

ABSTRAK

Clarissa Angelia:

Skripsi

Penerapan *Rule-Based* dan *Intent Classification* dalam Chatbot Konsultasi Proposal
Tugas Akhir Mahasiswa Universitas Kristen Petra

Dalam proses penyusunan skripsi, mahasiswa sering menghadapi berbagai kendala seperti keterbatasan pemahaman dan kesulitan dalam konsultasi dengan dosen pembimbing. Penelitian ini mengusulkan penggunaan chatbot sebagai solusi untuk membantu proses konsultasi skripsi mahasiswa program studi Informatika di Universitas Kristen Petra. Chatbot yang dikembangkan menggunakan metode Intent Classification berbasis LSTM dengan model Pre-trained Word2Vec untuk meningkatkan akurasi dalam memahami input user. Percobaan dilakukan dengan membandingkan tiga jenis chatbot: Rule-Based, Hybrid, dan Scripted Chatbot. Hasil penelitian menunjukkan bahwa model Word2Vec terbaik memiliki dimensi 100 dengan epoch optimal 30. Model LSTM dengan dua LSTM layer dengan parameter 128 dan 64 hidden unit cell, masing-masing menggunakan dropout dan recurrent_dropout 0.2 sebagai hyperparameter menghasilkan akurasi tertinggi sebesar 95,60% pada dataset kecil. Meskipun Hybrid Model memiliki kepuasan pengguna paling rendah, model ini menunjukkan keunggulan dalam hal efisiensi dan pemahaman pertanyaan yang lebih kompleks. Dari uji kepuasan pengguna, diketahui bahwa Scripted Chatbot lebih disukai karena kemudahannya dalam penggunaan. Chatbot yang dikembangkan diharapkan dapat digunakan secara fungsional untuk membantu proses konsultasi proposal skripsi mahasiswa.

Kata kunci: metode hibrida, klasifikasi intent, LSTM, chatbot, konsultasi proposal skripsi

ABSTRACT

Clarissa Angelia:

Undergraduate Thesis

Implementation of Rule-Based and Intent Classification in a Chatbot for Consultation on Thesis Proposals for Petra Christian University Students

In the process of thesis preparation, students often face various challenges such as limited understanding and difficulty in consulting with their thesis supervisors. This study proposes the use of a chatbot as a solution to aid in the thesis consultation process for students in the Informatics program at Petra Christian University. The developed chatbot utilizes Intent Classification based on LSTM with Pre-trained Word2Vec models to enhance accuracy in understanding user inputs. The experiment compares three types of chatbots: Rule-Based, Hybrid, and Scripted Chatbots. The research findings reveal that the optimal Word2Vec model has a dimension of 100 with 30 epochs. The LSTM model with two layers, featuring 128 and 64 hidden units each, along with dropout and recurrent_dropout set at 0.2 as hyperparameters, achieved the highest accuracy of 95.60% on a small dataset. Despite the Hybrid Model exhibiting the lowest user satisfaction, it demonstrates advantages in efficiency and comprehension of complex queries. User satisfaction testing indicates a preference for Scripted Chatbots due to their ease of use. The developed chatbot is expected to functionally assist in the thesis consultation process for students.

Keywords: hybrid method, intent classification, LSTM, chatbot, thesis proposal consultation

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN.....	ii
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	iii
KATA PENGANTAR.....	iv
ABSTRAK.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR SEGMENT PROGRAM.....	xii
1. PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan Penelitian.....	3
1.4. Ruang Lingkup.....	3
1.5. Metodologi Penelitian.....	5
1.6. Manfaat Penelitian.....	6
1.7. Sistematika Penulisan.....	7
2. LANDASAN TEORI.....	8
2.1. Tinjauan Pustaka.....	8
2.1.1. Chatbot.....	8
2.1.2. Rule-Based Chatbot.....	8
2.1.3. Scripted Chatbot.....	8
2.1.4. Intent Classification.....	8
2.1.5. Long Short Term Memory (LSTM).....	9
2.1.6. Word2Vec.....	9
2.1.7. Word2Vec Metode Dieka Nugraha (2018).....	10
2.1.8. Word2Vec Metode Eko Setiawan (2023).....	10
2.2. Tinjauan Studi.....	12
2.2.1. Sistem Chatbot dengan Metode Naive Bayes Untuk Menanggapi Pertanyaan Mahasiswa Tentang Pra Skripsi (Giovanni, 2023).....	12
2.2.2. Intent Classification in Question-Answering Using LSTM Architectures (Gennaro et.al, 2020).....	12
2.2.3. Implementasi Long Short-Term Memory pada Chatbot Informasi Akademik Teknik Informatika Universitas Lampung (Wintoro et.al, 2022).....	13
2.2.4. Rule-Based Chatbot for Student Enquiries (Singh et.al, 2019).....	13
3. ANALISA DAN DESAIN SISTEM.....	15
3.1. Analisis Permasalahan dan Kebutuhan.....	15

3.2. Pengolahan dan Analisis Data.....	16
3.3. Desain Sistem.....	21
3.3.1. Desain Intent Classification LSTM.....	21
3.3.2. Text Preprocessing.....	22
3.3.3. Word Embedding.....	23
3.3.4. Model LSTM.....	26
3.3.5. Desain Rule-Based Chatbot.....	26
3.3.6. Hierarchical Rule Structure.....	27
3.4. Desain UI.....	49
3.4.1. Desain UI Pengguna.....	51
4. IMPLEMENTASI SISTEM.....	52
4.1. Implementasi Text Preprocessing.....	52
4.2. Implementasi Word2Vec.....	53
4.2.1. Pre-Trained Word2Vec.....	53
4.2.2. Fine-tuning Word2Vec Model.....	55
4.3. Implementasi Model LSTM.....	58
4.4. Implementasi Rule-Based Model.....	63
4.5. Implementasi Hybrid Model Chatbot.....	65
5. PENGUJIAN.....	68
5.1. Pengujian Model Pre-Trained Word2Vec.....	68
5.1.1. Pengujian menggunakan Referensi Dieka Nugraha (2018).....	68
5.1.2. Pengujian menggunakan Referensi Eko Setiawan (2023).....	69
5.1.3. Pengujian menggunakan Metode Gabungan (Dieka Nugraha dan Eko Setiawan)...	69
5.2. Pengujian Fine-Tuning Word2Vec.....	71
5.2.1. Fine-Tuning Percobaan 1.....	72
5.2.2. Fine-Tuning Percobaan 2.....	75
5.3. Pengujian Model LSTM.....	76
5.4. Pengujian Survey Pengguna.....	78
5.4.1. Hybrid Chatbot.....	79
5.4.2. Rule-Based Chatbot.....	80
5.4.3. Scripted Chatbot.....	81
5.5. Pengujian Running Time.....	96
5.5.1. Pengujian Running Time Rule Based.....	96
5.5.2. Pengujian Running Time Hybrid Model.....	100
6. KESIMPULAN DAN SARAN.....	103
6.1. Kesimpulan.....	103
6.2. Saran.....	104
DAFTAR REFERENSI.....	105

DAFTAR TABEL

Tabel 4.1 Implementasi Desain ke Pemrograman.....	52
Tabel 5.1 Hasil Pengujian Menggunakan Metode Dieka Nugraha.....	68
Tabel 5.2 Hasil Pengujian Menggunakan Metode Eko Setiawan.....	69
Tabel 5.3 Hasil Pengujian Menggunakan Metode Gabungan.....	70
Tabel 5.4 Hasil Pengujian Fine Tuning Menggunakan Percobaan 1.....	72
Tabel 5.5 Lanjutan Hasil Pengujian ada Tabel 5.4 dengan Modifikasi Program.....	73
Tabel 5.6 Lanjutan Hasil Pengujian pada Tabel 5.5 dengan Modifikasi Dataset.....	73
Tabel 5.7 Hasil Pengujian Menggunakan Dataset yang Telah Dimodifikasi.....	74
Tabel 5.8 Hasil PengujianFine-Tuning Menggunakan Percobaan 2.....	75
Tabel 5.9 Hasil Pengujian Akurasi pada Model 1.....	76
Tabel 5.10 Hasil Pengujian Akurasi pada Model 2.....	77
Tabel 5.11 Hasil Pengujian Akurasi pada Model 3.....	77
Tabel 5.12 Hasil Pengujian Akurasi pada Model 4.....	77
Tabel 5.13 Rata-Rata Penilaian Survei Jenis Chatbot.....	81
Tabel 5.14 Data Prediksi Intent pada Hybrid Chatbot.....	82
Tabel 5.15 Data Analisa Pertanyaan Tidak Terjawab pada Hybrid Chatbot.....	90
Tabel 5.16 Data Prediksi Intent pada Rule-Based Chatbot.....	91
Tabel 5.17 Data Analisa Pertanyaan Tidak Terjawab pada Rule-Based Chatbot.....	94

DAFTAR GAMBAR

Gambar 2.1 Visualisasi LSTM.....	9
Gambar 2.2 Perbandingan model CBOW dan Skip-gram.....	10
Gambar 3.1 Dataset Chatbot Worksheet 'Dataset Chat UNFILTERED'.....	18
Gambar 3.2 Dataset Chatbot Worksheet 'Dataset chat'.....	19
Gambar 3.3 Dataset Chatbot Worksheet 'Dataset ss'.....	20
Gambar 3.4 Dataset Chatbot Worksheet 'Intent List'.....	20
Gambar 3.5 Dataset Chatbot Worksheet 'Data Augment'.....	20
Gambar 3.6 Dataset Chatbot pada file JSON.....	21
Gambar 3.7 Desain Sistem Intent Classification Training using LSTM.....	22
Gambar 3.8 Flowchart Text Preprocessing.....	23
Gambar 3.9 Desain Sistem Word2Vec Training.....	24
Gambar 3.10 Desain Sistem Fine-Tuning Word2Vec.....	25
Gambar 3.11 Struktur Model LSTM.....	26
Gambar 3.12 Desain Sistem Rule Based Chatbot.....	27
Gambar 3.13 Desain Sistem Hybrid Chatbot.....	50
Gambar 3.14 Desain tampilan chatbot.....	51
Gambar 5.1 Perbandingan Struktur Model LSTM.....	76
Gambar 5.2 Hasil Pertanyaan 1 pada Hybrid Chatbot.....	79
Gambar 5.3 Hasil Pertanyaan 2 pada Hybrid Chatbot.....	79
Gambar 5.4 Hasil Pertanyaan 3 pada Hybrid Chatbot.....	79
Gambar 5.5 Hasil Pertanyaan 1 pada Rule-Based Chatbot.....	80
Gambar 5.6 Hasil Pertanyaan 2 pada Rule-Based Chatbot.....	80
Gambar 5.7 Hasil Pertanyaan 3 pada Rule-Based Chatbot.....	80
Gambar 5.8 Hasil Pertanyaan 1 pada Scripted Chatbot.....	80
Gambar 5.9 Hasil Pertanyaan 2 pada Scripted Chatbot.....	81
Gambar 5.10 Hasil Pertanyaan 3 pada Scripted Chatbot.....	81

DAFTAR SEGMENT PROGRAM

Segment Program 4.1 Text Preprocessing.....	52
Segment Program 4.2 Fetch data, create, and merge corpus from wikidump file.....	53
Segment Program 4.3 Case folding, Cleaning, and tokenization data.....	54
Segment Program 4.4 Train Word2Vec Model.....	54
Segment Program 4.5 Testing Word2Vec model.....	55
Segment Program 4.6 Change dataset format from CSV to JSON.....	55
Segment Program 4.7 Preprocess chat dataset.....	56
Segment Program 4.8 Concatenate intent with designated pattern.....	56
Segment Program 4.9 Load Pretrained Word2Vec Model.....	57
Segment Program 4.10 Fine Tuning Word2Vec.....	57
Segment Program 4.11 Testing FineTuned Word2Vec.....	58
Segment Program 4.12 Fetch dataset and building tokenizer vocabulary.....	58
Segment Program 4.13 Change text data to sequence.....	59
Segment Program 4.14 Apply padding to sequences.....	59
Segment Program 4.15 Format Dataset to X and Y.....	59
Segment Program 4.16 Splitting the dataset.....	60
Segment Program 4.17 Creating embedding matrix using fine-tuned Word2Vec model.....	61
Segment Program 4.18 Building LSTM Model.....	61
Segment Program 4.19 Training LSTM Model.....	61
Segment Program 4.20 Accuracy and loss of the model.....	62
Segment Program 4.21 Evaluate the model.....	63
Segment Program 4.22 Format response dataset from CSV to JSON.....	63
Segment Program 4.23 Prepare rules for rule based model.....	63
Segment Program 4.24 Word embedding function.....	64
Segment Program 4.25 Cosine similarity matching.....	64
Segment Program 4.26 Preparing rules for specified top level intent.....	65
Segment Program 4.27 Change in code segment 4.26 for hybrid chatbot usage.....	65
Segment Program 4.28 Load LSTM Model.....	66
Segment Program 4.29 Hybrid Chatbot implementation.....	66