



***BOOTSTRAPPING HPO UNTUK MENANGANI
OVERFITTING PADA KLASIFIKASI SENTIMEN UNTUK
DATA BERDIMENSI TINGGI DARI REVIEWS INDIE GAME
MOBILE LOKAL***

MUHAMMAD SYIDIQ ABDJANUR



**PROGRAM STUDI STATISTIKA DAN SAINS DATA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “*Bootstrapping* HPO untuk Menangani *Overfitting* pada Klasifikasi Sentimen untuk Data Berdimensi Tinggi dari *Reviews Indie Game Mobile Lokal*” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2024

Muhammad Syidiq Abdjanur
G1501211009

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

RINGKASAN

MUHAMMAD SYIDIQ ABDJANUR. *Bootstrapping* HPO untuk Menangani *Overfitting* pada Klasifikasi Sentimen untuk Data Berdimensi Tinggi dari *Reviews Indie Game Mobile* Lokal. Dibimbing oleh I MADE SUMERTAJAYA dan ANWAR FITRIANTO.

Penggunaan *smartphone* telah mendorong popularitas permainan seluler, memungkinkan perkembangan sub-sektor pengembang aplikasi dan *game* di Indonesia. Namun, *indie game developers* menghadapi kekurangan sumber daya promosi, yang mengakibatkan kurangnya data ulasan untuk membangun model klasifikasi yang baik. Mengklasifikasikan ulasan *game* dengan dimensi tinggi menimbulkan tantangan statistik yang signifikan, seperti *overfitting*. Penelitian ini bertujuan untuk mengevaluasi metode *optimisasi hyperparameter* terbaik dalam klasifikasi *support vector machines* (SVM) untuk menentukan model klasifikasi sentimen terbaik. Secara khusus, tujuannya adalah mengevaluasi metode optimisasi *hyperparameter* terbaik berdasarkan nilai akurasi hasil klasifikasi SVM pada data berdimensi tinggi dan menerapkan *bootstrapping* pada metode *optimisasi hyperparameter* terbaik untuk memperoleh nilai akurasi model terbaik. Hasil analisis menunjukkan bahwa *optuna optimization* dan *genetic algorithm* memiliki nilai akurasi dan *F1 score* yang lebih baik dibandingkan dengan metode optimisasi lainnya. Dari hasil evaluasi, *optuna optimization* dianggap sebagai metode terbaik dalam konteks ini tanpa menggunakan *bootstrapping*. Hasil analisis menunjukkan bahwa *optuna optimization* memiliki nilai akurasi yang stabil, dengan nilai *F1 score* yang menurun ketika menggunakan *bootstrapping*.

Kata kunci: *Bootstrapping, Mobile Games, Sentiment Classification, Hyperparameter Optimization, Support Vector Machines* (SVM)

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

SUMMARY

MUHAMMAD SYIDIQ ABDJANUR. *HPO Bootstrapping to Handle Overfitting in Sentiment Classification for High-Dimensional Data from Local Mobile Indie Game Reviews*. Supervised by I MADE SUMERTAJAYA and ANWAR FITRIANTO.

The use of smartphones has driven the popularity of mobile games, enabling the development of the sub-sector of application and game developers in Indonesia. However, indie game developers face a lack of promotional resources, which results in a lack of review data to build a good classification model. Classifying game reviews with high dimensions poses significant statistical challenges, such as overfitting. This study aims to evaluate the best hyperparameter optimization method in support vector machines (SVM) classification to determine the best sentiment classification model. Specifically, the goal is to evaluate the best hyperparameter optimization method based on the accuracy value of SVM classification results on high-dimensional data and apply bootstrapping to the best hyperparameter optimization method to obtain the best model accuracy value. The analysis results show that optuna optimization and genetic algorithm have better accuracy values and F1 scores compared to other optimization methods. From the evaluation results, optuna optimization is considered the best method in this context without using bootstrapping. The analysis results show that optuna optimization has a stable accuracy value, with an F1 score value that decreases when using bootstrapping.

Keywords: *Bootstrapping, Mobile Games, Sentiment Classification, Hyperparameter Optimization, Support Vector Machines (SVM)*

© Hak Cipta milik IPB, tahun 2024
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



***BOOTSTRAPPING HPO UNTUK MENANGANI
OVERFITTING PADA KLASIFIKASI SENTIMEN UNTUK
DATA BERDIMENSI TINGGI DARI REVIEWS INDIE GAME
MOBILE LOKAL***

MUHAMMAD SYIDIQ ABDJANUR

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI STATISTIKA DAN SAINS DATA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Tesis:
Dr. Bagus Sartono, M.Si



Judul Tesis : *Bootstrapping* HPO untuk Menangani *Overfitting* pada Klasifikasi Sentimen untuk Data Berdimensi Tinggi dari *Reviews Indie Game Mobile Lokal*

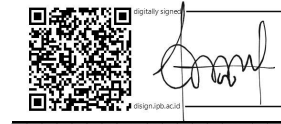
Nama : Muhammad Syidiq Abdjanur
NIM : G11501211009

Disetujui oleh

Pembimbing 1:
Dr. Ir. I Made Sumertajaya, M.Si.



Pembimbing 2:
Dr. Anwar Fitrianto, S.Si, M.Sc.



Diketahui oleh

Ketua Program Studi:
Dr. Agus Mohamad Soleh, S.Si., M.T.
NIP 197503151999031004



Dekan Fakultas Matematika dan Ilmu
Pengetahuan Alam
Dr. Berry Juliandi, S.Si., M.Si.
NIP 197807232007011001



Tanggal Ujian: 30 Mei 2024

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanahu wa taala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian ini ialah klasifikasi sentimen dengan judul “*Bootstrapping* HPO untuk Menangani *Overfitting* pada Klasifikasi Sentimen untuk Data Berdimensi Tinggi dari *Reviews Indie Game Mobile* Lokal”.

Penulis menyadari bahwa penyusunan karya ilmiah ini terselesaikan dengan baik karena adanya bantuan, dorongan, dan bimbingan dari banyak pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Ir. I Made Sumertajaya, M.Si dan Bapak Dr. Anwar Fitrianto, S.Si, M.Sc. selaku komisi pembimbing yang senantiasa bersedia meluangkan waktunya untuk membimbing, memberikan masukan dan saran, serta memotivasi penulis sehingga penelitian ini dapat terselesaikan dengan baik.
2. Pimpinan, dosen pengajar, dan pegawai Departemen Statistika yang telah membantu terkait penelitian ini.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan. Penulis menyadari masih ada kekurangan dalam penyusunan penelitian ini, untuk itu penulis mengharapkan kritik dan saran yang membangun. Akhir kata, semoga Allah SWT. selalu memberikan rahmat dan hidayah-Nya kepada kita semua. *Aamiin*.

Bogor, Juli 2024

Muhammad Syidiq Abdjanur

DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 Manfaat	3
II TINJAUAN PUSTAKA	4
2.1 <i>High Dimensionality</i>	4
2.2 <i>Support Vector Machines (SVM)</i>	4
2.3 Optimasi Bayes	6
2.4 Algoritma Genetika	7
2.5 <i>Grid Search</i>	9
2.6 Hyperopt	9
2.7 <i>Optuna</i>	10
2.8 Metode <i>Bootstrapping</i>	10
2.9 <i>Preprocessing Data Text</i>	11
2.10 Evaluasi Model	12
III METODE	13
3.1 Data	13
3.2 Prosedur Analisis	13
3.2.1 Data Pra-proses	13
3.2.2 Metode Optimasi SVM	14
3.2.3 Metode Optimasi SVM Terbaik dengan dan Tanpa Menerapkan <i>Bootstrapping</i>	18
IV HASIL DAN PEMBAHASAN	20
4.1 Data	20
4.2 Data Pra-proses	21
4.2.1 <i>Case Folding</i> atau <i>Lowercasing Text</i>	22
4.2.2 <i>Text Normalization</i>	22
4.2.3 <i>Stemming</i>	22
4.2.4 <i>Stopword Removal</i>	23
4.2.5 <i>Bag of Words</i>	23
4.3 Data <i>Splitting</i> dan <i>Labeling</i>	24
4.4 Penerapan Metode Optimasi Parameter pada Model <i>Support Vector Machine</i>	24
4.5 <i>Bootstrapping</i> Optimasi Parameter Terbaik pada Model SVM	27
4.6 Klasifikasi Ulasan <i>Game Rating 3</i>	28
V KESIMPULAN DAN SARAN	33
5.1 Kesimpulan	33
5.2 Saran	34

35
37
40

DAFTAR PUSTAKA	35
LAMPIRAN	37
RIWAYAT HIDUP	40

© Hak cipta milik IPB University

IPB University



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR TABEL

1	Contoh data <i>bag of words</i> (Dandale 2021)	12
2	Matriks <i>confusion</i> untuk klasifikasi biner	12
3	Jumlah ulasan setiap <i>game</i>	13
4	Evaluasi metode optimasi hyperparameter SVM pada setiap data	17
5	Evaluasi metode optimasi <i>hyperparameter</i> SVM menggunakan <i>bootstrapping</i>	19
6	Contoh <i>record</i> data hasil <i>scraping</i>	20
7	Summary data yang di <i>scraping</i>	21
8	Hasil proses <i>lowercasing text</i>	22
9	Hasil proses <i>normalization text</i>	22
10	Hasil proses <i>stemming text</i>	22
11	Hasil proses <i>stopword removal text</i>	23
12	Hasil proses <i>bag of words</i>	23
13	Data setelah pembersihan	24
14	Hasil klasifikasi SVM menggunakan optimasi parameter	25
15	Hasil <i>report</i> akurasi	26
16	Hasil klasifikasi SVM menggunakan optimasi parameter <i>bootstrapping</i>	27
17	Hasil report akurasi menggunakan <i>bootstrapping</i>	27
18	Parameter SVM pada setiap data	28

DAFTAR GAMBAR

1	<i>Hyperplane</i> yang memisahkan dua kelas positif (+1) dan negatif(-1) (Samsudiney 2019)	5
2	<i>Flowchart Bayesian Optimization</i>	7
3	GA-SVM <i>modeling flow</i> (Ji et al. 2021)	8
4	<i>Flowchart</i> pra-proses data	14
5	<i>Flowchart</i> metode optimasi SVM terbaik	17
6	<i>Flowchart</i> optimasi SVM terbaik dengan dan tanpa menerapkan <i>bootstrapping</i>	18
7	Data hasil transformasi dari <i>json</i> ke bentuk tabel	20
8	Visualisasi efek parameter (García 2021)	25
9	<i>Wordclouds</i> dari hasil prediksi ulasan rating 3 data tahu bulat 1, (a) kelas positif, (b) kelas negatif	28
10	<i>Wordclouds</i> dari hasil prediksi ulasan rating 3 data tahu bulat 2, (a) kelas positif (b) kelas negatif	29
11	<i>Wordclouds</i> dari hasil prediksi ulasan rating 3 data nasi goreng, (a) kelas positif (b) kelas negatif	30
12	<i>Wordclouds</i> dari hasil prediksi ulasan rating 3 data bakso simulator, (a) kelas positif (b) kelas negatif	31
13	<i>Wordclouds</i> dari hasil prediksi ulasan rating 3 data warnet simulator, (a) kelas positif (b) kelas negatif	32

DAFTAR LAMPIRAN

1	Hasil <i>output</i> pada setiap model dan data	37
2	Hasil <i>output</i> pada setiap model dan data menggunakan <i>bootstrapping</i>	39

© Hak cipta milik IPB University

IPB University



- Hak Cipta Dilindungi Undang-undang
- Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 - Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.