IMPLEMENTASI KLASIFIKASI *CATBOOST* DENGAN MENGGUNAKAN *HYPER-PARAMETER TUNING BAYESIAN SEARCH* UNTUK MEMPREDIKSI PENYAKIT DIABETES

Skripsi

Untuk Memenuhi Persyaratan

Dalam Menyelesaikan Strata-1 Ilmu Komputer

Oleh

ARIF DARMAWAN

NIM 171101621002

PROGRAM STUDI S-1 ILMU KOMPUTER

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS LAMBUNG MANGKURAT

BANJARBARU

SEPTEMBER 2023

IMPLEMENTASI KLASIFIKASI *CATBOOST* DENGAN MENGGUNAKAN *HYPER-PARAMETER TUNING BAYESIAN SEARCH* UNTUK MEMPREDIKSI PENYAKIT DIABETES

Skripsi

Untuk Memenuhi Persyaratan

Dalam Menyelesaikan Strata-1 Ilmu Komputer

Oleh

ARIF DARMAWAN

NIM 1711016210002

PROGRAM STUDI S-1 ILMU KOMPUTER

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS LAMBUNG MANGKURAT

BANJARBARU

SEPTEMBER 2023

SKRIPSI

IMPLEMENTASI KLASIFIKASI *CATBOOST* DENGAN MENGGUNAKAN *HYPER-PARAMETER TUNING BAYESIAN SEARCH* UNTUK MEMPREDIKSI PENYAKIT DIABETES

Oleh:
ARIF DARMAWAN
NIM 1711016210002

Telah dipertahankan di depan Dosen Penguji pada tanggal 06 September 2023.

Susunan Dosen Penguji:

Pembimbing I	Dosen Penguji I
Muliadi, S.Kom., M.Cs.	**Triando Hamonangan Saragih, S.Kom., M.Kom.**
NIP. 197804222010121002	**NIP. 199308242019031012**
Pembimbing II	Dosen Penguji II
Dwi Kartini, S.Kom., M.Kom	**Radityo Adi Nugroho, S.T., M.Kom.**
NIP. 198704212012122003	**NIP. 198212042008011006**

Banjarbaru 06 September 2023
Koordinator Program Studi Ilmu Komputer

Irwan Budiman, S.T., M.Kom
NIP. 197703252008121001

ii

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yangpernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskahini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, Juli 2022

Yang Menyatakan,

Arif Darmawan

NIM.1711016210002

ABSTRAK

IMPLEMENTASI KLASIFIKASI *CATBOOST* DENGAN MENGGUNAKAN *HYPER-PARAMETER TUNING BAYESIAN SEARCH* UNTUK MEMPREDIKSI PENYAKIT DIABETES

(Oleh : Arif Darmawan; Pembimbing: Muliadi S.Kom, M.Cs dan Dwi Kartini S.Kom., M.Kom.; 2022; 77 halaman)

Diabetes merupakan masalah kesehatan masyarakat dunia dengan prevalensi yang selalu meningkat setiap tahun. Penyakit Diabetes ini perlu didiagnosis sejak dini menggunakan algoritma klasifikasi. Dataset yang digunakan yaitu *PIMA Indians Diabetes Database* dari *Kaggle* dengan 768 data dan 8 fitur. Metode pengklasifikasi yang digunakan yaitu *CatBoost*. Klasifikasi *CatBoost* dapat bekerja baik dalam menangani ketidak seimbangan data, namun kinerja algoritma ini masih bisa ditingkatkan lagi. Untuk mengatasi permasalahan tersebut peneliti menggunakan solusi *Hyper-parameter tuning*. *CatBoost* memiliki beberapa *Hyper-parameter* yang dapat dikonfigurasi untuk meningkatkan kinerja dari model. Untuk mengidentifikasi nilai yang baik untuk *Hyper-parameter* disebut *Hyper-parameter tuning*. Metode *Hyper-parameter tuning* yang digunakan yaitu *Bayesian Search* yang kemudian divalidasi menggunakan *10-Fold Cross Validation* sebanyak 10 iterasi. *Hyper-parameter CatBoost* yang dikonfigurasi antara lain *depth (1,9), learning_rate (0.01, 1.00) dan Iterations (1,500)*. Penelitian ini bertujuan untuk mengetahui AUC dan presisi pada masing masing model. Pengujian pada *CatBoost* tanpa *Hyper-parameter tuning* memperoleh nilai AUC sebesar 0,859 dan presisi sebesar 62,5%. Untuk pengujian *CatBoost* dengan *Hyper-parameter tuning* memperoleh AUC sebesar 0,906 dan presisi sebesar 63,46% .Menambahkan *Hyper-parameter tuning Bayesian Search* pada metode klasifikasi *CatBoost* dapat meningkatkan hasil nilai AUC dan nilai presisi.

Kata Kunci : Klasifikasi; *CatBoost*; AUC; *Hyper-parameter Tuning*; *Bayesian Search*

ABSTRACT

IMPLEMENTATION OF CATBOOST CLASSIFICATION USING BAYESIAN SEARCH HYPER-PARAMETER TUNING TO PREDICT DIABETES

(*By* : Arif Darmawan; *Supervisor*: Muliadi S.Kom, M.Cs *and* Dwi Kartini S.Kom., M.Kom.; 2022; 77 *pages*)

Diabetes is a global public health problem with a continuously increasing prevalence every year. Early diagnosis of this disease using classification algorithms is crucial. The dataset used is the PIMA Indians Diabetes Database from Kaggle, which consists of 768 data points and 8 features. The classification method employed is *CatBoost*. *CatBoost* classification performs well in handling imbalanced data, but its algorithm's performance can still be further improved. To address this issue, researchers utilized Hyper-parameter tuning as a solution. *CatBoost* has several Hyper-parameters that can be configured to enhance the model's performance. Determining good values for these Hyper-parameters is known as Hyper-parameter tuning. *Bayesian Search* was the Hyper-parameter tuning method employed, which was then validated using 10-Fold Cross Validation over 10 iterations. The configured Hyper-parameters for *CatBoost* included depth (1,9), learning_rate (0.01, 1.00), and Iterations (1,500). The aim of this study was to evaluate the AUC, accuracy, and precision for each model. Testing *CatBoost* without Hyper-parameter tuning resulted in an AUC of 0.859, an accuracy of 71.43%, and a precision of 62.5%. On the other hand, testing *CatBoost* with Hyper-parameter tuning yielded an AUC of 0.906, an accuracy of 74.03%, and a precision of 63.46%. Adding *Bayesian Search* Hyper-parameter tuning to the *CatBoost* classification method can improve the AUC value, accuracy, and precision

Kata Kunci : *Classification*; *CatBoost*; *Accuracy*; AUC; *Hyper-parameter Tuning*; *Bayesian Search*

PRAKATA

Puji syukur penulis panjatkan ke Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul **"IMPLEMENTASI KLASIFIKASI *CATBOOST* DENGAN MENGGUNAKAN HYPER-PARAMETER TUNING *BAYESIAN SEARCH* UNTUK MEMPREDIKSI PENYAKIT DIABETES"** untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

1. Keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini, terutama untuk Bapak, Ibu, Dea dan Viona.
2. Bapak Muliadi, S.Kom, M.Cs. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalampenyelesaian skripsi ini.
3. Ibu Dwi Kartini, S.Kom, M.Kom. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2017 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
7. Ucapan terima kasih kepada Intan, Ryan, Barok, Shega, Bowo, Aldo, Irvan, Ubai, Ilham ,Zamzam, Yudhit, Fauzi, Haikal, Said, Eddy, Azmi, Luthfi, Didit, Rizqon, Reza,

Muflih, Noval, Apri, lastri, yang memberikan dukungan dalam proses mengerjakan skripsi.

8. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan serupa berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembacakhususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 06 September 2023

Arif Darmawan

DAFTAR ISI

DAFTAR TABEL

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR GAMBAR

DAFTAR LAMPIRAN

Lampiran