.

Kemudian dari hasil pengujian juga dianalisa berdasarkan kurva ROC yang menunjukkan *trade-off* antara *true positive rate* (proporsi *tuple* positif yang teridentifikasi dengan benar) dan *false positive rate* (proporsi *tuple* negatif yang teridentifikasi salah sebagai positif) dalam suatu model. Garis merah untuk *true* positif dan biru untuk *threshold* seperti pada gambar 4.3 dibawah ini :



**Gambar 4. 2 Kurva ROC**

Menurut Gorunescu (Gorunescu, 2010) tingkat keakurasian AUC dapat diklasifikasikan menjadi lima kelompok yaitu :

1. 0.90 - 1.00 = *excellent classification*
2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*

Nilai AUC yang didapatkan dari pengujian berdasarkan kurva ROC menggunakan metode *Naive Bayes* sebesar 0.930 sehingga dari hasil tersebut berdasarkan kriteria diatas menunjukkan klasifikasi yang dihasilkan termasuk kedalam kelompok *excellent classification* atau klasifikasi yang sangat baik.

### Evaluasi Dan Validasi Hasil

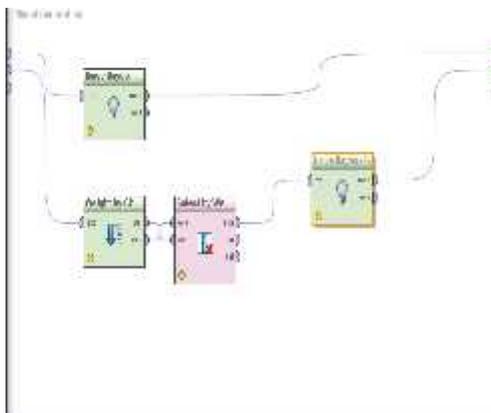
Dari tabel terlihat penggunaan *X-Validation* dengan *sampling type stratified* memiliki tingkat akurasi yang sedikit lebih baik daripada *sampling type* yang lain meskipun dalam hal kehandalan klasifikasi masih sedikit lebih unggul penggunaan *sampling type stratified*.

Kemudian percobaan selanjutnya akan dilakukan pengujian algoritma *Naive Bayes* dengan teknik *folds cross validation* dengan pengujian data mulai 2,3,4,5,6,7,8,9 dan 10. Hasil validasi percobaan algoritma *Naive Bayes* tersebut

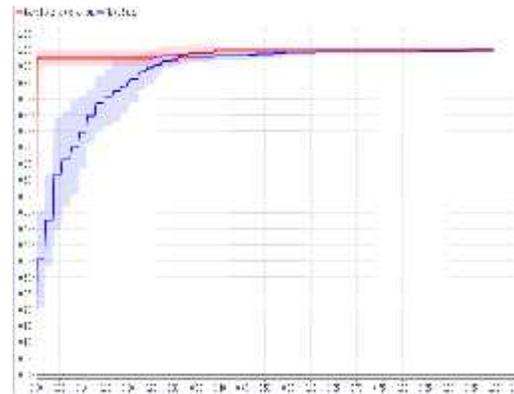
Dari hasil percobaan didapatkan pengulangan pengujian sebanyak 4 kali dengan hasil pengukuran berupa nilai rata-rata 4 kali pengujian memiliki tingkat akurasi tertinggi sebesar 86.28% dengan nilai AUC sebesar 0.940. Namun hasil dari berbagai percobaan yang ekstensif dan pembuktian teoritis, menunjukkan bahwa penggunaan *10-fold cross-validation* adalah pilihan terbaik untuk mendapatkan hasil validasi yang akurat. Hal tersebut karena *10-fold cross-validation* akan mengulang pengujian sebanyak 10 kali dan hasil

pengukuran adalah nilai rata-rata dari 10 kali pengujian sehingga dalam evaluasi hasil perbandingan *performance* dengan mengukur seberapa keakuratan akurasi yang dihasilkan menggunakan pengukuran *10-fold cross-validation* dan pengujian *performance* dengan menggunakan *Confusion Matrix* sehingga dapat diketahui hasil akurasi dan nilai AUC untuk menentukan hasil klasifikasi kedalam klasifikasi sangat baik, klasifikasi baik, klasifikasi cukup, klasifikasi buruk dan klasifikasi salah. Hasil akurasi yang dihasilkan sebesar 85.80 % sudah cukup baik namun untuk meningkatkan hasil akurasi yang telah dihasilkan dan untuk mendapatkan penggunaan atribut terbaik maka pengujian selanjutnya dengan melakukan seleksi atribut menggunakan *Naive Bayes* berbasis *Chi Squared*.

Ketika diimplementasi menghasilkan komparasi kurva ROC seperti gambar 4.9 dibawah ini :



**Gambar 4. 2 Model Generation Menggunakan Compare ROC**



**Gambar 4. 3 Perbandingan Kurva ROC Menggunakan Compare ROC**

### Implementasi Penelitian

Dari beberapa hasil pengujian tersebut disimpulkan bahwa algoritma *Naive Bayes* dalam mengklasifikasi lancar dan tidak lancar. Dengan demikian, adanya penerapan *Naive Bayes* mampu memberikan solusi untuk mengidentifikasi pembayar lancar dan tidak lancar.

### PENUTUP DAN KESIMPULAN

Pada penelitian ini ternyata dapat mengoptimalkan algoritma *Naive Bayes* dengan pembobotan atribut terlebih dahulu sehingga akurasi dapat meningkat. Dapat dibuktikan dengan akurasi algoritma *Naive Bayes* adalah 82,19% dan untuk algoritma *Naive Bayes* akurasi meningkat menjadi 89,70%, yang berarti akurasinya meningkat 7,51%.

Sedangkan perbandingan dengan penelitian sebelumnya yaitu penelitian [10] yang menggunakan datase yang sama namun menggunakan algoritma yang berbeda yaitu algoritma *C4.5* mempunyai akurasi sebesar 88,83%

dan C4.5 memperoleh akurasi sebesar 89,26% dan untuk Naive baiyes memperoleh akurasi sebesar 82,19% sedangkan algoritma Naive Baiyes memperoleh akurasi sebesar 89,70%, sehingga dapat disimpulkan pada penelitian ini bahwa algoritma Naive Baiyes merupakan algoritma terbaik yang digunakan untuk memprediksi dataset pembayaran angsuran lancar dan tidak lancar yang digunakan untuk memprediksi nasabah yang berpotensi membayar angsuran Perusahaan leasing.

Penelitian ini membandingkan antara algoritma Naive Baiyes dengan Naive Baiyes, untuk penelitian selanjutnya dapat dikembangkan dengan membandingkan algoritma klasifikasi lainnya seperti algoritma *Support Vector Machine*, *Decision Tree*, *Neural Network*, yang dapat dikombinasikan dengan algoritma optimasi lain seperti Genetic Algorithm (GA), Ant Colony Optimization (ACO), Adaboost dan algoritma optimasi yang lainnya.

#### DAFTAR PUSTAKA

- [1] Lembaga Penjamin Simpanan, "Distribusi Simpanan Bank Umum," Lembaga Penjamin Simpanan, 2013.
- [2] Dewan Perwakilan Rakyat Republik Indonesia, "Undang-Undang Republik Indonesia Nomor 10 Tahun 1998," Dewan Perwakilan Rakyat Republik Indonesia, 1998.
- [3] S. Y. Baswara, "Analisa Nilai Hasil Investasi Deposito Rupiah, Deposito Dolar Amerika, dan Dinar Emas Dengan Emas Sebagai Alat Ukur," Skripsi Akuntansi Universitas Diponegoro, 2012.
- [4] D. T. Larose, *Discovering Knowledge In Data*. United States of America: John Wiley & Sons, Inc., 2005.
- [5] F. Gorunescu, *Data Mining Concepts, Models and Techniques*. Chennai, India: Springer, 2011.
- [6] E. Prasetyo, *Data Mining : Konsep dan Aplikasi menggunakan MATLAB*, 1st ed. Yogyakarta, Indonesia: Andi, 2012.
- [7] Jie Lin dan Jiankun Yu dari Yunnan University of Finance and Economics telah melakukan penelitian mengenai penggunaan Weighted Naive Bayes berbasis algoritma Particle Swarm Optimization (PSO) yang digunakan untuk pembobotan atribut
- [8] Nur Aeni Widiastuti., Stefanus Santosa, Catur Supriyanto dengan judul penelitian Algoritma Klasifikasi Data

Mining *Naïve Bayes* Berbasis  
*Particle Swarm Optimization*.

- [9] Mutia Hafilizara dan Julio Adisantoso. dengan penelitian Metode Smoothing dalam *Naïve Bayes* untuk Klasifikasi Email Spam *Naïve Bayes Smoothing Method for Spam Email Classification*