



Implementasi Penerapan Metode C4.5 dan Naïve Bayes Dalam Tingkat Kelulusan Akreditasi Lembaga PAUD Pada Badan Akreditasi Nasional

Lenggo Genisa, Dadang Iskandar Mulyana*

Prodi Teknik Informatika, STIKOM CKI, Jakarta, Indonesia
Email: ¹lenggo.genisa@gmail.com, ^{2,*}mahvin2012@gmail.com
Email Penulis Korespondensi: mahvin2012@gmail.com

Abstrak—Pendidikan sejak usia dini merupakan salah satu cara untuk menstimulasi potensi anak. Hal ini dijelaskan dalam Undang-Undang Republik Indonesia Nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional menyatakan bahwa Pendidikan Anak Usia Dini (PAUD) adalah suatu upaya pembinaan yang ditujukan kepada anak sejak lahir sampai dengan usia enam tahun yang dilakukan melalui pemberian rangsangan pendidikan untuk membantu pertumbuhan dan perkembangan jasmani dan rohani agar anak memiliki kesiapan dalam memasuki pendidikan lebih lanjut. Badan Akreditasi Nasional Pendidikan Anak Usia Dini dan Pendidikan Non Formal yang selanjutnya disebut BAN PAUD dan PNF adalah badan evaluasi mandiri yang menetapkan kelayakan program dan/ atau satuan PAUD dan PNF. BAN PAUD dan PNF dibentuk berdasarkan Permendikbud Nomor 52 tahun 2015 tentang Badan Akreditasi Nasional Pendidikan Anak Usia Dini dan Pendidikan Non Formal yang merupakan pengganti dari Permendikbud 59 tahun 2012. Peningkatan kualitas pelaksanaan Akreditasi PAUD dan PNF dapat dilakukan dengan cara meningkatkan ketersediaan layanan akreditasi pendidikan nonformal. Hal lain yang dapat dilakukan Meningkatkan kualitas pelaksanaan akreditasi PAUD adalah dengan memberikan kepastian dan keterjaminan memperoleh layanan akreditasi pendidikan nonformal serta meningkatkan sistem tata kelola yang handal dalam menjamin terselenggaranya layanan akreditasi pendidikan nonformal. Penelitian ini menggunakan teknik data mining dalam memprediksi status akreditasi satuan pendidikan PAUD. Pertama, preprocessing digunakan untuk mendapatkan dataset yang berkualitas. Kedua, data diproses untuk mendapatkan serangkaian prediksi. Dilangkah ini kedua algoritma data mining diterapkan, yaitu Algoritma Naïve Bayes dan Algoritma C4.5 dengan tujuan untuk mengetahui kinerja dari kedua algoritma dengan tingkat akurasi yang lebih besar akan direkomendasikan dalam menyelesaikan masalah prediksi terakreditasinya satuan pendidikan PAUD di BAN PAUD dan PNF Provinsi DKI Jakarta. Lalu yang Ketiga, hasilnya akan di Confusion Matrix untuk memvalidasi nilai akurasi hasil prediksi. Dan hasil penilaian meunjukkan bahwa metode Algoritma C4.5 dan Naïve Bayes sama sama bisa di gunakan untuk memprediksi status akreditasi satuan pendidikan PAUD dengan akurasi 99,00%.

Kata Kunci: Akreditasi; Algoritma; C4.5; Naïve Bayes; Implementasi

Abstract— Education from an early age is one way to stimulate children's potential. This is explained in the Law of the Republic of Indonesia Number 20 of 2003 concerning the National Education System which states that Early Childhood Education (PAUD) is a coaching effort aimed at children from birth to the age of six which is carried out through the provision of educational stimuli to help growth and physical and spiritual development so that children have readiness to enter further education. The National Accreditation Board for Early Childhood Education and Non-Formal Education, hereinafter referred to as BAN PAUD and PNF, is an independent evaluation body that determines the feasibility of PAUD and PNF programs and/or units. BAN PAUD and PNF were formed based on Permendikbud Number 52 of 2015 concerning the National Accreditation Board for Early Childhood Education and Non-Formal Education which is a substitute for Permendikbud 59 of 2012. Improving the quality of the implementation of PAUD and PNF Accreditation can be done by increasing the availability of non-formal education accreditation services. Other things that can be done to improve the quality of the implementation of PAUD accreditation are by providing certainty and guarantee of obtaining non-formal education accreditation services and improving a reliable governance system in ensuring the implementation of non-formal education accreditation services. This study uses data mining techniques in predicting the accreditation status of PAUD education units. First, preprocessing is used to get a quality dataset. Second, the data is processed to get a series of predictions. In this step, two data mining algorithms are applied, namely the Naïve Bayes Algorithm and the C4.5 Algorithm with the aim of knowing the performance of the two algorithms with a greater level of accuracy will be recommended in solving the problem of predicting the accreditation of PAUD education units in BAN PAUD and PNF DKI Jakarta Province. Then the third, the results will be in the Confusion Matrix to validate the accuracy of the prediction results. And the results of the assessment show that the C4.5 and Naïve Bayes Algorithm methods can be used to predict the accreditation status of PAUD education units with an accuracy of 99.00%.

Keywords: Accreditation; Algorithm; C4.5; Naive Bayes; Implementation

1. PENDAHULUAN

Pendidikan sejak usia dini merupakan salah satu cara untuk menstimulasi potensi anak. Hal ini dijelaskan dalam Undang-Undang Republik Indonesia Nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional pada bab I pasal 1 ayat 14 menyatakan bahwa Pendidikan Anak Usia Dini (PAUD) adalah suatu upaya pembinaan yang ditujukan kepada anak sejak lahir sampai dengan usia enam tahun yang dilakukan melalui pemberian rangsangan pendidikan untuk membantu pertumbuhan dan perkembangan jasmani dan rohani agar anak memiliki kesiapan dalam memasuki pendidikan lebih lanjut. Hal ini menjelaskan bahwa tujuan Pendidikan Anak Usia Dini adalah mengembangkan berbagai potensi anak sejak dini sebagai persiapan untuk hidup dan dapat menyesuaikan diri dengan lingkungannya.



Upaya pengembangan potensi yang dimiliki oleh setiap anak pada usia emasnya (golden age) diperlukan sebuah bentuk layanan yang dapat memberikan stimulasi, perawatan, dan pengasuhan yang dapat membantu pertumbuhan dan perkembangan anak. Pendidikan usia dini merupakan wahana pendidikan yang sangat fundamental dalam memberikan kerangka dasar terbentuk dan berkembangnya dasar-dasar pengetahuan, sikap, dan keterampilan pada anak. Pendidikan anak usia dini (PAUD) menjadi harapan baru bagi terbentuknya generasi penerus bangsa yang berkualitas dan berkarakter sebagai generasi dan calon pemimpin masa depan. Dalam mewujudkan harapan tersebut, bentuk penyelenggaraan pendidikan anak usia dini pun harus dilaksanakan secara serius sesuai dengan kebutuhan dan tahapan perkembangan anak.

Pemerintah melalui Kemendikbud sebagai lembaga yang mewadahi penyelenggaraan Pendidikan Anak Usia Dini (PAUD) telah memberikan standar bagi penyelenggaraan Pendidikan Anak Usia Dini. Melalui Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 137 Tahun 2014, pemerintah telah menetapkan standar Nasional Pendidikan Anak Usia Dini yang terdiri dari Delapan Standar, yaitu 1) Standar Tingkat Pencapaian Perkembangan Anak; 2) Standar Isi; 3) Standar Proses; 4) Standar Penilaian; 5) Standar Pendidik dan Tenaga Kependidikan; 6) Standar Sarana dan Prasarana; 7) Standar Pengelolaan; dan 8) Standar Pembiayaan. Delapan Standar yang telah ditetapkan melalui peraturan Menteri Pendidikan dan Kebudayaan merupakan dasar dan acuan dari setiap satuan dan aktivitas PAUD untuk mewujudkan tujuan pendidikan dan penjaminan kualitas mutu PAUD.

Badan Akreditasi Nasional Pendidikan Anak Usia Dini dan Pendidikan Non Formal yang selanjutnya disebut BAN PAUD dan PNF adalah badan evaluasi mandiri yang menetapkan kelayakan program dan/ atau satuan PAUD dan PNF. BAN PAUD dan PNF dibentuk berdasarkan Permendikbud Nomor 52 tahun 2015 tentang Badan Akreditasi Nasional Pendidikan Anak Usia Dini dan Pendidikan Non Formal yang merupakan pengganti dari Permendikbud 59 tahun 2012.

Peningkatan kualitas pelaksanaan Akreditasi PAUD dan PNF dapat dilakukan dengan cara meningkatkan ketersediaan layanan akreditasi pendidikan nonformal.

Hal lain yang dapat dilakukan Meningkatkan kualitas pelaksanaan akreditasi PAUD adalah dengan memberikan kepastian dan keterjaminan memperoleh layanan akreditasi pendidikan nonformal serta meningkatkan sistem tata kelola yang handal dalam menjamin terselenggaranya layanan akreditasi pendidikan nonformal.

Dengan memanfaatkan ilmu pengetahuan yang semakin maju, berdasarkan permasalahan tersebut peneliti melakukan penelitian, yaitu Implementasi Penerapan Metode C4.5 dan Naïve Bayes dalam tingkat kelulusan akreditasi Lembaga PAUD pada Badan Akreditasi Nasional. Dengan demikian peneliti akan mengimplementasikan metode C4.5 dan Naïve Bayes dalam melihat hasil akurasi dari kedua metode tersebut.

Teknik *data mining* merupakan cara yang mudah dan *relatif* cepat untuk memperoleh pengetahuan secara otomatis (Suyanto, 2017) dan pengetahuan abstrak dari sebuah *database* yang besar (Mulya, 2019)[1] yang meliputi bentuk dan/atau hubungan antar data [2].

Data mining disebut sebagai proses ekstraksi pengetahuan dari data yang besar, sesuai fungsinya *data mining* adalah proses pengambilan keputusan dari *volume* data yang besar yang disimpan dalam basis data, *data warehouse*, atau informasi yang disimpan dalam *repository* (Wu et al., 2008)[3].

Data mining merupakan tahap dalam pengklasifikasian terhadap data dengan menghubungkan masing-masing pola pada setiap data set yang berukuran besar dengan jumlah data yang besar pula. *Data mining* dapat juga didefinisikan menggali data dari banyaknya informasi yang akan dicari sehingga data yang perlu diketahui akan lebih mudah dicari dengan adanya sistem pola yang dibuat berdasarkan titik terdekat dengan informasi yang sering di perlukan[4], [5]. Istilah *data mining* berasal dari kemiripan antara pencarian informasi yang bernilai dari *basis data* yang besar dengan menambang sebuah gunung untuk memperoleh sesuatu yang bernilai (Sumathi & Sivanandam, 2007)[3]. Secara teknis, *data mining* dapat disebut sebagai proses untuk menemukan aturan atau pola dari ratusan atau ribuan data dari sebuah relasi basis data yang sangat besar[6]

Menurut Han dan Kamber (2011), *data mining* adalah proses menemukan pola yang menarik dan pengetahuan dari data yang berjumlah besar [7]. Salah Satu metode yang dapat digunakan untuk menganalisa database tersebut adalah metode *data mining*. Menurut Jing dan Marceron (2017) data yang berlimpah membuka peluang diterapkannya *data mining* untuk pengelolaan pendidikan yang lebih baik dan *data mining* untuk pelaksanaan pembelajaran[8]

Bagian penting dalam *data mining* adalah teknik klasifikasi, yaitu cara digunakan dalam mempelajari dataset agar didapatkan hubungan antara data yang membentuk *pattern* (pola) sehingga dapat diperoleh *knowledge*. Ada beberapa jenis metode pada penerapan *data mining* yang bisa diterapkan untuk klasifikasi, diantaranya Artificial Neural Network, Algoritma C4.5, Nearest Neighbour Rule, Naïve Bayes, K-Mean, dan lain-lain.

Peneliti mengambil topik tentang klasifikasi dan penerapan *data mining* telah banyak dilakukan sebelumnya. Sebagian peneliti sebelumnya juga mengukur tingkat akurasi pada masing-masing metode *data mining*. Menurut Mulya (2019) RapidMiner adalah sebuah alat *data mining* yang digunakan untuk menganalisis informasi [1]. Pada penelitiannya Supriyanti dkk (2016) menyatakan bahwa RapidMiner merupakan sebuah tool yang dapat digunakan untuk membantu menyelesaikan masalah prediksi, proses *data mining* dan *text mining*[9]. RapidMiner adalah sebuah *tools* yang digunakan dalam teknik yang berada di lingkungan *machine learning*, data



mining, text mining dan *predictive analytics* [10] RapidMiner merupakan software atau perangkat lunak untuk pengolahan data. Dengan menggunakan prinsip dan algoritma data mining. RapidMiner mengesktrak pola dari dataset yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database. RapidMiner memudahkan penggunaanya dalam melakukan perhitungan data yang sangat banyak dengan menggunakan operator. Operator berfungsi untuk memodifikasi data. Data dihubungkan dengan node-node pada operator, lalu dihubungkan ke node hasil untuk melihat hasilnya. Hasil yang ditampilkan RapidMiner dapat ditampilkan secara virtual dengan grafik. Menjadikan RapidMiner adalah salah satu software pilihan untuk melakukan ekstraksi data dengan metode-metode data mining [11].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada penelitian ini menggunakan tahap penelitian sebagai berikut:

1. Studi Pustaka (*Study Literature*)

Studi Pustaka adalah proses pengumpulan data yang bersumber dari informasi tertulis atau bisa dari data sekunder yang didapat melalui referensi buku dan internet. Pada tahap ini peneliti melakukan studi literatur untuk mendapatkan referensi untuk mendukung penelitian ini menjadi semakin relevan.

2. Pengamatan Lapangan (Observasi)

Observasi adalah metode pengumpulan data melalui pengamatan langsung atau peninjauan secara cermat dan langsung di lapangan. Dalam hal ini perlu mengunjungi lokasi penelitian untuk mengamati langsung berbagai hal atau kondisi yang ada di lapangan. Pada Langkah ini peneliti melakukan observasi untuk mendapatkan data penelitian yang akan diolah. Peneliti melakukan pengamatan dengan mendatangi ke BAN PAUD dan PNF Provinsi DKI Jakarta. Dengan datang langsung ke kantor BAN PAUD dan PNF Provinsi DKI Jakarta, penulis mengamati sistem yang sudah berjalan. Sistem yang sudah berjalan di BAN PAUD dan PNF Provinsi DKI Jakarta sudah lumayan efektif, karena sudah menggunakan sistem aplikasi.

3. Wawancara

Wawancara adalah percakapan dengan maksud-maksud tertentu. Pada metode ini penulis dan responden berhadapan langsung (face to face) untuk mendapatkan informasi secara lisan dengan tujuan mendapatkan data yang dapat menjelaskan permasalahan penelitian yang berkaitan dengan masalah yang diangkat dalam penelitian ini. Penulis melakukan wawancara tanya-jawab dengan mengajukan beberapa pertanyaan secara langsung dengan salah seorang karyawan di bagian pemegang data dan sistem akreditasi.

4. Pengumpulan Data

Pada penelitian ini, peneliti mengambil data dokumen berupa SK Akreditasi BAN PAUD dan PNF tahun 2019, berguna untuk peneliti mengetahui satuan lembaga PAUD mendapatkan peringkat akreditasi berapa, dan juga mengambil data 8 standar tiap 100 satuan Lembaga PAUD yang ikut akreditasi di tahun 2019.

5. Analisis Data

Setelah mendapatkan data penelitian, peneliti melakukan tahap analisis data dimana data tersebut disesuaikan dengan kebutuhan penelitian saat ini. Tahap ini dilakukan pengolahan data awal untuk mendapatkan data yang baik sebelum data diolah menggunakan metode Algoritma C4.5 dan *Naïve Bayes*.

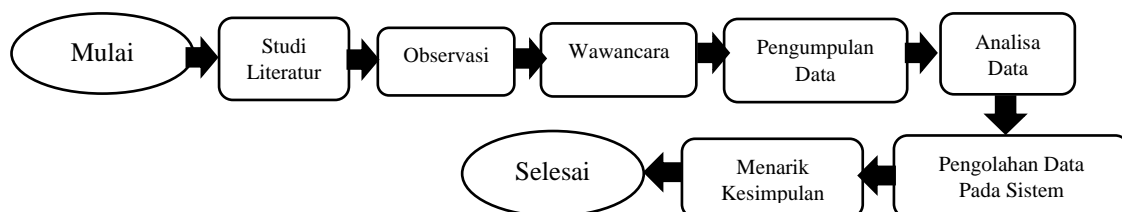
6. Pengolahan Data Pada Sistem

Tahap berikutnya dilakukan proses mining menggunakan metode Algoritma C4.5 dan *Naïve Bayes* pada tools RapidMiner.

7. Menarik Kesimpulan

Tahap selanjutnya hasil pengujian dari metode Algoritma C4.5 dan *Naïve Bayes* akan dibandingkan dengan tujuan untuk melihat akurasi klasifikasi yang mana yang bagus Ketika diimplementasikan dengan menggunakan kedua metode tersebut, yaitu metode C4.5 dan *Naïve Bayes* dan juga bisa melihat persyaratan yang paling berpengaruh dengan akreditasi.

Agar lebih jelas desain pada penelitian yang penulis gunakan, maka dapat dilihat pada gambar 1



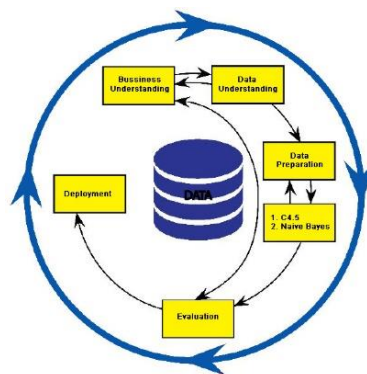
Gambar 1. Tahapan Penelitian



2.2 CRISP-DM

Metodologi CRISP-DM menyediakan standar proses baku untuk *data mining* yang dapat diterapkan ke dalam strategi pemecahan masalah umum pada unit penelitian. CRISP-DM membandingkan metodologi *data mining* lain lebih lengkap dan terdokumentasi dengan baik. Tiap fase terstruktur dan terdokumentasi dengan jelas sehingga mudah diaplikasikan.

CRISP-DM – awalnya dibangun oleh 3 perusahaan yaitu SPSS, Daimler Chrysler, dan NCR yang kemudian dikembangkan ratusan organisasi dan perusahaan secara bersama-sama. CRISP-DM merupakan singkatan dari *Cross-Industry Standard Process for Data Mining* dan memiliki 6 tahapan yaitu *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment*. Selain penggunaan metode yang tepat, pemilihan Algoritma yang tepat adalah yang penting dilakukan dalam *data mining*. Peneliti memilih Algoritma C4.5 dan Naïve Bayes pada Implementasi penerapan dalam tingkat kelulusan akreditasi lembaga paud pada badan akreditasi nasional[12]. Proses CRISP-DM pada penelitian ini dapat dilihat pada gambar 2:



Gambar 2. Proses CRISP-DM menggunakan metode C4.5 dan Naïve Bayes

2.3 Pengumpulan Data

Teknik pengumpulan data merupakan tahap yang paling strategis dalam penelitian, karena tujuan utama dari penelitian adalah mendapatkan data. Pengumpulan data dilakukan langsung dilapangan yaitu data akreditasi lembaga satuan PAUD tahun 2019 yang telah mengikuti tahap akreditasi yang diperoleh dari tim bagian pemegang data akreditasi di BAN PAUD dan PNF Provinsi DKI Jakarta. Data yang diperoleh yaitu 100 dataset lembaga satuan PAUD dengan 9 variabel. Variabel yang digunakan antara lain adalah Standar 1 Standar Tingkat Pencapaian Perkembangan Anak, Standar 2 Standar Isi, Standar 3 Standar Proses, Standar 4 Standar Pendidikan dan Tenaga Kependidikan, Standar 5 Standar Sarana dan Prasarana, Standar 6 Standar Pengelolaan, Standar 7 Standar Pembiayaan, Standar 8 Standar Penilaian, dan Peringkat Akreditasi. Adapun contoh data yang digunakan dapat dilihat pada tabel 1:

Tabel 1. Contoh Data

No	NPSN	Nama Lembaga	Standar Tingkat Pencapaian Perkembangan Anak	Standar 2 Standar Isi	Standar 3 Standar Proses	Standar 4 Standar Pendidikan dan Tenaga Kependidikan	Standar 5 Standar Sarana dan Prasarana	Standar 6 Standar Pengelolaan	Standar 7 Standar Pembiayaan	Standar 8 Standar Penilaian	Peringkat Akreditasi
1	69929608	PAUD PELITA HATI	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Tidak Lengkap	Tidak Lengkap	Lengkap	Terakreditasi
2	69857264	PAUD MATAHARI 2	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Tidak Lengkap	Lengkap	Lengkap	Terakreditasi
3	69821033	AYAH BUNDA	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Tidak Lengkap	Lengkap	Lengkap	Terakreditasi
4	69820933	BKB PAUD RIZKI	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Tidak Lengkap	Lengkap	Lengkap	Terakreditasi
5	20112056	TK MARGA LAKSANA	Lengkap	Lengkap	Tidak Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Terakreditasi
6	69844981	PAUD NUSA INDAH	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Tidak Lengkap	Lengkap	Lengkap	Terakreditasi
7	69821006	bkb PAUD WIJAYA KUSUMA	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Terakreditasi
8	20177886	TK LINTANG HARAPAN	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Tidak Lengkap	Lengkap	Terakreditasi
9	69732552	RA Al Lathifiyah	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Tidak Lengkap	Lengkap	Terakreditasi
10	69945470	PAUD OXFORD	Lengkap	Tidak Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Lengkap	Terakreditasi



No	NPSN	Nama Lembaga	Standar Tingkat Pencapaian Perkembangan Anak	Standar 2 Standar Isi	Standar 3 Standar Proses	Standar 4 Standar Pendidikan dan Tenaga Kependidikan	Standar 5 Standar Sarana dan Prasarana	Standar 6 Standar Pengelolaan	Standar 7 Standar Pembiayaan	Standar 8 Standar Penilaian	Peringkat Akreditasi
UTOMO PRESCHOOL											

2.5 Analisis Data

Setelah mendapatkan data penelitian, peneliti melakukan tahap analisis data dimana data tersebut disesuaikan dengan kebutuhan penelitian saat ini. Tahap ini dilakukan pengolahan data awal untuk mendapatkan data yang baik sebelum data diolah menggunakan metode Algoritma C4.5 dan *Naive Bayes*. Hal ini dimaksudkan agar dapat diketahui nilai akurasi hasil prediksi berdasarkan penerapan dari dua metode *data mining* yang digunakan.

Hasil dari pengumpulan data didapatkan *record* sebanyak 100 dataset Lembaga satuan PAUD yang telah menyelesaikan Akreditasi tahun 2019 dengan 11 atribut. Lembaga PAUD yang dikategorikan terakreditasi adalah Lembaga PAUD yang dapat melengkapi 8 persyaratan yang diminta untuk syarat akreditasi satuan PAUD. Sedangkan Lembaga PAUD yang tidak melengkapi 8 persyaratan, maka dikategorikan tidak terakreditasi.

Tetapi dari hasil pengumpulan data, data *record* dan atribut tidak seluruhnya bisa digunakan karena perlu dilakukan Analisis Data untuk mendapatkan data yang baik. Adapun rincian atribut yang belum dilakukan Analisis Data seperti tabel 2. Atribut berikut:

Tabel 2. Atribut Data Sebelum Analisis Data

No	Nama	Jenis Data
1	NPSN	Karakter
2	Nama Lembaga	Karakter
3	Standar 1	Karakter
4	Standar 2	Karakter
5	Standar 3	Karakter
6	Standar 4	Karakter
7	Standar 5	Karakter
8	Standar 6	Karakter
9	Standar 7	Karakter
10	Standar 8	Karakter
11	Peringkat Akreditasi	Kategorikal

Dari beberapa penelitian yang telah dilakukan menyatakan bahwa melakukan Analisis data perlu dilakukan untuk mendapatkan dataset dengan kualitas yang baik. Untuk mendapatkan data dengan kualitas baik, beberapa teknis yang dapat dilakukan antara lain *validation, integration and transformation, size reduction/discretization*. Maka pada penelitian ini, untuk Analisis data dilakukan untuk mendapatkan data dengan kualitas baik. Analisis data yang penulis gunakan diantaranya:

- a. Memperbaiki keterangan data pada atribut peringkat akreditasi, yang awalnya peringkat akreditasi terdiri dari A (Sangat Baik), B (Baik), C (Cukup), dan TT (Tidak Terakreditasi), dan berubah menjadi Terakreditasi dan Tidak Terakreditasi. Pada tahap ini dipakai untuk menghindarkan adanya *missing value* dalam dataset.
- b. Reduksi data dilakukan untuk mendapatkan dataset dengan *record* dan jumlah atribut yang bersifat hanya Informasinya saja. Contohnya pada atribut “NPSN” dan “Nama Lembaga” tidak digunakan pada proses mining karena tidak relevan. Jadi atribut yang dipakai pada proses mining hanya atribut yang bersifat informasi, yaitu “Standar 1”, “Standar 2”, “Standar 3”, “Standar 4”, “Standar 5”, “Standar 6”, “Standar 7”, “Standar 8”, dan “Peringkat Akreditasi”. Setelah melakukan proses analisis data, maka dataset yang digunakan pada proses mining adalah 100 data satuan akreditasi dengan 9 atribut yang telah dinormalisasi dan *missing value* tidak ada pada dataset tersebut. Adapun atribut yang digunakan pada proses data mining seperti pada tabel 3:

Tabel 3. Atribut Data Sesudah Analisis Data

No	Nama	Jenis Data
1	Standar 1	Karakter
2	Standar 2	Karakter
3	Standar 3	Karakter
4	Standar 4	Karakter
5	Standar 5	Karakter
6	Standar 6	Karakter
7	Standar 7	Karakter



No	Nama	Jenis Data
8	Standar 8	Karakter
9	Peringkat Akreditasi	Kategorikal

2.6 Algoritma C4.5

Algoritma C4.5 adalah metode yang bisa diterapkan untuk menyelesaikan masalah klasifikasi data dengan atribut kategori [13]. Sedangkan (Anam & Santoso 2018) berpendapat bahwa Algoritma C4.5 diterapkan guna membentuk sebuah pohon keputusan yang mempresentasikan aturan dalam klasifikasi [14]. C4.5 adalah ekstensi dari algoritma *decision-tree ID3*. Algoritma ID3/C4.5 ini secara rekursif membuat sebuah decision tree berdasarkan training data yang telah disiapkan [15]. Elemen penting dalam Algoritma C4.5 antara lain Entropy dan Gain. Tahapan dalam membangun pohon keputusan diantaranya:

1. Memilih atribut untuk dijadikan node/akar
2. Membuat cabang dari setiap nilai
3. Membagi kasus pada setiap cabang
4. Mengulang proses tiap cabang hingga seluruh kasus pada cabang berada pada kelas yang sama.

Pemilihan atribut sebagai node/akar yakni berdasarkan nilai Gain tertinggi dari semua atribut. Perhitungan metode algoritma C4.5 menggunakan konsep information gain atau entropy reduction untuk memilih pembagian yang optimal. Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5 yaitu [16]:

1. Mempersiapkan data training, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu
2. Menentukan akar dari pohon dengan menghitung nilai gain yang tertinggi dari masing-masing atribut atau berdasarkan nilai indeks entropi terendah.

$$\text{Gaint (S, A)} = \text{Entropy (S)} \sum_{i=1}^n \frac{|S_i|}{|S|} * \quad (1)$$

Keterangan:

- S : Himpunan Kasus
- A : Atribut
- n : Jumlah Partisi Atribut A
- | Si | : Jumlah Kasus Pada Partisi Ke – i
- | S | : Jumlah Kasus Dalam S

Nilai Entropy dapat dihitung menggunakan persamaan berikut:

$$\text{Entropy (S)} = \sum_{i=1}^n P_i * \text{Log}_2 P_i \quad (2)$$

Keterangan :

- S : Himpunan Kasus
- A : Fitur
- n : Jumlah Partisi S
- Pi : Proporsi Dari Si Terhadap S

2.8 Naïve Bayes

Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistic yang dikemukakan oleh ilmuan Inggris Thomas Bayes. Naïve Bayes telah banyak digunakan dalam prediksi di berbagai hal. Naïve Bayes telah digunakan untuk prediksi di bidang perbankan/keuangan, bidang sosial kemasyarakatan, serta bidang pendidikan[17]. Naïve bayes digunakan untuk Algoritma perhitungan untuk menyelesaikan masalah prediksi. Metode ini menggunakan teorema Bayes, yang bekerja berdasarkan probabilitas sederhana dinyatakan persamaan berikut:

$$P (H|E) = \frac{P(E|H)P(H)}{P(E)} \quad (3)$$

Keterangan:

- E : Bukti
- H : Hipotensi
- P(H|E) : Hipotesis H benar untuk bukti E
- P(E|H) : Kemungkinan sebuah bukti E terjadi akan memengaruhi hipotesis H atau kata lain kemungkinan bahwa bukti E benar untuk hipotesis H
- P(H) : Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun
- P(E) : Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/bukti yang lain.

Maksud dari aturan bayes yakni berdasarkan bukti-bukti (E) yang diamati maka hasil hipotesis (H) dapat diprediksi.

2.9 Evaluasi Metode Klasifikasi



Evaluasi ini bertujuan untuk menganalisis perbandingan pada metode klasifikasi yang digunakan. Pada penelitian ini, hasil implementasi metode C4.5 akan dibandingkan dengan Naïve Bayes. Untuk memvalidasi nilai akurasi model ini dibagun menggunakan metode *Confusion Matrix*.

a. Confusion Matrix

Tujuan menganalisis menggunakan metode *Confusion Matrix* adalah untuk menganalisis kualitas kinerja model klasifikasi dalam mengetahui variabel dari seluruh kelas. *Confusion Matrix* berisi mengenai informasi kelas sebenarnya dan kelas prediksi dari suatu proses klasifikasi. Misalnya dataset terbagi menjadi kelas A dan B, maka kelas A dikategorikan menjadi variabel positif dan kelas B dikategorikan menjadi variabel negatif. Nilai *accuracy*, *reccal* dan *preccision* didapat dari hasil evaluasi menggunakan *Confusion Matrix*. Contoh *Confusion Matrix* di lihat pada gambar 2:

		Ya	Tidak	Jumlah
Kelas Aktual	Ya	TP	FN	P
	Tidak	FP	TN	N
	Jumlah	P	N	P + N

Gambar 2. Confusion Matrix

Berikut perhitungan nilai akurasi, *preccision* dan *reccal* dinyatakan dalam persamaan:

$$\text{Accuracy} = \frac{TP+TN}{P+N} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Reccal} = \frac{TP}{P} \quad (6)$$

Keterangan:

TP (True Positive) : Jumlah variabel positif yang dilabeli dengan benar oleh *classifier*, sebagai contoh variabel dengan label status kelulusan = tepat waktu

TN (True Negative) : Jumlah variabel negatif yang dilabeli dengan benar oleh *classifier*

FP (False Positive) : Jumlah variabel negatif yang salah dilabeli oleh *classifier*

FN (False Negative) : Jumlah variabel positif yang salah dilabeli oleh *classifier*

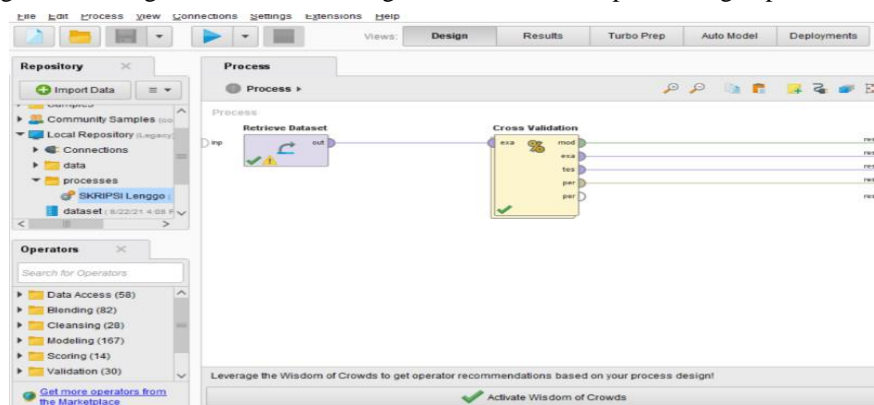
P : Jumlah sampel positif

3. HASIL DAN PEMBAHASAN

3.1 Pengolahan Data Pada Sistem

Data yang digunakan pada proses training dan testing telah melewati proses Analisis data, sehingga data telah siap diolah. Dataset yang dipakai dalam proses training ataupun testing sejumlah 100 record. Sedangkan untuk atribut yang digunakan dalam proses training dan testing terdiri dari atribut Standar 1 Standar Tingkat Pencapaian Perkembangan Anak, Standar 2 Standar Isi, Standar 3 Standar Proses, Standar 4 Standar Pendidikan dan Tenaga Kependidikan, Standar 5 Standar Sarana dan Prasarana, Standar 6 Standar Pengelolaan, Standar 7 Standar Pembiayaan, Standar 8 Standar Penilaian, Peringkat Akreditasi. Proses ini diterapkan pada *tool* RapidMiner 9.9, dan dengan memakai metode perhitungan yaitu Algoritma C4.5 dan Naïve bayes. Selanjutnya hasil pengujian dapat dilihat dari Confussion Matrik seperti Gambar 4 dan Gambar 6.

Tahap selanjutnya dilakukan pembuktian hasil penelitian menggunakan software RapidMiner Studio, atribut yang digunakan sebagai *Class* adalah Peringkat Akreditasi. Tahap modeling dapat dilihat pada gambar 3



Gambar 3. Proses Metode Algoritma C4.5



3.2 Implementasi Metode Algoritma C4.5

Pada hasil implementasi dengan menggunakan metode Algoritma C4.5 didapatkan nilai akurasi sebesar 99,00% yang telah dievaluasi menggunakan *Confusion Matrix*. Pada Gambar 4 menunjukkan detail hasil yang didapatkan pada proses mining yang dikerjakan dengan menggunakan Algoritma C4.5:

Table View Plot View

accuracy: 99.00% +/- 3.16% (micro average: 99.00%)

	true Terakreditasi	true Tidak Terakreditasi	class precision
pred. Terakreditasi	99	1	99.00%
pred. Tidak Terakreditasi	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 4. Nilai Akurasi Metode Algoritma C4.5

Pada gambar 3, menampilkan hasil perhitungan akurasi dataset dengan menggunakan metode C4.5 berdasarkan *Confussion matrix*. Pada gambar tersebut dapat dilihat nilai *Accuracy* rata rata sebesar 99,00% nilai *Reccal* Terakreditasi sebesar 100,00%, Nilai *Reccal* Tidak Terakreditasi sebesar 0,00%. Nilai *Preccision* Terakreditasi sebesar 99,00%, dan nilai *Preccision* Tidak Terakreditasi sebesar 0,00%. Dari 100 dataset terdapat 99 data yang sesuai dengan prediksi yaitu “Terakreditasi”, dan 1 diprediksi “Terakreditasi” ternyata “Tidak Terakreditasi”. Dan sebanyak 0 data yang diprediksi “Tidak Terakreditasi” ternyata termasuk klasifikasi “Terakreditasi” dan sebanyak 0 data prediksi yaitu “Tidak Terakreditasi”.

Dari hasil implementasi Algoritma C4.5 menggunakan *tool* RapidMiner, maka terbentuklah pohon keputusan berdasarkan nilai gain tertinggi, dapat dilihat di gambar 5:



Gambar 5. Pohon Keputusan Berdasarkan Informasi Gain

Pada gambar 4, dapat dilihat bahwa variabel atau kriteria yang berpengaruh dalam prediksi Terakreditasi pada BAN PAUD dan PNF Provinsi DKI Jakarta adalah di Standar 4 Standar Pendidikan dan Tenaga Pendidikan. Variabel standar 4 adalah nilai kumulatif terakhir yang diambil selama proses akreditasi pada satuan pendidikan PAUD. Data tersebut merupakan data paling dekat yang menggambarkan data prediksi terakreditasi satuan pendidikan PAUD dibandingkan variabel variabel lain.

3.3 Implementasi Metode Naïve Bayes

Selanjutnya pada hasil implementasi dengan menggunakan metode Naïve Bayes didapatkan nilai akurasi sebesar 99,00% yang telah dievaluasi menggunakan *Confusion Matrix*. Pada Gambar 6 menunjukkan detail hasil yang didapatkan pada proses mining yang dikerjakan dengan menggunakan Naïve Bayes:

Table View Plot View

accuracy: 99.00% +/- 3.16% (micro average: 99.00%)

	true Terakreditasi	true Tidak Terakreditasi	class precision
pred. Terakreditasi	99	1	99.00%
pred. Tidak Terakreditasi	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 6. Nilai Akurasi Metode Naïve Bayes

Pada Gambar 4, menampilkan hasil perhitungan akurasi dataset dengan menggunakan metode Naïve Bayes berdasarkan *Confussion matrix*. Pada gambar tersebut dapat dilihat nilai *Accuracy* rata rata sebesar 99,00% nilai *Reccal* Terakreditasi sebesar 100,00%, Nilai *Reccal* Tidak Terakreditasi sebesar 0,00%. Nilai *Preccision* Terakreditasi sebesar 99,00%, dan nilai *Preccision* Tidak Terakreditasi sebesar 0,00%. Dari 100 dataset terdapat 99 data yang sesuai dengan prediksi yaitu “Terakreditasi”, dan 1 diprediksi “Terakreditasi” ternyata “Tidak Terakreditasi”. Dan sebanyak 0 data yang diprediksi “Tidak Terakreditasi” ternyata termasuk klasifikasi “Terakreditasi” dan sebanyak 0 data prediksi yaitu “Tidak Terakreditasi”.



3.4 Perbandingan Hasil Akurasi Metode Algoritma C4.5 dan Naïve Bayes

Hasil dari Implementasi yang telah dilakukan, perbandingan tingkat akurasi antara metode Algoritma C4.5 dan Naïve Bayes.

Tabel 4. Perbandingan Hasil Akurasi Metode Algoritma C4.5 dan Naïve Bayes

No	Metode	Nilai Akurasi
1	Algoritma C4.5	99,00%
2	Naïve Bayes	99,00%

Dari hasil tabel 4, prediksi Terakreditasi satuan pendidikan PAUD menggunakan metode C4.5 memiliki nilai akurasi sama dengan nilai akurasi metode Naïve Bayes, yaitu 99,00%. Akhir pada pengujian ini, yaitu hasil prediksi Terakreditasi satuan pendidikan PAUD pada BAN PAUD dan PNF Provinsi DKI Jakarta berdasarkan dataset yang telah diimplementasikan dengan metode Algoritma C4.5 dan Naïve Bayes sama-sama menunjukkan nilai Accuracy 99,00%. Dari pohon keputusan hasil implementasi Algoritma C4.5 disimpulkan bahwa variabel atau kriteria yang berpengaruh dalam prediksi Terakreditasi satuan pendidikan PAUD di BAN PAUD dan PNF Provinsi DKI Jakarta adalah Standar 4 Standar Pendidikan dan Tenaga Pendidikan.

4. KESIMPULAN

Dari permasalahan di Badan Akreditasi Nasional Pendidikan Anak Usia Dini dan Pendidikan Nonformal (BAN PAUD dan PNF Provinsi DKI Jakarta) untuk kesimpulan dari penelitian ini bahwa prediksi Nilai Akreditasi pada Lembaga Satuan PAUD berdasarkan data set yang diimplementasikan dengan metode Naïve Bayes menunjukkan nilai Accuracy sama dengan nilai Accuracy pada metode Algoritma C4.5 yaitu dengan nilai Accuracy 99,00%. Karena hasil nilai Accuracy kedua metode sama, maka Metode Algoritma C4.5 dan Naïve Bayes sama-sama direkomendasikan untuk digunakan dalam menyelesaikan masalah prediksi menentukan nilai akreditasi pada lembaga satuan PAUD pada Badan Akreditasi Nasional Pendidikan Anak Usia Dini dan Pendidikan Nonformal Provinsi DKI Jakarta (BAN PAUD dan PNF Provinsi DKI Jakarta). Dari pohon keputusan hasil implementasi Algoritma C4.5, dapat disimpulkan bahwa variabel atau kriteria yang berpengaruh dalam prediksi nilai akreditasi Lembaga satuan PAUD di BAN PAUD dan PNF Provinsi DKI Jakarta adalah standar 4 Standar Pendidikan dan Tenaga Pendidikan.

UCAPAN TERIMAKASIH

Ucapan Terimakasih peneliti sampaikan kepada Bpk. Dadang Iskandar Mulyana, selaku Dosen Pembimbing dan Kepala Program Studi Teknik Informatika di Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, yang telah memberikan saran, masukan dan arahan sehingga penelitian ini dapat diselesaikan dengan baik. Selain itu terimakasih kepada Bpk. Hidayat Darwis selaku Sekretaris dan Amirul Hilman selaku Sekretariat Bagian SIMA di BAN PAUD dan PNF Provinsi DKI Jakarta yang telah memberikan referensi data untuk bisa melakukan penelitian ini. Serta terimakasih kepada Orang tua dan serta Teman teman seperjuangan di STIKOM CKI yang senantiasa memberikan semangat dan masukan kepada peneliti dalam menyelesaikan penelitian ini.

REFERENCES

- [1] D. P. Mulya, "Analisa Dan Implementasi Association Rule Dengan Algoritma Fp-Growth Dalam Seleksi Pembelian Tanah Liat (Studi Kasus Di Pt. Anveve Ismi Berjaya)," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 1, no. 1, pp. 47–57, 2019.
- [2] E. Etriyanti, D. Syamsuar, and N. Kunang, "Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa," *Telematika*, vol. 13, no. 1, pp. 56–67, 2020, doi: 10.35671/telematika.v13i1.881.
- [3] H. Dhika, F. Destiwati, and A. Fitriansyah, "Implementasi Algoritma C4. 5 terhadap Kepuasan Pelanggan," pp. 80–86, 2018, doi: 10.31227/osf.io/fgc7a.
- [4] A. Amrin, "Perbandingan Metode Neural Network Model Radial Basis Function Dan Multilayer Perceptron Untuk Analisa Risiko Kredit Mobil," *Paradig. Komput. dan Inform.*, vol. 20, no. 1, pp. 31–38, 2018.
- [5] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Deepublish, 2020.
- [6] D. Sianto and E. Mulyanto, "Perbandingan K-Nearest Neighbor Dan Naive Bayes Untuk Klasifikasi Tanah Layak Tanam Pohon Jati," *Techno. Com*, vol. 15, no. 3, pp. 241–245, 2016.
- [7] Septiani, "Komparasi Metode Klasifikasi Data Mining Algoritma C4. 5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis," *J. Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 76–84, 2017.
- [8] L. Setiyani and M. W. D. A. S. P. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes: Systematic Review," *Fakt. Exacta*, vol. 13, no. 1, pp. 38–47, 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [9] W. Supriyanti, K. Kusriani, and A. Amborowati, "Perbandingan Kinerja Algoritma C4. 5 dan Naïve Bayes Untuk Ketepatan Pemilihan Konsentrasi Mahasiswa," *J. Inf. J. Penelit. dan Pengabd. Masy.*, vol. 1, no. 3, pp. 61–67, 2016.
- [10] W. Muslehatin and Ibnu, "Penerapan Naïve Bayes Classification untuk Klasifikasi Tingkat Kemungkinan Obesitas



- Mahasiswa Sistem Informasi UIN Suska Riau,” in *Seminar Nasional Teknologi Informasi Komunikasi dan Industri*, 2017, pp. 250–256.
- [11] I. et al Brilian Rahmat C.T.I, “Implemetasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan,” in *Prosiding Seminar Nasional Riset Kuantitatif Terapan*, 2017, pp. 58–62.
- [12] A. P. Fadillah, “Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ),” *J. Tek. Inform. dan Sist. Inf.*, vol. 1, no. 3, pp. 260–270, 2015, doi: 10.28932/jutisi.v1i3.406.
- [13] S. A. Prakoso and E. T. Tutik, “KOMPARASI ALGORITMA C4. 5 DENGAN NAÏVE BAYES UNTUK KLASIFIKASI KELULUSAN MAHASISWA TEPAT WAKTU DI PTS ‘KZX,’” *KOMPUTAKI*, vol. 3, no. 1, 2017.
- [14] C. Anam and Santoso, “Perbandingan Kinerja Algoritma C4. 5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa,” *ENERGY*, vol. 8, no. 1, pp. 13–19, 2018.
- [15] Deteksi et al, “Eksperimen Naïve Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia Naïve Bayes’s Experiment On Hoax News Detection In Indonesian Language,” *vol*, vol. 23, pp. 1–15, 2019.
- [16] H. Amalia and E. Evicienna, “PENENTUAN PROSES PERSALINAN IBU MELAHIRKAN MENGGUNAKAN ALGORITMA C4. 5,” in *PROSIDING SEMINAR NASIONAL CENDEKIAWAN*, 2017, pp. 101–107.
- [17] B. Bahar and D. W. Pratama, “Penerapan Teorema Bayes Dalam Sistem Pakar Untuk Konsultasi Siswa Bermasalah,” *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 1529–1540, 2018.