

ABSTRAK

IMPELEMENTASI ALGORITMA *NAÏVE BAYES* UNTUK MENGUKUR TINGKAT MINAT BACA MASYARAKAT BENGKULU TERHADAP PROGRAM AKSARAYA SEMESTA

Nama : Vony Kurnia Safitri
NPM : 2255201031
Pembimbing : Muhammad Husni Rifqo, S.Kom., M.Kom.

Minat baca masyarakat merupakan salah satu aspek penting dalam meningkatkan kualitas sumber daya manusia. Tingkat minat baca yang berbeda pada masyarakat menjadi salah satu hal yang perlu diketahui sebagai bahan evaluasi terhadap pelaksanaan program literasi. Penelitian ini bertujuan untuk mengimplementasikan algoritma *Naïve Bayes* dalam mengklasifikasikan tingkat minat baca masyarakat Bengkulu terhadap Program Komunitas Literasi Aksaraya Semesta serta mengetahui kinerja model yang dihasilkan. Penelitian ini menggunakan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM) yang meliputi tahapan *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Data penelitian diperoleh melalui kuesioner digital berbasis *Google Form* dan setelah melalui proses *data preparation* diperoleh sebanyak 538 data yang digunakan dalam penelitian. Data dibagi menjadi 376 data training dan 162 data testing. Variabel penelitian terdiri atas sepuluh fitur yang berkaitan dengan aktivitas dan minat baca responden, sedangkan variabel target diklasifikasikan menjadi tiga kategori, yaitu Rendah, Sedang, dan Tinggi. Implementasi algoritma *Naïve Bayes* dilakukan menggunakan bahasa pemrograman *Python* pada platform *Google Colab* dengan library *Scikit-learn*. Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* menghasilkan *accuracy* sebesar 80,25%, *precision* sebesar 60,21%, *recall* sebesar 68,09%, dan *F1-score* sebesar 62,44%. Pengujian menggunakan 10-Fold Cross Validation menghasilkan rata-rata *accuracy* sebesar 80,64%. Hasil klasifikasi menunjukkan bahwa model memiliki performa terbaik dalam mengklasifikasikan kategori minat baca Tinggi, sedangkan kategori Sedang masih menjadi kelas yang paling sulit diklasifikasikan. Berdasarkan hasil tersebut, algoritma *Naïve Bayes* dapat digunakan untuk mengklasifikasikan tingkat minat baca masyarakat Bengkulu terhadap Program Komunitas Literasi Aksaraya Semesta dan memberikan gambaran mengenai distribusi tingkat minat baca berdasarkan data hasil kuesioner.

Kata Kunci: *Naïve Bayes*, CRISP-DM, klasifikasi, minat baca, literasi, Google Colab.

ABSTRACT

IMPLEMENTATION OF THE NAÏVE BAYES ALGORITHM FOR MEASURING THE READING INTEREST LEVEL OF THE BENGKULU COMMUNITY TOWARD THE AKSARAYA SEMESTA PROGRAM

Name : Vony Kurnia Safitri
NPM 2255201031
Advisor : Muhammad Husni Rifqo, S.Kom., M.Kom.

Reading interest is an important aspect in improving the quality of human resources. Differences in reading interest among communities need to be identified as a basis for evaluating the implementation of literacy programs. This study aims to implement the Naïve Bayes algorithm to classify the reading interest level of the Bengkulu community toward the Aksaraya Semesta Literacy Community Program and to evaluate the performance of the resulting classification model. This study applies the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, consisting of the stages of Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The research data were collected through a digital questionnaire using Google Forms. After the data preparation process, a total of 538 data were used in the study, consisting of 376 training data and 162 testing data. The research variables consisted of ten features related to respondents' reading activities and reading interest, while the target variable was classified into three categories: Low, Moderate, and High. The Naïve Bayes algorithm was implemented using the Python programming language on the Google Colab platform with the Scikit-learn library. The results showed that the Naïve Bayes algorithm achieved an accuracy of 80.25%, precision of 60.21%, recall of 68.09%, and F1-score of 62.44%. Evaluation using 10-Fold Cross Validation resulted in an average accuracy of 80.64%. The classification results showed that the model performed best in classifying the High reading interest category, while the Moderate category was the most difficult class to classify. Based on these results, the Naïve Bayes algorithm can be used to classify the reading interest level of the Bengkulu community toward the Aksaraya Semesta Literacy Community Program and provide an overview of the distribution of reading interest levels based on questionnaire data.

Keywords: Naïve Bayes, CRISP-DM, classification, reading interest, literacy, Google Colab

BAB I

PENDAHULUAN

1.1 Latar Belakang

Minat baca merupakan salah satu indikator penting dalam pembangunan kualitas sumber daya manusia karena berperan dalam meningkatkan kemampuan berpikir kritis, memperluas wawasan, serta mendukung kemajuan pendidikan dan kehidupan sosial masyarakat. Individu dengan minat baca yang baik cenderung lebih aktif dalam memperoleh informasi dan mengembangkan pengetahuan. Rendahnya minat baca dapat berdampak pada kualitas sumber daya manusia, sedangkan peningkatan minat baca dapat dilakukan melalui berbagai kegiatan, penyediaan bahan bacaan, tempat membaca, serta pemanfaatan teknologi sebagai sarana pendukung Literasi (Anidi & Anlianna, 2024).

Meskipun tingkat melek huruf di Indonesia tergolong tinggi, kondisi tersebut belum sepenuhnya diikuti oleh tingginya minat baca masyarakat. Hal ini menunjukkan adanya kesenjangan antara kemampuan membaca dan kebiasaan membaca yang berkelanjutan. Rendahnya minat baca dipengaruhi oleh faktor internal seperti kemampuan membaca, motivasi, kebiasaan membaca, dan kurangnya dorongan untuk membaca, serta faktor eksternal seperti ketersediaan buku atau bahan bacaan, lingkungan keluarga, lingkungan sekolah, penggunaan teknologi, dan teman sebaya (Yuniar & Pamujo, 2024).

Dalam konteks daerah, termasuk Provinsi Bengkulu, rendahnya partisipasi masyarakat dalam aktivitas literasi masih menjadi perhatian. Penelitian mengenai pembinaan komunitas Taman Bacaan Masyarakat (TBM) di Kota Bengkulu

menunjukkan bahwa pemanfaatan dan pembinaan kegiatan literasi masih perlu ditingkatkan agar masyarakat dapat memanfaatkan fasilitas literasi secara lebih optimal. Pembinaan komunitas TBM melalui literasi digital juga menunjukkan adanya peningkatan minat baca dan keterampilan literasi digital masyarakat. Hal ini menunjukkan bahwa penyediaan sarana literasi saja belum cukup, tetapi perlu disertai pembinaan dan kegiatan yang mendorong keterlibatan masyarakat dalam aktivitas literasi (Syahril & Astria, 2025).

Dalam upaya meningkatkan budaya membaca, berbagai program literasi mulai berkembang melalui pendekatan berbasis komunitas. Komunitas literasi hadir sebagai ruang sosial yang mendorong masyarakat untuk terlibat dalam kegiatan membaca, berdiskusi, serta berbagi pengetahuan secara lebih fleksibel dan partisipatif. Pendekatan komunitas dinilai mampu menciptakan lingkungan belajar yang lebih dekat dengan masyarakat karena kegiatan yang dilakukan tidak hanya berfokus pada aktivitas membaca, tetapi juga membangun interaksi sosial yang mendukung tumbuhnya minat literasi. Kegiatan komunitas literasi juga dapat dikembangkan melalui penyediaan sumber belajar, kegiatan membaca, pendampingan, dan aktivitas lain yang mendorong masyarakat untuk terlibat dalam kegiatan literasi.

Salah satu komunitas yang aktif menjalankan kegiatan literasi di Bengkulu adalah Komunitas Literasi Aksaraya Semesta. Komunitas ini menyelenggarakan berbagai kegiatan, seperti diskusi buku, penyediaan bahan bacaan, serta aktivitas literasi yang melibatkan masyarakat secara langsung. Kehadiran program tersebut diharapkan mampu meningkatkan ketertarikan masyarakat terhadap kegiatan

membaca. Namun demikian, pelaksanaan program literasi berbasis komunitas masih memerlukan evaluasi yang terukur agar efektivitas program dapat diketahui secara lebih objektif. Selama ini, penilaian terhadap keberhasilan program umumnya hanya didasarkan pada tingkat partisipasi masyarakat tanpa didukung analisis data yang mendalam mengenai tingkat minat baca peserta kegiatan. Oleh karena itu, diperlukan pengolahan data yang mampu memberikan gambaran mengenai tingkat minat baca masyarakat secara lebih terstruktur dan objektif. Penelitian mengenai pembinaan komunitas TBM di Kota Bengkulu juga menunjukkan pentingnya strategi pembinaan yang terarah dalam upaya meningkatkan minat baca masyarakat (Syahril & Astria, 2025).

Perkembangan teknologi komputasi memungkinkan dilakukannya analisis data secara lebih sistematis melalui pendekatan data mining. Data mining merupakan teknik pengolahan data yang digunakan untuk menemukan pola atau informasi tertentu dari sekumpulan data sehingga dapat menghasilkan informasi yang bermanfaat bagi proses pengambilan keputusan. Salah satu pendekatan yang banyak digunakan dalam data mining adalah klasifikasi. Metode klasifikasi dapat dimanfaatkan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan karakteristik yang dimiliki. Dalam konteks penelitian ini, klasifikasi digunakan untuk mengelompokkan tingkat minat baca masyarakat ke dalam kategori rendah, sedang, dan tinggi berdasarkan data hasil kuesioner.

Algoritma *Naïve Bayes* merupakan salah satu metode klasifikasi yang menggunakan pendekatan probabilitas untuk menentukan kelas suatu data berdasarkan karakteristik atau atribut yang dimiliki. Algoritma ini memiliki proses

perhitungan yang relatif sederhana serta dapat diterapkan pada data hasil kuesioner. Penerapan *Naïve Bayes* pada data berbasis kuesioner juga telah dilakukan pada penelitian yang menggunakan skala Likert sebagai sumber data evaluasi. Hasil penelitian tersebut menunjukkan bahwa *Naïve Bayes* dapat diterapkan untuk melakukan proses klasifikasi berdasarkan atribut yang diperoleh melalui kuesioner (Erinsyah et al., 2024).

Beberapa penelitian sebelumnya menunjukkan bahwa algoritma *Naïve Bayes* mampu diterapkan pada analisis perilaku pengguna, sistem rekomendasi, serta pengelompokan minat baca berdasarkan data survei. Penelitian Utari & Ulfah (2021) menerapkan metode *Naïve Bayes* untuk prediksi minat baca berdasarkan usia dan menunjukkan bahwa metode tersebut dapat digunakan untuk melakukan prediksi minat baca berdasarkan karakteristik data responden. Selain itu, penelitian yang dilakukan oleh Hafiz et al (2026) menggunakan *Naïve Bayes Classifier* untuk mengklasifikasikan tingkat minat baca berdasarkan 20 butir pernyataan kuesioner skala Likert yang merepresentasikan kebiasaan, frekuensi, motivasi, dan preferensi membaca. Penelitian tersebut menggunakan tiga kategori tingkat minat baca, yaitu tinggi, sedang, dan rendah, dan memperoleh akurasi terbaik sebesar 98,18% pada pembagian data 70:30.

Meskipun demikian, penerapan algoritma *Naïve Bayes* dalam kajian literasi dan minat baca masih memiliki perbedaan dari sisi objek dan konteks penelitian. Penelitian Utari & Ulfah (2021) membahas prediksi minat baca berdasarkan usia, sedangkan Hafiz et al (2026) berfokus pada siswa MTsN 1 Payakumbuh menggunakan data kuesioner skala Likert. Penelitian-penelitian tersebut

menunjukkan bahwa *Naïve Bayes* telah digunakan untuk klasifikasi minat baca, tetapi konteks objek penelitian masih berbeda dengan penelitian yang berfokus pada masyarakat yang mengikuti program literasi berbasis komunitas. Dengan demikian, penelitian yang menerapkan algoritma *Naïve Bayes* untuk mengklasifikasikan tingkat minat baca masyarakat dalam konteks program literasi komunitas nonformal, khususnya di Bengkulu, masih memiliki ruang untuk dikembangkan (Utari & Ulfah, 2021).

Kondisi tersebut menunjukkan adanya peluang penelitian untuk mengintegrasikan pendekatan machine learning dengan evaluasi program literasi masyarakat berbasis komunitas. Penelitian ini tidak hanya diarahkan untuk mengetahui kategori tingkat minat baca masyarakat, tetapi juga untuk mengetahui performa algoritma yang digunakan melalui proses evaluasi model. Dengan pendekatan tersebut, hasil klasifikasi diharapkan dapat memberikan informasi yang lebih objektif mengenai karakteristik tingkat minat baca masyarakat yang mengikuti kegiatan literasi. Penggunaan algoritma klasifikasi pada data kuesioner juga telah menunjukkan bahwa data hasil survei dapat dimanfaatkan sebagai dasar untuk menghasilkan informasi yang lebih terstruktur (Erinsyah et al., 2024).

Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk mengimplementasikan algoritma *Naïve Bayes* dalam mengukur tingkat minat baca masyarakat Bengkulu terhadap Program Komunitas Literasi Aksaraya Semesta. Data penelitian diperoleh melalui penyebaran kuesioner digital kepada peserta kegiatan komunitas. Hasil penelitian diharapkan mampu memberikan gambaran mengenai tingkat minat baca masyarakat secara lebih objektif serta dapat

dimanfaatkan sebagai bahan evaluasi bagi pengembangan Program Komunitas Literasi Aksaraya Semesta.

1.2 Pertanyaan Penelitian

Berdasarkan latar belakang yang telah diuraikan, maka pertanyaan penelitian dalam penelitian ini adalah sebagai berikut:

1. Bagaimana tingkat minat baca masyarakat Bengkulu terhadap Program Komunitas Literasi Aksaraya Semesta?
2. Bagaimana penerapan algoritma *Naïve Bayes* dalam mengklasifikasikan tingkat minat baca masyarakat Bengkulu ke dalam kategori rendah, sedang, dan tinggi?
3. Bagaimana kinerja algoritma *Naïve Bayes* dalam mengklasifikasikan tingkat minat baca masyarakat Bengkulu berdasarkan confusion matrix, accuracy, precision, recall, F1-score, dan 10-fold cross validation?

1.3 Tujuan penelitian

Berdasarkan pertanyaan penelitian yang telah dirumuskan, tujuan penelitian ini adalah sebagai berikut:

1. Menganalisis tingkat minat baca masyarakat Bengkulu terhadap Program Komunitas Literasi Aksaraya Semesta berdasarkan data kuesioner yang diperoleh dari peserta kegiatan literasi.
2. Mengimplementasikan algoritma *Naïve Bayes* untuk mengklasifikasikan tingkat minat baca masyarakat Bengkulu ke dalam tiga kategori, yaitu

rendah, sedang, dan tinggi berdasarkan karakteristik aktivitas literasi yang digunakan sebagai variabel input.

3. Mengevaluasi kinerja algoritma *Naïve Bayes* dalam mengklasifikasikan tingkat minat baca masyarakat Bengkulu menggunakan accuracy, precision, recall, F1-score, confusion matrix, dan 10-fold cross validation.
4. Menganalisis hasil klasifikasi sebagai dasar dalam memberikan rekomendasi untuk mendukung evaluasi dan pengembangan Program Komunitas Literasi Aksaraya Semesta.

1.4 Kerangka Penelitian

Kerangka penelitian digunakan untuk menggambarkan tahapan penelitian yang dilakukan secara sistematis dalam mengimplementasikan algoritma *Naïve Bayes* untuk mengklasifikasikan tingkat minat baca masyarakat Bengkulu terhadap Program Komunitas Literasi Aksaraya Semesta. Penelitian ini menggunakan pendekatan Cross-Industry Standard Process for Data Mining (CRISP-DM) yang terdiri atas tahapan *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*.

Setiap tahapan memiliki aktivitas, metode atau teknik yang digunakan, serta keluaran yang menjadi dasar bagi tahapan berikutnya. Kerangka penelitian yang digunakan dalam penelitian ini disajikan pada Tabel 1.1.

Tabel 1.1 Framework Penelitian

Tahap Penelitian	Aktivitas	Metode/ Teknik	Output
Business Understanding	Mengidentifikasi permasalahan dan menentukan tujuan penelitian terkait tingkat minat baca masyarakat serta evaluasi Program Komunitas Literasi Aksaraya Semesta	Analisis permasalahan dan kebutuhan penelitian	Rumusan masalah & tujuan penelitian
Data Understanding	Mengumpulkan dan memahami data hasil kuesioner masyarakat yang mengikuti kegiatan Komunitas Literasi Aksaraya Semesta	Kuesioner digital menggunakan Google Form dengan skala Likert 1–5	Data set awal hasil kusioner
Data Preparation	Melakukan pembersihan data, menangani missing value, melakukan encoding, membentuk label tingkat minat baca, menentukan fitur, dan membagi data menjadi training dan testing	Data cleaning, encoding, pembentukan label, dan random split 70:30	Dataset bersih dengan fitur X dan label Y

Tahap Penelitian	Aktivitas	Metode/ Teknik	Output
Modeling	Membangun model klasifikasi tingkat minat baca berdasarkan data training	Algoritma <i>Naïve Bayes</i>	Model klasifikasi tingkat minat baca
Evaluation	Menguji dan mengevaluasi kemampuan model dalam melakukan klasifikasi terhadap data testing	Confusion matrix, accuracy, precision, recall, F1-score, dan 10-fold cross validation	Nilai performa dan konsistensi model
Deployment	Memanfaatkan hasil klasifikasi sebagai dasar interpretasi dan penyusunan rekomendasi untuk pengembangan Program Komunitas Literasi Aksaraya Semesta.	Interpretasi hasil klasifikasi dan analisis deskriptif.	Informasi tingkat minat baca dan rekomendasi strategis untuk Program Komunitas Literasi Aksaraya Semesta

Berdasarkan Tabel 1.1, penelitian diawali dengan memahami permasalahan dan tujuan penelitian pada tahap *Business Understanding*. Selanjutnya, data hasil kuesioner dipahami dan diperiksa pada tahap *Data Understanding*. Data kemudian dipersiapkan melalui proses *data cleaning*, penanganan *missing value*, *encoding*, pembentukan label, serta pembagian data menjadi data training dan testing. Pada

tahap *Modeling*, algoritma *Naïve Bayes* digunakan untuk membangun model klasifikasi tingkat minat baca. Model kemudian dievaluasi menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan F1-score serta 10-fold cross validation untuk mengetahui performa dan konsistensi model. Tahap terakhir adalah *Deployment*, yaitu memanfaatkan hasil klasifikasi sebagai dasar penyusunan rekomendasi bagi pengembangan Program Komunitas Literasi Aksaraya Semesta.

