



**IMPLEMENTASI PENGGUNAAN ALGORITMA  
CATEGORICAL BOOSTING (CATBOOST) DENGAN  
OPTIMISASI HIPERPARAMETER DALAM MEMPREDIKSI  
PEMBATALAN PESANAN KAMAR HOTEL**

**SKRIPSI**

**JOHANNES CHRISTIAN**

**1810511068**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN  
JAKARTA**

**FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA**

**2022**



**IMPLEMENTASI PENGGUNAAN ALGORITMA  
CATEGORICAL BOOSTING (CATBOOST) DENGAN  
OPTIMISASI HIPERPARAMETER DALAM MEMPREDIKSI  
PEMBATALAN PESANAN KAMAR HOTEL**

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar  
Sarjana Komputer**

**JOHANNES CHRISTIAN**

**1810511068**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN  
JAKARTA**

**FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA**

## PERNYATAAN ORSINALITAS

Tugas Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk saya nyatakan dengan benar.

Nama : Johannes Christian

NIM : 1810511068

Tanggal : 3 Juni 2022

Bilamana dikemudian hari ditemukan ketidak sesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 3 Juni 2022

Yang Menyatakan,



(Johannes Christian)

## PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan dibawah ini:

Nama : Johannes Christian

NIM : 1810511068

Fakultas : Ilmu Komputer

Program Studi : S1-Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (*Non-Exchange Royalty Free Right*) atas karya ilmiah saya yang berjudul:

Implementasi Penggunaan Algoritma *Categorical Boosting (CatBoost)* dengan Optimisasi Hiperparameter dalam Memprediksi Pembatalan Pesanan Kamar Hotel

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti di Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formalitas, mengelola dalam bentuk pengkalan data (Basis Data), merawat dan mempublikasi Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta, Demikian Pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada Tanggal : 3 Juni 2022

Yang Menyatakan,



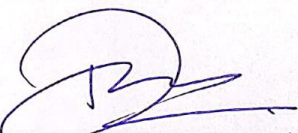
(Johannes Christian)

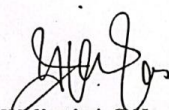
## LEMBAR PENGESAHAN


Dengan ini menyatakan bahwa proposal berikut:


Nama : Johannes Christian  
NIM : 1810511068  
Program Studi : S1 Informatika  
Judul : Implementasi Penggunaan Algoritma *Categorical Boosting*  
(*CatBoost*) dengan Optimisasi Hiperparameter dalam Memprediksi  
Pembatalan Pesanan Kamar Hotel

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.


  
Dr. Didit Widiyanto, S.Kom., M.Si.  
Penguji 1

  
Yuni Widiastwi, S.Kom., M.Si.  
Penguji 2

  
Iin Ernawati, S.Kom., M.Si.  
Pembimbing 1

  
Nurul Chamidah, S.Kom., M.Kom.  
Pembimbing 2

  
  
Dr. Ermatita, M.Kom.  
Dekan

  
Desta Sandya Pravista, S.Kom.,  
M.Kom.  
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal pengesahan : 3 Juni 2022



**IMPLEMENTASI PENGGUNAAN ALGORITMA  
CATEGORICAL BOOSTING (CATBOOST) DENGAN  
OPTIMISASI HIPERPARAMETER DALAM MEMPREDIKSI  
PEMBATALAN PESANAN KAMAR HOTEL**

**JOHANNES CHRISTIAN**

**ABSTRAK**

*Online Travel Agent (OTA)* yang tumbuh menjadi sebuah aplikasi untuk memudahkan masyarakat dalam melakukan pemesanan kamar hotel secara daring memberikan dampak yang signifikan terhadap manajemen hotel. Kemudahan tersebut membuat pengunjung melakukan *multiple-booking* yang mengakibatkan tingginya tingkat pembatalan pesanan kamar hotel. Strategi *overbooking* yang diterapkan oleh manajemen hotel memerlukan tingkat keakuratan yang tinggi dalam memperkirakan pengunjung yang melakukan pembatalan pesanan. Maka dari itu penelitian ini akan berfokus pada masalah tersebut menggunakan model pembelajaran mesin dengan algoritma *CatBoost*. Dalam proses pengklasifikasian data perlu dilakukan pembersihan data melalui proses *data preparation*. Setelah itu data akan dilakukan ekstraksi dan seleksi atribut sehingga data siap digunakan untuk melatih *machine learning*. Untuk meningkatkan performa dari model, optimisasi hiperparameter *RandomizedSearchCV* diterapkan terhadap model *CatBoost*. Hasil evaluasi dengan *confusion matrix* yaitu akurasi sebesar 88% dan *precision* sebesar 86%. Dengan menerapkan visualisasi SHAP pada *CatBoost* berhasil dihasilkan karakteristik – karakteristik pesanan kamar hotel yang akan berpeluang besar dibatalkan.

**Kata Kunci:** Hotel, *Overbooking*, *CatBoost*, *Hyperparameter Tuning*

# **IMPLEMENTATION OF USING CATEGORICAL BOOSTING (CATBOOST) ALGORITHM WITH HYPERPARAMETER TUNING IN PREDICTING CANCELLATION OF HOTEL ROOM BOOKINGS**

**JOHANNES CHRISTIAN**

## **ABSTRAK**

*Online Travel Agent (OTA) which has grown into an application to make it easier for people to book hotel rooms online has a significant impact on hotel management. This convenience makes visitors do multiple-booking which results in a high rate of cancellation of hotel room orders. The overbooking strategy implemented by hotel management requires a high level of accuracy in estimating visitors who cancel orders. Therefore, this research will focus on this problem using a machine learning model with the CatBoost algorithm. In the process of classifying data, it is necessary to clean the data through the data preparation process. After that, the data will be extracted and selected so that the data is ready to be used to train machine learning. To improve the performance of the model, the RandomizedSearchCV hyperparameter optimization was applied to the CatBoost model. The results of the evaluation with the confusion matrix are 88% accuracy and 86% precision. By using SHAP from CatBoost, produced the characteristics of hotel room orders that have a high chance of being cancelled.*

**Kata Kunci:** Hotel, Overbooking, CatBoost, Hyperparameter Tuning

## KATA PENGANTAR

Puji Syukur penulis panjatkan ke Tuhan Yang Maha Esa karena rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi dengan baik sehingga penulis dapat menyelesaikan studi tepat waktu. Penulisan skripsi ini bertujuan untuk memenuhi salah satu prasyarat untuk memperoleh gelar Sarjana Komputer, Jurusan Informatika.

Dalam penulisan skripsi ini, penulis mendapat banyak dukungan serta bantuan dari berbagai pihak, baik berupa materi, spiritual, dan informasi. Pada kesempatan kali ini, penulis mengucapkan terimakasih kepada:

1. Tuhan Yang Maha Esa yang telah memberikan kesehatan kepada penulis sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
2. Kedua orangtua dan keluarga tercinta yang selalu mendoakan serta dukungan kepada penulis sehingga dapat menyelesaikan Skripsi ini.
3. Ibu Iin Ernawati, S.Kom., M.Si. dan Nurul Chamidah, S.Kom, M.Kom., selaku dosen pembimbing yang berjasa dan memberikan bimbingan hingga terselesaikannya skripsi ini.
4. Bapak Desta Sandya Prasvita, S.Kom., M.Kom selaku kaprodi Informatika dan pembimbing akademik yang selalu memberikan informasi serta dukungan selama ini.
5. Ibu Dr. Ermatita, M.Kom. selaku Dekan Fakultas Ilmu Komputer.
6. Seluruh jajaran Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta yang telah membantu dalam perizinan dan administrasi.
7. Teman – teman Informatika yang berjuang bersama selama perkuliahan, memberikan semangat dan dorongan untuk dapat menyelesaikan kuliah dan skripsi ini.
8. Sahabat – sahabat perjuangan penulis Hilda, Hasan, Taufik, dan Devina yang selalu membantu dan memberikan masukan serta doa kepada penulis agar terselesaikannya skripsi ini dengan baik.
9. Terima kasih banyak kepada semua pihak atas bantuan, dukungan, semangat, dan doa yang tidak dapat penulis sebutkan satu persatu



Penyusun menyadari bahwa penyusunan skripsi ini masih jauh dari sempurna, oleh karena itu penyusun mengharapkan kritik dan saran yang bersifat membangun untuk kesempurnaan skripsi ini. Akhir kata penyusun mengharapkan semoga skripsi ini dapat bermanfaat bagi semua pihak.

Jakarta, 3 Juni 2022

Yang Menyatakan,

A handwritten signature in black ink, appearing to read 'Johannes Christian', with a stylized flourish at the end.

(Johannes Christian)

## DAFTAR ISI

|   |      |
|---|------|
| <b>SKRIPSI</b> .....                          | i    |
| <b>SKRIPSI</b> .....                          | ii   |
| <b>PERNYATAAN ORSINALITAS</b> .....           | iii  |
| <b>PERNYATAAN PERSETUJUAN PUBLIKASI</b> ..... | iv   |
| <b>LEMBAR PENGESAHAN</b> .....                | v    |
| <b>KATA PENGANTAR</b> .....                   | viii |
| <b>DAFTAR ISI</b> .....                       | x    |
| <b>DAFTAR TABEL</b> .....                     | xiv  |
| <b>DAFTAR GAMBAR</b> .....                    | xv   |
| <b>DAFTAR SIMBOL</b> .....                    | xvi  |
| <b>BAB I PENDAHULUAN</b> .....                | 1    |
| 1.1 Latar Belakang .....                      | 1    |
| 1.2 Rumusan Masalah .....                     | 3    |
| 1.3 Tujuan Penelitian.....                    | 3    |
| 1.4 Ruang Lingkup .....                       | 3    |
| 1.5 Luaran Yang Diharapkan .....              | 4    |
| 1.6 Manfaat Penelitian.....                   | 4    |
| 1.7 Sistematika Penulisan.....                | 4    |
| <b>BAB II TINJAUAN PUSTAKA</b> .....          | 7    |
| 2.1 <i>Overbooking</i> .....                  | 7    |
| 2.2 <i>Online Travel Agent (OTA)</i> .....    | 8    |
| 2.3 <i>Data Mining</i> .....                  | 9    |
| 2.3.1 Definisi <i>Data Mining</i> .....       | 9    |
| 2.3.2 Tahapan <i>Data Mining</i> .....        | 10   |

|                                |  |    |
|--------------------------------|--|----|
| 2.4                            | <i>Exploratory Data Analysis (EDA)</i> .....                   | 12 |
| 2.4.1                          | Analisis Univariat.....  | 13 |
| 2.4.2                          | Analisis Bivariat.....   | 15 |
| 2.5                            | Praproses Data .....   | 16 |
| 2.5.1                          | Pembersihan Data ( <i>Data Cleaning</i> ) .....                | 17 |
| 2.5.2                          | Ekstraksi Fitur .....  | 20 |
| 2.5.3                          | Seleksi Atribut ( <i>Feature Selection</i> ).....              | 20 |
| 2.6                            | Prediksi.....  | 21 |
| 2.7                            | Klasifikasi.....   | 22 |
| 2.7.1                          | Definisi Klasifikasi.....                                      | 22 |
| 2.7.2                          | Jenis Klasifikasi .....  | 22 |
| 2.7.3                          | Tahapan Klasifikasi.....                                       | 23 |
| 2.8                            | Ensemble Learning.....   | 24 |
| 2.8.1                          | Bagging .....  | 25 |
| 2.8.2                          | Boosting .....   | 25 |
| 2.8.3                          | Stacking.....  | 26 |
| 2.9                            | CatBoost (Categorical Boosting).....                           | 26 |
| 2.9.1                          | Definisi <i>CatBoost</i> .....                                 | 26 |
| 2.9.2                          | Categorical Features.....                                      | 27 |
| 2.9.3                          | Tahapan <i>CatBoost</i> .....                                  | 29 |
| 2.10                           | <i>Hyperparameter Tuning</i> (Optimisasi Hiperparameter) ..... | 32 |
| 2.11                           | Evaluasi.....  | 33 |
| 2.12                           | Studi Literatur .....  | 35 |
| BAB III METODE PENELITIAN..... |  | 37 |
| 3.1                            | Alur Penelitian.....   | 37 |
| 3.2                            | Identifikasi Masalah .....                                     | 38 |

|                                   |  |    |
|-----------------------------------|--|----|
| 3.3                               | Studi Pustaka .....  | 38 |
| 3.4                               | Metode Pengumpulan Data .....                                    | 38 |
| 3.4.1                             | Data Sekunder .....  | 39 |
| 3.4.2                             | Karakteristik Data .....   | 39 |
| 3.5                               | Exploratory Data Analysis (EDA) .....                            | 40 |
| 3.5.1                             | Analisis Univariat.....  | 40 |
| 3.5.2                             | Analisis Bivariat.....   | 41 |
| 3.6                               | Pengolahan Data ( <i>Preprocessing</i> ) .....                   | 41 |
| 3.6.1                             | Pembersihan Data ( <i>Data Cleaning</i> ) .....                  | 41 |
| 3.6.2                             | Ekstraksi Atribut ( <i>Feature Extraction</i> ) .....            | 41 |
| 3.6.3                             | Seleksi Atribut ( <i>Feature Selection</i> ).....                | 42 |
| 3.7                               | Pembagian Data.....  | 43 |
| 3.8                               | Pemodelan Data.....  | 43 |
| 3.8.1                             | Optimisasi Hiperparameter ( <i>Hyperparameter-tuning</i> ) ..... | 43 |
| 3.8.2                             | Pembangunan Model.....   | 44 |
| 3.9                               | Evaluasi .....   | 45 |
| 3.10                              | Analisis Model Klasifikasi.....                                  | 45 |
| 3.11                              | Alat Bantu Penelitian .....                                      | 46 |
| 3.12                              | Jadwal Penelitian .....  | 46 |
| BAB IV HASIL DAN PEMBAHASAN ..... |  | 48 |
| 4.1                               | Data .....   | 48 |
| 4.2                               | <i>Exploratory Data Analysis</i> .....                           | 50 |
| 4.2.1                             | <i>Statistical Summary</i> .....                                 | 50 |
| 4.2.2                             | Analisis Univariat.....  | 53 |
| 4.2.3                             | Analisis Bivariat.....   | 58 |
| 4.3                               | <i>Data Preprocessing</i> .....                                  | 59 |

|                           |                                    |    |
|---------------------------|------------------------------------|----|
| 4.3.1                     | Pembersihan Data.....              | 59 |
| 4.3.2                     | Ekstraksi Fitur / Atribut .....    | 60 |
| 4.3.3                     | Seleksi Fitur / Atribut.....       | 61 |
| 4.4                       | Pembagian Data.....                | 62 |
| 4.5                       | Pemodelan Data.....                | 63 |
| 4.5.1                     | <i>Hyperparameter Tuning</i> ..... | 63 |
| 4.5.2                     | Pembangunan Model.....             | 65 |
| 4.6                       | Evaluasi .....                     | 70 |
| 4.7                       | Analisis Model Klasifikasi .....   | 73 |
| BAB V PENUTUP.....        |                                    | 80 |
| 5.1                       | Kesimpulan.....                    | 80 |
| 5.2                       | Saran.....                         | 81 |
| DAFTAR PUSTAKA .....      |                                    | 82 |
| DAFTAR RIWAYAT HIDUP..... |                                    | 86 |
| LAMPIRAN.....             |                                    | 88 |

## DAFTAR TABEL

|  |    |
|--|----|
| Tabel 1. Hasil Imputasi Data.....  | 17 |
| Tabel 2. Penghapusan data invalid.....                                   | 18 |
| Tabel 3. Penghapusan data redundan .....                                 | 18 |
| Tabel 4. Penghapusan data yang bernilai ekstrim.....                     | 20 |
| Tabel 5. Hasil Ekstraksi Fitur.....                                      | 20 |
| Tabel 6. Metrik Pengukuran Evaluasi Model Klasifikasi .....              | 33 |
| Tabel 7. Definisi dan Contoh Istilah pada <i>Confusion Matrix</i> .....  | 34 |
| Tabel 8. Tabel <i>Confusion Matrix</i> .....                             | 35 |
| Tabel 9. Parameter yang akan diuji berulang.....                         | 43 |
| Tabel 10. Jadwal Penelitian.....   | 46 |
| Tabel 11. Karakteristik data .....                                       | 48 |
| Tabel 12. Hasil Seleksi Atribut .....                                    | 61 |
| Tabel 13. Atribut <i>Independent</i> yang Menjadi <i>Predictor</i> ..... | 62 |
| Tabel 14. Atribut <i>Dependent</i> yang Menjadi Target .....             | 63 |
| Tabel 15. Hasil Pembagian Data Latih dan Uji.....                        | 63 |
| Tabel 16. Variasi Parameter .....  | 64 |
| Tabel 17. Atribut Kategorik .....  | 65 |
| Tabel 18. Atribut Kategorik dengan Permutasi Acak .....                  | 65 |
| Tabel 19. Hasil Pengubahan Atribut Kategorik .....                       | 67 |
| Tabel 20. Kombinasi <i>Hyperparameter</i> dan <i>Score Akurasi</i> ..... | 68 |
| Tabel 21. Performa Prediksi Data Latih dan Data Uji .....                | 72 |

## DAFTAR GAMBAR

|   |    |
|---|----|
| Gambar 1. Proses CRISP-DM (Sumber: Huber et al., 2019).....   | 10 |
| Gambar 2. Visualisasi <i>Boxplot</i> .....  | 13 |
| Gambar 3. Visualisasi <i>Histplot</i> .....   | 14 |
| Gambar 4. Visualisasi <i>Countplot</i> .....  | 14 |
| Gambar 5. Visualisasi <i>Heatmap Pearson Correlation</i> .....  | 16 |
| Gambar 6. Penghitungan <i>outliers</i> dengan metode <i>Interquartile Range (IQR)</i><br>(Sumber: <a href="https://www.ichi.pro">https://www.ichi.pro</a> ) ..... | 19 |
| Gambar 7. Proses Operasi Klasifikasi.....   | 24 |
| Gambar 8. Algoritma Catboost (Sumber: Dewi, 2021) .....   | 30 |
| Gambar 9. Alur Penelitian.....  | 37 |
| Gambar 10. Proporsi label 1 dan 0.....  | 51 |
| Gambar 11. Statistik Deskriptif pada numerik.....   | 51 |
| Gambar 12. Statistik Deskriptif pada kategorik.....   | 52 |
| Gambar 13. Visualisasi <i>Boxplot</i> .....   | 53 |
| Gambar 14. Visualisasi histogram plot .....   | 55 |
| Gambar 15. Visualisasi <i>Countplot</i> .....   | 56 |
| Gambar 16. Visualisasi <i>Countplot Country dan Company</i> .....   | 58 |
| Gambar 17. Visualisasi heatmap korelasi <i>pearson</i> .....  | 59 |
| Gambar 18. Jumlah dan Persentase <i>Missing Data</i> .....  | 60 |
| Gambar 19. Atribut Kategorik sebelum <i>Encoding</i> .....  | 67 |
| Gambar 20. Atribut Kategorik setelah <i>Encoding</i> .....  | 68 |
| Gambar 21. Visualisasi <i>Log-loss training CatBoost</i> .....  | 69 |
| Gambar 22. Sampel Visualisasi Tree dari Model .....   | 70 |
| Gambar 23. <i>Confussion Matrix</i> Data Uji .....  | 70 |
| Gambar 24. <i>Confussion Matrix</i> Data Latih .....  | 71 |
| Gambar 25. Distribusi Probabilitas Prediksi Data Uji .....  | 72 |
| Gambar 26. <i>Features Importance</i> .....   | 73 |
| Gambar 27. Visualisasi <i>Beeswarm</i> .....  | 74 |
| Gambar 28. Visualisasi <i>Beeswarm</i> dengan <i>One-Hot-Encoding</i> .....   | 76 |
| Gambar 29. Visualisasi <i>Waterfall</i> .....   | 77 |