



**PERBANDINGAN METODE *CATBOOST* DAN *RANDOM FOREST* UNTUK
KLASIFIKASI PENYAKIT KANKER PARU-PARU MENGGUNAKAN
HYPERPARAMETER-TUNING BERBASIS *BAYESIAN OPTIMIZATION***

Skripsi

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

Oleh

YRA FATRIA ZAMZAM

NIM 2011016220012

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
MARET 2024**



**PERBANDINGAN METODE *CATBOOST* DAN *RANDOM FOREST* UNTUK
KLASIFIKASI PENYAKIT KANKER PARU-PARU MENGGUNAKAN
HYPERPARAMETER-TUNING BERBASIS *BAYESIAN OPTIMIZATION***

Skripsi

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

Oleh

YRA FATRIA ZAMZAM

NIM 2011016220012

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

MARET 2024

SKRIPSI

PERBANDINGAN METODE *CATBOOST* DAN *RANDOM FOREST* UNTUK KLASIFIKASI PENYAKIT KANKER PARU-PARU MENGGUNAKAN *HYPERPARAMETER-TUNING* BERBASIS *BAYESIAN OPTIMIZATION*

Oleh:

Yra Fatria Zamzam

NIM. 2011016220012

Telah dipertahankan di depan Dosen Penguji pada tanggal 21 Maret 2024.

Susunan Dosen Penguji:

Pembimbing I



Triando Hamonangan Saragih, S.Kom., M.Kom.

NIP. 199308242019031012

Dosen Penguji I



Muliadi, S.Kom., M.Sc.

NIP. 197804222010121002


Pembimbing II



Rudy Hexteno, S.Kom., M.Kom.

NIP. 198809252022031003

Dosen Penguji II



Dodon Turianto Nugrahadi., S.Kom, M.Eng.

NIP. 198001122009121002

Banjarnbaru, 21 Maret 2024

Koordinator Program Studi Ilmu Komputer



Iwan Budiman, S. T., M. Kom.

NIP. 197703252008121001

SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa dalam jurnal 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, 21 Maret 2024

Yang Menyatakan,



Yra Fatria Zamzam

NIM. 2011016220012

ABSTRAK

PERBANDINGAN METODE *CATBOOST* DAN *RANDOM FOREST* UNTUK KLASIFIKASI PENYAKIT KANKER PARU-PARU MENGGUNAKAN *HYPERPARAMETER-TUNING* BERBASIS *BAYESIAN OPTIMIZATION*

(Oleh: Yra Fatria Zamzam; Pembimbing: Triando Hamonangan Saragih, S.Kom., M.Kom. dan Rudy Herteno, S.Kom., M.Kom.; 2024; halaman)

Kanker Paru-paru adalah penyakit yang memiliki tingkat kematian yang tinggi dan seringkali sulit dideteksi hingga mencapai stadium yang sangat parah. Data menunjukkan bahwa kasus kanker paru biasanya terlambat didiagnosis, sehingga menimbulkan tantangan yang signifikan terhadap pengobatan yang efektif. Upaya deteksi dini menawarkan potensi peluang pemulihan yang lebih baik. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan metode identifikasi dan klasifikasi kanker paru dengan harapan dapat memberikan pengetahuan lebih lanjut tentang cara yang efektif untuk mendeteksi kondisi ini pada tahap awal. Salah satu pendekatan yang diteliti adalah dengan menggunakan teknik klasifikasi pembelajaran mesin, yang diharapkan dapat berfungsi sebagai alat penting dalam deteksi dini penyakit dan meningkatkan tingkat kelangsungan hidup pasien. Penelitian ini melibatkan lima tahap: pengumpulan data, prapemrosesan data, partisi data untuk pelatihan dan pengujian menggunakan *10-fold cross validation*, pelatihan model, dan analisis hasil evaluasi. Dalam penelitian ini, empat percobaan terdiri dari penerapan dua metode klasifikasi, *CatBoost* dan *Random Forest*, yang masing-masing diuji dengan menggunakan *hyperparameter default* dan *tuning hyperparameter* menggunakan *Bayesian Optimization*. Ditemukan bahwa model *Random Forest* menggunakan *tuning hyperparameter Bayesian Optimization* mengungguli model lainnya dengan akurasi (0.97106), presisi (0.97339), *recall* (0.97185), *f-measure* (0.97011), dan AUC (0.99974) untuk data kanker paru-paru. Temuan ini menyoroti bahwa Optimasi *Bayesian* untuk penyetelan *hyperparameter* dalam model klasifikasi dapat meningkatkan prediksi klinis kanker paru-paru dari rekam medis pasien. Integrasi *Bayesian Optimization* dalam penyetelan *hyperparameter* merupakan langkah maju yang signifikan dalam menyempurnakan akurasi dan efektivitas model klasifikasi, sehingga berkontribusi pada peningkatan diagnostik medis dan strategi perawatan kesehatan yang sedang berlangsung.

Kata kunci: Kanker Paru-paru, *CatBoost*, *Random Forest*, *Bayesian Optimization*

ABSTRACT

COMPARISON OF CATBOOST AND RANDOM FOREST METHODS FOR LUNG CANCER CLASSIFICATION USING HYPERPARAMETER TUNING BAYESIAN OPTIMIZATION- BASED

(By: Yra Fatria Zamzam; Supervisors: Triando Hamonangan Saragih, S.Kom., M.Kom. and Rudy Herteno, S.Kom., M.Kom.; 2024; page)

Lung Cancer is a disease that has a high mortality rate and is often difficult to detect until it reaches a very severe stage. Data indicates that lung cancer cases are typically diagnosed late, posing significant challenges to effective treatment. Early detection efforts offer the potential for better recovery chances. Therefore, this research aims to develop methods for the identification and classification of lung cancer in the hope of providing further knowledge on effective ways to detect this condition at an early stage. One approach under scrutiny involves employing machine learning classification techniques, anticipated to serve as a pivotal tool in early disease detection and enhancing patient survival rates. This study involves five stages: data collection, data preprocessing, data partitioning for training and testing using 10-fold cross validation, model training, and analysis of evaluation results. In this research, four experiments consist of applying two classification methods, CatBoost and Random Forest, each tested using default hyperparameter and hyperparameter tuning using Bayesian Optimization. It was found that the Random Forest model using hyperparameter tuning Bayesian Optimization outperformed the other models with accuracy (0.97106), precision (0.97339), recall (0.97185), f-measure (0.97011), and AUC (0.99974) for lung cancer data. These findings highlight that Bayesian Optimization for hyperparameter tuning in classification models can improve clinical prediction of lung cancer from patient medical records. The integration of Bayesian Optimization in hyperparameter tuning represents a significant step forward in refining the accuracy and effectiveness of classification models, thus contributing to the ongoing enhancement of medical diagnostics and healthcare strategies.

Keywords: Lung Cancer, CatBoost, Random Forest, Bayesian Optimization

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT, yang atas rahmat dan karunia-Nya penulis dapat menyelesaikan jurnal yang berjudul “*Comparison of CatBoost and Random Forest Methods for Lung Cancer Classification using Hyperparameter Tuning Bayesian Optimization-based*” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Proses penyusunan skripsi ini telah menjadi perjalanan yang tak terlupakan bagi saya, penuh dengan tantangan, pengalaman, dan pembelajaran yang mendalam.

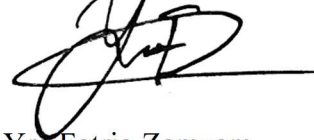
Dalam perjalanan ini, penulis tidak sendirian. Penulis ingin mengucapkan terima kasih yang tulus kepada semua pihak yang telah memberikan dukungan, bimbingan, dan motivasi kepada penulis dalam pembuatan dan penyusunan jurnal ini, adapun yang dimaksud adalah sebagai berikut:

1. Diri saya sendiri yang tidak patah semangat dan tetap bertanggung jawab untuk menyelesaikan apa yang telah dimulai walaupun banyak menemui kesulitan baik disebabkan oleh diri sendiri maupun hal lain.
2. Keluarga besar yang selalu memberikan bantuan, semangat, doa, dan dukungan yang menjadi motivasi tersendiri bagi penulis.
3. Bapak Triando Hamonangan Saragih, S.Kom., M.Kom. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian jurnal ini.
4. Bapak Rudy Herteno, S.Kom., M.Kom. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian jurnal ini.
5. Bapak Irwan Budiman, S.T., M. Kom. selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau jurnal ini dapat diselesaikan.
6. Seluruh Dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu, bimbingan, dan dukungan yang tak ternilai selama ini
7. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan jurnal.

8. Serta semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian jurnal ini.

Akhir kata, penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna. Namun, penulis berharap dengan tulus bahwa hasil penelitian ini dapat memberikan manfaat yang berarti bagi perkembangan ilmu pengetahuan serta menjadi sumber inspirasi bagi pembaca, serta mendapat keridhaan Allah SWT.

Banjarbaru, 21 Maret 2024



Yra Fatria Zamzam