



**IMPLEMENTASI *CATBOOST* DENGAN *HYPER-PARAMETER TUNING*
RANDOM SEARCH UNTUK PREDIKSI PENYAKIT GINJAL KRONIS**

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

REZA PERDANA

NIM 1711016210028

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

SEPTEMBER 2023



**IMPLEMENTASI *CATBOOST* DENGAN *HYPER-PARAMETER TUNING*
RANDOM SEARCH UNTUK PREDIKSI PENYAKIT GINJAL KRONIS**

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

**REZA PERDANA
NIM 1711016210028**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
SEPTEMBER 2023**

SKRIPSI

**IMPLEMENTASI CATBOOST DENGAN HYPER-PARAMETER TUNING
RANDOM SEARCH UNTUK PREDIKSI PENYAKIT GINJAL KRONIS**

Oleh:

REZA PERDANNA

NIM. 1711016210028

Telah dipertahankan di depan Dosen Penguji pada tanggal 1 September 2023.

Susunan Dosen Penguji:

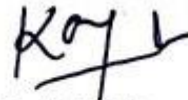
Pembimbing I

Dosen Penguji I



Muliadi, S.Kom., M.Cs.

NIP. 197804222010121002



Dwi Kartini S.Kom., M.Kom.

NIP. 198704212012122003

Pembimbing II

Dosen Penguji II



Irwan Budiman, S.T., M.Kom

NIP. 197703252008121001



Andi Farmadi., S.Si., M.T.

NIP. 197307252008011006



1 September 2023

Program Studi Ilmu Komputer

Irwan Budiman, S.T., M.Kom

NIP. 197703252008121001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah 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 naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 1 September 2023

Yang Menyatakan,



Reza Perdana

NIM.1711016210028

ABSTRAK

IMPLEMENTASI *CATBOOST* DENGAN *HYPER-PARAMETER TUNING* *RANDOM SEARCH* UNTUK *PREDIKSI PENYAKIT GINJAL KRONIS*

(Oleh : Reza Perdana; Pembimbing: Muliadi S.Kom, M.Cs dan Irwan Budiman, S.T., M.Kom.; 2022; 63 halaman)

Penyakit ginjal kronis (PGK) adalah masalah kesehatan masyarakat dunia dengan prevalensi dan insiden gagal ginjal kronis yang sering selalu meningkat setiap tahun. Penyakit ginjal kronis ini perlu didiagnosis sejak dini menggunakan algoritma klasifikasi. Dataset yang digunakan yaitu *Chronic Kidney Disease Dataset* dari *UCI Repository* dengan 400 data dan 25 fitur. Metode pengklasifikasi yang digunakan yaitu *Catboost*. *Catboost* baik untuk melakukan klasifikasi terhadap data dengan fitur kategorikal, namun kinerja algoritma ini masih bisa ditingkatkan. Untuk mengatasi permasalahan tersebut peneliti menggunakan solusi *Hyper-parameter tuning*. *Catboost* memiliki beberapa *hyper-parameter* yang dapat dikonfigurasi untuk meningkatkan kinerja dari model. Masalah mengidentifikasi nilai yang baik untuk *Hyper-parameter* disebut *Hyper-parameter tuning*. Metode *Hyper-parameter tuning* yang digunakan yaitu *Random Search* yang kemudian divalidasi menggunakan *10-Fold Cross Validation* sebanyak 10 iterasi. *Hyper-parameter Catboost* yang dikonfigurasi antara lain *depth*, *learning_rate* dan *Iterations*. Pengujian pada *Catboost* tanpa *Hyper-parameter tuning* memperoleh akurasi sebesar 96,6%, Recall 96,1% dan Presisi 98,6%. Untuk pengujian *Catboost* dengan *hyper-parameter tuning* memperoleh akurasi sebesar 97,5%, Recall 97,3% dan Presisi 98,6%. Menambahkan *Hyper-parameter tuning* pada metode klasifikasi *Catboost* dapat meningkatkan akurasi dan kinerja.

Kata kunci: *Cronic Kidney Disease* , *Catboost*, *Hyperparameter Tuning*, *Random Search*

ABSTRACT

IMPLEMENTATION CATBOOST IMPLEMENTATION WITH HYPER-PARAMETER TUNING RANDOM SEARCH FOR CHRONIC KIDNEY DISEASE PREDICTION

(By : Reza Perdana; Supervisor: Muliadi S.Kom, M.Cs dan Irwan Budiman, S.T., M.Kom.; 2022; 63 pages)

Chronic kidney disease (CKD) is a global public health problem of the prevalence and incidence of chronic kidney failure constantly increasing every year. Chronic kidney disease needs to be diagnosed early using a classification algorithm. The dataset used is a Chronic kidney disease dataset from the UCI Repository with 400 data and 25 features. The classifier method used is Catboost. Catboost is good for classifying data with categorical features, but the performance of this algorithm can still be improved. To overcome these problems, researchers used a Hyper-parameter tuning solution. Catboost has several Hyper-parameters that can be configured to improve the performance of the model . The problem of identifying good values for Hyper-parameters is called Hyper-parameter tuning. The Hyper-parameter tuning method used is Random Search which is then validated using 10-Fold Cross Validation for 30 iterations. The configured Catboost Hyper-parameters include depth, learning_rate and iterations. Testing on Catboost without hyper- parameter tuning obtained an accuracy of 96,6%, Recall 96,1% dan Presisi 98,6%. Testing on Catboost with hyper-parameter tuning obtained an accuracy of 97,5%, Recall 97,3% dan Presisi 98,6%. Adding Hyper-parameter tuning to the Catboost classification method can improve accuracy and performance.

Keywords: *Classification, Catboost, Hyperparameter Tuning, Random 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 *CATBOOST* DENGAN *HYPER-PARAMETER TUNING RANDOM SEARCH* UNTUK PREDIKSI PENYAKIT GINJAL KRONIS” 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
2. Bapak Muliadi S.Kom., M.Cs.selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Irwan Budiman S.T., M.Kom selaku dosen pembimbing pendamping serta Ketua Program Studi Ilmu Komputer FMIPA ULM yang turut serta membantu dan meluangkan waktu demi kelancarandalam penyelesaian skripsi ini.
4. Seluruh Dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
5. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2017 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
6. Ucapan terima kasih kepada, Shega, Yudhit, Aldo, Said, Bowo, Edho, Ipang, Gotai, Arif, Lutfi, Ryan, Rahmat, Mufleh, Yanche,

Ilham, Mubarak, Apri, Uzix, Zamzam, Nuzhan, Sartika, Eddy dan teman teman yang memberikan dukungan dalam proses mengerjakan skripsi.

7. 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 pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 1 September 2023

Reza Perdana

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	Error! Bookmark not defined.
HALAMAN PERNYATAAN	iii
ABSTRAK	iv
ABSTRACT	v
PRAKATA	vi
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan	3
1.5 Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA	5
2.1 Kajian Terdahulu	5
2.2 Landasan Teori	8
2.2.1 Data Mining	8
2.2.2 Ginjal Kronis	9
2.2.3 Decision Tree	10
2.2.4 Preprocessing	11
2.2.5 Boosting	13
2.2.6 Categorical Boosting	13
2.2.7 Hyper-parameter	15
2.2.8 Hyper-parameter Tuning	17
2.2.9 Confusion Matrix	19
BAB III METODE PENELITIAN	21
3.1 Bahan Penelitian	21
3.2 Alat Penelitian	21
3.3 Variabel Penelitian	21
3.4 Prosedur Penelitian	21
BAB IV HASIL DAN PEMBAHASAN	22
4.1 Hasil	22
4.1.1 Pengumpulan Dataset	22
4.1.2 Preprocessing Data	24
4.1.2.1 Missing Value	24
4.1.2.2 Label Encoder	28
4.1.3 Klasifikasi Catboost tanpa Hyper-parameter Tuning	30
4.1.4 Klasifikasi <i>Catboost</i> dengan <i>Hyper-parameter Tuning</i>	34

4.1.5.1 Hasil Akurasi <i>Catboost</i>	44
4.1.5.2 Hasil Akurasi Catboost dengan Hyper-parameter tuning.....	44
4.2 Pembahasan.....	45
BAB V PENUTUP	50
5.1 Kesimpulan.....	50
5.2 Saran	50
DAFTAR PUSTAKA	51