

Prediksi Pembatalan Reservasi Hotel Menggunakan Algoritma Naive Bayes

Farah Husna Qani'ah⁻¹, Rafiq Ramadhan⁻², Annisa Cintakami Firdaus⁻³, Ionia Veritawati⁻⁴
Program Studi Teknik Informatika^{1,2,3,4}
Fakultas Teknik Universitas Pancasila, Jakarta, Indonesia^{1,2,3,4}
4519210066@univpancasila.ac.id¹, 4519210019@univpancasila.ac.id²,
4519210073@univpancasila.ac.id³, ionia.veritawati@univpancasila.ac.id⁴

Abstrak—Pembatalan reservasi hotel dapat mempengaruhi manajemen industri dalam mengambil keputusan manajerialnya. Untuk mengatasi permasalahan dalam hal pembatalan reservasi hotel dapat diatasi dengan menerapkan prediksi dengan menggunakan metode *machine learning* dengan mempelajari data yang sudah dimiliki. Salah satu algoritma *machine learning* yang dapat digunakan dalam mengklasifikasikan data yaitu Naive bayes. Melalui hasil perhitungan pada 36.275 data yang memiliki 19 atribut dapat diketahui bahwa dalam menggunakan algoritma naive bayes untuk memprediksi pembatalan reservasi hotel mendapatkan hasil yang nilai ketepatan hanya mencapai 41% yang tidak memenuhi standar untuk pengambilan keputusan prediksi pembatalan reservasi hotel.

Kata Kunci—*machine learning*, naive bayes, prediksi, reservasi hotel

I. PENDAHULUAN

Mekanisme reservasi hotel secara *online* telah secara dramatis mengubah kemungkinan pemesanan dan perilaku pelanggan. Tidak sedikit terjadi pembatalan atas pemesanan hotel yang telah dilakukan dengan berbagai alasan, seperti adanya perubahan rencana, jadwal yang tidak sesuai dan lain sebagainya. Dalam industri perhotelan, pembatalan reservasi berpengaruh signifikan terhadap pendapatan dan reputasi kinerja bisnis sebuah tempat penginapan [1]. Selain itu, pembatalan reservasi juga sangat mempengaruhi manajemen industri dalam mengambil keputusan manajerialnya [2].

Para pemilik industri rata-rata ingin melihat keuntungan bisnisnya dengan melakukan prediksi berdasarkan data yang sudah ada [3]. Selain itu, prediksi berdasarkan data juga dapat membantu seseorang dalam mengambil keputusan. Maka, untuk mengatasi masalah di atas, dapat diterapkan prediksi terhadap pembatalan reservasi menggunakan metode *machine learning* dengan mempelajari data yang sudah dimiliki.

Salah satu algoritma *machine learning* yang dapat digunakan untuk memprediksi dengan mempelajari data adalah naive bayes. Naive bayes merupakan algoritma klasifikasi probabilistik yang sederhana, algoritma ini merupakan algoritma *supervised classification* yang berarti membutuhkan *data training* sebelum melakukan klasifikasi [4]. Naive bayes juga memiliki performa yang cukup baik dalam mengklasifikasikan data.

Terdapat penelitian yang mengulas perbandingan kinerja antara algoritma naive bayes dan algoritma KNN dalam mengklasifikasikan artikel berbahasa Indonesia [4]. Didapatkan hasil naive bayes memiliki kinerja yang lebih baik dengan tingkat akurasi sebesar 70% dan 40% untuk tingkat akurasi metode KNN. kemudian ada penelitian lain yang meneliti pengujian performa algoritma naive bayes untuk memprediksi masa studi mahasiswa [5]. Pada penelitian tersebut, pengujian dengan metode *10-Fold Cross Validation* dan *Confusion Matrix* menghasilkan nilai akurasi sebesar 68%. Selain itu terdapat penelitian lain yang melakukan komparasi akurasi antara algoritma C4.5 dan naive bayes untuk mendeteksi dini gangguan autisme pada anak [6]. Penelitian tersebut menghasilkan perbedaan yang tipis antara kedua algoritma, yaitu nilai akurasi sebesar 72% untuk algoritma C4.5 dan 73,33% untuk algoritma naive bayes.

Kajian yang telah ada, menunjukkan bahwa algoritma naive bayes memiliki performa yang cukup baik dalam melakukan prediksi dan klasifikasi. Maka pada penelitian ini akan diimplementasikan algoritma naive bayes untuk memprediksi pembatalan reservasi hotel, yang hasilnya dapat digunakan untuk mengambil keputusan manajerial hotel terkait dengan pembatalan reservasi di masa yang akan datang.

II. DASAR TEORI

A. Machine Learning

Machine learning merupakan sebuah mesin yang dibangun dengan sedemikian rupa agar dapat belajar tanpa harus diprogram secara eksplisit untuk menjalankannya [7].

Machine learning dapat dikatakan sebagai pendekatan Artificial Intelligence (AI). Dalam penyelesaian masalah, program dibentuk sedemikian rupa agar dapat mengambil sebuah keputusan menyerupai manusia. Program harus melakukan data training sebelum melakukan analisis terhadap data yang diuji.

B. Naive Bayes

Naive Bayes merupakan suatu metode pengklasifikasian probabilistik sederhana dengan cara menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang ada. Dengan menggunakan algoritma teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas [8]. Persamaan dari teorema bayes dapat dituliskan seperti berikut [5][6]:

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

Dengan:

- $P(B)$ = Hipotesis data B merupakan suatu kelas spesifik.
- $P(A)$ = Data dengan kelas yang belum diketahui.
- $P(A|B)$ = Probabilitas hipotesis A pada kondisi B (posterior probability).
- $P(B|A)$ = Probabilitas B berdasarkan kondisi pada hipotesisi A (likelihood).
- $P(A)$ = Probabilitas hipotesis A (class prior probability).
- $P(B)$ = probabilitas dari B (predictor prior probability).

Meskipun naive bayes dapat mengklasifikasikan data dengan baik, namun asumsinya terhadap atribut yang tidak saling bergantung dapat menurunkan akurasi prediksi dari algoritma tersebut. Berikut ini merupakan beberapa kelebihan dan kekurangan dari algoritma naive bayes [9]:

Kelebihan Naive Bayes:

1. Tidak membutuhkan jumlah data training yang besar,
2. Perhitungan cepat dan efisien,
3. Mudah dipahami dan dibuat.

Kekurangan Naive Bayes:

1. Apabila probabilitas kondisional bernilai nol, maka probabilitas prediksi juga akan bernilai nol,
2. Asumsi terhadap setiap atribut adalah independen mengurangi akurasi, karena biasanya terdapat korelasi antara atribut yang satu dengan yang lainnya.

III. METODOLOGI PENELITIAN

A. Dataset

Pada penelitian ini digunakan data reservasi hotel yang diperoleh dari website Kaggle dengan nama dataset yaitu Hotel Reservation Dataset yang dapat diakses pada <https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset>. Dataset tersebut memiliki 19 atribut dengan jumlah data sebanyak 36.275 data.

B. Preprocessing

Sebelum dilakukan pengolahan dengan algoritma naive bayes, dataset reservasi hotel yang telah diperoleh dilakukan preprocessing terlebih dahulu. Dataset hotel reservation memiliki data dengan jumlah atribut sebanyak 19 atribut.

Tabel 1. Deskripsi Atribut

| Typo Data | Atribut |
|-----------|---|
| Object | Booking_ID, type_of_meal_plan, room_type_reserved, market_segment_type, booking_status |
| Int | no_of_adults, no_of_children, no_of_weekend_nights, no_of_week_nights, required_car_parking_space, lead_time, arrival_year, arrival_month, arrival_date, repeated_guest, no_of_previous_cancellations, no_of_previous_booking_not_canceled, no_of_special_request |
| Float | avg_price_per_room |

Dari 19 atribut tersebut, tidak semua atribut akan digunakan pada tahap pengolahan. Maka beberapa atribut yang tidak berpengaruh akan dikecualikan pada proses pengolahan atau tidak digunakan. Seperti Booking_ID, atribut tersebut hanya berperan sebagai identitas dari data reservasi. Kemudian atribut arrival_year, arrival_month, dan arrival_date tidak diperlukan karena waktu atau tanggal reservasi tidak berpengaruh terhadap pembatalan reservasi.

Selanjutnya, atribut data memiliki tipe data yang berbeda-beda. Atribut yang memiliki tipe data objek atau kategori akan dikonversi ke dalam bentuk angka dengan metode One-Hot-Encoding. Atribut yang akan dilakukan konversi adalah type_of_meal_plan, room_type_reserved, dan market_segment_type. Atribut booking_status tidak dikonversi karena merupakan atribut label yang akan menjadi class target dari prediksi yang dilakukan. Ketiga

atribut yang diubah menggunakan metode *One-Hot-Encoding* akan dibentuk ke dalam variabel baru berdasarkan kategorinya, dan setiap variabel tersebut akan diisi dengan data biner.

Setelah data sudah dilakukan persiapan, selanjutnya adalah membagi data menjadi *data training* dan *data testing*. Pada penelitian ini *data training* dan *data testing* dibagi dengan rasio 0,7:0,3, yang berarti 70% data digunakan sebagai *data training* dan 30% digunakan untuk *data testing*.

C. Proses dengan Naive Bayes

Langkah-langkah prediksi data menggunakan algoritma naive bayes yang pertama adalah mencari peluang target atau label dari setiap atribut berdasarkan kondisi yang diberikan, atau menghitung *likelihood* dari persamaan naive bayes. Kemudian, hasil perhitungan *likelihood* dari setiap atribut yang telah diperoleh dikalikan berdasarkan masing-masing labelnya.

Variabel $P(B)$ atau probabilitas dari B pada persamaan naive bayes seringkali diabaikan karena nilainya yang selalu tetap. Hasil dari perkalian *likelihood* sebelumnya, akan dijadikan sebagai acuan untuk menentukan label dari data yang diprediksi. Label dengan hasil perhitungan tertinggi yang akan dijadikan sebagai label untuk data yang diprediksi.

Pada penelitian ini, proses pengolahan data dengan algoritma naive bayes dilakukan menggunakan bahasa pemrograman Python, dengan modul Gaussian Naive Bayes dari *library Scikit-learn* (Sklearn).

D. Metode Evaluasi

Proses prediksi pembatalan hotel yang sudah dilakukan akan dievaluasi menggunakan metode *confusion matrix*. *Confusion matrix* merupakan metode evaluasi yang digunakan untuk mengukur performa dari hasil klasifikasi. *Confusion matrix* berbentuk tabel dengan 4 kombinasi dari nilai prediksi dan nilai aktual yang biasa disebut juga dengan *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). *Confusion matrix* dari hasil prediksi pembatalan hotel menggunakan algoritma naive bayes dapat dilihat pada Tabel 2.

Tabel 2. *Confusion Matrix* Prediksi Pembatalan Reservasi Hotel

| Aktual | Prediksi | |
|-----------------|-----------------|---------------------|
| | <i>Canceled</i> | <i>Not_Canceled</i> |
| <i>Canceled</i> | 3467 | 145 |

| | | |
|---------------------|------|-----|
| <i>Not_Canceled</i> | 6316 | 955 |
|---------------------|------|-----|

Dari nilai pada tabel *confusion matrix*, dapat diperoleh perhitungan dengan *output* yang dapat mengukur performa dari hasil klasifikasi, di antaranya adalah:

$$\frac{TP}{TP+FP} \times 100\% \dots (2)$$

$$\frac{TN}{TN+FN} \times 100\% \dots (3)$$

$$1 - \frac{FP+FN}{TP+TN+FP+FN} = 2 \times \frac{TP \times TN}{TP+TN+FP+FN} \times 100\% \dots (4)$$

$$\frac{TP+TN}{TP+FP+TN+FN} \times 100\% \dots (5)$$

Hasil dari perhitungan 4 nilai di atas terhadap prediksi pembatalan hotel menggunakan algoritma naive bayes yang sudah dilakukan dapat dilihat pada Tabel 3.

Tabel 3. Hasil *Precision*, *Recall*, *F1-score*, dan *Accuracy* Dari Prediksi Pembatalan Hotel Dengan Naive Bayes

| | <i>Precision</i> | <i>Recall</i> | <i>F1-score</i> |
|------------------------|------------------|---------------|-----------------|
| <i>Canceled</i> | 35% | 96% | 52% |
| <i>Not_Canceled</i> | 87% | 13% | 23% |
| <i>Accuracy</i> | | | 41% |

IV. HASIL DAN PEMBAHASAN

Hasil dari metode yang sudah dilakukan membuktikan bahwa algoritma naive dapat digunakan untuk memprediksi pembatalan reservasi hotel. Berdasarkan Tabel 2, naive bayes mampu memprediksi pelanggan yang akan membatalkan reservasinya dengan baik, namun tidak dengan pelanggan yang akan meneruskan reservasi hotel. Pada tabel tersebut dapat dilihat prediksi *canceled* terhadap aktual *not_canceled* memiliki angka yang lebih besar dibandingkan dengan prediksi *not_canceled*.

Hal ini dibuktikan juga pada Tabel 3, dapat dilihat nilai pada kolom *recall* yang menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi, untuk label *canceled* didapatkan nilai 96% yang mana jauh lebih tinggi dari pada label *not_canceled* yang mendapat nilai sebesar 13%. Akurasi yang didapatkan dari seluruh proses

komputasi dengan algoritma naive bayes juga tidak terlalu tinggi, yaitu sebesar 41%.

Algoritma naive bayes mungkin memiliki performa yang baik dalam mengklasifikasi dan memprediksi data lain, namun dalam penelitian ini, naive bayes tidak menghasilkan performa yang cukup baik. Maka hasil dari prediksi menggunakan naive bayes ini tidak dapat digunakan untuk mengambil keputusan manajerial hotel terkait pembatalan reservasi hotel, karena nilai akurasi yang didapatkan dari prediksi yang telah dilakukan tidak cukup baik. Beberapa perbandingan dari prediksi yang diperoleh menggunakan naive bayes dengan nilai aktual dapat dilihat pada Tabel 4.

Tabel 4. Perbandingan Status Aktual Dengan Status Prediksi Pembatalan Reservasi Hotel

| | <i>Actual Status</i> | <i>Predicted Status</i> |
|-------|----------------------|-------------------------|
| 25479 | <i>Not_Canceled</i> | <i>Canceled</i> |
| 13623 | <i>Canceled</i> | <i>Canceled</i> |
| 2800 | <i>Not_Canceled</i> | <i>Canceled</i> |
| 30897 | <i>Not_Canceled</i> | <i>Canceled</i> |
| 26067 | <i>Canceled</i> | <i>Canceled</i> |
| ... | ... | ... |
| 476 | <i>Not_Canceled</i> | <i>Not_Canceled</i> |
| 34404 | <i>Not_Canceled</i> | <i>Canceled</i> |
| 9843 | <i>Canceled</i> | <i>Canceled</i> |
| 25320 | <i>Canceled</i> | <i>Canceled</i> |
| 2128 | <i>Not_Canceled</i> | <i>Not_Canceled</i> |

V. KESIMPULAN

Penggunaan algoritma naive bayes untuk memprediksi pembatalan reservasi hotel tidak mendapatkan hasil yang diinginkan. Jauhnya nilai ketepatan prediksi yang hanya mencapai 41%, tidak memenuhi standar untuk pengambilan keputusan prediksi pembatalan reservasi hotel. Maka

penggunaan algoritma *naive bayes* kurang direkomendasikan sebagai alat uji untuk pengujian prediksi seperti pada kasus pengujian hasil prediksi dari pembatalan reservasi hotel ini.

DAFTAR PUSTAKA

- [1] N. A. Putro, R. Septian, Widiastuti, M. Maulidah, and H. F. Pardede. (2021, March). Prediction of Hotel Booking Cancellation Using Deep Neural Network and Logistic Regression Algorithm. *Jurnal TECHNO Nusa Mandiri*. 18 (1), pp. 1-8. doi: <https://doi.org/10.33480/techno.v18i1.2056>.
- [2] Y. Azhar, G. A. Mahesa, and M. C. Mustaqim. (2020, Dec). Prediksi Pembatalan Pemesanan Hotel Menggunakan Optimalisasi Hiperparameter Pada Algoritme Random Forest. *Jurnal Teknologi dan Sistem Komputer*. [Online]. 9 (1), pp. 15-21. Available: <https://jtsiskom.undip.ac.id/index.php/jtsiskom/article/viewFile/13790/12654>.
- [3] S. Kitamori, H. Sakai, and H. Sakaji. (2017, Dec). Extraction of Sentences Concerning Business Performance Forecast and Economic Forecast from Summaries of Financial Statements by Deep Learning. *IEEE Symposium Series on Computational Intelligence*, Honolulu, USA, pp. 1-7. doi: <https://doi.org/10.1109/SSCI.2017.8285335>.
- [4] R. N. Devita, H. W. Herwanto, and A. P. Wibawa. (2018, Sep). Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*. 5 (4), pp. 427-434. doi: <http://dx.doi.org/10.25126/jtiik.201854773>.
- [5] I. W. Saputro, and B. W. Sari. (2019, Jan). Uji Performa Algoritma Naive Bayes Untuk Prediksi Masa Studi Mahasiswa. *Citec Journal*. 6 (1), pp 1-11. Doi: <https://doi.org/10.24076/citec.2019v6i1.178>.
- [6] B. Sugara, D. Adidarma, and S. Budilaksono. (2019, March). Perbandingan Akurasi Algoritma C4.5 dan Naive Bayes Untuk Deteksi Dini Gangguan Autisme Pada Anak. *Jurnal IKRA-ITH Informatika*. [Online]. 3 (1), pp 119-128. Available: <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/308>.
- [7] F. Rabbi. (2019, Jan). A Review of The Recent Trends in The Use of Machine Learning in Business. *International Journal of Artificial Intelligence and Machine Learning*. [Online]. 1 (1), pp. 1-6. Available:

- <https://www.ijaiml.com/wp-content/uploads/2019/04/Volume1-Issue1-Paper-1.pdf>.
- [8] W. Ananda, M. Safii, and M. Fauzan. (2021, March). Prediksi Jumlah Hasil Panen Sawit Menggunakan Algoritma Naive Bayes. *TIN: Terapan Informatika Nusantara*. [Online]. 1 (10), pp 513-519. Available: <https://ejournal.seminar-id.com/index.php/tin/article/download/698/471>.
- [9] M. H. Widiyanto. (2019, Dec). Algoritma Naive Bayes. *Binus University*. [Online]. Available: <https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/>.