

TUGAS AKHIR

STUDI EKSPLORASI DAN PEMANFAATAN SISTEM AUTOML



Fachri Mohamad Soetisna

NPM: 6181801017

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2024

FINAL PROJECT

**EXPLORATION STUDY AND APPLICATION OF AUTOML
SYSTEM**



Fachri Mohamad Soetisna

NPM: 6181801017

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2024**

LEMBAR PENGESAHAN

STUDI EKSPLORASI DAN PEMANFAATAN SISTEM AUTOML

Fachri Mohamad Soetisna

NPM: 6181801017

Bandung, 21 Juni 2024

Menyetujui,

Pembimbing

Digitally signed
by Veronica Sri
Moertini

Dr. Veronica Sri Moertini

Ketua Tim Penguji

Digitally signed
by Natalia

Natalia, M.Si.

Anggota Tim Penguji

Digitally signed
by Raymond
Chandra Putra

Raymond Chandra Putra, M.T.

Mengetahui,

Ketua Program Studi

Digitally signed
by Lionov

Lionov, Ph.D.

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa tugas akhir dengan judul:

STUDI EKSPLORASI DAN PEMANFAATAN SISTEM AUTOML

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 21 Juni 2024



Fachri Mohamad Soetisna
NPM: 6181801017

ABSTRAK

Bidang pembelajaran mesin atau *machine learning* (ML) telah mengalami kemajuan signifikan namun keahlian yang diperlukan untuk membangun dan mengoptimalkan model ML tetap menjadi hambatan bagi banyak pengguna khususnya yang bukan informatikawan. Sistem Pembelajaran Mesin Otomatis (AutoML) bertujuan untuk mempermudah akses ke ML dengan mengotomatisasi beberapa proses penerapan ML pada masalah dunia nyata. Skripsi ini menyajikan studi eksplorasi tentang sistem AutoML H2O, PyCaret, dan AutoGluon dengan fokus pada kinerja, kecepatan, dan aplikasi praktisnya.

Skripsi dimulai dengan tinjauan mengenai sistem AutoML H2O, PyCaret, dan AutoGluon dengan mengkaji ide dasar, keunggulan, dan keterbatasannya. Skripsi kemudian dilanjutkan dengan melakukan serangkaian eksperimen untuk mengevaluasi sistem-sistem ini pada berbagai dataset dan membandingkan kinerjanya dengan model ML yang dibuat secara manual. Metrik utama seperti performa, waktu komputasi, dan kemudahan penggunaan dianalisis untuk menilai efektivitas solusi AutoML. Hasil dari eksplorasi AutoML H2O, PyCaret, dan AutoGluon menunjukkan bahwa ketiga AutoML tersebut mudah digunakan dan menghasilkan model dengan kualitas yang setara dengan teknik manual.

Pada tahap pemanfaatan sistem AutoML diterapkan pada data dunia nyata untuk menyelesaikan masalah kompleks tepatnya pada teks berita politik. Hasilnya menunjukkan potensi AutoML untuk mengurangi waktu dan keahlian yang dibutuhkan dalam mengembangkan model berkinerja tinggi pada data tidak terstruktur. Hasil dari pembuatan model menunjukkan bahwa AutoML H2O, PyCaret, dan AutoGluon mampu menghasilkan model dengan kualitas yang sangat baik dengan metrik presisi di tingkat $>80\%$. Di akhir skripsi juga dibuat beberapa perangkat lunak untuk menguji tahap peluncuran model yang dihasilkan oleh sistem AutoML H2O, PyCaret, dan AutoGluon. Pengujian menunjukkan bahwa model-model tersebut dapat diluncurkan dengan baik.

Kata-kata kunci: Pembelajaran mesin, pembelajaran mesin otomatis, pembelajaran, otomatis, eksplorasi, H2O, PyCaret, AutoGluon

ABSTRACT

The field of machine learning (ML) has made significant advancements, yet the expertise required to build and optimize ML models remains a barrier for many users, especially those who are not computer scientists. Automated Machine Learning (AutoML) systems aim to make ML more accessible by automating several processes involved in applying ML to real-world problems. This thesis presents an exploratory study of the AutoML systems H2O, PyCaret, and AutoGluon, focusing on their performance, speed, and practical applications.

The thesis begins with a review of the AutoML systems H2O, PyCaret, and AutoGluon, examining their fundamental concepts, advantages, and limitations. It then continues with a series of experiments to evaluate these systems on various datasets and compare their performance with manually created ML models. Key metrics such as performance, computational time, and ease of use are analyzed to assess the effectiveness of AutoML solutions. The results of the exploration of AutoML H2O, PyCaret, and AutoGluon show that all three AutoML systems are easy to use and produce models of comparable quality to manual techniques.

In the application phase, the AutoML systems are applied to real-world data to solve complex problems, specifically on political news texts. The results demonstrate the potential of AutoML to reduce the time and expertise required to develop high-performing models on unstructured data. The model-building results indicate that AutoML H2O, PyCaret, and AutoGluon are capable of producing high-quality models with precision metrics above 80%. At the end of the thesis, several software tools are developed to test the deployment phase of models generated by the AutoML systems H2O, PyCaret, and AutoGluon. Testing shows that these models can be deployed successfully.

Keywords: Machine learning, automated machine learning, learning, automated, exploration, H2O, PyCaret, AutoGluon

*Saya persembahkan skripsi ini untuk kedua orang tua, seluruh kelas
pekerja, dan kaum tertindas di dunia.*

KATA PENGANTAR

Dengan segala kerendahan hati, penulis mempersembahkan tugas akhir ini yang berjudul "Studi Eksplorasi dan Pemanfaatan Sistem AutoML." Tugas akhir ini disusun sebagai salah satu syarat untuk menyelesaikan program sarjana di Universitas Parahyangan pada jurusan Informatika.

Proses penyusunan tugas akhir ini tidak terlepas dari dukungan dan bantuan berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Dr. Veronica Sri Moertini, selaku dosen pembimbing yang telah memberikan bimbingan, saran, dan motivasi yang sangat berharga selama proses penyusunan tugas akhir ini.
2. Kristopher David Harjono, M.T., selaku mentor yang membimbing dan membantu selama masa perkuliahan.
3. Cevas Bungaran, Widad Prio Utomo, Syahdan Riyantyo Putro, dan Edward Tjahyadi yang sudah meluangkan waktu untuk membantu dan menemani selama masa pengerjaan skripsi.
4. Mohammad Farhan Mazhari, Velya Putridita, Daffa Arianda, Imara Sungkar, Rafaelia Risha Doodoh, dan Alberto Tom yang sudah memberikan dukungan dalam berbagai bentuk hingga dapat menyelesaikan skripsi ini.
5. Kedua orang tua dan keluarga yang sudah memberikan dukungan moral dan materiil, serta doa yang tiada henti.

Penulis menyadari bahwa tugas akhir ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran sangat diharapkan demi perbaikan di masa mendatang. Semoga tugas akhir ini dapat memberikan manfaat sebesar-besarnya bagi sebanyak-banyaknya pihak yang membacanya dan dapat menjadi sumbangsih yang berguna bagi perkembangan ilmu pengetahuan, khususnya dalam bidang AutoML.

Akhir kata, penulis mengucapkan terima kasih.

Bandung, Juni 2024

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xxi
DAFTAR TABEL	xxvii
DAFTAR KODE PROGRAM	xxx
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Batasan Masalah	3
1.5 Metodologi	3
1.6 Sistematika Pembahasan	4
DAFTAR NOTASI	1
2 LANDASAN TEORI	5
2.1 <i>Data Science</i> [1]	5
2.2 <i>Machine Learning</i>	6
2.2.1 Model Linier [2] [3]	7
2.2.2 <i>Tree-Based</i>	10
2.2.3 Algoritma <i>Boosting</i>	12
2.2.4 <i>Neural Network</i> [3]	15
2.2.5 <i>Support Vector Machine</i> [3]	16
2.2.6 <i>K-Nearest Neighbor</i> [3]	17
2.2.7 <i>Naive Bayes</i> [3]	17
2.3 Metrik Evaluasi Model	19
2.3.1 <i>R-Squared</i> [4]	19
2.3.2 <i>Error Measures</i> [3]	21
2.3.3 <i>Confusion Matrix</i> [3]	21
2.3.4 Akurasi [3]	22
2.3.5 Metrik untuk <i>Text Retrieval: Recall dan Precision</i>	23
2.4 Analisis Data Teks dan <i>Information Retrieval</i> [3]	23
2.4.1 Metode <i>Text Retrieval</i>	24
2.4.2 <i>Vector Space Model</i>	25
2.4.3 Analisis Klasifikasi Dokumen	26
2.5 AutoML [5]	27
2.6 H2O AutoML [6]	29
2.6.1 <i>Data Pre-processing</i>	29

2.6.2	Model	29
2.6.3	<i>Leaderboard</i>	30
2.6.4	Analisis Data Teks	30
2.6.5	Instalasi dan Pengaksesan H2O Flow	30
2.6.6	Parameter	31
2.7	PyCaret [7]	31
2.7.1	Fitur-fitur dalam PyCaret	31
2.7.2	Instalasi dan Pengaksesan PyCaret	33
2.8	AutoGluon [8]	34
2.8.1	<i>Data Pre-processing</i>	34
2.8.2	Model	34
2.8.3	Strategi Pelatihan	36
2.8.4	Instalasi dan Pengaksesan AutoGluon	36
2.8.5	Parameter	37
3	STUDI EKSPLORASI	39
3.1	Deskripsi Masalah	39
3.1.1	Regresi	39
3.1.2	Klasifikasi	45
3.2	Kinerja Pembandingan	48
3.3	Pembuatan Model H2O	49
3.4	Pembuatan Model PyCaret	67
3.5	Pembuatan Model AutoGluon	83
3.6	Analisis Kinerja	95
3.6.1	Perbandingan Performa	96
3.6.2	Jumlah Sampel vs <i>Runtime</i>	98
3.7	Kesimpulan	100
4	PEMBUATAN MODEL STUDI KASUS	105
4.1	Latar Belakang Masalah dan Usulan Solusi	105
4.2	Pengumpulan Data	106
4.3	Pelabelan Data	113
4.4	Analisis Sifat Kata Kunci Berita <i>Related</i> dan <i>Not Related</i>	115
4.5	Pembersihan Data	116
4.6	Penyiapan Data	117
4.6.1	Tokenisasi	117
4.6.2	Vektorisasi	118
4.7	Pembuatan Model H2O	119
4.8	Pembuatan Model PyCaret	126
4.9	Pembuatan Model AutoGluon	129
4.10	Evaluasi Hasil Model	131
4.10.1	H2O	132
4.10.2	PyCaret	133
4.10.3	AutoGluon	136
4.11	Kesimpulan	137
5	PEMBANGUNAN PERANGKAT LUNAK DAN PENGUJIAN	139
5.1	Perancangan Perangkat Lunak	139
5.1.1	Desain Perangkat Lunak	139
5.1.2	Fitur-Fitur Perangkat Lunak	141
5.2	Implementasi Perangkat Lunak Eksperimen Sederhana H2O	147
5.3	Implementasi Perangkat Lunak Eksperimen Sederhana PyCaret	151

5.4	Implementasi Perangkat Lunak Eksperimen Sederhana AutoGluon	153
5.5	Implementasi Perangkat Lunak Eksperimen Studi Kasus Bagian <i>Preprocessing</i> . .	154
5.6	Implementasi Perangkat Lunak Eksperimen Studi Kasus Bagian Prediksi H2O . .	156
5.7	Implementasi Perangkat Lunak Eksperimen Studi Kasus Bagian Prediksi PyCaret	158
5.8	Implementasi Perangkat Lunak Eksperimen Studi Kasus Bagian Prediksi AutoGluon	158
5.9	Tampilan Antar Muka dan Pengujian Perangkat Lunak	159
5.9.1	Perangkat Lunak Eksperimen Data Sederhana H2O, PyCaret, dan AutoGluon	159
5.9.2	Perangkat Lunak Eksperimen Data Studi Kasus Bagian <i>Preprocessing</i> . . .	161
5.9.3	Perangkat Lunak Eksperimen Data Studi Kasus Bagian Prediksi H2O, Pycaret, dan AutoGluon	162
5.10	Pengujian Perangkat Lunak	163
5.10.1	Pengujian Perangkat Lunak Eksperimen Data Sederhana	164
5.10.2	Pengujian Perangkat Lunak Eksperimen Studi Kasus	167
5.11	Kesimpulan	175
6	KESIMPULAN DAN SARAN	177
6.1	Kesimpulan	177
6.2	Saran	177
	DAFTAR REFERENSI	179
	A KODE PROGRAM	181

DAFTAR GAMBAR

2.1 Tahapan <i>Data Science</i>	5
2.2 Contoh Nilai <i>R-Squared</i> Negatif	20
2.3 Contoh Nilai <i>R-Squared</i> Sama Dengan Nol	20
2.4 Contoh Nilai <i>R-Squared</i> Positif	21
2.5 Contoh <i>Confusion Matrix</i> Untuk Positif Dan Negatif Tuple	22
2.6 Hubungan Kumpulan Dokumen Relevan dan Dokumen yang Terambil	23
2.7 <i>Flowchart</i> tahapan ML	28
2.8 Contoh tampilan evaluasi hasil model dengan plot <i>feature importance</i>	33
3.1 Dialog <i>Assistance</i>	50
3.2 Dialog Import	51
3.3 Dialog Konfirmasi <i>Import</i>	51
3.4 Dialog <i>Parse</i> 1	52
3.5 Dialog <i>Parse</i> 2	52
3.6 Dialog Konfirmasi <i>Parse</i>	53
3.7 Dialog Hasil <i>Parse</i>	53
3.8 Dialog Hasil <i>Parsing</i>	54
3.9 Dialog <i>Split</i>	54
3.10 Dialog Konfirmasi <i>Split</i>	55
3.11 Dialog Parameter H2O 1	56
3.12 Dialog Parameter H2O 2	56
3.13 Dialog Parameter H2O 3	57
3.14 Dialog Parameter H2O 4	58
3.15 Dialog Konfirmasi Pembuatan Model <i>Car Price</i>	59
3.16 Dialog Konfirmasi Pembuatan Model <i>Housing</i>	59
3.17 Dialog Konfirmasi Pembuatan Model <i>Life Expectancy</i>	59
3.18 Dialog Konfirmasi Pembuatan Model <i>Insurance Payout</i>	59
3.19 Dialog Konfirmasi Pembuatan Model <i>Credit Score</i>	60
3.20 Dialog Konfirmasi Pembuatan Model <i>Person's Happiness</i>	60
3.21 Dialog Konfirmasi Pembuatan Model <i>Heart Attack</i>	60
3.22 Dialog Konfirmasi Pembuatan Model <i>Iris Species</i>	60
3.23 Dialog Konfirmasi Pembuatan Model <i>NBA Player's Longevity</i>	60
3.24 <i>Leaderboard</i> Model Masalah <i>Car Price</i>	61
3.25 Dialog Detil Model Masalah <i>Car Price</i>	61
3.26 <i>Report</i> Model Masalah <i>Car Price</i>	62
3.27 <i>Variable Importance</i> Model Masalah <i>Car Price</i>	62
3.28 <i>Leaderboard</i> Model Masalah <i>Crab Age</i>	63
3.29 <i>Report</i> Model Masalah <i>Crab Age</i>	63
3.30 <i>Variable Importance</i> Model Masalah <i>Crab Age</i>	64
3.31 <i>Leaderboard</i> Model Masalah <i>Insurance Payout</i>	64
3.32 <i>Report</i> Model Masalah <i>Insurance Payout</i>	64
3.33 <i>Variable Importance</i> Model Masalah <i>Insurance Payout</i>	65

3.34	<i>Leaderboard Model Masalah Person's Happiness</i>	65
3.35	<i>Report Model Masalah Person's Happiness</i>	65
3.36	<i>Variable Importance Model Masalah Person's Happiness</i>	66
3.37	<i>Leaderboard Model Masalah Heart Attack</i>	66
3.38	<i>Report Model Masalah Person's Happiness</i>	66
3.39	<i>Confusion Matrix Model Masalah Heart Attack</i>	67
3.40	Hasil Eksekusi Kode Pengecekan <i>Import Dataset Car Price</i>	68
3.41	Hasil Eksekusi Kode Pengecekan Tipe Data pada Kolom Dataset <i>Car Price</i>	69
3.42	Hasil Eksekusi Kode Pengubahan Tipe Data pada Kolom Target Dataset <i>Car Price</i>	69
3.43	Keluaran Perintah Setup Beserta Daftar Parameter untuk Dataset <i>Car Price</i>	70
3.44	<i>Leaderboard Dataset Car Price</i>	70
3.45	<i>Pipeline Model Terbaik Dataset Car Price</i>	71
3.46	<i>Plot Variable Importance Model Terbaik Dataset Car Price</i>	71
3.47	Keluaran Perintah Penyimpanan Model Terbaik Dataset <i>Car Price</i>	71
3.48	<i>Leaderboard Dataset Crab Age</i>	72
3.49	<i>Pipeline Model Terbaik Dataset Crab Age</i>	73
3.50	<i>Plot Variable Importance Model Terbaik Dataset Crab Age</i>	73
3.51	<i>Leaderboard Dataset Housing</i>	73
3.52	<i>Pipeline Model Terbaik Dataset Housing</i>	74
3.53	<i>Plot Variable Importance Model Terbaik Dataset Housing</i>	74
3.54	<i>Leaderboard Dataset Life Expectancy</i>	75
3.55	<i>Pipeline Model Terbaik Dataset Life Expectancy</i>	75
3.56	<i>Plot Variable Importance Model Terbaik Dataset Life Expectancy</i>	75
3.57	<i>Leaderboard Dataset Insurance Payout</i>	76
3.58	<i>Pipeline Model Terbaik Dataset Insurance Payout</i>	76
3.59	<i>Plot Variable Importance Model Terbaik Dataset Insurance Payout</i>	76
3.60	<i>Leaderboard Dataset Credit Score</i>	77
3.61	<i>Pipeline Model Terbaik Dataset Credit Score</i>	77
3.62	<i>Confusion Matrix Model Terbaik Dataset Credit Score</i>	78
3.63	<i>Leaderboard Dataset Person's Happiness</i>	78
3.64	<i>Pipeline Model Terbaik Dataset Person's Happiness</i>	78
3.65	<i>Confusion Matrix Model Terbaik Dataset Person's Happiness</i>	79
3.66	<i>Leaderboard Dataset Heart Attack</i>	79
3.67	<i>Pipeline Model Terbaik Dataset Heart Attack</i>	80
3.68	<i>Confusion Matrix Model Terbaik Dataset Heart Attack</i>	80
3.69	<i>Leaderboard Dataset Iris Species</i>	81
3.70	<i>Pipeline Model Terbaik Dataset Iris Species</i>	81
3.71	<i>Confusion Matrix Model Terbaik Dataset Iris Species</i>	81
3.72	<i>Leaderboard Dataset NBA</i>	82
3.73	<i>Pipeline Model Terbaik Dataset NBA</i>	82
3.74	<i>Confusion Matrix Model Terbaik Dataset NBA</i>	83
3.75	Hasil Eksekusi Kode Pencetakan Lima Data Teratas	84
3.76	Keluaran Setelah Menjalankan <i>Method Fit</i> untuk Dataset <i>Car Price</i>	85
3.77	<i>Leaderboard Dataset Car price</i>	86
3.78	<i>Variable Importance Model Terbaik Dataset Car Price</i>	86
3.79	<i>Leaderboard Dataset Crab Age</i>	87
3.80	<i>Variable Importance Model Terbaik Dataset Crab Age</i>	88
3.81	<i>Leaderboard Dataset Housing</i>	88
3.82	<i>Variable Importance Model Terbaik Dataset Housing</i>	89
3.83	<i>Leaderboard Dataset Life Expectancy</i>	89
3.84	<i>Variable Importance Model Terbaik Dataset Life Expectancy</i>	90

3.85	<i>Leaderboard</i> Dataset <i>Insurance Payout</i>	90
3.86	<i>Variable Importance</i> Model Terbaik Dataset <i>Insurance Payout</i>	90
3.87	<i>Leaderboard</i> Dataset <i>Credit Score</i>	91
3.88	<i>Confusion Matrix</i> Model Terbaik Dataset <i>Credit Score</i>	91
3.89	<i>Leaderboard</i> Dataset <i>Credit Score</i>	92
3.90	<i>Confusion Matrix</i> Model Terbaik Dataset <i>Person's Happiness</i>	92
3.91	<i>Leaderboard</i> Dataset <i>Heart Attack</i>	93
3.92	<i>Confusion Matrix</i> Model Terbaik Dataset <i>Heart Attack</i>	93
3.93	<i>Leaderboard</i> Dataset <i>Iris Species</i>	94
3.94	<i>Confusion Matrix</i> Model Terbaik Dataset <i>Iris Species</i>	94
3.95	<i>Leaderboard</i> Dataset <i>NBA</i>	95
3.96	<i>Confusion Matrix</i> Model Terbaik Dataset <i>NBA</i>	95
3.97	Perbandingan Kinerja Masalah Regresi	97
3.98	Perbandingan Kinerja Masalah <i>Crab Age</i>	97
3.99	Perbandingan Kinerja Masalah Klasifikasi	98
3.100	Jumlah Sampel VS <i>Runtime</i> untuk Masalah Regresi	99
3.101	Jumlah Sampel VS <i>Runtime</i> untuk Masalah Klasifikasi	100
4.1	Situs berita Detik	106
4.2	Visualisasi 10 kata dengan selisih kemunculan terbesar	115
4.3	Visualisasi selisih jumlah kemunculan <i>stop word</i> antara kelas <i>related</i> dan <i>not related</i>	117
4.4	Dialog <i>Assist</i>	119
4.5	Dialog Memilih <i>File</i> pada Perintah <i>Import Files</i>	119
4.6	Dialog Konfirmasi <i>Import Files</i>	120
4.7	Dialog <i>Setup Parsing</i>	120
4.8	Dialog Konfirmasi <i>Parsing</i>	121
4.9	Dialog <i>View Data</i>	121
4.10	Dialog <i>Setup Train-Test Split</i>	122
4.11	Dialog Konfirmasi <i>Train-Test Split</i>	122
4.12	Dialog Parameter AutoML H2O	123
4.12	Dialog Parameter AutoML H2O	124
4.13	Dialog Konfirmasi <i>Run</i> AutoML	125
4.14	Dialog <i>Leaderboard</i>	125
4.15	Dialog Model Terbaik	126
4.16	Keluaran <i>Setup</i> Eksperimen	127
4.17	<i>Leaderboard</i> Model–Model yang Dibuat	128
4.18	Keluaran Perintah Penyimpanan Model Terbaik	128
4.19	Keluaran <i>Run</i> AutoML AutoGluon	130
4.20	<i>Leaderboard</i> Model-Model Terbaik AutoGluon	131
4.21	<i>Confusion Matrix</i> Model Terbaik H2O	132
4.22	<i>Confusion Matrix</i> Model Terbaik PyCaret	134
4.23	<i>Classification Report</i> Model Terbaik PyCaret	134
4.24	<i>Confusion Matrix</i> Model Terbaik AutoGluon	136
4.25	<i>Classification Report</i> Model Terbaik AutoGluon	136
5.1	Diagram Desain Perangkat Lunak untuk Model Eksperimen Sederhana	140
5.2	Diagram Desain Perangkat Lunak untuk Model Studi Kasus	141
5.3	Diagram <i>Use Case</i> PL Eksperimen Sederhana	142
5.4	Diagram <i>Use Case</i> PL Eksperimen Studi Kasus <i>Preprocessing</i>	144
5.5	Diagram <i>Use Case</i> PL Eksperimen Studi Kasus Prediksi	146
5.6	Tampilan perangkat lunak eksperimen data sederhana H2O	160
5.7	Tampilan perangkat lunak eksperimen data sederhana PyCaret	160

5.8	Tampilan perangkat lunak eksperimen data sederhana AutoGluon	161
5.9	Tampilan perangkat lunak eksperimen studi kasus bagian <i>preprocessing</i>	161
5.10	Tampilan perangkat lunak eksperimen studi kasus bagian prediksi H2O	162
5.11	Tampilan perangkat lunak eksperimen studi kasus bagian prediksi PyCaret	163
5.12	Tampilan perangkat lunak eksperimen studi kasus bagian prediksi AutoGluon	163
5.13	Tampilan awal perangkat lunak eksperimen data sederhana H2O	164
5.14	Tampilan data pengujian terunggah pada perangkat lunak eksperimen data sederhana H2O	165
5.15	Tampilan hasil prediksi pada perangkat lunak eksperimen data sederhana H2O	165
5.16	Tampilan awal perangkat lunak eksperimen data sederhana PyCaret	165
5.17	Tampilan data pengujian terunggah pada perangkat lunak eksperimen data sederhana PyCaret	166
5.18	Tampilan hasil prediksi pada perangkat lunak eksperimen data sederhana PyCaret	166
5.19	Tampilan awal perangkat lunak eksperimen data sederhana AutoGluon	166
5.20	Tampilan data pengujian terunggah pada perangkat lunak eksperimen data sederhana AutoGluon	167
5.21	Tampilan hasil prediksi pada perangkat lunak eksperimen data sederhana AutoGluon	167
5.22	Tampilan awal pada perangkat lunak eksperimen studi kasus bagian <i>preprocessing</i>	168
5.23	Tampilan berita terunggah pada perangkat lunak eksperimen studi kasus bagian <i>preprocessing</i> kelas <i>related</i>	168
5.24	Tampilan berita terunggah pada perangkat lunak eksperimen studi kasus bagian <i>preprocessing</i> kelas <i>not related</i>	168
5.25	Tampilan hasil <i>preprocess</i> pada perangkat lunak eksperimen studi kasus bagian <i>preprocessing</i> kelas <i>related</i>	169
5.26	Tampilan hasil <i>preprocess</i> pada perangkat lunak eksperimen studi kasus bagian <i>preprocessing</i> kelas <i>not related</i>	169
5.27	Tampilan penyimpanan vektor pada perangkat lunak eksperimen studi kasus bagian <i>preprocessing</i> kelas <i>related</i>	169
5.28	Tampilan penyimpanan vektor pada perangkat lunak eksperimen studi kasus bagian <i>preprocessing</i> kelas <i>not related</i>	170
5.29	Tampilan dialog konfirmasi pada perangkat lunak eksperimen studi kasus bagian <i>preprocessing</i>	170
5.30	Tampilan awal pada perangkat lunak eksperimen studi kasus bagian prediksi H2O	170
5.31	Tampilan vektor terunggah kelas <i>related</i> pada perangkat lunak eksperimen studi kasus bagian prediksi H2O	171
5.32	Tampilan vektor terunggah kelas <i>not related</i> pada perangkat lunak eksperimen studi kasus bagian prediksi H2O	171
5.33	Tampilan hasil prediksi untuk kelas <i>related</i> pada perangkat lunak eksperimen studi kasus bagian prediksi H2O	171
5.34	Tampilan hasil prediksi untuk kelas <i>not related</i> pada perangkat lunak eksperimen studi kasus bagian prediksi H2O	172
5.35	Tampilan awal pada perangkat lunak eksperimen studi kasus bagian prediksi PyCaret	172
5.36	Tampilan vektor terunggah kelas <i>related</i> pada perangkat lunak eksperimen studi kasus bagian prediksi PyCaret	172
5.37	Tampilan vektor terunggah kelas <i>not related</i> pada perangkat lunak eksperimen studi kasus bagian prediksi PyCaret	173
5.38	Tampilan hasil prediksi untuk kelas <i>related</i> pada perangkat lunak eksperimen studi kasus bagian prediksi PyCaret	173
5.39	Tampilan hasil prediksi untuk kelas <i>not related</i> pada perangkat lunak eksperimen studi kasus bagian prediksi PyCaret	173

5.40	Tampilan awal pada perangkat lunak eksperimen studi kasus bagian prediksi AutoGluon	174
5.41	Tampilan vektor terunggal kelas <i>related</i> pada perangkat lunak eksperimen studi kasus bagian prediksi AutoGluon	174
5.42	Tampilan vektor terunggal kelas <i>not related</i> pada perangkat lunak eksperimen studi kasus bagian prediksi AutoGluon	174
5.43	Tampilan hasil prediksi untuk kelas <i>related</i> pada perangkat lunak eksperimen studi kasus bagian prediksi AutoGluon	175
5.44	Tampilan hasil prediksi untuk kelas <i>not related</i> pada perangkat lunak eksperimen studi kasus bagian prediksi AutoGluon	175

DAFTAR TABEL

2.1	Contoh data pemasukan dengan fitur umur	11
2.2	Tabel Contoh Kasus Pembelian Komputer	18
2.3	Matriks Kemunculan Term di Setiap Dokumen	26
3.1	Lima data teratas dari <i>dataset</i> Housing.csv	40
3.2	Lima data teratas <i>CarPrice.csv</i>	42
3.3	Lima data teratas <i>CrabAgePrediction.csv</i>	43
3.11	Benchmark masalah–masalah regresi	49
3.12	<i>Benchmark</i> masalah–masalah klasifikasi	49
3.13	Informasi <i>Train/Test Split</i> Masalah Regresi	55
3.14	Informasi <i>Train/Test Split</i> Masalah Klasifikasi	55
3.15	Daftar Parameter Eksperimen H2O	58
3.16	Tabel Hasil Eksperimen	96
3.4	Lima data teratas <i>Life Expectancy.csv</i>	102
3.5	Lima data teratas <i>Medical expenses.csv</i>	102
3.6	Lima data teratas <i>dataset</i> IRIS.csv	103
3.7	Lima data teratas {Credit Score Classification Dataset.csv}	103
3.8	Lima data teratas <i>happydata.csv</i>	103
3.9	Lima data teratas <i>heart.csv</i>	103
3.10	Lima data teratas <i>nbaplayers.csv</i>	104
4.1	Lima data teratas data <i>link_berita.xlsx</i>	112
4.2	Lima data teratas <i>dataset_news_labelled_indexed.csv</i>	113
4.3	Contoh judul berita dari kelas Related dan Not Related	114
4.4	Tabel hasil eksperimen studi kasus	131
5.1	Deskripsi Pengguna pada Diagram Use Case	142
5.2	Deskripsi Use Case pada Diagram Use Case	142
5.3	Skenario use case mengunggah data	143
5.4	Skenario use case prediksi	143
5.5	Deskripsi Pengguna pada Diagram <i>Use Case</i> PL Eksperimen Studi Kasus <i>Preprocessing</i>	144
5.6	Deskripsi Use Case pada Diagram Use Case	144
5.7	Skenario use case mengunggah data teks	145
5.8	Skenario use case <i>preprocessing</i>	145
5.9	Skenario use case menyimpan hasil <i>preprocessing</i>	145
5.10	Deskripsi Pengguna pada Diagram <i>Use Case</i> PL Eksperimen Studi Kasus Prediksi	146
5.11	Deskripsi Use Case pada Diagram Use Case	146
5.12	Skenario use case mengunggah data PL eksperimen studi kasus	146
5.13	Skenario use case prediksi PL eksperimen studi kasus	147
5.14	Data pengujian untuk perangkat lunak eksperimen data sederhana	164

DAFTAR KODE PROGRAM

3.1	Kode Untuk <i>Import</i> PyCaret Regresi	67
3.2	Kode Untuk <i>Import</i> PyCaret Klasifikasi	67
3.3	Kode Untuk <i>Import library</i> Pendukung	67
3.4	Kode Untuk <i>Import</i> Dataset	68
3.5	Kode Untuk Melihat 5 Data Teratas <i>Car Price</i>	68
3.6	Kode Untuk Melihat Tipe Data Kolom–Kolom Dataset <i>Car Price</i>	68
3.7	Kode Untuk Mengubah Tipe Data Kolom <i>price</i> Dataset <i>Car Price</i>	69
3.8	Kode Untuk Menginisiasi <i>method setup</i> PyCaret	69
3.9	Kode Untuk Menjalankan AutoML PyCaret	70
3.10	Kode Untuk Mengevaluasi Model Terbaik	71
3.11	Kode Untuk Menyimpan Model Terbaik	71
3.12	Kode Untuk <i>Import Library</i> AutoGluon	83
3.13	Kode Untuk <i>Import Library</i> Pendukung	83
3.14	Kode Untuk <i>Import</i> Dataset	84
3.15	Kode Untuk Mencetak Lima Data Teratas	84
3.16	Membuat Variabel X dan y	84
3.17	Melakukan <i>Split</i> Dataset	84
3.18	Membuat Variabel Parameter	85
3.19	Pemanggilan <i>Method Fit</i>	85
3.20	Kode Pemanggilan <i>Leaderboard</i>	85
3.21	Kode untuk Melihat <i>Variable Importance</i>	86
3.22	Kode untuk Membuat <i>Confusion Matrix</i>	87
4.1	Kode untuk meng- <i>import library</i> yang diperlukan	106
4.2	Kode untuk membuat <i>method scroll</i>	107
4.3	Kode untuk membuat <i>method scrape_page</i>	108
4.4	Kode untuk membuat <i>method</i> untuk menangani <i>pagination</i>	109
4.5	Kode untuk menginisialisasi tautan berita dan <i>WebDriver</i>	109
4.6	Kode untuk membuat <i>loop</i> untuk mengambil judul dan tautan berita	110
4.7	Kode untuk meng- <i>export</i> judul dan tautan berita dalam format <i>csv</i>	110
4.8	Kode untuk melakukan <i>scraping</i> konten berita	110
4.9	Kode untuk membuang berita yang tidak memiliki teks	111
4.10	Kode untuk meng- <i>export</i> berita dalam format <i>txt</i>	112
4.11	Kode Untuk Membersihkan Teks	116
4.12	Kode Untuk Melakukan Tokenisasi	118
4.13	Kode Untuk Melakukan Vektorisasi	118
4.14	Kode untuk Meng- <i>import file</i>	126
4.15	Kode untuk <i>Setup</i> Eksperimen	126
4.16	Kode untuk Pelatihan Model	127
4.17	Kode untuk Menyimpan Model	128
4.18	Kode untuk Meng- <i>import</i> Dataset	129
4.19	Kode untuk <i>Split</i> Dataset	129

4.20	Kode untuk Pelatihan Model	129
4.21	Kode untuk Memanggil <i>Leaderboard</i>	130
5.1	Kode Untuk Mengimport Library yang Diperlukan	147
5.2	Kode Untuk Membuat <i>Method Load Model</i>	147
5.3	Kode Untuk Membuat Method Prediksi	148
5.4	Kode untuk inialisasi kelas <i>Application</i>	148
5.5	Kode untuk membuat tampilan	148
5.6	Kode untuk mendefinisikan <i>method upload file</i>	149
5.7	Kode untuk mendefinisikan <i>method predict</i> dalam kelas <i>Application</i>	150
5.8	Kode untuk mendefinisikan <i>method predict</i> dalam kelas <i>Application</i>	150
5.9	Kode Untuk Mengimport Library yang Diperlukan	151
5.10	Kode Untuk Membuat <i>Method</i> Pemuatan Model	151
5.11	Kode Untuk Membuat Method Prediksi	151
5.12	Kode untuk inialisasi kelas <i>Application</i>	152
5.13	Kode untuk membuat tampilan	152
5.14	Kode untuk mendefinisikan <i>method upload file</i>	152
5.15	Kode untuk mendefinisikan <i>method predict</i> dalam kelas <i>Application</i>	153
5.16	Kode untuk mendefinisikan <i>method predict</i> dalam kelas <i>Application</i>	153
5.17	Kode untuk meng- <i>import library</i> yang diperlukan	153
5.18	Kode untuk membuat <i>method</i> pemuatan model	154
5.19	Kode untuk melakukan <i>import library</i> yang diperlukan	154
5.20	Kode untuk membuat <i>method</i> tokenisasi	154
5.21	Kode untuk membuat <i>method</i> pemuatan <i>vectorizer</i>	154
5.22	Kode untuk membuat <i>method preprocessing</i>	154
5.23	Kode untuk membuat tampilan perangkat lunak	155
5.24	Kode untuk membuat <i>method</i> pemuatan <i>file, preprocessing, dan menyimpan vektor</i>	155
5.25	Kode untuk memuat <i>vectorizer</i> dan menjalankan perangkat lunak	156
5.26	Kode Untuk Mengimport Library yang Diperlukan	156
5.27	Kode Untuk Membuat <i>Method Load Model</i>	156
5.28	Kode Untuk Membuat Method Prediksi	156
5.29	Kode untuk inialisasi kelas <i>Application</i>	156
5.30	Kode untuk membuat tampilan	157
5.31	Kode untuk membuat <i>method</i> pengunggahan <i>file</i> dan <i>predict</i>	157
5.32	Kode untuk memuat model dan menjalankan perangkat lunak	157
5.33	Kode Untuk Mengimport Library yang Diperlukan	158
5.34	Kode Untuk Membuat <i>Method</i> Pemuatan Model	158
5.35	Kode untuk mendefinisikan <i>method predict</i> dalam kelas <i>Application</i>	158
5.36	Kode untuk memuat model dan menjalankan perangkat lunak	158
5.37	Kode untuk meng- <i>import library</i> yang diperlukan	159
5.38	Kode untuk membuat <i>method</i> pemuatan model	159
5.39	Kode untuk memuat model dan menjalankan perangkat lunak	159
A.1	Kode untuk mengambil sampel berita di tiap kuadran <i>confusion matrix</i>	181

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Menurut Karmaker et al. (2021), AutoML adalah sebuah paradigma untuk mengotomatisasi seluruh proses *machine learning* agar dapat diterapkan ke masalah dunia nyata [5]. AutoML dibuat dengan harapan dapat mengurangi beban pekerjaan dalam melakukan proses *data science*. AutoML juga dapat membantu *domain expert* yang memiliki kemampuan teknis minimal untuk dapat memanfaatkan kemampuan *data science*.

Dalam lima tahun terakhir, menurut observasi yang dilakukan menggunakan *Google Trends*, pencarian pengguna internet dunia terhadap kata kunci “*machine learning*” mengalami peningkatan. Hal tersebut diduga berhubungan dengan perkembangan yang dialami oleh bidang *machine learning* (ML) serta semakin tingginya tingkat adopsi ML. Dengan semakin tingginya adopsi ML maka praktisi-praktisi bidang lain (*domain expert*) yang ingin mengaplikasikan ML ke bidangnya masing-masing juga meningkat. Untuk menjembatani celah antara *domain expert* yang kekurangan pengetahuan tentang ML dan pengaplikasiannya maka diciptakan sebuah konsep bernama *automated machine learning* atau disingkat AutoML.

AutoML lazim digunakan karena adanya masalah di mana meningkatnya permintaan berbagai industri untuk *insight* dan pengambilan keputusan berbasis data namun kurangnya pemahaman praktisi *data science* terhadap domain spesifik. AutoML diharapkan dapat menutup celah tersebut dengan memfasilitasi ahli-ahli dari berbagai industri akses terhadap *data science* tanpa perlu memiliki kemampuan *machine learning* yang mendalam sehingga dapat memanfaatkan keunggulan yang disediakan oleh model pembelajaran mesin. AutoML juga lazim digunakan di kalangan praktisi *data science* karena AutoML mengatasi masalah biaya komputasi dan biaya waktu yang dihadapi ketika melakukan *tuning* parameter secara manual. Hal tersebut dikarenakan AutoML melakukan otomatisasi pada proses pencarian konfigurasi parameter paling optimal dengan cara yang lebih efisien.

Sebagai suatu sistem yang dapat memudahkan pengaplikasian ML, AutoML biasa digunakan dalam pengerjaan suatu proyek *data science*. *Data science* adalah bidang yang telah menjadi disiplin ilmu penting yang memanfaatkan kemampuan data dengan baik untuk menarik *insight* dan membantu pengambilan keputusan. *Data science* bertujuan untuk menganalisis data, mengungkap pola, dan membangun model prediktif, deskriptif, maupun preskriptif. Praktisi *data science* melakukan hal tersebut dengan menggabungkan berbagai teknik dari berbagai bidang seperti statistik, ilmu komputer, dan pengetahuan terhadap domain spesifik. Kemampuan *data science* ini membuatnya menjadi lazim digunakan di berbagai industri seperti kesehatan, keuangan, pemasaran, dan teknologi karena dapat mengoptimasi cara kerja organisasi-organisasi pada industri tersebut.

Data science memiliki beberapa tahapan yang dilakukan pada pembuatan model ML yang layak. Karmaker et al. (2021) merumuskan tahapan-tahapan tersebut adalah Task Formulation (TF), Data Visualization, Cleaning, and Curation (DCC), Prediction Engineering (PE), Feature Engineering (FE), Machine Learning (ML), Alternative Models Exploration, Testing, and Validation (ATV), dan Result Summary and Recommendation (RSR) [5]. Tahapan-tahapan tersebut membutuhkan pemahaman yang mendalam terhadap domain masalah yang akan dipecahkan, data

yang digunakan untuk dianalisis, serta algoritma pembelajaran mesin yang digunakan. Oleh sebab itu, dalam melakukan tahapan-tahapan pembuatan model dibutuhkan keahlian dalam berbagai bidang sekaligus.

Seorang praktisi *data science* dapat melakukan tahap yang bersifat teknis seperti FE, ML, dan ATV. Akan tetapi, karena luasnya domain penerapan *data science*, praktisi *data science* dituntut untuk menguasai banyak bidang sekaligus. Sebaliknya, seorang ahli suatu bidang atau biasa disebut *domain expert*, memiliki pemahaman yang mendalam terhadap suatu bidang sehingga menguasai tahapan yang bersifat strategis seperti TF, PE, dan RSR. Tahap DCC memiliki sifat berbeda di mana dibutuhkan pemahaman teknis dan strategis untuk melakukan visualisasi, pembersihan, dan kurasi data. Namun, masalah yang dihadapi *domain expert* adalah kurangnya pemahaman terhadap sisi teknis dari tahapan pembuatan model. Untuk itu, sebuah alat berupa AutoML dibuat untuk membantu keduanya sehingga diharapkan dapat meminimalisir kekurangan keduanya dan memaksimalkan kelebihan masing-masing.

Meski dengan kemajuan AutoML beberapa tahun terakhir yang pesat, tetap ada kekurangan yang penting dalam sistem AutoML. Beberapa tahapan pembuatan model seperti TF, PE, DCC, dan RSR sulit untuk dapat diotomatisasi oleh AutoML sehingga masih dilakukan secara manual oleh para praktisi *data science* maupun *domain expert*. Namun untuk tahapan yang lebih bersifat sedikit bergantung pada pemahaman domain spesifik, sedikit membutuhkan interpretasi manusia, dan membutuhkan banyak pengulangan seperti FE, ML, dan ATV dapat dengan baik diotomatisasi oleh AutoML sehingga dapat mengurangi pekerjaan dalam pembuatan model. Untuk *domain expert*, AutoML bisa membantu pada saat melakukan tahapan yang teknis seperti FE, ML, dan ATV yang mana mereka kurang memiliki kemampuan untuk melakukan tahapan-tahapan tersebut.

Pada penelitian ini, akan dipelajari dan digunakan tiga *library* AutoML yaitu H2O [6], PyCaret [7], dan AutoGluon [8]. Ketiga *library* tersebut dipilih sebab masing-masing memiliki algoritma ML dan metrik evaluasi yang tidak dimiliki satu sama lain. Pada ketiga *library* tersebut akan dilakukan studi eksplorasi untuk mengetahui cara penggunaan, perbandingan hasil, dan pemanfaatan *library* AutoML H2O, PyCaret, dan AutoGluon.

Untuk mengetahui pemanfaatan dari *library* AutoML H2O, PyCaret, dan AutoGluon, akan dilakukan studi kasus menggunakan data asli dari masalah dunia nyata. Masalah yang dipilih adalah masalah klasifikasi untuk pemberitaan politik di Indonesia. Masalah tersebut berpusat pada bagaimana menyaring berita-berita politik tentang kandidat politik yang mengandung substansi (kebijakan, program kerja, janji kampanye, dan rekam jejak). Data yang digunakan adalah teks berita politik tentang pemilihan presiden Indonesia tahun 2024 yang diambil melalui situs berita Detik. Oleh karena data tersebut berbentuk teks, maka data bersifat tidak terstruktur.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dibahas pada bagian sebelumnya, berikut merupakan masalah-masalah yang akan diselesaikan:

1. Bagaimana cara menggunakan *library* AutoML H2O, PyCaret, dan AutoGluon?
2. Bagaimana perbandingan kualitas model yang dihasilkan masing-masing *library* AutoML H2O, PyCaret, dan AutoGluon?
3. Bagaimana cara memanfaatkan *library* AutoML H2O, PyCaret, dan AutoGluon pada studi kasus data berita politik di Indonesia?
4. Bagaimana meluncurkan model yang telah dibuat menggunakan AutoML H2O, PyCaret, dan AutoGluon?

1.3 Tujuan

Berdasarkan masalah-masalah yang sudah disebutkan di bagian sebelumnya, tujuan dari dibuatnya skripsi ini adalah sebagai berikut:

1. Melakukan studi eksplorasi dan eksperimen analisis data menggunakan *library* AutoML H2O, PyCaret, dan AutoGluon
2. Melakukan perbandingan kinerja *library* AutoML H2O, PyCaret, dan AutoGluon dengan serangkaian eksperimen analisis data
3. Memanfaatkan *library* AutoML H2O, PyCaret, dan AutoGluon untuk melakukan analisis data studi kasus data berita politik di Indonesia
4. Membuat perangkat lunak yang memuat model hasil dari AutoML H2O, PyCaret, dan AutoGluon

1.4 Batasan Masalah

Ada beberapa batasan masalah pada pengerjaan skripsi ini, yaitu sebagai berikut:

1. Masalah yang digunakan untuk menguji *library* AutoML hanya regresi dan klasifikasi karena hanya PyCaret yang mendukung pembuatan model *clustering*
2. Pada bagian studi kasus hanya memecahkan masalah klasifikasi karena beban pekerjaan terlalu besar apabila mengikutsertakan masalah regresi

1.5 Metodologi

Metodologi yang digunakan dalam melakukan skripsi ini adalah sebagai berikut:

1. Melakukan studi literatur tentang *Machine Learning*
Topik utama dalam skripsi ini adalah sistem AutoML sehingga langkah pertama yang dilakukan adalah studi literatur tentang *machine learning*
2. Melakukan studi literatur tentang AutoML.
Topik utama dalam skripsi ini adalah sistem AutoML sehingga langkah kedua setelah mempelajari *machine learning* adalah studi literatur tentang AutoML
3. Melakukan studi literatur tentang AutoML H2O, PyCaret, dan AutoGluon
Setelah mempelajari AutoML maka dipilih 3 *library* yang akan dicoba dipelajari dan dimanfaatkan
4. Melakukan pengumpulan data sederhana untuk studi eksplorasi pada ketiga *library* AutoML H2O, PyCaret, dan AutoGluon.
Melakukan pengumpulan data sederhana untuk digunakan dalam studi eksplorasi dan mempelajari cara kerja AutoML H2O, PyCaret, dan AutoGluon
5. Melakukan *benchmarking* untuk studi eksplorasi data sederhana
Mencari hasil eksperimen pada data yang dikumpulkan untuk dibandingkan dengan hasil eksperimen oleh AutoML H2O, PyCaret, dan AutoGluon
6. Melakukan studi eksplorasi AutoML H2O, PyCaret, dan AutoGluon pada data sederhana
Melakukan studi eksplorasi pada data sederhana untuk mempelajari cara kerja AutoML H2O, PyCaret, dan AutoGluon serta membandingkan hasil eksperimen satu sama lain dan membandingkan hasil eksperimen dengan nilai *benchmark*
7. Melakukan studi kasus AutoML H2O, PyCaret, dan AutoGluon pada data teks berita
Melakukan studi kasus pada data teks berita untuk dapat memanfaatkan kemampuan AutoML H2O, PyCaret, dan AutoGluon
8. Membuat perangkat lunak
Membuat perangkat lunak yang dapat meluncurkan model-model hasil AutoML H2O, PyCaret, dan AutoGluon
9. Mendokumentasikan studi eksplorasi dan pemanfaatan sistem AutoML
Seluruh hasil dan proses dalam studi eksplorasi dan pemanfaatan sistem AutoML didokumentasikan untuk pengembangan lebih lanjut

1.6 Sistematika Pembahasan

Sistematika pembahasan pada penulisan skripsi ini adalah berikut:

- Bab 1 Pendahuluan
Bab ini membahas bagaimana latar belakang masalah sehingga dibuatnya skripsi ini, rumusan masalah, tujuan penelitian, batasan masalah, metodologi penelitian, dan sistematika pembahasan.
- Bab 2 Landasan Teori
Bab ini membahas macam-macam teori yang melandasi skripsi ini seperti *data science*, *machine learning*, metrik evaluasi model, analisis data teks dan *information retrieval*, AutoML, H2O, PyCaret, dan AutoGluon.
- Bab 3 Studi Eksplorasi
Bab ini membahas proses mempelajari AutoML H2O, PyCaret, dan AutoGluon menggunakan data sederhana serta membandingkan hasil dari eksperimen.
- Bab 4 Pembuatan Model Studi Kasus
Bab ini membahas pemanfaatan AutoML H2O, PyCaret, dan AutoGluon pada kasus data teks.
- Bab 5 Pembangunan Perangkat Lunak dan Pengujian
Bab ini membahas proses pembuatan perangkat lunak yang mampu meluncurkan model-model yang sudah dibuat.
- Bab 6 Kesimpulan dan Saran
Bab ini membahas mengenai kesimpulan yang didapatkan selama skripsi dibuat dan saran untuk penelitian di bidang serupa kedepannya.