

Klasifikasi Kesiapan Anak Taman Kanak-Kanak Masuk Sekolah Dasar Menggunakan Metode Naive Bayes

Yuli Praptomo Pamungkas Hari Sungkowo¹, Deni Ambarwati²

¹Sistem Informasi, STMIK El Rahma Yogyakarta, 55153, Indonesia

²Informatika, STMIK El Rahma Yogyakarta, 55153, Indonesia
y.praptomo@gmail.com, deniambar53@gmail.com

Abstrak

Sesuai dengan peraturan pemerintah bahwa calon peserta didik baru kelas 1 (satu) SD harus memenuhi persyaratan usia 7 (tujuh) tahun; atau paling rendah 6 (enam) tahun pada tanggal 1 Juli tahun berjalan. Selain faktor usia, terdapat aspek lain yang memengaruhi kesiapan anak TK masuk SD yaitu aspek kognitif, motorik kasar, motorik halus, bahasa, dan sosial emosional. Namun banyak peserta didik yang memasuki sekolah dasar meskipun masih ada beberapa aspek yang belum bisa dipenuhi. Oleh sebab itu, dibutuhkan klasifikasi untuk mengetahui apakah peserta didik tersebut memiliki kesiapan untuk memasuki jenjang sekolah dasar atau tidak. Klasifikasi kesiapan anak untuk memasuki sekolah dasar dalam penelitian ini menggunakan metode naive bayes, dengan atribut yang digunakan antara lain usia, kemampuan membaca, menulis, berhitung, dan kemandirian. Hasil dari penelitian ini adalah untuk mengetahui kesiapan peserta didik guna memasuki sekolah dasar agar nantinya peserta didik maupun pengajar nantinya tidak kewalahan dalam mengikuti kegiatan belajar mengajar. Berdasarkan hasil pengujian, diperoleh nilai akurasi sebesar 94,23%. Dari hasil pengujian tersebut dapat disimpulkan bahwa metode Naive Bayes dapat digunakan untuk mengklasifikasi kesiapan siswa memasuki sekolah dasar.

Kata kunci: klasifikasi, sekolah dasar, data mining, naive bayes.

Abstract

In accordance with government regulations that prospective new students in grade 1 (one) elementary school must meet the age requirements of 7 (seven) years; or at least 6 (six) years on July 1 of the current year. In addition to the age factor, there are other aspects that affect the readiness of kindergarten children to enter elementary school, namely cognitive, gross motor, fine motor, language, and social emotional aspects. However, many students enter elementary school although there are still some aspects that cannot be fulfilled. Therefore, a classification is needed to find out whether these students have readiness to enter the elementary school level or not. The classification of children's readiness to enter elementary school in this study uses the naive Bayes method, with attributes used include age, ability to read, write, count, and independence. The results of this study are to determine the readiness of students to enter elementary school so that later students and teachers will not be overwhelmed in participating in teaching and learning activities. Based on the test results, obtained an accuracy value of 94.23%. From the test results, it can be concluded that the Naive Bayes method can be used to classify students' readiness to enter elementary school.

Keywords: Classification, Elementary school, Data mining, Nave Bayes.

1. PENDAHULUAN

Taman kanak-kanak (TK) merupakan pendidikan formal untuk anak usia dini dengan rentang usia empat sampai dengan enam tahun. Pada taman kanak-kanak dibedakan menjadi dua kelompok yaitu TK A dan TK B. TK A berisi anak-anak mulai dari usia empat tahun, sedangkan TK B berisi anak dengan rentang usia lima sampai dengan enam tahun.

Masa pendidikan pada TK A dan TK B masing-masing adalah satu tahun. Setelah siswa menempuh pendidikan di TK, selanjutnya anak-anak akan melanjutkan ke jenjang pendidikan berikutnya yaitu sekolah dasar (SD). Sesuai dengan Pasal 4 Ayat 1 Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 1 Tahun 2021 Tentang Penerimaan Peserta Didik Baru pada Taman Kanak-kanak,

Sekolah Dasar, Sekolah Menengah Pertama, Sekolah Menengah Atas, dan Sekolah Menengah Kejuruan bahwa calon peserta didik baru kelas 1 (satu) SD harus memenuhi persyaratan usia 7 (tujuh) tahun; atau paling rendah 6 (enam) tahun pada tanggal 1 Juli tahun berjalan. Selain faktor usia, terdapat aspek lain yang memengaruhi kesiapan anak TK masuk SD yaitu aspek kognitif, motorik kasar, motorik halus, bahasa, dan sosial emosional.

Banyak penelitian sebelumnya telah dilakukan, di antaranya Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa (Jananto 2013), Penerapan Data Mining Pada Penjualan Barang Menggunakan Metode Naive Bayes Classifier untuk Optimasi Strategi Pemasaran (Nurdiawan and Salim 2018), Sistem Informasi Peramalan Harga Pangan Dengan Menggunakan Metode Naive Bayes Di Kota Makassar (Eden, Asrul, and Zuhriyah 2018), Penerapan Data Mining Untuk Memprediksi Kesiapan dan Kematangan Anak Masuk Sekolah Dasar Menggunakan Metode Naive Bayes (Hartinah 2020), Sistem Pendukung Keputusan Bantuan Sosial Dengan Menggunakan Metode Naive Bayes (Anam 2021).

Untuk mengetahui kesiapan anak untuk masuk SD tentunya dibutuhkan cara untuk mengukur kesiapan tersebut. Beberapa cara untuk mendeteksi kesiapan anak masuk SD adalah dengan menggunakan *Nijmeegse Schoolbekwaamheid Test* (NST) dan *Children Personal Matrix Test* (CPM).

Hal yang mendasari dilakukannya penelitian ini adalah untuk mengetahui bagaimana kesiapan masuk SD pada siswa TK dengan melihat aspek-aspek tersebut dengan melakukan teknik data mining dengan metode naive bayes. Teknik data mining yang digunakan pada penelitian ini adalah klasifikasi.

Penelitian ini diharapkan dapat memberikan tambahan informasi kesiapan masuk SD pada TK ABA Gunturgeni, Poncosari, Srandakan, Bantul dan sebagai langkah antisipasi bagi pengajar dan orangtua kedepannya.

2. METODE

Klasifikasi

Berdasarkan (Penyusun 2008), definisi dari kata klasifikasi adalah penyusunan bersistem dalam kelompok atau golongan menurut kaidah atau

standar yang ditetapkan. Klasifikasi menurut (Werdingisih, I., Nuqoba, B. 2020) adalah penggalian data yang menetapkan item dalam koleksi untuk menargetkan kelas tertentu. Klasifikasi dimulai dengan mengumpulkan data di mana data tersebut terdapat kelas yang sudah diketahui. Sebagai contoh, klasifikasi yang untuk mengidentifikasi peringkat kredit, berupa pemohon pinjaman sebagai risiko kredit rendah, sedang, atau tinggi. Atribut yang digunakan dapat berupa data pemohon yang terdiri dari historis kredit pemohon, riwayat pekerjaan pemohon, kepemilikan atau sewa rumah pemohon, tahun tinggal pemohon, jumlah dan jenis investasi pemohon, dan sebagainya. Peringkat kredit akan menjadi target/kelas dan data setiap pemohon menjadi kasusnya.

(Yendrizal 2022) menyatakan bahwa klasifikasi adalah tugas yang sangat sering terjadi dalam kehidupan sehari-hari. Dasarnya melibatkan membagi objek sehingga masing-masing ditugaskan ke salah satu nomor kategori saling lengkap dan eksklusif hanya berarti bahwa setiap objek harus ditugaskan untuk tepat satu kelas, yaitu tidak pernah lebih dari satu dan tidak pernah ada kelas sama sekali. Hasil klasifikasi dapat disajikan ke dalam berbagai bentuk yaitu aturan klasifikasi (*If-Then*), *decision trees* atau pohon keputusan, formula matematis dan *neural networks* atau jaringan syaraf tiruan.

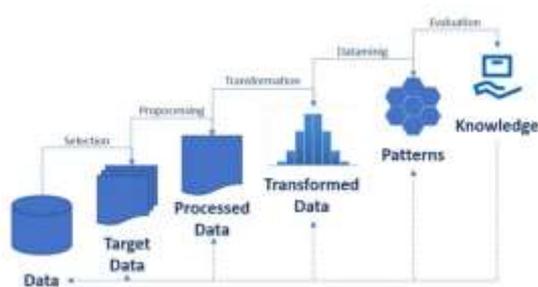
Secara umum proses klasifikasi dapat dilakukan dalam dua tahap, yaitu proses belajar dari data pelatihan dan klasifikasi kasus baru. Pada proses belajar, algoritma klasifikasi mengolah data pelatihan untuk menghasilkan sebuah model. Setelah model diuji dan diterima, pada tahap klasifikasi, model tersebut digunakan untuk memprediksi kelas dari kasus baru untuk membantu proses pengambilan keputusan.

Teknik klasifikasi dalam data mining dikelompokkan ke dalam teknik pohon keputusan, Bayesian (*Naive Bayesian* dan *Bayesian Belief Networks*), Jaringan Saraf Tiruan (*Backpropagation*), teknik yang berbasis konsep dari penambangan aturan-aturan asosiasi dan teknik lain (*K-Nearest Neighbor*, algoritma genetik, teknik dengan pendekatan himpunan *rough* dan *fuzzy*). Setiap teknik memiliki kelebihan dan kekurangannya sendiri. Data dengan profil tertentu mungkin paling optimal jika diklasifikasi dengan teknik tertentu atau dengan kata lain. Profil data tertentu dapat mendukung termanfaatkannya kelebihan dari teknik ini. Dengan demikian, sebaiknya DBMS

menyediakan semua teknik-teknik klasifikasi bagi pengguna, agar pengguna dapat memilih sesuai dengan kebutuhan.

Data Mining

Menurut (Bustami 2014) data mining adalah teknik yang memanfaatkan data dalam jumlah yang besar untuk memperoleh informasi berharga yang sebelumnya tidak diketahui dan dapat dimanfaatkan untuk pengambilan keputusan penting. Sedangkan definisi sederhana data mining menurut (Siregar AM 2017) adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di database yang besar. Dalam jurnal ilmiah, data mining juga dikenal dengan nama *Knowledge Discovery in Database* (KDD).



Gambar 1 Tahapan dalam Data Mining

Tahapan yang dilakukan pada proses data mining diawali dari seleksi data dari data sumber ke data target, tahap *preprocessing* untuk memperbaiki kualitas data, transformasi, data mining serta tahap interpretasi dan evaluasi yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik. Secara detail dijelaskan sebagai berikut (Ziegel, E. R., Fayyad, U. M., Piatetski-Shapiro, G., Smyth, P., & Uthurusamy 1998).

1. **Data selection.** Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.
2. **Pre-processing** atau *cleaning*. Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.
3. **Coding** adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut

sesuai untuk proses data mining. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. **Data mining.** Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
5. **Interpretation** atau *evaluation*. Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

Data mining didefinisikan sebagai satu set teknik yang digunakan secara otomatis untuk mengeksplorasi secara menyeluruh dan membawa ke permukaan relasi-relasi yang kompleks pada set data yang sangat besar. Set data yang dimaksud disini adalah set yang berbentuk tabulasi, seperti yang banyak diimplementasikan dalam teknologi manajemen basis data relasional. Akan tetapi, teknik-teknik data mining dapat juga diaplikasikan pada representasi data yang lain, seperti domain data spatial, berbasis teks, dan multimedia (citra).

Naive Bayes

Menurut (Saputro and Sari 2020), *Naive Bayes* merupakan algoritma yang digunakan untuk klasifikasi yang menggunakan teorema bayes dan berasumsi bahwa nilai antar variabel saling bebas (independen) pada suatu nilai output. Dalam hal ini, diasumsikan bahwa kehadiran atau ketiadaan dari suatu variabel tertentu tidak berhubungan dengan kehadiran atau ketiadaan dari variabel lainnya.

Menurut (Sanubari T, Prianto C 2020) terdapat beberapa kelebihan dari penggunaan metode algoritma Naive Bayes yaitu hanya memerlukan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena data yang akan dijadikan bahan penelitian yaitu berupa variabel independen, sehingga hanya varians dari suatu variabel dalam

sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians.

Tahapan proses algoritma *Naïve Bayes* adalah:

1. Menghitung jumlah kelas/label.
2. Menghitung jumlah kasus per kelas.
3. Kalikan semua variable kelas.
4. Bandingkan hasil per kelas.

Keblebihan dan kekurangan algoritma *Naïve Bayes* adalah sebagai berikut.

1. Mudah untuk dibuat.
2. Hasilnya yang lebih baik.

Kekurangan dari algoritma *Naïve Bayes* adalah asumsi independence antar atribut membuat akurasi berkurang (karena biasanya ada keterkaitan).

Persamaan Teorema Bayes

Menurut (Saputro and Sari 2020), *teorema bayes* dapat ditulis menggunakan persamaan berikut.

$$P(A|B) = \frac{P(B | A)P(A)}{P(B)} \quad (1)$$

Dimana:

"P(A|B)" = Probabilitas posterior dari A pada kondisi B (posterior probability).

"P(B|A)" = Probabilitas posterior dari B pada kondisi A (likelihood).

P(A) = Probabilitas prior dari A (class prior probability).

P(B) = Probabilitas prior dari B (predictor prior probability).

Proses untuk menghitung probabilitas kelas suatu data dimulai dengan menentukan likelihood berdasarkan dataset yang digunakan, menggunakan metode yang sesuai dengan bentuk dari data yang digunakan. *Likelihood* yang diperoleh akan dikalikan dengan probabilitas dari masing-masing kelas. Hasil dari proses tersebut akan digunakan sebagai acuan untuk mengklasifikasi data baru.

Cross Validation

Menurut (Menarianti 2015), setiap kelas pada kelompok data harus diwakili dalam proporsi yang tepat antara data training dan data testing. Data dibagi secara acak pada masing-masing kelas dengan perbandingan yang sama. Untuk mengurangi bias yang disebabkan oleh sampel tertentu, seluruh proses training dan testing diulangi beberapa kali dengan sampel yang berbeda. Tingkat kesalahan pada iterasi yang berbeda akan dihitung rata-ratanya untuk menghasilkan error rate secara keseluruhan. Model yang memberikan rata-rata kesalahan terkecil adalah model yang terbaik.

Confusion Matrix

Menurut (Saputro and Sari 2020) confusion matrix Merupakan tabel yang menggambarkan performa dari sebuah model atau algoritma secara spesifik. Setiap baris dari matriks tersebut, merepresentasikan kelas aktual dari data, dan setiap kolom merepresentasikan kelas prediksi dari data (atau sebaliknya). Matriks tersebut dijelaskan pada Tabel 1.

Tabel 1 *Confusion matrix*

	Predicted Negative	Predicted Positive
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

Keterangan:

TP = data aktual kelasnya positif dan model memprediksi positif.

TN = data aktual kelasnya negatif dan model memprediksi negatif.

FP = data aktual kelasnya negatif, namun model memprediksi positif.

FN = data aktual kelasnya positif, namun model memprediksi negatif.

Melalui 4 data tersebut, dapat diperoleh data data lain yang sangat berguna untuk mengukur performa sebuah model, diantaranya:

Accuracy = Total keseluruhan seberapa sering model benar mengklasifikasi. Formula *accuracy* dapat ditulis menggunakan persamaan 2.

$$Accuracy = \frac{TP+TN}{Total\ Data\ Set} \quad (2)$$

Precision = Ketika model memprediksi positif, seberapa sering prediksi itu benar. Formula *precision* dapat ditulis menggunakan persamaan 3.

$$Precision = \frac{TP}{FP+TP} \quad (3)$$

Recall (Sensitivity / True Positive Rate) = Ketika kelas aktualnya positif, seberapa sering model memprediksi positif. Formula *recall* dapat ditulis menggunakan persamaan 4.

$$Recall = \frac{TP}{FN+TP} \quad (4)$$

F1-Score = Merupakan rata-rata harmonik dari *Precision* dan *Recall*. Formula *f1-score* dapat ditulis menggunakan persamaan 5

$$F1\ Score = 2 * \frac{precision * recall}{precision + recall} \quad (5)$$

3. HASIL DAN PEMBAHASAN

Dataset

Dalam menganalisa kesiapan dan kematangan anak memasuki sekolah dasar ada beberapa atribut dataset yang digunakan oleh peneliti, dataset yang digunakan terdapat pada Tabel 2.

Perhitungan *Naïve Bayes*

Dataset akan dibagi menjadi dua bagian yaitu data training dan data testing. Data training diperoleh dari pembagian sumber data yang berjumlah 52 data, dimana dua per tiga dari sumber data yaitu 34 data dialokasikan untuk data training dan 18 data sisanya sebagai data testing. Pada tabel 3 merupakan sampel data training yang berjumlah 34 data.

Berdasarkan tabel 4 menjelaskan bahwa siswa yang telah memiliki kesiapan untuk memasuki sekolah dasar terdapat 29 siswa, sedangkan 23 siswa lainnya belum siap memasuki sekolah dasar. Tabel 5 menjelaskan bahwa 19 siswa laki-laki dan 10 siswa perempuan telah siap untuk memasuki sekolah dasar, sedangkan 15 siswa laki-laki dan 8 siswa perempuan belum siap memasuki sekolah dasar.

Berdasarkan tabel 6 terdapat 13 siswa dengan usia 5 tahun yang belum siap masuk sekolah dasar. Siswa usia 6 tahun terdapat 17 siswa yang sudah siap masuk sekolah dasar, namun 10 siswa belum siap masuk sekolah dasar. Terdapat 11 siswa yang berusia 7 tahun yang memiliki kesiapan masuk sekolah dasar dan 1 siswa berusia 8 tahun sudah siap memasuki sekolah dasar.

Tabel 7 menjelaskan bahwa siswa yang sudah memiliki kemampuan membaca dan sudah siap memasuki sekolah dasar terdapat 24 siswa, sedangkan siswa yang dapat membaca namun belum siap masuk sekolah dasar terdapat 5 siswa, dan siswa yang belum memiliki kemampuan membaca namun sudah siap memasuki sekolah dasar terdapat 3 siswa. 20 siswa belum memiliki kemampuan membaca dan belum memiliki kesiapan untuk memasuki sekolah dasar.

Berdasarkan tabel 8, siswa yang sudah memiliki kemampuan menulis dan sudah siap memasuki sekolah dasar terdapat 28 siswa, namun 1 siswa belum bisa menulis tapi sudah siap untuk memasuki sekolah dasar. Terdapat 6 siswa yang memiliki kemampuan menulis namun belum siap untuk memasuki sekolah dasar. Sedangkan 17 siswa yang belum bisa menulis belum siap untuk memasuki sekolah dasar.

Tabel 2. Atribut dataset penelitian

Nama	Nama Siswa
Jenis Kelamin	Laki-laki dan Perempuan
Usia	5 Tahun, 6 Tahun, 7 Tahun, 8 Tahun
Baca	Ya dan Tidak
Tulis	Ya dan Tidak
Hitung	Ya dan Tidak
Kemandirian	Ya dan Tidak
Kesiapan	Ya dan Tidak

Tabel 4 Atribut kelas

Kelas	
Siap	29
Belum Siap	23

Tabel 3 Sampel data training

1	2	3	4	5	6	7	8	9	10
2020009	Nabila Fahrani	P	A1	5,4	Y	T	Y	T	T
2020010	Rafa Azka R	L	A1	6,0	T	T	Y	Y	T
2020011	Rahmad Burhan	L	A1	6,0	T	T	Y	Y	T
2020012	Ringga Arya Bima	L	A1	5,2	T	T	Y	Y	T
2020013	Sofyan Wicaksono	L	A1	5,6	T	T	Y	Y	T
2020014	Syafira Anindya P	P	A1	5,4	T	T	Y	Y	T
2020015	Wahyu Rafanda	L	A1	5,6	T	T	Y	Y	T
2020016	Widayat	L	A1	5,2	Y	T	Y	Y	T
2020017	Yuan Raka H	L	A1	5,8	T	Y	Y	Y	Y

Tabel 5 Atribut jenis kelamin

Jenis Kelamin	Laki-laki	Perempuan
Siap	19	10
Belum Siap	15	8

Tabel 6 Atribut usia

USIA	5	6	7	8
Siap	0	17	11	1
Belum Siap	13	10	0	0

Tabel 7 Atribut kemampuan membaca

Membaca	Siap	Belum Siap
Ya	24	3
Tidak	5	20

Tabel 8 Atribut kemampuan menulis

Menulis	Siap	Belum Siap
Ya	28	6
Tidak	1	17

Tabel 9 Atribut kemampuan berhitung

Berhitung	Siap	Belum Siap
Ya	28	19
Tidak	1	4

Tabel 10 Atribut kemandirian

Mandiri	Siap	Belum Siap
Ya	28	16
Tidak	1	7

Tabel 11 Probabilitas kelas

P(Siap)	P(Belum Siap)
29/52	23/52
0,5577	0,4423

Berdasarkan tabel 9, siswa yang sudah memiliki kemampuan menghitung dan sudah siap memasuki sekolah dasar terdapat 28 siswa, namun terdapat 1 siswa yang belum bisa berhitung tetapi sudah siap memasuki sekolah dasar. Terdapat 19 siswa yang memiliki kemampuan menghitung namun belum siap untuk memasuki sekolah dasar. Sedangkan 4 siswa yang belum bisa menghitung belum siap untuk memasuki sekolah dasar.

Berdasarkan tabel 10, siswa yang sudah memiliki kemandirian dan sudah siap untuk memasuki sekolah dasar terdapat 28 siswa, sedangkan terdapat 1 siswa yang belum mandiri namun sudah siap untuk masuk sekolah dasar. Lalu terdapat 16 siswa yang memiliki kemandirian namun belum siap untuk memasuki sekolah dasar. Terdapat 7 siswa yang belum memiliki kemandirian dan belum siap memasuki sekolah dasar. Dari tabel atribut diatas dapat diperoleh data siswa yang siap atau belum siap memasuki sekolah dasar.

Menghitung Probabilitas Kelas

Tahap selanjutnya adalah menghitung probabilitas setiap kelas dengan menentukan kelas yaitu “Siap” dan “Belum Siap”. Untuk

menghitung probabilitas kelas adalah dengan mencari jumlah data yang sudah siap dan belum siap dari jumlah keseluruhan data training lalu membaginya dengan jumlah keseluruhan data. Hasil perhitungan dapat dilihat pada tabel 11.

Menghitung Probabilitas Setiap Atribut

Untuk menghitung probabilitas masing-masing atribut adalah dengan membandingkan atribut data testing dengan data training. Hitung jumlah atribut kelas yang sudah siap pada data training lalu bagi dengan probabilitas kelas siap dan juga probabilitas kelas belum siap, seperti terlihat pada tabel 12.

Menghitung Probabilitas Akhir Setiap Kelas

Probabilitas akhir diperoleh dari data training dan mengubahnya menjadi nilai yang sudah ditentukan. Probabilitas akhir dihitung dari pengalihan setiap atribut dengan nilai probabilitas. Kemudian bandingkan hasil akhir yang sudah ditentukan pada setiap kelas, jika hasil kelas “Siap” bernilai lebih tinggi daripada kelas “Tidak Siap”, maka hasil klasifikasi adalah “Siap”, begitu pula sebaliknya.

Perhitungan Menggunakan Naïve Bayes

Implementasi perhitungan Naïve Bayes yang akan diuji adalah data testing. Terdapat siswa bernama Intan Nurur Romadhoni dengan data seperti terlihat dalam tabel 13. Langkahnya yaitu: mencari nilai $P(C_i)$ tabel 14. Kemudian mencari nilai $P(X|C_i)$ tabel 15. Lalu mengalikan semua hasil atribut “Siap” dan “Belum Siap” lalu dikalikan dengan probabilitas masing-masing kelas. hasilnya seperti tabel 16.

Dapat dilihat pada tabel 16 bahwa hasil dari atribut “Siap” adalah 0,0030 sedangkan hasil dari atribut “Belum Siap” adalah 0,0003. Karena hasil dari atribut “Siap” lebih besar dari hasil atribut “Belum Siap” maka hasil klasifikasi kesiapan siswa atas nama Intan Nur Romadhoni adalah “Siap”.

Memasukkan Laplace Correction

Ketika probabilitas bernilai 0, maka atribut atau tupel yang terdapat nilai 0 tersebut ditambah dengan 1 data dan jumlah kelas ditambah dengan banyaknya atribut. Penerapan *laplace correction* dapat dilihat pada tabel 17 dan tabel 18.

Tabel 14. Nilai $P(C_i)$

$P(C_i)$	
P(Siap)	P(Belum Siap)
29/52	23/52
0,5577	0,4423

Tabel 12 Probabilitas Kelas

		Siap	Belum Siap
Jenis Kelamin	Laki-laki	19/29 0,6552	15/23 0,6522
	Perempuan	10/29 0,3448	8/23 0,3478
Usia	5 Tahun	0/29 0	13/23 0,5652
	6 Tahun	17/29 0,5862	10/23 0,4348
	7 Tahun	11/29 0,3793	0/23 0
	8 Tahun	1/29 0,0345	0/23 0
Membaca	Ya	24/29 0,8276	3/23 0,1304
	Tidak	5/29 0,1724	20/23 0,8696
Menulis	Ya	28/29 0,9655	6/23 0,2609
	Tidak	1/29 0,0345	17/23 0,7391
Berhitung	Ya	28/29 0,9655	19/23 0,8261
	Tidak	1/29 0,0345	4/23 0,1739
Mandiri	Ya	28/29 0,9655	16/23 0,6957
	Tidak	1/29 0,0345	7/23 0,3043

Tabel 13 Sampel data testing

1	2	3	4	5	6	7	8	9	10	11
202046	Intan Nurul	P	28/06/15	A2	6.0	Y	Y	T	Y	Y

Tabel 15. Nilai masing-masing atribut

		Siap	Belum Siap
Jenis Kelamin	Perempuan	29/52 0,5577	23/52 0,4423
Usia	6 Tahun	17/29 0,5862	10/23 0,4348
Membaca	Ya	24/29 0,8276	3/23 0,1304
Menulis	Ya	28/29 0,9655	6/23 0,2609
Berhitung	Tidak	1/29 0,0345	19/23 0,1739
Mandiri	Ya	28/29 0,9655	16/23 0,6957

Tabel 16. Perhitungan seluruh atribut

	Siap	Belum Siap
P(Ci)	0,5577	0,4423
P(X Ci)	0,3448	0,3478
	0,5862	0,4348
	0,8276	0,1304
	0,9655	0,2609
	0,0345	0,1739
	0,9655	0,6957
Hasil	0,0030	0,0003
Kesiapan		> Siap

Tabel 17. Sampel data testing

1	2	3	4	5	6	7	8	9	10	11
202052	Zahra Khoirunisa	P	21/07/15	A2	5,9	T	Y	Y	T	T

Tabel 18 Hasil probabilitas setiap atribut

Atribut	Siap	Atribut	Belum Siap
P	29/52	P	23/52
JK “P”	10/29	JK “P”	8/23
Usia “5 Tahun”	0/29	Usia “5 Tahun”	13/23
Baca “Tidak”	5/29	Baca “Tidak”	20/23
Tulis “Ya”	28/29	Tulis “Ya”	6/23
Hitung “Ya”	28/29	Hitung “Ya”	19/23
Mandiri “Tidak”	1/29	Mandiri “Tidak”	7/23

Pada Tabel 18, probabilitas “Usia 5 Tahun” pada kelas “Siap” adalah 0/29 atau hasilnya adalah 0. Jika suatu record bernilai 0, maka semua data yang terdapat probabilitas “Usia 5 Tahun” juga akan bernilai 0. Maka di sinilah fungsi dari laplace correction yaitu membangkitkan probabilitas yang bernilai 0. Cara kerja *Laplace Correction* adalah sebagai berikut.

1. Menambahkan 1 tupel atau record pura-pura dengan nilai kelas “Usia 5 Tahun” = “Siap”. Karena probabilitas yang memiliki nilai 0 adalah “Usia 5 Tahun” = “Siap”. Sehingga kita tidak perlu mengubah kelas “Belum Siap”.
2. Terdapat 6 atribut yaitu “Jenis Kelamin = Perempuan”, “Usia = 5 Tahun”, “Baca = Tidak”, “Tulis = Ya”, “Hitung = Ya”, “Mandiri = Tidak” sehingga tupel yang ditambahkan adalah 6.
3. Perlu diingat bahwa tupel yang dibuat adalah tupel pura-pura, sehingga hanya untuk menghindari probabilitas 0.

Penyelesaian Kasus

Perbandingan penggunaan *Laplace Correction* terdapat pada Tabel 19.

Pada Tabel 19. terdapat perubahan data dari sebelumnya yaitu telah ditambahkan 1 data dan 6 tupel.

Hasil Pengujian

Berikut merupakan proses klasifikasi yang digunakan untuk mengetahui kesiapan anak masuk sekolah dasar. Diperoleh nilai confusion matrix, accuracy, precision, recall, dan F1 score sebagai berikut.

a. Confusion Matrix

Berdasarkan Tabel 20 terdapat 28 data aktual kelasnya positif dan model memprediksi positif yang disebut sebagai TP atau true positive. Lalu terdapat 2 data aktual kelasnya negatif, namun model memprediksi positif yang disebut sebagai FP atau false positive. Sedangkan, 1 data aktual kelasnya positif, namun model memprediksi negatif, disebut sebagai FN atau false negative. Dan terdapat 21 data aktual kelasnya negatif dan model memprediksi negatif yang disebut sebagai TN atau true negative.

b. Accuracy

Total keseluruhan seberapa sering model benar mengklasifikasi disebut accuracy. Nilai accuracy dihitung menggunakan rumus berikut.

Tabel 19. Perbandingan penggunaan *Laplace Correction*

Sebelum <i>Laplace Correction</i>		Setelah <i>Laplace Correction</i>	
Sampel	Siap	Sampel	Belum Siap
P	29/52	P	29/52
JK “P”	10/29	JK “P”	11/35
Usia “5 Tahun”	0/29	Usia “5 Tahun”	1/35
Baca “Tidak”	5/29	Baca “Tidak”	6/35
Tulis “Ya”	28/29	Tulis “Ya”	29/35
Hitung “Ya”	28/29	Hitung “Ya”	29/35
Mandiri “Tidak”	1/29	Mandiri “Tidak”	2/35

Tabel 20. *Confusion matrix*

Data Aktual	Data Prediksi	
	Ya	Tidak
Ya	28	1
Tidak	2	21
Jumlah	52	

$$Accuracy = \frac{TP+TN}{Total\ Data\ Set}$$

$$= \frac{28+21}{52} = 94,23\%$$

$$F1\ Score = 2 * \frac{precision * recall}{precision + recall}$$

$$= 2 * \frac{93,33 * 96,55}{93,33 + 96,55} = 94,92\%$$

c. Precision

Ketika model memprediksi positif, seberapa sering prediksi itu benar. Nilai precision diperoleh dengan rumus berikut.

$$Precision = \frac{TP}{FP+TP}$$

$$= \frac{28}{2+28} = 93,33\%$$

d. Recall

Ketika kelas aktualnya positif, seberapa sering model memprediksi positif. Nilai recall dapat dihitung menggunakan rumus berikut.

$$Recall = \frac{TP}{FN+TP}$$

$$= \frac{28}{1+28} = 96,55\%$$

e. F1 Score

F1 score merupakan rata-rata harmonik dari precision dan recall. Untuk mengetahui nilai F1 score dapat menggunakan rumus berikut.

Semakin tinggi persentase akurasi, presisi, recall, dan F1 score, maka semakin baik kinerja sistem klasifikasi. Dengan demikian, dapat ditarik kesimpulan bahwa metode naïve bayes dapat digunakan untuk mengklasifikasi kesiapan anak taman kanak-kanak masuk sekolah dasar.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan untuk mengetahui klasifikasi kesiapan anak masuk sekolah dasar, maka dapat ditarik kesimpulan sebagai berikut.

Dengan adanya sistem klasifikasi ini maka dapat menghasilkan informasi apakah peserta didik sudah siap atau belum siap untuk memasuki sekolah dasar.

Metode Naive Bayes dapat digunakan untuk mengklasifikasi data kesiapan anak TK untuk memasuki sekolah dasar. Hal ini dibuktikan dengan hasil penelitian yang menghasilkan nilai akurasi sebesar 94,23%. Nilai tersebut terbilang sangat tinggi yang berarti sangat baik dalam mengklasifikasi kesiapan siswa memasuki sekolah dasar.

DAFTAR PUSTAKA

- Anam, Moh. Syaiful. 2021. "Social Assistance Decision Support System Using the Naive Bayes Method." *JTECS: Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem dan Komputer* 1(1): 85.
- Bustami. 2014. "Penerapan Algoritma Naive Bayes." *Jurnal Informatika* 8(1): 884-98.
- Eden, Billy, William Asrul, and Sitti Zuhriyah. 2018. "Sistem Informasi Peramalan Harga Pangan Dengan Menggunakan Metode Naive Bayes Di Kota Makassar." *e-Jurnal JUSITI (Jurnal Sistem Informasi dan Teknologi Informasi)* 7-2(2): 163-71.
- Hartinah, S. 2020. "Penerapan Data Mining Untuk Memprediksi Kesiapan Dan Kematangan Anak Masuk Sekolah Dasar Menggunakan Metode Naive Bayes. Universitas Pelita Bangsa." 14(1): 67-73.
- Jananto, Arief. 2013. "Algoritma Naive Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa." *Jurnal Teknologi Informasi DINAMIK* 18(1): 09-16.
- Menarianti, Ika. 2015. "Klasifikasi Data Mining Dalam Menentukan Pemberian Kredit Bagi Nasabah Koperasi." *Jurnal Ilmiah Teknosains* 1(1): 1-10. <http://e-jurnal.upgrismg.ac.id/index.php/JITEK/article/view/836>.
- Nurdiawan, Odi, and Noval Salim. 2018. "Penerapan Data Mining Pada Penjualan Barang Menggunakan Metode Metode Naive Bayes Classifier Untuk Optimasi Strategi Pemasaran." *Jurnal Teknologi Informasi dan Komunikasi* (April): 84-95.
- Penyusun, Tim. 2008. *Kamus Bahasa Indonesia*. Jakarta: Pusat Bahasa. <https://oldi.lipi.go.id/public/KamusIndonesia.pdf>.
- Sanubari T, Prianto C, Riza N. 2020. *Noviana Riza Odol (One Desa One Product Unggulan Online) Penerapan Metode Naive Bayes Pada Pengembangan Aplikasi e-Commerce Menggunakan Codeigniter*. 1st ed. Bandung: Kreatif. https://books.google.co.id/books?hl=id&lr=&id=s4j_DwAAQBAJ&oi=fnd&pg=PP7&ots=TVLcYQDI-1&sig=xg3I0iyGFoinmbTrlr8JMBmBT8g&redir_esc=y#v=onepage&q&f=false.
- Saputro, Irkham Widhi, and Bety Wulan Sari. 2020. "Uji Performa Algoritma Naive Bayes Untuk Prediksi Masa Studi Mahasiswa." *Creative Information Technology Journal* 6(1): 1.
- Siregar AM, Puspabhuana A. 2017. *Data Mining: Pengolahan Data Menjadi Informasi Dengan RapidMiner*. Jakarta: CV Kekata Group. https://books.google.co.id/books?id=rTImDwAAQBAJ&printsec=frontcover&hl=id&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false.
- Werdiningsih, I., Nuqoba, B., & Muhammadun. 2020. *Data Mining Menggunakan Android, Weka, Dan SPSS*. 1st ed. Surabaya: Airlangga University Press. <https://scholar.unair.ac.id/en/publications/data-mining-menggunakan-android-weka-dan-spss>.
- Yendrizar, S. 2022. *Monograf Algoritma C4.5 Pada Teknik Klasifikasi Penyusutan Volume Pupuk*. 1st ed. ed. Suardi M. Sumatera Barat.
- Ziegel, E. R., Fayyad, U. M., Piatetski-Shapiro, G., Smyth, P., & Uthurusamy, R. 1998. "Advances in Knowledge Discovery and Data Mining. In *Technometrics*." *Technometrics* 40(1): 83. <https://www.jstor.org/stable/1271410>.