

**IMPLEMENTASI METODE *TERMS FREQUENCY-INVERS DOCUMENT  
FREQUENCY (TF-IDF)* DAN *SUPPORT VECTOR MACHINE (SVM)*  
DALAM PEMBUATAN SISTEM PROGRAM *CHATBOT* PARIYATA  
(Studi Kasus: Informasi Objek Wisata Daerah Istimewa Yogyakarta)**

**TUGAS AKHIR**



**Fadhilah Khairunnisa**

**14 611 145**

**JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA  
2018**

**IMPLEMENTASI METODE *TERMS FREQUENCY-INVERS DOCUMENT  
FREQUENCY (TF-IDF)* DAN *SUPPORT VECTOR MACHINE (SVM)*  
DALAM PEMBUATAN SISTEM PROGRAM *CHATBOT* PARIYATA  
(Studi Kasus: Informasi Objek Wisata Daerah Istimewa Yogyakarta)**

**TUGAS AKHIR**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana  
Jurusan Statistika**



**Fadhilah Khairunnisa**

**14 611 145**

**JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA**

**2018**

**ii**

**HALAMAN PERSETUJUAN PEMBIMBING**

**TUGAS AKHIR**

Judul : IMPLEMENTASI METODE *TERMS FREQUENCY-  
INVERS DOCUMENT FREQUENCY (TF-IDF)* DAN  
*SUPPORT VECTOR MACHINE (SVM)* DALAM  
PEMBUATAN SISTEM PROGRAM *CHATBOT*  
PARIYATA  
(Studi Kasus: Informasi Objek Wisata Daerah Istimewa  
Yogyakarta)


Nama Mahasiswa : Fadhilah Khairunnisa

Nomor Mahasiswa : 14611145

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK  
DIUJIKAN**

Yogyakarta, 14 September 2018

Pembimbing



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

**HALAMAN PENGESAHAN**

**TUGAS AKHIR**

**IMPLEMENTASI METODE *TERMS FREQUENCY-INVERS DOCUMENT  
FREQUENCY (TF-IDF) DAN SUPPORT VECTOR MACHINE (SVM)*  
DALAM PEMBUATAN SISTEM PROGRAM *CHATBOT PARIYATA*  
(Studi Kasus: Informasi Objek Wisata Daerah Istimewa Yogyakarta)**

Nama Mahasiswa : Fadhilah Khairunnisa

Nomor Mahasiswa : 14611145

**TUGAS AKHIR INI TELAH DIAJUKAN  
PADA TANGGAL, 11 Oktober 2018**

**Nama Penguji**

**Tanda Tangan**

1. Andrie Pasca Hendradewa, S.T.,

M.T.

2. Muhammad Muhajir, S.Si., M.Sc.

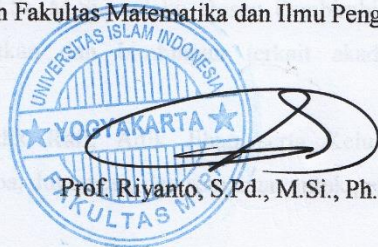
3. Ayundyah Kesumawati, S.Si.,

M.Si.

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

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



## KATA PENGANTAR



*Assalamu'alaikum Wr.Wb.*

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya berupa keimanan, kekuatan, kesabaran, kelancaran serta keselamatan selama melaksanakan penelitian hingga laporan ini dapat terselesaikan. Shalawat serta salam tercurah kepada Nabi Muhammad SAW beserta keluarga dan para pengikut-pengikutnya. Penelitian ini tersusun sebagai hasil Tugas Akhir (TA) untuk memenuhi syarat untuk memperoleh gelar sarjana jurusan statistika.

Penelitian ini berisi tentang “**Implementasi Metode *Terms Frequency-Invers Document Frequency (TF-IDF)* dan *Support Vector Machine (SVM)* dalam Pembuatan Sistem Program *Chatbot* Pariwisata (Studi Kasus: Informasi Objek Wisata Daerah Istimewa Yogyakarta)**”. Selama menyusun laporan, penulis telah banyak mendapat bimbingan dan bantuan dari berbagai pihak. Pada kesempatan ini penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Riyanto, S.pd., M.Si., Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, Yogyakarta.
2. Bapak Dr. Edy Widodo, S.Si., M.Si., selaku Ketua Jurusan Statistika beserta seluruh jajarannya.
3. Ibu Ayundyah Kesumawati, S.Si., M.Si., selaku Dosen Pembimbing Tugas Akhir yang telah memberi bimbingan selama penyusunan Laporan Tugas Akhir ini.
4. Ibu Kariyam, S.Si, M.Si., selaku Dosen Pembimbing Akademik yang telah memberi masukan dan bimbingan terkait akademik selama menjalani perkuliahan.
5. Ayah, Ibu, Adik Rafi, Adik Jihan serta Keluarga Besar yang selalu mencurahkan doa, dukungan, dan semangat untuk penulis.

6. Sahabat PTL yaitu Nanda, Julia, Yusi, Zarmeila, Septi, Ella, Tiwi, Hanna, Tista, Feby, Hafizhan, Sendhy, Samsudin, Husni, Alan, Aufa, dan Febrian yang sudah menemani dan memberi banyak pengalaman serta banyak memberikan semangat serta bantuan dalam memulai dan mengakhiri tugas akhir dan masa kuliah ini.
7. Teman – teman seperjuangan seimbang yaitu Yayan, Samsudin, Hary, dan Afifah yang sudah menemani dan banyak memberikan bantuan dalam mengerjakan tugas akhir ini.
8. Putri Ayuni, sebagai sahabat yang telah banyak menemani dan memberikan semangat serta doa untuk peneliti dalam segala hal.
9. Sahabat Enggo yaitu Himelda, Inayatun, Hani, Resti, Nur, dan Eren yang telah menemani dan mengajari banyak hal selama ini.
10. Novinda Widya Putri, sebagai sahabat berbagi cerita saat Kerja Praktik, yang telah memberikan supportnya.
11. Teman-teman KKN Unit 72 yaitu, Tamara, Dias, Arnetta, Naya, Izzatulah, Vandi, Bimo dan Ilham yang telah memberi pelajaran dalam berbagai hal.
12. Teman XIX Statistika angkatan 2014 yang telah membantu dan memberikan pengalaman selama masa kuliah.
13. Semua pihak yang tidak dapat peneliti sebutkan satu per satu, terima kasih.

Peneliti menyadari sepenuhnya bahwa laporan tugas akhir ini masih jauh dari kesempurnaan, Oleh karena itu segala kritik dan saran yang sifatnya membangun selalu peneliti harapkan. Semoga tugas akhir ini dapat bermanfaat bagi peneliti 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, 14 September 2018

Penulis

## DAFTAR ISI

<b>HALAMAN PERSETUJUAN PEMBIMBING .....</b>	<b>ii</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>iii</b>
<b>KATA PENGANTAR.....</b>	<b>v</b>
<b>DAFTAR ISI.....</b>	<b>vii</b>
<b>DAFTAR TABEL .....</b>	<b>x</b>
<b>DAFTAR GAMBAR.....</b>	<b>xi</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xiii</b>
<b>PERNYATAAN.....</b>	<b>xiv</b>
<b>INTISARI .....</b>	<b>xv</b>
<b>ABSTRACT .....</b>	<b>xvi</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah.....	6
1.3. Batasan Masalah .....	6
1.4. Tujuan Penelitian .....	7
1.5. Manfaat Penelitian .....	7
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>9</b>
2.1. Penelitian Terdahulu .....	9
<b>BAB III LANDASAN TEORI.....</b>	<b>13</b>
3.1. Definisi Pariwisata .....	13
3.1.1. Kategori Wisata Alam.....	13
3.1.2. Kategori Wisata Budaya dan Sejarah.....	14
3.1.3. Kategori Wisata Museum.....	14
3.1.4. Kategori Wisata Pantai.....	14
3.1.5. Kategori Wisata Minat Khusus .....	14
3.2. Definisi Statistika.....	14
3.2.1. Statistika Deskriptif.....	15
3.2.2. Statistika Inferensial.....	15
3.3. <i>Artificial Intelligence</i> (Kecerdasan Buatan).....	16

3.3.1.	<i>Machine Learning</i> .....	18
3.3.2.	<i>Deep Learning</i> .....	19
3.4.	<i>Natural Language Processing (NLP)</i> .....	21
3.5.	<i>Chatbot</i> .....	24
3.6.	<i>Artificial Intelligence Markup Language (AIML)</i> .....	25
3.7.	<i>Graphmaster Pattern Matching</i> .....	27
3.8.	<i>Terms Frequency-Inverse Document Frequency (TF-IDF)</i> .....	29
3.8.1.	Simulasi Pembobotan Kata dengan <i>Terms Frequency-Inverse Document Frequency (TF-IDF)</i> .....	31
3.9.	Klasifikasi .....	34
3.9.1.	Model Klasifikasi .....	34
3.9.2.	<i>Support Vector Machine (SVM)</i> .....	35
3.9.3.	SVM Data Terpisah Secara Linear .....	36
3.9.4.	SVM Data Tidak Terpisah Secara Linear .....	38
3.9.5.	<i>Kernel Trick</i> dan <i>Non-Linear Classification</i> pada SVM .....	39
<b>BAB IV METODOLOGI PENELITIAN .....</b>		<b>41</b>
4.1.	Populasi dan Sampel .....	41
4.2.	Sumber Data.....	41
4.3.	Variabel Penelitian.....	41
4.4.	Metode Pengumpulan Data.....	43
4.5.	Alat dan Cara Organisir Data.....	43
4.6.	Diagram Alir .....	44
<b>BAB V HASIL DAN PEMBAHASAN .....</b>		<b>45</b>
5.1.	Perhitungan Pembobotan Kata Menggunakan <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	46
5.1.1.	Mentransfer <i>Dataset</i> ke dalam <i>Corpus</i> .....	46
5.1.2.	<i>Case Folding</i> .....	46
5.1.3.	<i>Tokenizing (Parsing)</i> .....	47
5.1.4.	<i>Stopwords</i> .....	47
5.1.5.	<i>Filtering</i> .....	48



5.1.6. Analisis Pembobotan <i>Term Frequency-Invers Document Frequency</i> (TF-IDF) .....	48
5.2. Klasifikasi Menggunakan <i>Support Vector Machine</i> (SVM).....	53
5.3. Rancangan Sistem Program <i>Chatbot</i> .....	56
5.3.1. Analisa Representasi Pengetahuan.....	57
5.3.2. Analisa Sistem.....	64
5.3.3. Implementasi Sistem .....	68
5.3.4. Pengujian Sistem Program .....	68
<b>BAB VI KESIMPULAN DAN SARAN.....</b>	<b>74</b>
6.1. Kesimpulan .....	74
6.2. Saran .....	75
<b>DAFTAR PUSTAKA .....</b>	<b>76</b>
<b>LAMPIRAN</b>	

## DAFTAR TABEL

<b>Nomor</b>	<b>Judul</b>	<b>Halaman</b>
Tabel 3.1	Menghitung nilai TF	32
Tabel 3.2	Menghitung nilai DF	32
Tabel 3.3	Menghitung nilai IDF	33
Tabel 3.4	Menghitung nilai TF-IDF	33
Tabel 3.5	<i>Confusion Matrix</i>	35
Tabel 4.1	Definisi Operasional Variabel <i>Dataset 1</i>	41
Tabel 4.2	Definisi Operasional Variabel <i>Dataset 2</i>	42
Tabel 5.1	Sampel <i>Dataset 1</i>	45
Tabel 5.2	Hasil <i>Case Folding</i>	46
Tabel 5.3	Hasil <i>Tokenizing (Parsing)</i>	47
Tabel 5.4	Menghitung nilai TF	51
Tabel 5.5	Menghitung nilai DF	52
Tabel 5.6	Menghitung nilai IDF	52
Tabel 5.7	Menghitung nilai TF-IDF	52
Tabel 5.8	Perbandingan data latih dan data uji pada <i>Dataset 1</i>	54
Tabel 5.9	<i>Confusion Matrix</i> pada data uji	55
Tabel 5.10	Sampel <i>Dataset 2</i>	57
Tabel 5.11	Percakapan <i>chatbot</i> Pariyata	66
Tabel 5.12	Pengujian Sistem Program <i>chatbot</i> Pariyata	69

## DAFTAR GAMBAR

Nomor	Judul	Halaman
Gambar 1.1	Grafik jumlah orang yang berwisata dalam waktu satu tahun	3
Gambar 1.2	Presentase orang memperoleh informasi tentang pariwisata DI Yogyakarta berdasarkan sumbernya	4
Gambar 1.3	Presentase orang yang pernah menggunakan alat bantu penyedia informasi yang interaktif	4
Gambar 1.4	Presentase orang yang mengetahui apa itu <i>chatbot</i>	5
Gambar 1.5	Presentase orang yang pernah mendapatkan informasi pariwisata DI Yogyakarta menggunakan <i>chatbot</i>	5
Gambar 3.1	Cara kerja sistem <i>Artificial Intelligence</i>	18
Gambar 3.2	Visualisasi perbedaan melalui gambar	20
Gambar 3.3	Tag AIML <i>Category</i>	26
Gambar 3.4	Tag AIML <i>Pattern</i>	26
Gambar 3.5	Tag AIML <i>Template</i>	26
Gambar 3.6	Elemen AIML <i>Srai</i>	27
Gambar 3.7	Visualisasi SVM berusaha menemukan <i>hyperplane</i> terbaik yang memisahkan kedua class -1 dan +1	36
Gambar 4.1	<i>Flowchart</i> Penelitian	44
Gambar 5.1	Proses transfer <i>dataset</i> ke <i>Corpus</i>	46
Gambar 5.2	Proses <i>Case Folding</i>	46
Gambar 5.3	Proses <i>Tokenizing (Parsing)</i>	47
Gambar 5.4	Proses <i>Stopwords</i>	48
Gambar 5.5	Proses <i>Filtering</i>	48
Gambar 5.6	Mengubah data menjadi dokumen matriks	49
Gambar 5.7	Sampel kata yang telah dilakukan proses <i>preprocessing</i>	49

Gambar 5.8	Mengurangi dimensi dokumen matriks	49
Gambar 5.9	Sampel kata yang telah dilakukan pengurangan dimensi matriks	50
Gambar 5.10	Perhitungan TF-IDF	50
Gambar 5.11	Hasil perhitungan nilai TF-IDF	50
Gambar 5.12	<i>Reviews</i> 100 pertanyaan	51
Gambar 5.13	Nilai hasil pembobotan TF-IDF	51
Gambar 5.14	Pembagian data latih dan data uji	53
Gambar 5.15	Pembuatan model SVM	55
Gambar 5.16	Pengujian model data <i>training</i> dan <i>testing</i>	55
Gambar 5.17	Pengujian <i>Confusion Matrix</i>	55
Gambar 5.18	Tag AIML <i>Category</i>	58
Gambar 5.19	Tag AIML <i>Pattern</i>	58
Gambar 5.20	Tag AIML <i>Template</i>	59
Gambar 5.21	Elemen AIML <i>Srai</i>	59
Gambar 5.22	Representasi pengetahuan <i>chatbot</i> Pariyata	64
Gambar 5.23	<i>Flowchart Graphmaster Pattern Matching</i>	65
Gambar 5.24	Contoh <i>script</i> untuk Analisa Sistem	66
Gambar 5.25	<i>Script</i> untuk kernel <i>chatbot</i>	68
Gambar 5.26	Halaman Interaksi <i>chatbot</i> Pariyata	68

## DAFTAR LAMPIRAN

Lampiran	1	<i>Dataset 1</i>
Lampiran	2	<i>Script dan Output Term Frequency-Invers Document Frequency (TF-IDF)</i>
Lampiran	3	<i>Script dan Output Support Vector Machine (SVM)</i>
Lampiran	4	<i>Dataset 2</i>
Lampiran	5	<i>Script Chatbot Pariyata</i>

IMPLEMENTAN METOD **PERNYATAAN** FREQUENCY-INVERS DOCUMENT  
FREQUENCY (TF-IDF) DAN SUPPORT VECTOR MACHINE (SVM)

DA Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya 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 diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

INTISARI

Yogyakarta, 11 Oktober 2018



Perkembangan pariwisata di Indonesia saat ini dapat dilihat dari meningkatnya jumlah wisatawan, salah satu cara adalah dengan menggunakan teknologi yang dapat membantu dalam menyediakan sarana pelayanan pariwisata yang akurat sehingga dapat meningkatkan kualitas layanan yang erat dalam bidang ini. Untuk mendapatkan informasi mengenai perkembangan pariwisata yang lebih cepat, maka perlu dilakukan penelitian mengenai teknologi yang dapat membantu dalam menyediakan sarana pelayanan pariwisata yang akurat sehingga dapat meningkatkan kualitas layanan yang erat dalam bidang ini. Penelitian ini menggunakan metode Natural Language Processing (NLP) dengan menggunakan teknologi Artificial Intelligence (AI). Agar data yang digunakan merupakan informasi yang akurat maka dilakukan klasifikasi dan penentuan nilai serta perhitungan akurat pada dataset pariwisata yang akan digunakan untuk membangun sistem program chatbot menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) dan Support Vector Machine (SVM). Hasil yang diperoleh pada nilai bobot kosinus dokumen yang tertinggi adalah 1,529465 dan klasifikasi dengan nilai training sebesar 90% dan data test sebesar 88% diperoleh akurasi sebesar 92%. Maka dibangun sebuah sistem program chatbot yang diberi nama Perbata yang memuat terdapat informasi pariwisata Di Yogyakarta dengan menggunakan bahasa pemrograman AIML (Artificial Intelligence Markup Language) serta menggunakan algoritma Gradient Descent Method. Hasil pengujian sistem pada chatbot Perbata diperoleh akurasi sebesar 82% ber hasil yang lebih memuaskan yang diujikan pada sistem program chatbot.

Kata Kunci : Pariwisata, Di Yogyakarta, Chatbot, NLP, AI, Klasifikasi, TF-IDF, SVM, AIML, Gradient Descent Method.

**IMPLEMENTASI METODE *TERMS FREQUENCY-INVERS DOCUMENT*  
*FREQUENCY (TF-IDF) DAN SUPPORT VECTOR MACHINE (SVM)*  
DALAM PEMBUATAN SISTEM PROGRAM *CHATBOT* PARIYATA  
(Studi Kasus: Informasi Objek Wisata Daerah Istimewa Yogyakarta)**

Oleh : Fadhilah Khairunnisa

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Islam Indonesia

E-mail: [Fadhilahkhair34@gmail.com](mailto:Fadhilahkhair34@gmail.com)

**INTISARI**

*Perkembangan pariwisata di Indonesia saat ini dapat dilihat dari meningkatnya angka kunjungan wisatawan, salah satu daerah yang kaya akan potensi pariwisatanya adalah DI Yogyakarta yang memiliki potensi besar sebagai pengirim wisatawan yang dipengaruhi oleh beberapa faktor salah satunya adalah tersedianya sarana pendukung pariwisata yaitu sarana komunikasi. Peranan teknologi mempunyai kaitan yang erat dalam bidang informasi dan komunikasi. Informasi merupakan bagian terpenting dalam kehidupan sehari-hari yang diharapkan bisa di dapat dengan lebih cepat. Pada era ini untuk mendapatkan informasi tentunya membutuhkan alat bantu penyedia informasi yang lebih interaktif layaknya model diskusi seperti teknologi chatbot yang merupakan disiplin ilmu dari Natural Language Processing (NLP) dengan menggunakan teknologi Artificial Intelligence (AI). Agar data yang digunakan mempunyai landasan yang kuat maka dilakukan klasifikasi dan pembobotan nilai serta perhitungan akurasi pada dataset pertanyaan yang akan digunakan untuk membangun sistem program chatbot menggunakan Term Frequency-Invers Document Frequency (TF-IDF) dan Support Vector Machine (SVM). Hasil yang diperoleh pada nilai bobot kesamaan dokumen yang tertinggi adalah 1,5294468 dan klasifikasi dengan data training sebesar 90% dan data test sebesar 10% diperoleh akurasi sebesar 50%. Maka dibangun sebuah sistem program chatbot yang diberi nama Pariyata yang memuat tentang informasi pariwisata DI Yogyakarta dengan menggunakan bahasa pemrograman AIML (Artificial Intelligence Markup Language) serta menggunakan algoritma Graphmaster Pattern Machine. Hasil pengujian sistem pada chatbot Pariyata didapatkan akurasi sebesar 85% berhasil menjawab pertanyaan yang diujikan pada sistem program chatbot.*

**Kata Kunci :** *Pariwisata, DI Yogyakarta, Chatbot, NLP, AI, Klasifikasi, TF-IDF, SVM, AIML, Graphmaster Pattern Machine.*

**IMPLEMENTATION OF TERMS FREQUENCY-INVERS DOCUMENT  
FREQUENCY (TF-IDF) AND SUPPORT VECTOR MACHINE (SVM)  
METHODS IN THE CREATING CHATBOT PARIYATA PROGRAM  
SYSTEM**

***(Case Study: Tourism Information Special Region of Yogyakarta)***

By : Fadhilah Khairunnisa

*Faculty of Mathematics and Science, Department of Statistics*

*Universitas Islam Indonesia*

*E-mail: [Fadhilahkhair34@gmail.com](mailto:Fadhilahkhair34@gmail.com)*

**ABSTRACT**

*The development of tourism in Indonesia at this time can be seen from the increasing number of tourist visits, one of the areas that is lots of tourism potential is DI Yogyakarta which has great potential as a tourist sender which is influenced by several factors, one of which is the availability of tourism supporting facilities, namely means of communication. The role of technology has a close connection in the field of information and communication. Information is the most important part of everyday life that is expected to be able to be obtained more quickly. In this era, to get information, of course, requires more interactive information provider tools like discussion models such as chatbot technology which is a discipline of Natural Language Processing (NLP) using Artificial Intelligence (AI) technology. In order for the data used to have a strong foundation, the classification and weighting of the values and the calculation of accuracy in the dataset question will be used to build the chatbot program system using Term Frequency-Inverse Document Frequency (TF-IDF) and Support Vector Machine (SVM). The results obtained at the highest document similarity weighting value is 1.5294468 and classification with training data of 90% and 10% test data obtained accuracy of 50%. Then a chatbot program system was built named Pariyata which contains tourism information about DI Yogyakarta using the AIML programming language (Artificial Intelligence Markup Language) and using the Graphmaster Pattern Machine algorithm. The results of testing the system on Pariyata chatbot obtained an accuracy of 85% succeeded in answering the questions tested in the chatbot program system.*

***Keywords:*** *Tourism, Special Region of Yogyakarta, Chatbot, NLP, AI, Classification, TF-IDF, SVM, AIML, Graphmaster Pattern Machine.*