

ABSTRAK

Dalam penelitian ini akan dilakukan analisis sentiment terhadap review dan rating pada dataset yang berasal dari aplikasi Tripadvisor. Ada beberapa tahap untuk melakukan analisis sentiment, diantaranya adalah importing library, collecting dataset, dataset inspection, data preprocessing, training classifier model menggunakan *SGD Classifier*, *Random Forest Classifier*, *SVM Classifier*, *KNearest Neighbors Classifier*, *Naïve Bayes Classifier*, *Overall Classifier Accuracy*, dan juga data visualization. Hasil dari penelitian ini didapatkan bahwa hasil polaritas mendapatkan 73.66% untuk *sentiment positive*, 15.68% untuk *sentiment negative*. Kata yang sering keluar pada sentiment postive yaitu seperti “hotel”, “room”, “stay”, “staff”, “great”. . Kata yang sering keluar pada *sentiment negative* yaitu seperti “hotel”, “room”, “stay”, “service”, “bed”.

Kata kunci: analisis sentiment, dataset, tripadvisor.



ABSTRACT

In this study, sentiment analysis will be carried out on reviews and ratings on the dataset originating from the Tripadvisor application. There are several stages to carry out a sentiment analysis, including importing libraries, collecting datasets, inspecting datasets, preprocessing data, training classifier models using the SGD Classifier, Random Forest Classifier, SVM Classifier, KNearest Neighbors Classifier, Naïve Bayes Classifier, Overall Classifier Accuracy, and also data visualization. The results of this study show that the polarity results get 73.66% for positive sentiment, 15.68% for negative sentiment. Words that often come out in positive sentiments are like "hotel", "room", "stay", "staff", "great". . Words that often come out in negative sentiments are like "hotel", "room", "stay", "service", "bed".

Keywords: dataset, sentiment analysis, tripadvisor.



DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN ORISINALITAS LAPORAN PENELITIAN	ii
PERNYATAAN PUBLIKASI LAPORAN PENELITIAN	iii
PRAKATA	iv
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR NOTASI/ LAMBANG	xii
DAFTAR SINGKATAN	xiii
DAFTAR ISTILAH	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan Pembahasan	2
1.4 Ruang Lingkup	2
1.5 Sumber Data	2
1.6 Sistematika Penyajian	2
BAB 2 KAJIAN TEORI	4
2.1 <i>Sentiment Analysis</i>	4
2.2 <i>Natural Language Processing (NLP)</i>	4
2.2.1 <i>Tokenization</i>	4
2.2.2 <i>Stop Word Removal</i>	5

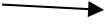
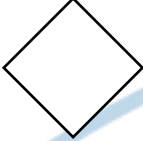
2.2.3 Normalizing	5
2.3 Python	5
2.4 Kaggle	6
2.5 Tripadvisor	6
BAB 3 METODOLOGI PENELITIAN	7
3.1 Tahapan Penelitian	7
3.2 Importing Library	8
3.3 Collecting Dataset	8
3.4 Dataset Inspection	8
3.5 Dataset Preprocessing	8
3.6 Training Classifier Model	8
3.7 Data Visualization	9
BAB 4 HASIL PENELITIAN	10
4.1 Importing Library	10
4.2 Collecting Dataset	10
4.3 Dataset Inspection	11
4.3.1 Top 5 Records	11
4.3.2 Bottom 5 Records	11
4.4 Dataset Preprocessing	12
4.4.1 Total Unique Word Overview After Data Preprocessing	12
4.4.2 Total Token Overview After Data Preprocessing	13
4.5 Training Classifier Model	13
4.5.1 SGD Classifier (Stochastic Gradient Descent Classifier)	14
4.5.2 Random Forest Classifier	14
4.5.3 SVM Classifier (Support Vector Machines Classifier)	15
4.5.4 KNearest Neighbors Classifier	15

4.5.5 Naive Bayes Classifier	16
4.5.6 Implementing Hyper Parameter Tunning Menggunakan Grid Search	16
4.5.7 Overall Classifier Accuracy	18
4.6 Data Visualization	18
4.6.1 Null Values Overview	18
4.6.2 Rating Overview (Pie Chart)	19
4.6.3 Rating Overview (Bar Chart)	20
4.6.4 Most Common Word Overview (Cloud)	20
4.6.5 Most Common Word Overview (Bar Chart)	21
4.7 Sentiment Analysis	21
4.7.1 Tingkat Polaritas (Polarity Level)	22
4.7.1.1 Positive Result by Polarity Level	22
4.7.1.2 Negative Result by Polarity Level	23
4.7.2 Custom Sentiment Analysis by Polarity Level	24
4.7.2.1 Positive Result Custom Sentiment Analysis by Polarity Level	24
4.7.2.2 Negative Result Custom Sentiment Analysis by Polarity Level	24
BAB 5 KESIMPULAN DAN SARAN	26
5.1 Kesimpulan	26
5.2 Saran	27
DAFTAR PUSTAKA	29
LAMPIRAN A NAMA LAMPIRAN	A-1
LAMPIRAN B NAMA LAMPIRAN	B-1
LAMPIRAN C NAMA LAMPIRAN	C-1

DAFTAR GAMBAR

Gambar 3.1 Tahap Penelitian	7
Gambar 4.1 <i>Importing Library</i>	10
Gambar 4.2 <i>Collecting Dataset</i>	10
Gambar 4.3 <i>Top 5 Records</i>	11
Gambar 4.4 <i>Bottom 5 Records</i>	11
Gambar 4.5 <i>Data Preprocessing</i>	12
Gambar 4.6 <i>Total Unique Word After Data Preprocessing</i>	13
Gambar 4.7 <i>Total Token After Data Preprocessing</i>	13
Gambar 4.8 SGD Classifier	14
Gambar 4.9 Random Forest Classifier	15
Gambar 4.10 SVM Classifier	15
Gambar 4.11 KNearest Neighbors Classifier	16
Gambar 4.12 Naive Bayes Classifier	16
Gambar 4.13 <i>Hyper Parameter Tunning Menggunakan Grid Search</i>	17
Gambar 4.14 <i>Hyper Parameter Tunning Menggunakan Grid Search</i>	17
Gambar 4.15 <i>Hyper Parameter Tunning Menggunakan Grid Search</i>	17
Gambar 4.16 <i>Overall Classifiers Accuracy</i>	18
Gambar 4.17 <i>Null Values Overview</i>	19
Gambar 4.18 <i>Rating Overview (Pie Chart)</i>	19
Gambar 4.19 <i>Rating Overview (Bar Chart)</i>	20
Gambar 4.20 <i>Most Common Word Cloud</i>	21
Gambar 4.21 <i>Most Common Word (Bar Chart)</i>	21
Gambar 4.22 <i>Polarity Level</i>	22
Gambar 4.23 <i>Positive Result Word Cloud</i>	23
Gambar 4.24 <i>Negative Result Word Cloud</i>	23
Gambar 4.25 <i>Custom Positive Result Word Cloud</i>	24
Gambar 4.26 <i>Custom Negative Result Word Cloud</i>	25

DAFTAR NOTASI/ LAMBANG

Jenis	Notasi/ Lambang	Nama	Arti
Flowchart		Proses	Mempresentasikan proses.
		Flow	Mempresentasikan alur proses.
		Keputusan (<i>Decision</i>)	Keputusan dalam sebuah proses

Referensi:

Notasi/ Lambang Flowchart dari Microsoft Visio



DAFTAR SINGKATAN

NLP	<i>Natural Language Processing</i>
SGD	<i>Stochastic Gradient Descent</i>
SVM	<i>Support Vector Machines</i>



DAFTAR ISTILAH

<i>Python</i>	<i>Python</i> merupakan bahasa pemrograman yang paling cocok dipelajari untuk <i>programmer</i> pemula, karena <i>python</i> memiliki <i>tools</i> yang canggih [1]
<i>Sentiment Analysis</i>	“ <i>Sentiment Analysis</i> adalah sejenis klasifikasi teks yang membuat katalog teks berdasarkan orientasi sentimen mengandung pendapat atau opini dari mereka” [2].

