

Analisa Klasifikasi Loyalitas Siswa Lembaga Pendidikan Tari dengan Metode Naïve Bayes

Indah Purnamasari

Program Studi Sistem Informasi
STMIK Nusa Mandiri Jakarta
E-mail: indah.ih@nusamandiri.ac.id

Abstrak

Tantangan kehidupan di masa mendatang semakin kompetitif, Pendidikan formal sudah menjadi hal umum yang dimiliki sebagian besar orang. Dibutuhkan suatu sarana untuk meningkatkan sumber daya manusia agar dapat berkompetisi maka para orang tua tidak hanya membekali putra putrinya dengan pendidikan formal akan tetapi juga melengkapi dengan pendidikan non formal. Tak ayal usaha pendidikan non formal pun semakin diminati. Berdasarkan data Sensus Ekonomi 2016 oleh Badan Pusat Statistik (BPS), jumlah usaha/perusahaan pendidikan di Indonesia mencapai 619.947 usaha. Jumlah ini setara dengan 2,32 persen dari jumlah usaha/perusahaan di Indonesia. Seiring dengan meningkatnya usaha pendidikan maka persaingan usaha akan semakin kompetitif. Pelanggan dapat memilih di antara beberapa usaha pendidikan yang ada dan dapat berpindah dari satu usaha pendidikan ke yang lainnya. Ketika seorang siswa suatu lembaga pendidikan keluar, dampak yang ditimbulkan tidak hanya mengurangi pendapatan, tetapi juga mempengaruhi sumber daya awal yang telah dikeluarkan untuk rekrutmen tenaga kerja baru, biaya publikasi dan diskon. Perusahaan perlu mengklasifikasikan siswa untuk mengetahui tingkat loyalitas siswa dan potensi siswa untuk *churn* sehingga perusahaan dapat menentukan langkah-langkah dalam meningkatkan pemeliharaan siswa. Dari data – data siswa dapat diperoleh informasi dan pengetahuan untuk dilakukan data mining klasifikasi. Algoritma Naïve Bayes (NB) merupakan salah satu metode/ algoritma data mining klasifikasi yang cepat, mudah diimplementasikan dengan struktur yang sederhana dan efektif dalam pengklasifikasian. Penelitian ini akan menerapkan data mining klasifikasi metode Naïve Bayes dari data- data siswa Lembaga Pendidikan Kursus Tari cabang Citayam untuk mengklasifikasikan loyalitas siswa dengan mengukur kinerja model menggunakan software RapidMiner. Hasil penelitian menunjukkan bahwa model dengan metode Naïve Bayes untuk data mining klasifikasi loyalitas siswa Lembaga Pendidikan Kursus Tari cabang Citayam memiliki nilai *accuracy* yaitu 80,42%, *classification error* 19,58% dan *AUC* yaitu 0.860 yang memiliki arti bahwa model ini dapat digunakan untuk mengklasifikasikan siswa untuk mengetahui tingkat loyalitas siswa dan potensi siswa untuk *churn*.

Kata Kunci: data mining, klasifikasi, Naïve Bayes

Pendahuluan

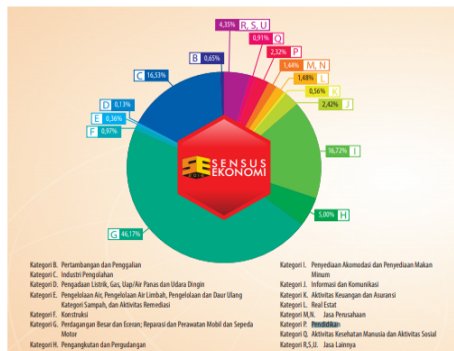
Saat ini kesadaran orang tua akan pentingnya pendidikan semakin tinggi. Tantangan kehidupan di masa mendatang semakin kompetitif, pendidikan formal sudah menjadi hal umum yang dimiliki sebagian besar orang. Dibutuhkan suatu sarana untuk meningkatkan sumber daya manusia agar dapat berkompetisi dengan baik kedepannya maka para orang tua tidak hanya membekali putra putrinya dengan pendidikan formal akan tetapi juga melengkapi dengan pendidikan non formal. Untuk memenuhi kebutuhan belajar masyarakat yang heterogen tersebut, tidak cukup hanya dilayani melalui satuan pendidikan formal (sekolah), akan tetapi membutuhkan peran satuan pendidikan non formal untuk dapat menjawab kebutuhan belajar

masyarakat yang dinamis tersebut [1].

Menurut UU No. 20 tahun 2003 tentang Sistem Pendidikan Nasional pasal 26 yaitu Pendidikan non formal terdiri atas lembaga kursus, lembaga pelatihan, kelompok belajar, pusat kegiatan belajar masyarakat, dan majelis taklim serta satuan pendidikan yang sejenis. Kursus dan pelatihan diselenggarakan bagi masyarakat yang memerlukan bekal pengetahuan, ketrampilan, kecakapan hidup dan sikap untuk mengembangkan diri, mengembangkan profesi, bekerja, usaha mandiri dan/atau melanjutkan ke jenjang yang lebih tinggi. Tak ayal usaha-usaha pendidikan baik formal maupun non formal pun semakin diminati. Berdasarkan data Sensus Ekonomi 2016 oleh Badan Pusat Statistik (BPS), jumlah usaha/perusahaan pendidikan di Indonesia mencapai 619.947 usaha. Jumlah ini setara

dengan 2,32 persen dari jumlah usaha/perusahaan di Indonesia menurut Badan Pusat Statistik pada tahun 2017.

Seiring dengan meningkatnya usaha pendidikan maka persaingan usaha akan semakin kompetitif. Banyak perusahaan industri memberi perhatian pada subjek retensi pelanggan, loyalitas, dan *churn*. *Churn* mengacu pada hilangnya pelanggan dari suatu perusahaan karena berpindah ke perusahaan lain. Pelanggan dapat memilih di antara beberapa penyedia layanan dan dapat berpindah dari satu penyedia layanan ke yang lainnya [2]. Ketika seorang pelanggan pergi, dampak yang ditimbulkan tidak hanya mengurangi pendapatan, tetapi juga sumber daya awal yang telah dikeluarkan untuk rekrutmen tenaga kerja baru, biaya publikasi dan diskon. Menariknya ribuan pelanggan baru tidak ada gunanya jika jumlah pelanggan yang sama pergi [3].



Sumber : (Badan Pusat Statistik, 2017)

Gambar 1: Distribusi Usaha menurut Lapangan Usaha

Perusahaan perlu mengklasifikasikan pelanggan untuk mengetahui tingkat loyalitas pelanggan dan potensi pelanggan untuk *churn*. Proses klasifikasi data dalam menemukan informasi dan pengetahuan dari data-data pelanggan yang tersedia dapat dilakukan dengan data mining. Data-data pelanggan tersebut dapat dimanfaatkan untuk mengklasifikasikan loyalitas pelanggan. Klasifikasi pelanggan dapat dilakukan dengan subjek *Customer Churn Prediction*. [4].

Prediksi *Churn* adalah aplikasi khas data mining klasifikasi dari data perilaku konsumen untuk memprediksi konsumen yang berpotensi untuk *churn*. *Churn* adalah masalah umum yang dihadapi oleh perusahaan dan penelitian menunjukkan bahwa biaya pengembangan pelanggan baru adalah sekitar 5 kali lebih tinggi dari biaya mempertahankan pelanggan lama [5]. Sebagian besar perusahaan akan menerapkan prediksi *churn* melalui data mining untuk meningkatkan pemeliharaan pelanggan. Perusahaan sebelumnya dapat mengenali konsumen yang tidak enggan berpindah ke perusahaan lain dengan prediksi *churn*, lalu melakukan pemasaran yang tepat untuk mempertahankan (retensi) pelanggan ini. Selain itu,

prediksi *churn* memungkinkan perusahaan untuk meningkatkan efisiensi retensi pelanggan dan untuk mengurangi dampak dari biaya yang terkait dengan *churn*. Jadi, bagaimana cara menggunakan data pelanggan dalam jumlah besar efektif dan meningkatkan kinerja model prediksi *churn* yang bersangkutan oleh perusahaan [6].

Penelitian data mining klasifikasi pernah dilakukan sebelumnya dengan menggunakan teknik resampling dan teknik ensemble AdaBoost untuk memperbaiki kinerja pengklasifikasi. Hasil penelitian menunjukkan bahwa model integrasi random oversampling, AdaBoost, dan Naïve Bayes memiliki kinerja yang lebih baik karena memiliki nilai AUC (Area Under the ROC (Receiver Operating Characteristic) Curve) yang lebih baik [7].

Penelitian lain menggunakan metode data mining untuk memprediksi potensi *churn* dan segmentasinya pada pelanggan dan hal ini dapat mendukung proses monitoring, pengendalian, serta penyusunan strategi pada management dengan Penelitian tersebut menggunakan metode Decision Tree dan Clustering yang merupakan metode data mining yang populer karena sangat mudah dipahami dan diinterpretasikan, sehingga dapat dengan mudah untuk digunakan sebagai teknik untuk melakukan prediksi *churn* dan segmentasinya [8].

Penelitian prediksi *churn* juga pernah dilakukan dengan menggunakan metode IG-KNN yang merupakan kombinasi dari algoritma pemilihan fitur information gain dengan algoritma klasifikasi KNN, kombinasi dari kedua algoritma ini dimaksudkan untuk meningkatkan akurasi dalam memprediksi customer *churn* telekomunikasi [9].

Pada penelitian ini akan menerapkan metode Naïve Bayes untuk klasifikasi loyalitas siswa suatu Lembaga Pendidikan Kursus Tari di Depok dengan mengukur kinerja model menggunakan software RapidMiner 5. Hal ini dilakukan untuk mengetahui kinerja model data mining dengan metode Naïve Bayes untuk klasifikasi loyalitas siswa terhadap lembaga pendidikannya sehingga selanjutnya lembaga pendidikan tersebut dapat mempersiapkan langkah – langkah pemeliharaan siswa-siswanya.

Data Mining Klasifikasi

Data Mining

Data mining merupakan proses menggali korelasi bermakna baru, pola dan trend oleh sejumlah data besar yang disimpan dalam repositori menggunakan teknologi pola serta teknik statistik dan matematika [10]. Data mining adalah proses analitik dirancang untuk mengeksplorasi sejumlah besar data mencari ilmu tersembunyi yang konsisten dan berharga, langkah pertama terdiri dalam eksplorasi dan data persiapan awal [11]. Data mining didefinisikan sebagai proses menemukan pola dalam data [12].

Proses data mining menggunakan teknik-teknik pembelajaran mesin (machine learning) untuk memeriksa data berukuran besar sebagai cara untuk menemukan pola baru yang berguna. Proses pencarian untuk menemukan pola atau model yang sah, baru, bermanfaat, dan dimengerti dilakukan secara iteratif dengan menggunakan perangkat lunak dalam menganalisis data untuk menemukan pola dan relasi data agar dapat menghasilkan prediksi yang lebih akurat. Meskipun tidak semua pekerjaan pencarian informasi dinyatakan sebagai data mining, seperti halnya pencarian situs tertentu menggunakan query search engine merupakan pekerjaan pencarian informasi yang berhubungan dengan information retrieval. Akan tetapi kemampuan sistem-sistem information retrieval dapat ditingkatkan dengan menggunakan teknik-teknik data mining.

Naïve Bayes

Naive Bayes Classifier merupakan satu dari beberapa algoritma klasifikasi data mining untuk menemukan model atau fungsi yang membedakan beberapa kelas data yang ada. Teorema Bayes dikemukakan oleh Reverend Thomas Bayes pada abad ke 18 [13]. Berdasarkan teorema Bayes dikembangkan Naive Bayes yaitu pengklasifikasi dengan metode probabilistik dan statistik istilah "naif" menunjukkan keindependenan antara atribut kelas [14]. Klasifikasi Naive Bayes diasumsikan bahwa ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Asumsi keindependenan atribut akan menghilangkan kebutuhan banyaknya jumlah data latih dari perkalian kartesius seluruh atribut yang dibutuhkan untuk mengklasifikasikan suatu data.

Algoritma Naïve Bayes (NB) merupakan algoritma yang cepat, mudah diimplementasikan dengan struktur yang sederhana dan efektif dalam pengklasifikasian. Selain itu NB terbukti memiliki akurasi yang tinggi saat diaplikasikan ke dalam database yang besar. Pada dasarnya metode NBC menempuh dua tahap klasifikasi, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses analisis terhadap sampel. Selanjutnya adalah penentuan probabilitas prior bagi tiap kategori berdasarkan sampel yang ada. Pada algoritma Naïve Bayes (NB) didasari oleh Persamaan Teorema Bayes, yang membedakan adalah asumsi keindependenan antara atribut. Persamaan Bayes adalah [11]:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Keterangan:

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

(H|X): Probabilitas hipotesis H berdasar kondisi X (posterior probability)

(H): Probabilitas hipotesis H (prior probability)

(X|H): Probabilitas X berdasarkan kondisi pada hipotesis H

(X): Probabilitas X

Dalam mendeskripsikan algoritma Naive Bayes, perlu diketahui bahwa proses klasifikasi memerlukan petunjuk- petunjuk dalam menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Oleh sebab itu, algoritma Bayes di atas disesuaikan sebagai berikut [11]:

$$P(H|X) = P(X|H) \cdot P(H) \quad (2)$$

Persamaan tersebut merupakan probabilitas dari masing masing kelas yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk data kontinyu dapat menggunakan rumus Densitas Gauss[11]:

$$P(X, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

Keterangan:

P: Peluang

X: Atribut

μ : Mean, menyatakan rata – rata dari seluruh atribut

σ : Deviasi standar, menyatakan varian dari seluruh atribut

$\pi = 3,1416$

$e = 2,7183$

$$\sigma^2 = \frac{1}{N-1} \sum (x_i - \mu)^2 \quad (4)$$

Akan tetapi pada kasus khusus untuk menghitung probabilitas posterior dimana $P(X|C) = 0$ selama $|X| = 0$ akan menghilangkan efek semua posterior $P(C)$. Untuk mengatasi hal ini, kita asumsikan dengan menambahkan nilai satu pada setiap perhitungan data pelatihan. Dalam perbedaan probabilitas perkiraan ini, perlu dibuat kesepakatan. Teknik ini dikenal sebagai koreksi Laplacian :

$$P(X|C) = \frac{1}{|C| + \text{jumlah kelas}} \quad (5)$$

Lembaga Pendidikan Kursus Tari

Ayodya Pala didirikan pada tanggal 24 April 1980 adalah lembaga pendidikan kesenian berbentuk Yayasan yang bergerak dalam bidang pelatihan pendidikan dan pelestarian berbagai seni budaya tradisional Indonesia.

Ayodya pala menerapkan konsep pelatihan dan pendidikan seni dimana memiliki kurikulum pada tiap semester yang dapat diterima oleh masyarakat pada umumnya. Hal ini dapat dilihat dari pertumbuhan cabang-cabang Ayodya Pala mencapai 31 cabang aktif dengan jumlah siswa sampai dengan 2.500 siswa berkisar antara anak usia 3 tahun s/d dewasa dan beberapa sekolah yang telah menjadi mitra ayodya pala melalui kegiatan ekstrakurikuler di sekolahnya.



Gambar 2: Lembaga Pendidikan Kursus Tari

Metode Penelitian



Gambar 3: Tahapan Penelitian

Pengumpulan Data

Tahap ini diawali dengan pengumpulan data dari lembaga pendidikan tari dan peralatan untuk eksperimen. Dataset digunakan mengidentifikasi masalah data, menemukan wawasan awal tentang data dan mendapatkan informasi tersembunyi dari dataset serta mempersiapkan alat berupa seperangkat komputer yang akan digunakan untuk eksperimen.

Sumber dalam penelitian terdiri dari sumber primer. Sumber primer merupakan sumber obyek penelitian yang langsung memberikan data kepada

pengumpul data penelitian sehingga data tersebut merupakan data primer. Data primer adalah data yang dikumpulkan pertama kali dan untuk melihat apa yang sesungguhnya terjadi [15].

Sumber sekunder merupakan sumber obyek penelitian yang tidak langsung memberikan data kepada pengumpul data penelitian melainkan melalui orang lain atau dokumen umum sehingga data tersebut merupakan data sekunder. Data sekunder adalah data yang sebelumnya pernah dibuat oleh seseorang baik data tersebut telah diterbitkan atau tidak [15].

Studi Literatur

Studi literatur dilakukan dengan mengumpulkan, mempelajari dan memahami permasalahan, faktor-faktor yang mempengaruhi dan algoritma data mining untuk klasifikasi yaitu Naive Bayes Classifier (NBC) dari referensi-referensi berupa jurnal penelitian, buku dan secara online.

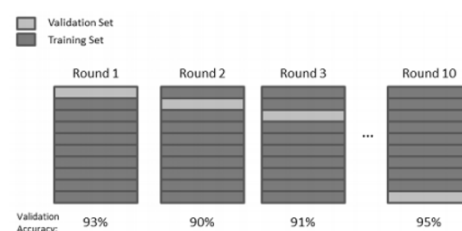
Algoritma Klasifikasi Naïve Bayes

Pada tahap ini pemilihan algoritma disesuaikan dengan permasalahan yang ada. Beberapa algoritma yang ada memiliki persyaratan yang spesifik pada bentuk data. Penelitian klasifikasi ini menggunakan algoritma data mining Naive Bayes Classifier (NBC) akan diterapkan pada atribut-atribut dataset dilakukan dengan cara :

1. Perhitungan probabilitas masing masing kelas
2. Perhitungan probabilitas prior
3. Perhitungan probabilitas posterior

Eksperimen

Dalam eksperimen yang dilakukan menggunakan spesifikasi hardware dan software sebagai alat bantu dalam pemrosesan data. Software yang digunakan adalah Rapid Miner 5 untuk mengolah proses data mining klasifikasi dengan metode Naive Bayes Classifier. Untuk pengukuran model menggunakan Cross Validation dan Performance sehingga diperoleh hasil akurasi, classification error dan AUC. Cross Validation merupakan proses evaluasi untuk mengukur keakuratan prediksi suatu model [13].



Gambar 4: Ten Cross Validation

Confusion Matrix merupakan alat untuk menganalisis seberapa baik classifier dapat mengenali tupel dari kelas yang berbeda. Confusion matrix berisi informasi aktual (actual) dan prediksi (predicted) kelas dari data pelanggan eksisting yang ada [11].

Tabel 1: Confusion Matrix

Classification	Predicted Class	
	Class=Yes	Class=No
Class=Yes	A (True Positive - tp)	B (False Negative - fn)
Class=No	C (False Positive - fp)	D (True Negative - tn)

Berikut adalah persamaan model confusion matrix untuk menghitung akurasi. Nilai akurasi adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan:

$$\text{Akurasi} = \frac{tp + tn}{tp + tn + fp + fn} \quad (6)$$

$$\text{Classification Error} = \frac{fp + fn}{tp + tn + fp + fn} \quad (7)$$

Curve ROC (Receiver Operating Characteristic) merupakan metode untuk mengevaluasi akurasi dari model klasifikasi secara visual dimana keakuratan tes menggunakan AUC disajikan di bawah ini [11] yaitu :

- Akurasi 0.90 – 1.00 = Excellent classification
- Akurasi 0.80 – 0.90 = Good classification
- Akurasi 0.70 – 0.80 = Fair classification
- Akurasi 0.60 – 0.70 = Poor classification
- Akurasi 0.50 – 0.60 = Failure

Hasil dan Kesimpulan

Setelah dilakukan pengukuran terhadap model datamining klasifikasi metode Naive Bayes Classifier (NBC) maka dari hasil pengujian dengan menggunakan akan diperoleh hasil akurasi, classification error dan AUC dari model tersebut sehingga dapat diambil kesimpulan tentang pemilihan metode Naive Bayes sebagai model untuk klasifikasi loyalitas siswa lembaga tari.

Dokumentasi

Dokumentasi dilakukan setelah tahapan demi tahapan diatas dilakukan.

Hasil dan Pembahasan

Data yang diperoleh dari lembaga pendidikan tari cabang Atsiri berupa data primer yang terdiri atas 85 data siswa dengan 7 atribut termasuk class label attribute (atribut output) yaitu atribut Aktif. Data-data ini merupakan data fakultatif yang telah

diketahui statusnya yaitu Ya (siswa yang masih aktif) atau Tidak (siswa yang churn). Dari 85 data siswa terdapat 37 siswa yang masih aktif (Ya) dan 48 siswa yang churn (Tidak). 6 atribut yang terdapat pada data siswa Lembaga Pendidikan Kursus Tari di Depok tersebut adalah :

1. Waktu
2. Pekerjaan ayah
3. Pekerjaan ibu
4. Anak ke
5. Minat
6. Ekskul sekolah
7. Aktif (class label atribut)

Tabel 2: Data Siswa Kursus Tari

No	Tgl Dftr	Kerja Ayah	Kerja Ibu	Anak ke	Minat	Ekskul	Aktif
1	24-Mar-13	Karyawan	IRT	1	Olahraga	Karate	Tidak
2	25-Mar-13	Karyawan	Karyawan	2	Olahraga	Silat	Tidak
3	13-Apr-13	Wira Usaha	Karyawan	1	Praktik	Pramuka	Tidak
4	20-Apr-13	PNS	PNS	1	Seni	Tari	Ya
5	24-Jun-13	PNS	Karyawan	3	Literasi	Menulis	Tidak
6	25-Aug-13	Wira Usaha	Karyawan	1	Praktik	Paskibra	Tidak
7	25-Aug-13	PNS	Karyawan	2	Praktik	Paskibra	Tidak
8	31-Aug-13	Karyawan	Karyawan	1	Olahraga	Karate	Tidak
9	29-Sep-13	Wira Usaha	Karyawan	1	Seni	Karate	Tidak
10	20-Oct-13	Karyawan	IRT	1	Praktik	Drumband	Tidak
11	27-Oct-13	PNS	IRT	2	Seni	Tari	Ya
12	15-Nov-13	PNS	IRT	1	Olahraga	Karate	Tidak
13	26-Nov-13	Karyawan	Wira Usaha	2	Olahraga	Karate	Tidak
14	5-Jan-14	PNS	PNS	2	Literasi	Menulis	Ya
15	22-Jan-14	Wira Usaha	Karyawan	2	Praktik	Pramuka	Tidak
16	9-Feb-14	PNS	Karyawan	3	Praktik	Pramuka	Tidak
17	16-Feb-14	PNS	PNS	2	Seni	Tari	Ya
18	16-Feb-14	PNS	PNS	2	Olahraga	Basket	Ya
19	2-Mar-14	PNS	IRT	1	Literasi	Menulis	Tidak
20	2-Mar-14	Karyawan	Karyawan	2	Seni	Tari	Ya
21	6-Apr-14	Karyawan	IRT	3	Seni	Tari	Ya
22	6-Apr-14	Karyawan	Wira Usaha	1	Seni	Pramuka	Tidak
23	21-Sep-14	Karyawan	IRT	1	Olahraga	Karate	Tidak
24	21-Sep-14	Karyawan	IRT	1	Olahraga	Silat	Tidak
25	22-Sep-14	Karyawan	PNS	1	Seni	Tari	Ya
26	11-Oct-14	Almarhum	PNS	1	Seni	Tari	Ya
27	26-Oct-14	Karyawan	Karyawan	2	Seni	Tari	Tidak
28	1-Nov-14	Karyawan	IRT	2	Seni	Tari	Ya
29	9-Nov-14	PNS	PNS	1	Praktik	Paskibra	Ya
30	28-Nov-14	PNS	PNS	2	Seni	Tari	Ya
31	26-Jul-15	Karyawan	IRT	1	Olahraga	Karate	Tidak
32	26-Jul-15	Karyawan	IRT	3	Seni	Tari	Ya
33	26-Jul-15	Karyawan	IRT	1	Seni	Tari	Ya
34	26-Jul-15	PNS	IRT	1	Praktik	Paskibra	Tidak

35	2-Aug-15	Karyawan	Karyawan	1	Praktik	Drumband	Tidak
36	23-Aug-15	Karyawan	PNS	1	Praktik	Paskibra	Tidak
37	24-Aug-15	Karyawan	PNS	2	Seni	Tari	Tidak
38	15-Nov-15	Karyawan	IRT	3	Olahraga	Karate	Ya
39	22-Nov-15	Karyawan	Guru	1	Literasi	Menulis	Tidak
40	22-Nov-15	Karyawan	IRT	1	Literasi	Menulis	Tidak
41	20-Dec-15	Karyawan	IRT	2	Seni	Tari	Ya
42	20-Dec-15	Karyawan	IRT	3	Olahraga	Karate	Tidak
43	6-Mar-16	Karyawan	IRT	2	Seni	Tari	Ya
44	24-Jul-16	Almarhum	Wira Usaha	4	Praktik	Pramuka	Tidak
45	24-Jul-16	Karyawan	PNS	1	Seni	Tari	Ya
46	31-Jul-16	Karyawan	Karyawan	1	Olahraga	Silat	Tidak
47	28-Aug-16	Karyawan	IRT	1	Literasi	Menulis	Tidak
48	11-Sep-16	Karyawan	Karyawan	1	Praktik	Drumband	Tidak
49	5-Nov-16	Karyawan	Karyawan	2	Seni	Lukis	Ya
50	29-Jan-17	Karyawan	Karyawan	1	Seni	Tari	Ya
51	29-Jan-17	Karyawan	IRT	2	Seni	Tari	Ya
52	18-Feb-17	Karyawan	Wira Usaha	2	Seni	Tari	Ya
53	16-Apr-17	Karyawan	IRT	1	Seni	Tari	Ya
54	7-May-17	Karyawan	Karyawan	2	Praktik	Drumband	Tidak
55	8-May-17	Karyawan	IRT	1	Seni	Tari	Ya
56	3-Sep-17	Wira Usaha	Guru	3	Praktik	Drumband	Tidak
57	17-Sep-17	PNS	PNS	2	Seni	Tari	Ya
58	17-Sep-17	Karyawan	Karyawan	2	Literasi	Menulis	Tidak
59	14-Jan-18	Karyawan	Karyawan	1	Seni	Tari	Ya
60	11-Feb-18	Karyawan	IRT	2	Literasi	Menulis	Tidak
61	4-Mar-18	Karyawan	Karyawan	1	Praktik	Pramuka	Tidak
62	4-Mar-18	Karyawan	IRT	3	Seni	Lukis	Tidak
63	3-Jun-18	Karyawan	IRT	2	Seni	Tari	Tidak
64	8-Jul-18	Karyawan	IRT	1	Olahraga	Silat	Tidak
65	8-Jul-18	Karyawan	Bidan	1	Praktik	Karate	Tidak
66	26-Aug-18	Karyawan	IRT	1	Olahraga	Drumband	Tidak
67	26-Aug-18	Karyawan	IRT	2	Olahraga	Drumband	Tidak
68	11-Sep-18	Karyawan	IRT	2	Seni	Tari	Ya
69	11-Sep-18	Pemusik	Karyawan	1	Literasi	Menulis	Tidak
70	18-Nov-18	TNI	Guru	1	Praktik	Drumband	Ya
71	6-Jan-19	Karyawan	IRT	1	Seni	Tari	Tidak
72	13-Jan-19	Karyawan	IRT	1	Olahraga	Karate	Tidak
73	27-Jan-19	Karyawan	Wira Usaha	2	Praktik	musik	Tidak
74	3-Feb-19	Karyawan	IRT	1	Praktik	Tari	Ya
75	3-Mar-19	PNS	PNS	1	Praktik	Tari	Ya
76	10-Mar-19	PNS	IRT	1	Seni	Tari	Tidak
77	4-Apr-19	Karyawan	Karyawan	2	Seni	Pramuka	Ya
78	27-Apr-19	Karyawan	Karyawan	2	Seni	Tari	Ya
79	30-Jun-19	PNS	PNS	1	Seni	Tari	Ya
80	4-Aug-19	PNS	PNS	1	Seni	Tari	Tidak
81	4-Aug-19	Karyawan	Wira Usaha	1	Literasi	Menulis	Ya
82	15-Sep-19	Karyawan	Wira Usaha	1	Praktik	Tari	Ya
83	22-Sep-19	Karyawan	Guru	2	Seni	Tari	Ya
84	22-Sep-19	Karyawan	IRT	1	Seni	Tari	Ya
85	22-Sep-19	Karyawan	IRT	1	Seni	Tari	Tidak

Diatas merupakan data – data siswa Lembaga Pendidikan Kursus Tari cabang Citayam. Selanjutnya data tersebut akan ditinjau kelengkapan datanya. Dari data yang diperoleh diketahui bahwa data lengkap dan tidak terdapat missing value pada atribut.

Data mining klasifikasi dengan menggunakan metode Naive Bayes Classifier dilakukan dengan menghitung probabilitas prior yang kemudian dilanjutkan dengan melakukan perhitungan probabilitas posterior.

Tahap awal Algoritma Naive Bayes dengan melakukan perhitungan nilai probabilitas hipotesis untuk masing-masing kelas P(H). Hipotesis data siswa Lembaga Pendidikan Kursus Tari cabang Citayam yang masih loyal memiliki atribut label AKTIF adalah YA. Data siswa Lembaga Pendidikan Kursus Tari di Depok yang masih sudah tidak loyal (*churn*) memiliki atribut label AKTIF adalah TIDAK.

Data yang digunakan telah diketahui statusnya yaitu YA (siswa yang loyal) atau decline (siswa yang *churn*). Dari 85 data siswa Lembaga Pendidikan Kursus Tari di Depok terdapat 37 siswa yang loyal (YA) dan 48 siswa yang *churn* (TIDAK) sehingga dengan menggunakan persamaan 2 diatas, maka :

$$P(YA) = 37 : 85 = 0,43529$$

$$P(TIDAK) = 48 : 85 = 0,56471$$

Perhitungan Probabilitas Prior untuk semua atribut adalah sebagai berikut :

Tabel 3: Perhitungan Probabilitas Prior

Atribut	Jml Kasus	Ya	Tidak	P(H X)	
				Ya	Tidak
Total	85	37	48	0.43529	0.56471
Waktu					
4	4	3	1	0.08108	0.02083
5	2	1	1	0.02703	0.02083
6	25	0	25	0	0.52083
7	1	1	0	0.02703	0
9	2	2	0	0.05405	0
10	1	1	0	0.02703	0
11	1	1	0	0.02703	0
12	12	0	12	0	0.25
13	1	1	0	0.02703	0
15	1	1	0	0.02703	0
18	2	0	2	0	0.04167
23	1	1	0	0.02703	0
24	4	0	4	0	0.08333
27	1	1	0	0.02703	0
31	1	1	0	0.02703	0
32	1	1	0	0.02703	0
34	1	1	0	0.02703	0
35	2	2	0	0.05405	0
36	2	0	2	0	0.04167
37	1	1	0	0.02703	0
41	1	1	0	0.02703	0
45	1	1	0	0.02703	0
48	2	1	1	0.02703	0.02083

49	1	1	0	0.02703	0
53	2	1	1	0.02703	0.02083
61	3	3	0	0.08108	0
62	1	1	0	0.02703	0
63	1	1	0	0.02703	0
68	1	1	0	0.02703	0
69	1	1	0	0.02703	0
70	2	2	0	0.05405	0
71	1	1	0	0.02703	0
74	1	1	0	0.02703	0
80	1	1	0	0.02703	0
Pekerjaan ayah					
Almarhum	2	1	1	0.02703	0.02083
Karyawan	58	25	33	0.67568	0.6875
Pemusik	1	0	1	0	0.02083
PNS	18	10	8	0.27027	0.16667
TNI	1	1	0	0.02703	0
Wira Usaha	5	0	5	0	0.10417
Pekerjaan ibu					
Bidan	1	1	0	0.02703	0
Guru	4	2	2	0.05405	0.04167
Ibu rumah Tangga	35	14	21	0.37838	0.4375
Karyawan	23	6	17	0.16216	0.35417
PNS	15	12	3	0.32432	0.0625
Wira Usaha	7	3	4	0.08108	0.08333
Minat					
Literasi	10	2	8	0.05405	0.16667
Olahraga	16	2	14	0.05405	0.29167
Praktik	21	5	16	0.13514	0.33333
Seni	38	28	10	0.75676	0.20833
Ekskul sekolah					
Basket	1	1	0	0.02703	0
Drumband	8	1	7	0.02703	0.14583
Karate	11	1	10	0.02703	0.20833
Lukis	2	1	1	0.02703	0.02083
Ménulis	10	2	8	0.05405	0.16667
Music	1	0	1	0	0.02083
Paskibra	5	1	4	0.02703	0.08333
Pramuka	7	1	6	0.02703	0.125
Silat	4	0	4	0	0.08333
Tari	36	29	7	0.78378	0.14583
Anak ke					
1	47	17	30	0.45946	0.625
2	29	17	12	0.45946	0.25
3	8	3	5	0.08108	0.10417
4	1	1	0	0.02703	0

Perhitungan probabilitas prior merupakan perhitungan kemungkinan setiap atribut dengan kelas YA dan kelas TIDAK. Perhitungan probabilitas prior ini berguna dalam menentukan kelas pada kasus baru berdasarkan data sebelumnya dengan cara menghitung probabilitas posteriornya. Sebagai contoh kasus data siswa seperti dibawah ini:

$$P(H|Aktif = Ya) = 0 \times 0.675657 \times 0.378378 \times 0.45946 \times 0.756757 \times 0.783784 = 0$$

$$P(H|Aktif = Tidak) = 0.5208333 \times 0.6875 \times 0.4375 \times 0.45946 \times 0.25 \times 0.28333 \times 0.1458333 = 0.00119$$

$$P(H|Aktif=Ya) P(Ya) = 0 \times 0.435294 = 0$$

$$P(H|Aktif=Tidak) P(Tidak) = 0.00119 \times 0.564706 = 0.00065$$

sehingga

$$P(H|Aktif=Ya)P(Ya) < P(H|Aktif=Tidak) P(Tidak) \ 0 < 0.00065$$

Tabel 4: Probabilitas Posterior

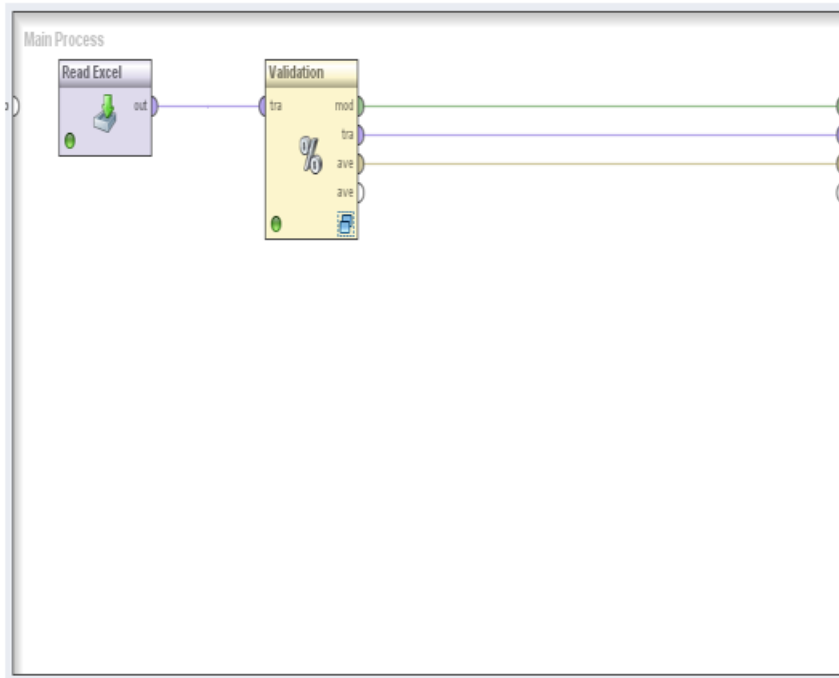
Atribut		P(H X)	
		Ya	Tidak
Waktu (bulan)	6	0	0.5208333
Pekerjaan Ayah	Karyawan	0.675657	0.6875
Pekerjaan Ibu	Ibu rumah Tangga	0.378378	0.4375
Anak ke	2	0.45946	0.25
Minat	Seni	0.756757	0.2083333
Elskul sekolah	Tari	0.783784	0.1458333

Dari hasil perhitungan probabilitas posterior diatas diketahui bahwa nilai P (H|Aktif=Ya) lebih kecil dari pada nilai P (H|Aktif=Tidak), sehingga dapat disimpulkan bahwa untuk kasus siswa tersebut masuk kedalam klasifikasi Tidak Aktif. Hasil klasifikasi ini sesuai dengan data yang tersedia. Hal ini menunjukkan bahwa model data mining klasifikasi dengan metode Naïve Bayes ini memiliki kriteria BAIK sebagai model data mining klasifikasi loyalitas siswa Lembaga Pendidikan Kursus Tari.

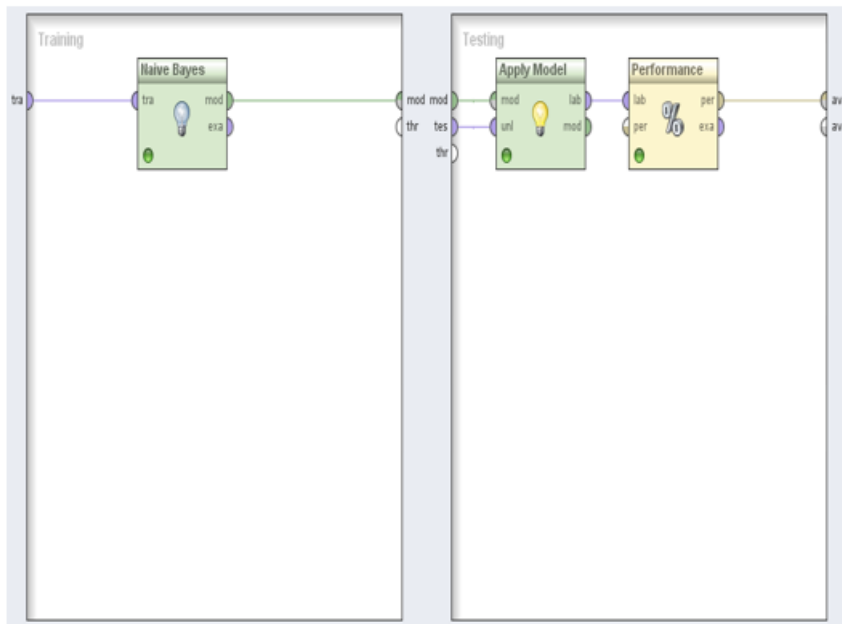
Untuk menguji kinerja model data mining klasifikasi dengan metode Naïve Bayes dilakukan eksperimen. Eksperimen yang dilakukan dengan pendekatan data mining algoritma Naive Bayes Classifier menggunakan perangkat lunak Rapidminer 5.3, lihat Gambar 5 dan 6.

Evaluasi dan validasi untuk menguji model pada penelitian ini menggunakan operator cross validation, confusion matriks dan ROC curve. Operator validation menggunakan Cross Validation dengan parameter number of validation 10 dimana data data siswa akan dibagi secara acak ke dalam 10 bagian dan masing-masing bagian tersebut akan dilakukan proses klasifikasi.

Metode Naïve Bayes yang digunakan dalam data mining klasifikasi loyalitas siswa Lembaga Pendidikan Kursus Tari di Depok. Hasil akurasi dengan menggunakan tools Rapid Miner menunjukkan 80,42%, lihat Gambar 7. Hasil classification error dengan menggunakan tools Rapid Miner menunjukkan 19,58%, lihat Gambar 8.



Gambar 5: Validasi Model dengan Rapid Miner 5.3



Gambar 6: Model Naïve Bayes dengan Rapid Miner 5.3

accuracy: 80.42% +/- 17.70% (mikro: 80.00%)			
	true Ya	true Tidak	class precision
pred. Ya	31	11	73.81%
pred. Tidak	6	37	86.05%
class recall	83.78%	77.08%	

Gambar 7: Hasil Accuracy Model Naïve Bayes

classification_error: 19.58% +/- 17.70% (mikro: 20.00%)			
	true Ya	true Tidak	class precision
pred. Ya	31	11	73.81%
pred. Tidak	6	37	86.05%
class recall	83.78%	77.08%	

Gambar 8: Hasil Classification Error

PerformanceVector

PerformanceVector:
 accuracy: 80.42% +/- 17.70% (mikro: 80.00%)
 ConfusionMatrix:
 True: Ya Tidak
 Ya: 31 11
 Tidak: 6 37
 classification_error: 19.58% +/- 17.70% (mikro: 20.00%)
 ConfusionMatrix:
 True: Ya Tidak
 Ya: 31 11
 Tidak: 6 37
 AUC: 0.860 +/- 0.144 (mikro: 0.860) (positive class: Tidak)

Gambar 9: Hasil Performance

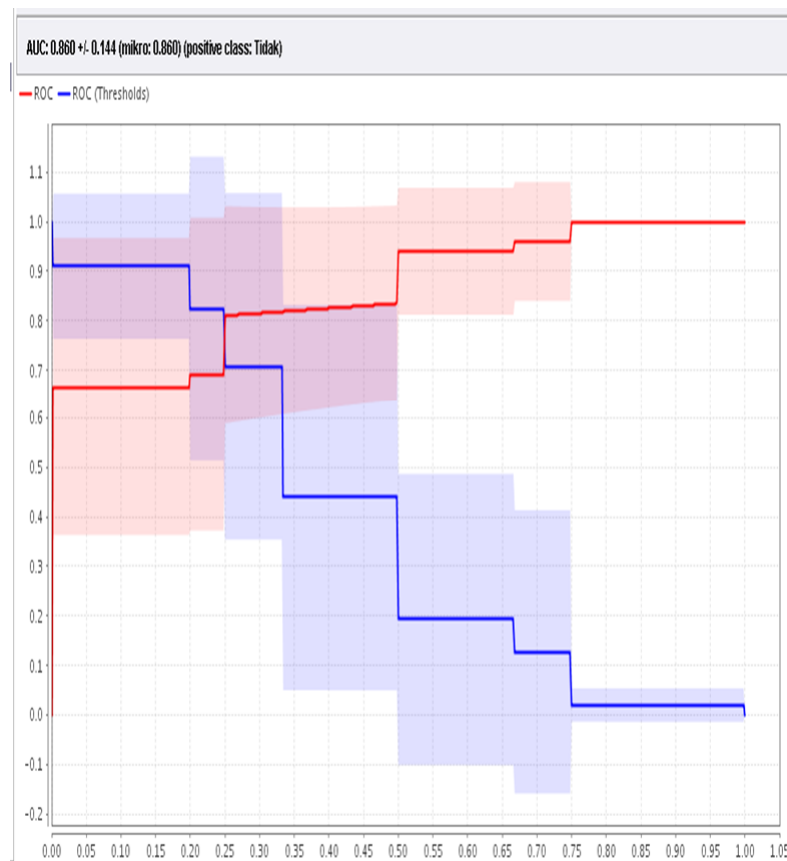
Gambar 9, memperlihatkan nilai akurasi dengan

pendekatan Naive Bayes Classifier adalah = 80,42% dapat dihitung dari hasil Confusion Matrix dimana true positif (tp) bernilai 31, false negatif (fn) bernilai 11, false positif (fp) bernilai 6 dan true negatif (tn) bernilai 37 sehingga:

$$Akurasi = \frac{31+37}{31+11+6+37} = 0,8042$$

$$Classification\ Error = \frac{11+6}{31+11+6+37} = 0,1958$$

Hasil kurva ROC menunjukkan arah garis melintang dari kiri bawah ke kanan atas didalam grafik dengan nilai AUC 0.860 bahwa nilai AUC 0.80 – 0.90 = Good classification sehingga model data mining dengan metode Naive Bayes ini merupakan model klasifikasi dengan kriteria BAIK, lihat Gambar 10.



Gambar 10: Hasil AUC dan Kurva ROC

Setelah melakukan eksperiment data mining klasifikasi dengan algoritma Naïve Bayes untuk mengklasifikasikan siswa dalam mengetahui tingkat loyalitas siswa dan potensi siswa *churn* sehingga lembaga dapat menentukan langkah-langkah dalam meningkatkan pemeliharaan siswa maka dihasilkan bahwa model klasifikasi dengan algoritma Naïve Bayes ini TEPAT bila akan diterapkan ke dalam aplikasi klasifikasi.

Penutup

Hasil perhitungan probabilitas prior dan probabilitas posterior data mining klasifikasi dengan metode Naïve Bayes dilakukan menggunakan data primer yang tersedia dimana menunjukkan bahwa model data mining klasifikasi dengan metode Naïve Bayes ini memiliki kriteria BAIK sebagai model data mining klasifikasi loyalitas siswa Lembaga Pendidikan Kursus Tari .

Kemudian untuk menguji kinerja model data mining klasifikasi ini menggunakan tools Rapid Miner5.3 diperoleh Hasil akurasi 80.42%, classification error 19.52%, kurva ROC menunjukkan arah garis melintang dari kiri bawah ke kanan atas didalam grafik dengan nilai AUC 0.860 dimana nilai AUC 0.80 – 0.90 = Good classification merupakan model yang BAIK.

Hasil penelitian dapat disimpulkan bahwa model data mining klasifikasi dengan metode Naïve Bayes adalah model yang memiliki kriteria BAIK untuk mengkalsifikasikan loyalitas siswa Lembaga Pendidikan Kursus Tari cabang Citayam sehingga lembaga dapat menentukan langkah-langkah dalam meningkatkan pemeliharaan siswa.

Penelitian selanjutnya dapat dilengkapi dengan metode seleksi fitur seperti Particle Swarm Optimization, Genetic Algorithm dan sebagainya untuk lebih meningkatkan akurasi dan mengetahui faktor utama yang dapat mempengaruhi klasifikasi loyalitas siswa.

Daftar Pustaka

- [1] M. A. Rizka and R. Hardiansyah, "Strategi pengembangan inovasi program pendidikan nonformal sebagai best practices bagi pusat kegiatan belajar masyarakat", *J. Pendidik. dan Pemberdaya. Masy.*, vol. 3, no. 2, p. 187, 2016.
- [2] W. Suharso and A. Djunaidy, "Analisis Customer Churn Menggunakan Bayesian Belief Network (Studi Kasus: Perusahaan Layanan Internet)", *Sisfo*, vol. 4, no. 5, pp. 323–335, 2013.
- [3] M. Divekar, S. Dashpute, and P. Kamble, "Analysis of Customer Churn in Mobile Industry using Data Mining", *International Journal of Emerging Technology and Advanced Engineering*, vol. 5, no. 3, pp. 1–6, 2015.
- [4] A. Nurzahputra, A. R. Safitri, and M. A. Muslim, "Klasifikasi Pelanggan pada Customer Churn Prediction Menggunakan Decision Tree", *Prosiding Seminar Nasional Mat. X 2016*, pp. 717–722, 2016.
- [5] A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari, and U. Abbasi, "Improved churn prediction in telecommunication industry using data mining techniques", *Appl. Soft Comput. J.*, vol. 24, pp. 994–1012, 2014.
- [6] Y. Liu and Y. Zhuang, "Research Model of Churn Prediction Based on Customer Segmentation and Misclassification Cost in the Context of Big Data", *J. Comput. Commun.*, vol. 3, no. 6, pp. 87–93, 2015.
- [7] S. Mulyati, Y. Yulianti, and A. Saifudin, "Penerapan Resampling dan Adaboost untuk Penanganan Masalah Ketidakeimbangan Kelas Berbasis Naïve Bayes pada Prediksi Churn Pelanggan", *J. Inform. Univ. Pamulang*, vol. 2, no. 4, p. 190, 2017.
- [8] N. Suryana, "Prediksi Churn Dan Segmentasi Pelanggan Tv Berlangganan (Studi Kasus Transvision Jawa Barat)", *Jurnal TEDC, Politeknik TEDC Bandung*, vol. 11, no. 2, pp. 185–192, ISSN 1978-0060, 2017.
- [9] M. Arifin, "Ig-Knn Untuk Prediksi Customer Churn Telekomunikasi", *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 6, no. 1, p. 1, 2015.
- [10] Daniel T. Larose, "Discovering Knowledge in Data", Wiley Interscience, A. John Wiley & Sons Inc. Publication, 2005.
- [11] Florin Gorunescu, "Data Mining", Springer, 2011.
- [12] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, "Data Mining: Practical Machine Learning Tools and Techniques", 3rd Edition, Morgan Kaufmann, ISBN : 9780080890364, 2011.
- [13] Max Bramer, "Principles of Data Mining", Springer, 2007.
- [14] E. D. Madyatmadja and M. Aryuni, "Comparative study of data mining model for credit card application scoring in bank", *J. Theor. Appl. Inf. Technol.*, vol. 59, no. 2, pp. 269–274, 2014.
- [15] C.R. Kothari, "Research Methodology: Methods and Techniques", 2nd Edition, New Age International Publisher, 2004.