INISIASI *NATURAL LANGUAGE PROCESSING* (NLP) DAN KLASIFIKASI JENIS WISATA KULINER UNTUK PROGRAM *CHATBOT*

(Studi Kasus: Informasi Wisata Kuliner Daerah Istimewa Yogyakarta)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana
Jurusan Statistika



Disusun Oleh:

Listari

15 611 064

JURUSAN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS ISLAM INDONESIA

YOGYAKARTA

2019

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul : INISIASI *NATURAL LANGUAGE PROCESSING* (NLP) DAN KLASIFIKASI JENIS WISATA KULINER UNTUK PROGRAM *CHATBOT*

Nama : Listari

Nomor Mahasiswa : 15 611 064

TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK DIUJIKAN

Yogyakarta, Januari 2019

Pembimbing

Ayundyah Kesumawati, S.Si., M.Si.

HALAMAN PENGESAHAN

TUGAS AKHIR

INISIASI *NATURAL LANGUANGE PROCESSING* (NLP) DAN KLASIFIKASI JENIS WISATA KULINER UNTUK PROGRAM *CHATBOT*

(Studi Kasus : Informasi Wisata Kuliner Daerah Istimewa Yogyakarta)

Nama Mahasiswa : Listari

Nomor Mahasiswa : 15 611 064

TUGAS AKHIR INI TELAH DIUJIKAN

PADA TANGGAL

Nama Penguji:	**Tanda Tangan**
1. Dr. Jaka Nugraha, S.Si., M.Si.	.....
2. Arum Handini Primandari, S.Pd.Si., M.Si.	.....
3. Ayundyah Kesumawati, S.Si., M.Si	.....

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Prof. Riyanto, S.Pd., M.Si., Ph.D.

KATA PENGANTAR



Assalamualaikum Wr.Wb

Alhamdulillah, puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayah-Nya berupa keimanan, kekuatan, kesabaran, kelancaran serta keselamatan selama melaksanakan penulisan hingga laporan ini dapat terselesaikan. Shalawat serta salam penulis haturkan kepada Nabi Muhammad SAW beserta keluarga, sahabat, dan umatnya.

Penulisan ini berjudul "**INISIASI *NATURAL LANGUAGE PROCESSING* (NLP) DAN KLASIFIKASI JENIS WISATA KULINER UNTUK PROGRAM *CHATBOT***". Penulis menyadari bahwa Penulisan laporan ini banyak memperoleh bantuan dari berbagai pihak, baik yang berupa saran, kritik, bimbingan maupun bantuan lainnya. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada:

1. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam.
2. Bapak Dr. Edy Widodo, M.Si., selaku Ketua Jurusan Statistika beserta jajarannya.
3. Ibu Ayundyah Kesumawati, S.Si., M.Si., selaku Dosen Pembimbing penulisan yang telah sabar membimbing.
4. Ibu Dr.techn. Rohmatul Fajriyah, S.Si., M.Si., selaku Dosen Pembimbing Akademik yang memberikan arahan dan motivasi dalam mengemban ilmu di Jurusan Statistika.
5. Seluruh staff pengajar Program Studi Statistika Universitas Islam Indonesia yang telah memberikan bekal ilmu selama menimba ilmu di universitas ini.
6. Ibu, Bapak, Kak Sofia Laura, Adik Muhammad Akbar dan Keluarga Besar yang selalu memberikan dukungan, semangat serta mendo'akan yang terbaik.

7. Sahabat Tralatrili sekaligus keluarga penulis yaitu Dita Apriani, Nita Tri Anggarini, Fitri Khirani, Vika A., Halimah S.R., yang senantiasa berbagi suka cita, pengalaman luar biasa, serta menemani selama perkuliahan.
8. Wilda Amalia L. & Yuli C. sahabat terbaik sejak kecil yang seantiasa memberikan dukungan dan do'a selama ini.
9. Sahabat AOB yaitu Taufan, Rafi'i, Rohmat, Dicky, Tihat, dan Tralatrili yang selalu memberi warna selama perkuliahan.
10. Teman-teman seperjuangan sekaligus sebimbingan yaitu Nita, Meymunah, Nida, Amdhu, Meila, Mimi, Hani dan Rhesa yang sudah menemani dan memberikan bantuan dalam mengerjakan tugas akhir ini
11. Pengurus Inti IKS (Ikatan Keluarga Statistika) periode 2017/2018, Wildan Al Azkia, Shelly Ila Amalia, Nilda Aulia, Meymunah, Diana, Salwa Y.G., Arie Syahputra, Tihat Jaya Kusuma, Galang Imanta, Sheila Diba dan Yusuf yang sudah berbagi pengalaman yang luar biasa dalam berorganisasi selama satu periode.
12. Pengurus Keilmuan IKS periode 2017/2018, Iqbal Fathur Rahman, Nurhikmah, Barlinda Titania, Bayu Galih Prianda, Qolbiyatul Lina, Salsabilla Novianingrum dan Tania Salsabila yang telah banyak memberikan pengalaman, waktu dan tenaga selama kepengurusan sekaligus senantiasa memotivasi penulis dalam menyelesaikan tugas akhir ini.
13. Sahabat AFFINISTIC (Statistika 2015) yang sudah banyak memberikan semangat dan bantuan dalam memulai dan mengakhiri tugas Tugas Akhir ini.
14. Semua pihak yang telah membantu dan tidak dapat penulis sebutkan satu per satu, terima kasih.

Penulis menyadari sepenuhnya bahwa tugas akhir ini masih jauh dari sempurna. Oleh karena itu segala kritik dan saran yang sifatnya membangun selalu penulis harapkan. Semoga tugas akhir dapat bermanfaat bagi penulis khususnya dan bagi semua yang membutuhkan. Akhir kata, semoga Allah SWT

selalu melimpahkan rahmat serta hidayah-Nya kepada kita semua, Amin amin ya robbal ‘alamiin

Wassalamu’alaikum, Wr.Wb.

Yogyakarta, Januari 2019

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN PEMBIMBING ..... **ii**
HALAMAN PENGESAHAN ..... **iii**
KATA PENGANTAR ..... **iv**
DAFTAR ISI ..... **vii**
DAFTAR TABEL ..... **x**
DAFTAR GAMBAR ..... **xi**
DAFTAR LAMPIRAN ..... **xiii**
PERNYATAAN ..... **xiv**
INTISARI ..... **xv**
ABSTRACT ..... **xvi**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 5
1.3 Batasan Masalah 5
1.4 Tujuan Penulisan 6
1.5 Manfaat Penulisan 6
BAB II TINJAUAN PUSTAKA **8**
BAB III LANDASAN TEORI **13**
3.1 Pengertian Wisata Kuliner 13
3.2 Analisis Deskriptif 13
3.3 *Artificial Intelligence* 14
3.3.1 *Machine Learning* 15
3.3.2 *Deep Learning* 16
3.3.2 *Natural Language Processing* 17
3.4 *Chatbot* 21
3.5 *Text Mining* 23
3.5.1 *Text Prepocessing* 23

3.5.2 Pembobotan Kata *Term Frequency–Inverse Document Frequency* 24
3.5.3 *Topic Modeling* 26
3.5.5 *Topic Coherence* 29
3.6 Data *Training* dan Data *Testing* 30
3.7 Klasifikasi 30
3.7.1 *Support Vector Machine* (SVM) 31
3.7.2 *Kernel Trick* dan *Non Linear* SVM 33
3.7.3 *Confusion Matrix* 34
3.8 *YAML Ain't Markup Language* (YAML) 35
BAB IV METODELOGI PENELITIAN 37
4.1 Populasi dan Sampel 37
4.2 Jenis dan Sumber Data 39
4.3 Variabel Penulisan 39
4.4 Metode Analisis Data 40
4.5 Tahapan Analisis / Diagram Alir 41
4.5.1 Pengumpulan Data Survei melalui Kuesioner *Google Form* 42
4.5.2 Tahapan Pemodelan Topik menggunakan LDA 44
4.5.3 Tahapan Klasifikasi dengan SVM 46
4.5.4 Tahapan Rancangan Sistem Program *Chatbot* 47
BAB V HASIL DAN PEMBAHASAN 50
5.1 Analisis Deskriptif (Hasil Studi Pendahuluan) 50
5.2 *Topic Modelling* 57
5.2.1 *Preprocessing* 57
5.2.2 *Lower Casing* 58
5.2.3 *Remove Punctuation* 58
5.2.4 *Stopwords Removal* 58
5.2.5 *Tokenizing* 59

5.2.6 Pembobotan *Term Frequency-Inverse Document Frequency* 59
5.2.7 Hasil *Topic Modeling* dengan *Latent Dirichlet Allocation* 62
5.3 Klasifikasi 67
5.3.1 Pembuatan Data *Training* dan Data *Testing* 68
5.3.2 Hasil Klasifikasi *Support Vector Machine (SVM)* 68
5.3.3 Prediksi Kategori untuk Data Pertanyaan Baru 73
5.4 *Chatbot* 74
5.4.1 Mekanisme Respon dari *Chatbot* 74
5.4.2 Pengujian *Chatbot* 77
5.43 Contoh Percakapan Program *Chatbot* 79
BAB VI KESIMPULAN DAN SARAN 80
6.1 Kesimpulan 80
6.2 Saran 81
DAFTAR PUSTAKA 82
LAMPIRAN 88

DAFTAR TABEL

Tabel 2. 1 Perbandingan dengan Penulisan Terdahulu 11
Tabel 3. 1 Kernel yang Umum Dipakai dalam SVM 34
Tabel 3. 2 *Confusion Matrix* 34
Tabel 4.1 Definisi Operasional Variabel Dataset 1 39
Tabel 4.2 Definisi Operasional Variabel Dataset 2 40
Tabel 4.3 Pengelompokan Tempat Kuliner 40
Tabel 4.4 Pertanyaan Kuesioner 42
Tabel 5. 1 Tabel Kontingensi Indentitas Responden 50
Tabel 5. 2 Tabel Kontingensi Pekerjaan dan Pendapatan Responden 51
Tabel 5. 3 Data Awal Penulisan 57
Tabel 5. 4 Hasil *Lower Casing* 58
Tabel 5. 5 Hasil *Remove Punctuation* 58
Tabel 5. 6 Hasil *Stopwords Removal* 59
Tabel 5. 7 Hasil *Tokenizing* 59
Tabel 5. 8 Sampel Hasil Pembobotan Kata TF-IDF 60
Tabel 5. 9 Sampel Hasil Nilai TF 61
Tabel 5. 10 Sampel Hasil Perhitungan TF-IDF Kata"asli" 62
Tabel 5. 11 *Coherence Score* 63
Tabel 5. 12 Rincian Data *Training* dan Data *Testing* 68
Tabel 5. 13 Perbandingan Hasil Klasifikasi SVM dengan Empat Kernel 72
Tabel 5. 14 Hasil Prediksi Kategori untuk Data Pertanyaan Baru 74
Tabel 5. 15 Hasil Pengujian *Chatbot* 77

DAFTAR GAMBAR

Gambar 1. 1 Persentase Persetujuan Pengadaan Program *Chatbot* Kulinet.....3
Gambar 3. 1 Bagian Utama dalam Artificial Intelligence (AI).....14
Gambar 3. 2 Struktur Kalimat Bahasa Indonesia.....18
Gambar 3. 3 Hubungan Beberapa Teknologi di Balik Program *Chatbot* 21
Gambar 3. 4 Contoh Percakapan dengan Rinna.....22
Gambar 3. 5 Metode *Latent Dirichlet Allocation* (LDA) (Blei, 2012).....27
Gambar 3. 6 Model representasi LDA (Blei, Ng, & Jordan, 2003).....28
Gambar 3. 7 SVM Berusaha Menemukan *Hyperplane* Terbaik.....32
Gambar 3. 8 Fungsi Φ Memetakan Data ke Ruang Vektor 33
Gambar 4. 1 Diagram Venn hubungan dataset 1 dan daraset 2 38
Gambar 4. 2 Tahapan Analisis Penelitian.....41
Gambar 4. 3 Tahapan Analisis *Topic Modeling* dengan LDA.....44
Gambar 4. 4 Tahapan Klasifikasi dengan SVM.....46
Gambar 4. 5 Tahapan Rancangan Sistem Program *Chatbot*.....48
Gambar 5. 1 Persentase Orang Suka Beriwisata dan Sering Beriwisata.....52
Gambar 5. 2 Persentase Responden Berwisata Kuliner 51
Gambar 5. 3 Persentase sumber perolehan informasi wisata kuliner.....52
Gambar 5. 4 Persentase Kendala Ketika Melakukaan Wisata Kuliner 54
Gambar 5. 5 Persentase Kriteria Kuliner yang Diinginkan.....554
Gambar 5. 6 Persentase Pengetahuan tentang *Chatbot*.....55
Gambar 5. 7 Persentase Penggunaan *Chatbot* tentang Wisata Kuliner.....56
Gambar 5. 8 Persentase Persetujuan Pengadaan Program *Chatbot* 56
Gambar 5. 9 Grafik *Coherence Score* terhadap Banyaknya Topik.....63
Gambar 5. 10 Visualisasi *Topic Modeling* dengan PyLDAvis 64
Gambar 5. 11 Visualisasi Topik 1.....65
Gambar 5. 12 Visualisasi topik 2 65
Gambar 5. 13 Visualisasi Topik 3.....66
Gambar 5. 14 *Wordcloud* Kata Terpopuler Pada Tiap Topik 67
Gambar 5. 15 *Confussion Matrix* SVM Kernel *Linear*.....69

Gambar 5. 16 *Confussion Matrix* SVM Kernel *Polynomial* 70
Gambar 5. 17 *Confussion Matrix* SVM Kernel *Sigmoid* 71
Gambar 5. 18 *Confussion Matrix* SVM Kernel *RBF* 72
Gambar 5. 19 Hasil Prediksi Kategori untuk Data Pertanyaan Baru 753
Gambar 5. 20 Pola Pertanyaan Jawaban untuk Kategori Perkenalan 75
Gambar 5. 21 Pola Pertanyaan Jawaban untuk Kategori *Cafe & Bakeries* 75
Gambar 5. 22 Pola Pertanyaan Jawaban untuk Kategori *Restaurants* 76
Gambar 5. 23 Contoh Percakapan Program *Chatbot* 79

DAFTAR LAMPIRAN

Lampiran 1 Dataset 1 (Data Pertanyaan) yang merupakan Hasil Survei……...…89
Lampiran 2 Dataset 2 (Data hasil *scrapping* di TripAdvisor)…….....90
Lampiran 3 *Script* dan *Output Topic Modeling*……...……92
Lampiran 4 *Output* Visualisasi *Topic Modeling* dengan LDA……...96
Lampiran 5 *Script* dan *Output* Klasifikasi SVM……99
Lampiran 6 *Script Chatbot* Teman Jajan……105
Lampiran 7 Pengujian Mesin *Chatbot* menggunakan 10 Pertanyaan baru……...106

PERNYATAAN

Dengan ini saya sebagai penulis menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan disuatu Perguruan Tinggi dan sepanjang pengetahuan penulis juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan diterbitkan dalam daftar pustaka.

Yogyakarta, 27 Januari 2019



Listari

INISIASI *NATURAL LANGUAGE PROCESSING* (NLP) DAN KLASIFIKASI JENIS WISATA KULINER UNTUK PROGRAM *CHATBOT*

(Studi Kasus : Informasi Wisata Kuliner Daerah Istimewa Yogyakarta)

Listari

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Islam Indonesia

INTISARI

Saat ini teknologi informasi berkembang pesat dalam berbagai bidang, termasuk di bidang kuliner. Wisata kuliner berpotensi menaikkan pendapatan daerah dan menjadi perhatian tersendiri oleh pemerintah maupun pihak-pihak pengelola yang berada dibidang tersebut. Yogyakarta memiliki ragam kuliner yang mampu manarik wisatawan, namun informasi kuliner tidak selalu diperoleh dengan mudah, sehingga dibutuhkan alat bantu yang mampu memberikan informasi tentang wisata kuliner secara interaktif dan cepat seperti chatbot. Chatbot merupakan program komputer yang memiliki kemampuan berkomunikasi dengan manusia, baik malalui teks maupun audio yang menggandeng Natural Language Processing (NLP) dalam memproses bahasa alami. Chatbot dinamai "Teman Jajan" dan dibangun dari data pertanyaan yang diperoleh dari hasil survei sebanyak 476 data. NLP juga diterapkan pada analisis Topic Modeling dengan metode Latent Dirichlet Allocation (LDA), hasilnya diperoleh 3 kelompok topik yang sering dibahas pada pertanyaan hasil survei, seperti kata "lokasi" yang memiliki keterkaitan antara satu topik dengan lainnya. Pada data yang sama, dilakukan analisis klasifikasi dengan Support Vector Machine (SVM), kernel Linear menghasilkan akurasi paling tinggi dibanding kernel Polynomial, Sigmoid, dan RBF yaitu sebesar 80%. Selanjutnya data tersebut diintegrasikani ke program chatbot, dihasilkan 70% chatbot menjawab dengan benar dan 30% menjawab salah dari 10 pertanyaan baru yang diajukan.

Kata Kunci: *Kuliner, Chatbot, NLP, Topic Modeling, SVM*

NATURAL LANGUAGE PROCESSING (NLP) INITIATION AND CLASSIFICATION OF CULINARY TOURISM FOR CHATBOT PROGRAMS

(Case Study: Culinary Tourism Information of the Special Region of Yogyakarta)

Listari

Department Statistics, Faculty of Mathematics and Natural Science

Islamic University of Indonesia

ABSTRACT

Nowadays information technology is growing rapidly in various fields, including in the culinary field. Culinary tourism has the potential to increase regional income and become a special concern by the government and the management parties in the field. Yogyakarta has a variety of culinary that is able to attract tourists, but culinary information is not always easily obtained, so it takes a tool that is able to provide information about culinary tourism interactively and quickly like chatbot. Chatbot is a computer program that has the ability to communicate with humans, both through text and audio that collaborates with Natural Language Processing (NLP) in processing natural languages. Chatbot is called "Teman Jajan" and is built from question data obtained from the results of a survey of 476 data. NLP is also applied to the Topic Modeling analysis with the Latent Dirichlet Allocation (LDA) method, the results of which are obtained 3 groups of topics which are often discussed in the survey results questions, such as the word "lokasi" which has a relationship between one topic and another. In the same data, classification analysis was carried out with Support Vector Machine (SVM), the Linear kernel produced the highest accuracy compared to the Polynomial, Sigmoid, and RBF kernels of 80%. Then the data is integrated into the chatbot program, resulting in 70% of the chatbot answering correctly and 30% answering incorrectly from the 10 new questions submitted.

Keywords: *Culinary, Chatbot, NLP, Topic Modeling, SVM*