

Analisis *Dataset* menggunakan *Sentiment Analysis* (Studi Kasus Pada Tripadvisor)

Haryo Bagas Assyafah^{#1}, Diana Trivena Yulianti S.Kom., M.T^{*2}

[#]Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Maranatha
Jalan Surya Sumantri No.65, Kota Bandung, Jawa Barat, 40164

¹1773031@maranatha.ac.id

²diana.trivena@it.maranatha.edu

Abstract — In this study, sentiment analysis will be carried out on reviews and ratings on datasets originating from the Tripadvisor application. There are several stages to carry out sentiment analysis, including importing libraries, collecting datasets, dataset inspection, preprocessing data, training classifier models using SGD Classifier, Random Forest Classifier, SVM Classifier, KNearest Neighbors Classifier, Naïve Bayes Classifier, Overall Classifier Accuracy, and also data visualization

Keywords— Dataset, Data Visualization, Preprocessing Data, Sentiment Analysis

I. PENDAHULUAN

Pada era saat ini, perkembangan teknologi semakin pesat sehingga untuk mendapatkan sebuah informasi sangatlah mudah. Salah satunya adalah untuk mencari informasi hotel di kota-kota besar. Pencarian hotel biasanya dilakukan untuk memudahkan pengguna melihat harga dan fasilitas pada hotel yang dicari. Sebelum memilih hotel, biasanya pengguna melihat terlebih dahulu *review* dari pengalaman konsumen yang pernah mengunjungi hotel tersebut.

Saat ini terdapat aplikasi yang dapat membantu kita dalam mendapatkan informasi mengenai berbagai hotel beserta *review* dan *rating* dari hotel tersebut yaitu aplikasi *Tripadvisors*. *Tripadvisors* ini digunakan untuk melakukan pemesanan kamar hotel serta membandingkan harga terendah untuk hotel-hotel yang tersedia pada aplikasi *Tripadvisors*. Pengguna juga dapat memberikan *review* pada hotel yang telah dikunjunginya.

Dataset yang digunakan dalam analisis ini yaitu data *Tripadvisors Hotel Review* yang didapatkan dari *Kaggle*. Fitur utama dari *Dataset Tripadvisor* ini adalah opini yang didapatkan dari customer yang telah melakukan pemesanan kamar hotel. Fitur utama *dataset* ini berupa *String* yang sangat cocok untuk membuat analisis mengenai kualitas dan layanan hotel menggunakan *sentiment analysis*. Dengan melakukan analisis pada *dataset Tripadvisors Hotel Review* menggunakan metode *sentiment analysis*, diharapkan dapat menyimpulkan data positif atau negatif yang dominan pada suatu produk / jasa yang terdapat didalam *dataset Tripadvisors Hotel Review* ini. Tujuannya agar dapat mengetahui kelebihan dan kelemahan pada suatu produk / jasa agar dapat meningkatkan kualitas pelayanannya.

II. KAJIAN TEORI

A. *Sentiment Analysis*

Untuk melakukan *analysis review* dibutuhkan sebuah metode yaitu metode *sentiment analysis*. *Sentiment Analysis* yang biasa disebut *opinion mining*, telah menjadi salah satu yang paling aktif pada area penelitian paling aktif pada *natural language processing* sejak awal tahun 2000 [3]. “*Sentiment Analysis* adalah bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya” [4]. “*Sentiment Analysis* adalah sejenis klasifikasi teks yang membuat katalog teks berdasarkan orientasi sentimen mengandung pendapat atau opini dari mereka” [2]. Dari kedua teori tersebut dapat disimpulkan, *Sentiment Analysis* merupakan analisis untuk mendapatkan nilai positif atau negatif yang dominan berdasarkan opini atau pendapat dari seseorang.

B. *Natural Language Processing (NLP)*

Sebelum melakukan analisis *review*, *Natural Language Processing* dibutuhkan untuk proses *sentiment analysis*. Menurut Jalaj Thanaki, *Natural Language Processing* merupakan kemampuan teknologi, komputasi atau linguistik komputasi untuk memproses bahasa alami manusia [5]. “*Natural Language Processing* adalah bidang *artificial intelligence*

yang difokuskan untuk menemukan interaksi antara komputer dan bahasa manusia” [6]. *Natural Language Processing* dapat digunakan untuk proses *tokenization*, *stop word removal*, dan masih banyak lainnya.

C. Tokenization

Dalam melakukan analisis *review* dengan *sentiment analysis* dibutuhkan proses *tokenization*. *Tokenization* membutuhkan *Natural Language Processing*. “*Tokenization* adalah sebuah kata yang dapat dipahami dan diproses oleh mesin, string teks apa pun tidak dapat diproses lebih lanjut tanpa melalui tokenisasi” [7]. Menurut teori Sumit Raj, *Tokenization* merupakan salah satu konsep dari *Natural Language Processing* (NLP) yang sederhana di mana *tokenization* membagi teks menjadi segmen yang bermakna [8]. Dari teori tersebut dapat disimpulkan bahwa *tokenization* merupakan proses memecah potongan teks menjadi potongan-potongan kecil berupa *list*.

D. Stop Word Removal

Stopword merupakan kata-kata yang mungkin penting dalam komunikasi, tetapi saat melakukan *sentiment analysis*, *stopword* nilainya sangat kecil untuk mesin. “*Stop Word Removal* bertujuan untuk menghapus beberapa kata yang lebih sering digunakan dalam *natural language processing* tetapi nilai pentingnya dalam mengekstraksi makna sangat besar tetapi hanya sedikit dalam hal ketika kita mempertimbangkan arti keseluruhan dari kalimat. Kata-kata seperti itu dinyatakan sebagai kata-kata berhenti dan dihapus” [9]. Oleh karena itu, kita perlu menghapus *Stopword* yang terdapat pada *dataset* tersebut.

E. Normalizing

Normalisasi merupakan metode untuk memproses *string* ke dalam *natural language*. Menurut Deepti Chopra, Nisheeth Joshi, Iti Mathur, untuk melakukan pemrosesan pada *natural language*, kita perlu melakukan normalisasi yang terutama melibatkan menghilangkan tanda baca, mengubah seluruh teks menjadi huruf kecil atau huruf besar, mengubah angka menjadi kata-kata, memperluas singkatan, kanonikalisasi teks dan sebagainya [10]. Terdapat beberapa karakter *unicode* dapat diwakili oleh lebih dari satu pengkodean *unicode*. Mereka akan terlihat sama, tetapi tidak akan membandingkan sama karena mereka memiliki urutan *byte internal* yang berbeda [11]. Dapat disimpulkan bahwa normalisasi dibutuhkan untuk melakukan *sentiment analysis*.

F. Python

Saat ini bahasa pemrograman yang sangat terkenal yaitu *Python*. “*Python* merupakan bahasa pemrograman yang memungkinkan untuk bekerja lebih cepat dan mengintegrasikan sistem dengan lebih efektif” [12]. *Python* merupakan bahasa pemrograman yang paling cocok dipelajari untuk *programmer* pemula, karena *python* memiliki *tools* yang canggih [1]. Fitur-fitur yang terdapat pada *python* yaitu *open source*, *portable*, *simple*, *support other technologies*, *extensible and embeddable*, *system programming*, *object oriented*, dan lain-lain. [13]. Dapat disimpulkan bahwa, *python* adalah bahasa pemrograman yang bersifat *open source* dan juga memudahkan *programming* dalam mengintegrasikan sistem yang lebih efektif.

G. Kaggle

Saat ini terdapat website yang memudahkan dalam menemukan *dataset* dan data tersebut dapat diolah menggunakan teknik *data science*. “*Kaggle* merupakan komunitas ilmu *data science* terbesar di dunia dengan *tools* dan sumber daya canggih untuk membantu dalam mencapai *data science goals*” [14]. Dalam *website Kaggle*, terdapat *data set*, kompetisi, komunitas *data science*, *courses*, dan masih banyak lainnya.

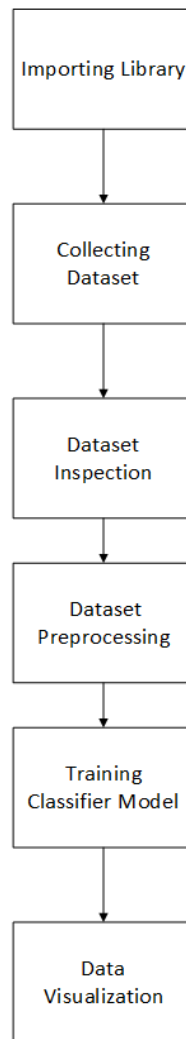
H. Tripadvisor

Terdapat aplikasi atau *website* yang membantu dalam pencarian hotel, pesawat, *restaurant* dan masih banyak lainnya yaitu *Tripadvisor*. “*Tripadvisor* merupakan *platform* panduan wisata terbesar di dunia yang membantu ratusan juta orang setiap bulannya dalam merencanakan, memesan, hingga melakukan perjalanannya” [15]. Pada *website Tripadvisor*, terdapat rekomendasi penginapan terdekat, *restaurant*, *travel forums*, *vacation rentals*, *blog* untuk para *traveler*, dan masih banyak lainnya.

III. METODOLOGI PENELITIAN

A. Tahapan Penelitian

Pada Gambar 1, merupakan tahapan penelitian dalam melakukan analisis data menggunakan *sentiment analysis*.



Gambar 1 Tahapan Penelitian

B. *Importing Library*

Importing Library berfungsi untuk melakukan pemanggilan pada *module – module* yang dibutuhkan untuk menjalankan fungsi pada proyek ini. *Module* dapat di definisikan sebagai fungsi, *class* atau *variable* yang dapat dirujuk pada *module Python* lainnya. *Module* dapat di akses dengan menggunakan *statement import*.

C. *Collecting Dataset*

Collecting Dataset adalah proses didalam sistem ketika melakukan pemanggilan untuk membaca sebuah *dataset* yang bentuk datanya bisa berupa .JSON, .CSV, .XLSX dan yang lainnya. Dalam hal ini, data yang di *collect* berbentuk .CSV dan berjumlah 20.492 rows. Data yang diambil adalah data *rating*, dan *review* Hotel *Tripadvisors*. *Dataset* ini berasal dari *website* dengan alamat URL <https://www.kaggle.com/andrewmvd/trip-advisor-hotel-reviews>

D. *Dataset Inspection*

Dataset Inspection adalah proses ketika melakukan pengecekan pada *dataset*. Misalnya melihat berapa *total value* yang *null*, atau melihat seperti apa bentuk datanya. Untuk mengetahui atribut dari *dataset* yang di pakai, sehingga dapat menentukan langkah-langkah yang harus diambil kedepan.

E. *Dataset Preprocessing*

Dataset Preprocessing adalah proses yang dijalankan sebelum menganalisis *dataset* yang dimana *dataset* tersebut akan diolah menjadi format yang ternormalisasi. Normalisasi adalah proses seperti ketika melakukan penghapusan tanda baca,

mengubah teks menjadi huruf kecil atau besar, mengubah angka menjadi kata-kata, dan lain-lain. Tujuannya agar data yang telah diolah, menjadi data dengan format standar bagi *machine learning model / classifier*.

F. Training Classifier Model

Classifier di dalam proyek ini berfungsi untuk mengukur tingkat keakuratan data dari sebuah *dataset*. *Training Classifier Model* dilakukan agar *classifier* dapat mempelajari karakteristik dari sebuah *dataset* dan mampu memberikan nilai dan kesimpulan terhadap keseluruhan data.

G. Data Visualization

Data Visualization dalam proyek ini dilakukan untuk memvisualisasikan data dari proses - proses yang bisa divisualisasikan. Data yang divisualisasikan berupa grafik-grafik seperti *line chart*, *histogram*, *bar chart*, dan lain-lain.

IV. HASIL PENELITIAN

A. Importing Library

Pada Gambar 2, *Importing Library* ini berfungsi untuk melakukan pemanggilan pada *module – module* yang dibutuhkan untuk menjalankan fungsi pada analisis ini.

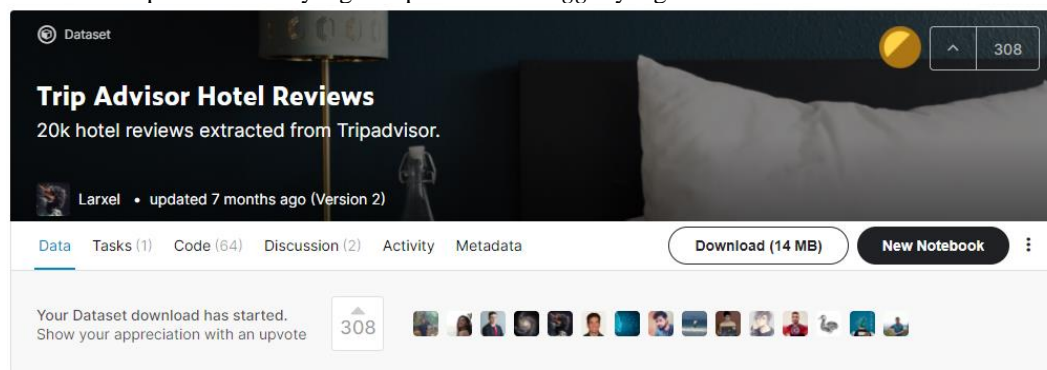
```
1. Importing Library
python -m spacy download en_core_web_sm

In [1]: import en_core_web_sm
import nltk
import re
import spacy
import bttext as bt
import numpy as np
import pandas as pd
from collections import Counter
from nltk.tokenize import RegexpTokenizer
from pyecharts.charts import Bar, Pie, WordCloud
from pyecharts import options as opts
from pyecharts.globals import ThemeType
from sklearn.metrics import classification_report
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import SGDClassifier as StochasticGradientDescentClassifier
from sklearn.ensemble import RandomForestClassifier as RandomForestClassifier
from sklearn import svm as SupportVectorMachinesClassifier
from sklearn.neighbors import KNeighborsClassifier as KNearestNeighborsClassifier
from sklearn.naive_bayes import MultinomialNB as NaiveBayesClassifier
from tqdm import tqdm
```

Gambar 2 *Importing Library*

B. Collecting Dataset

Pada Gambar 3 ini merupakan *dataset* yang didapatkan dari *Kaggle* yang berisi file CSV.



Gambar 3 *Collecting Dataset*

C. Dataset Inspection

1) *Top 5 Records* : Pada *Gambar 4*, merupakan 5 *review* terbaik yang terdapat di *dataset Tripadvisor* ini.

3.1 Top 5 Records

```
df.head()
```

	Review	Rating
0	nice hotel expensive parking got good deal sta...	4
1	ok nothing special charge diamond member hilto...	2
2	nice rooms not 4* experience hotel monaco seat...	3
3	unique, great stay, wonderful time hotel monac...	5
4	great stay great stay, went seahawk game aweso...	5

Gambar 4 Top 5 Records

2) Bottom 5 Records : Pada Gambar 5, merupakan 5 review terbawah yang terdapat di dataset Tripadvisor ini.

3.2 Bottom 5 Records

```
df.tail()
```

	Review	Rating
20486	best kept secret 3rd time staying charm, not 5...	5
20487	great location price view hotel great quick pl...	4
20488	ok just looks nice modern outside, desk staff ...	2
20489	hotel theft ruined vacation hotel opened sept ...	1
20490	people talking, ca n't believe excellent ratin...	2

Gambar 5 Bottom 5 Records

D. Data Preprocessing

Pada Gambar 6 ini menjelaskan mengenai proses dari *Data Preprocessing*. *Data preprocessing* mencakup beberapa proses yaitu *remove case*, *lower*, *convert_to_tokens*, *remove_stopword*, *convert_to_lemma*, *revert_token*. *remove_case* berfungsi untuk menghapus special text pada sebuah *text*. *lower* berfungsi untuk merubah menjadi huruf kecil, *convert_to_token* berfungsi untuk mengubah teks menjadi potongan-potongan kecil. *revert_token* berfungsi untuk merubah potongan-potongan kecil pada proses sebelumnya yaitu *convert_to_token* menjadi sebuah teks. *remove_stopword* berfungsi untuk menghapus kata seperti "a", "the", "an", "in". *convert_to_lemma* berfungsi untuk merubah *verb 2* menjadi *verb 1* seperti *ran* menjadi *run*.

4.1 Basic Function and Variable

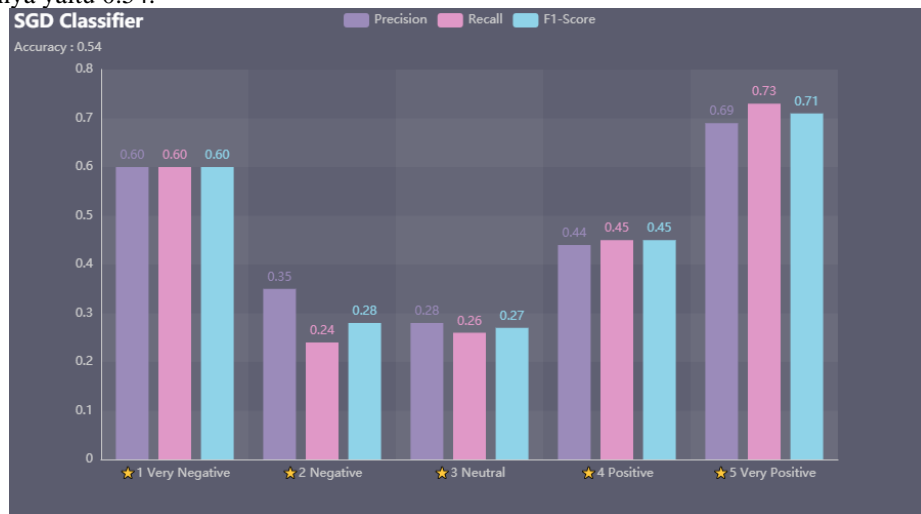
```
def preprocess_data(container, token_converter = en_core_web_sm.load()) :  
    def remove_scace(val) :  
        return bt.conslet(val)  
  
    def lower(val) :  
        return val.lower()  
  
    def convert_to_tokens(val) :  
        def filter_token(token) :  
            return str(token) != '' and len(str(token)) != 1  
  
        return [token for token in token_converter(val) if filter_token(token)]  
  
    def remove_stopword(tokens) :  
        return [token for token in tokens if not token.is_stop]  
  
    def convert_to_lemma(tokens) :  
        return [token.lemma_ for token in tokens]  
  
    def revert_token(tokens) :  
        return ' '.join(token for token in tokens)  
  
    output = {'data' : [], 'sum_token_before' : [], 'sum_token_after' : []}  
    append_data = output['data'].append  
    append_before = output['sum_token_before'].append  
    append_after = output['sum_token_after'].append
```

Gambar 6 Data preprocessing

E. Training Classifier Model

1) SGD Classifier

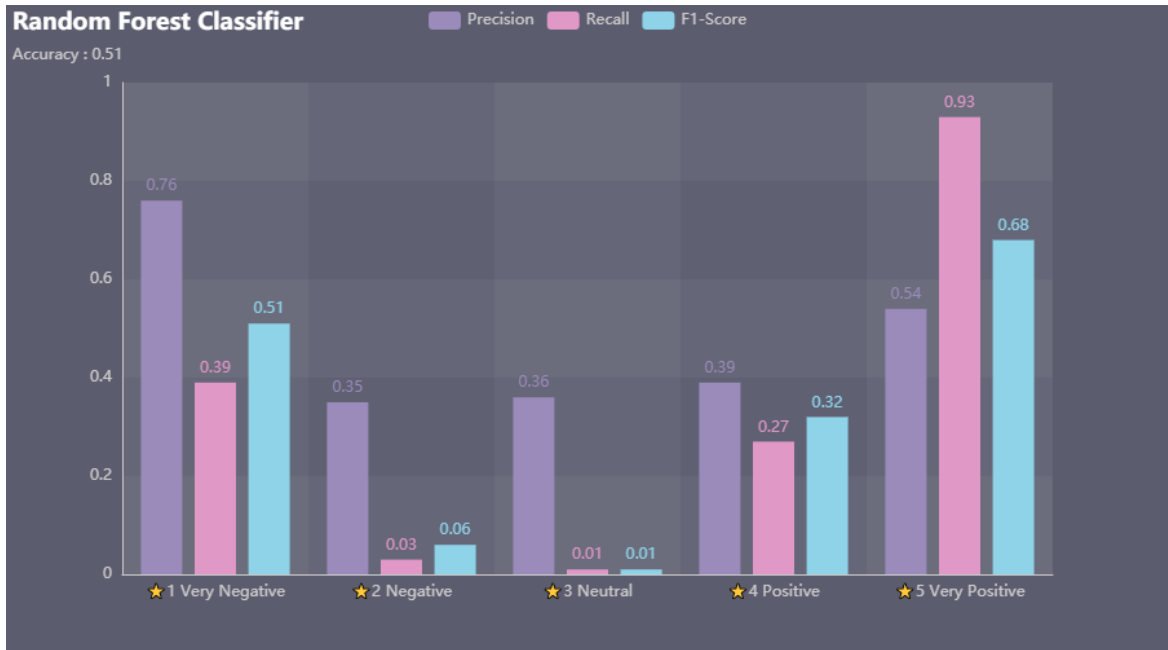
Pada Gambar 7 ini merupakan hasil dari *Classifier* dari SGD. Dengan melakukan metode ini, menghasilkan tingkat keakuratan datanya yaitu 0.54.



Gambar 7 SGD Classifier

2) Random Forest Classifier

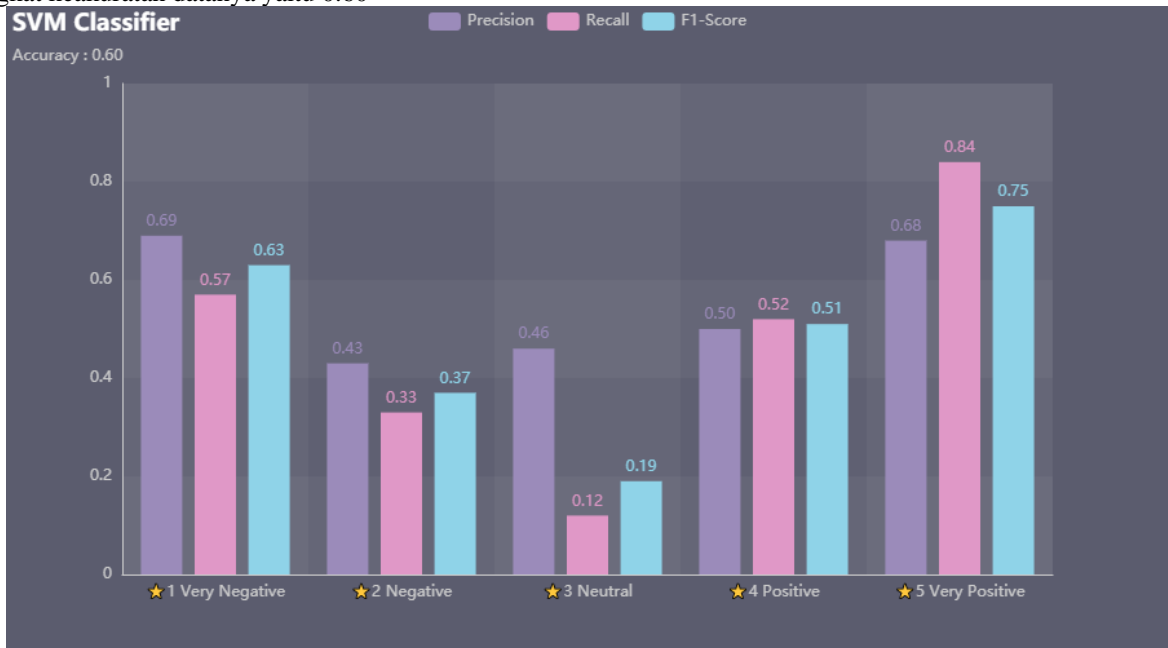
Pada Gambar 8 ini merupakan hasil dari *Classifier* dari *Random Forest Classifier*. *Random Forest* ini berfungsi untuk melakukan proses pemilihan fitur secara acak agar menghasilkan tingkat kesalahan yang relatif rendah. Dengan melakukan metode ini, menghasilkan tingkat keakuratan datanya yaitu 0.51.



Gambar 8 Random Forest Classifier

3) SVM Classifier (Support Vector Machines Classifier)

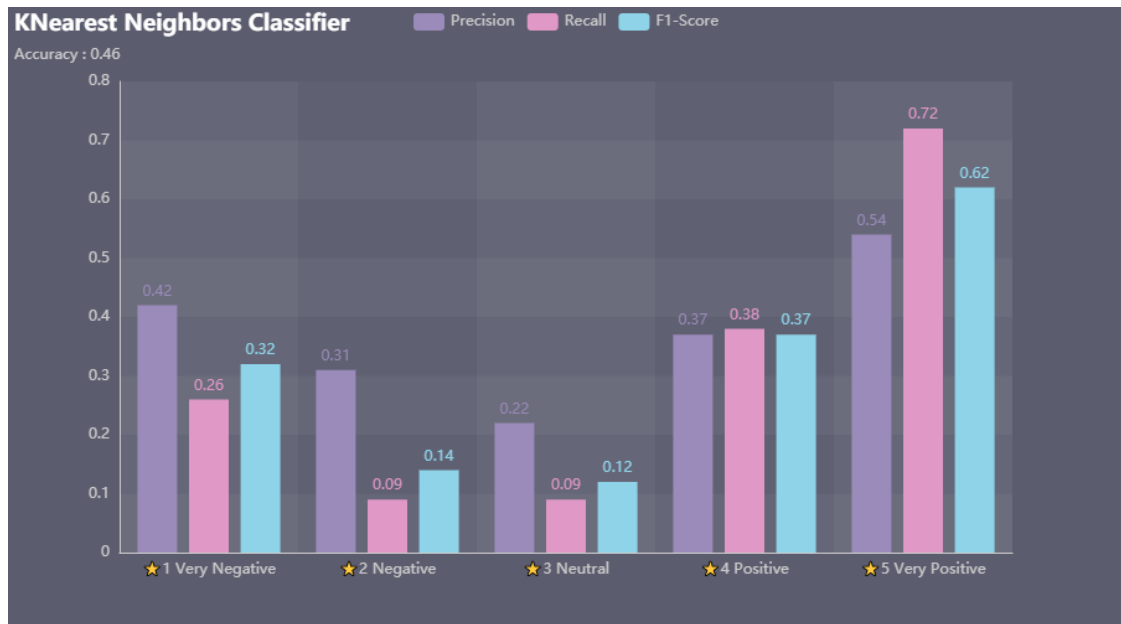
Pada Gambar 9 ini merupakan Classifier dari SVM Classifier. Dengan melakukan metode ini, menghasilkan tingkat keakuratan datanya yaitu 0.60



Gambar 9 SVM Classifier

4) KNearest Neighbors Classifier

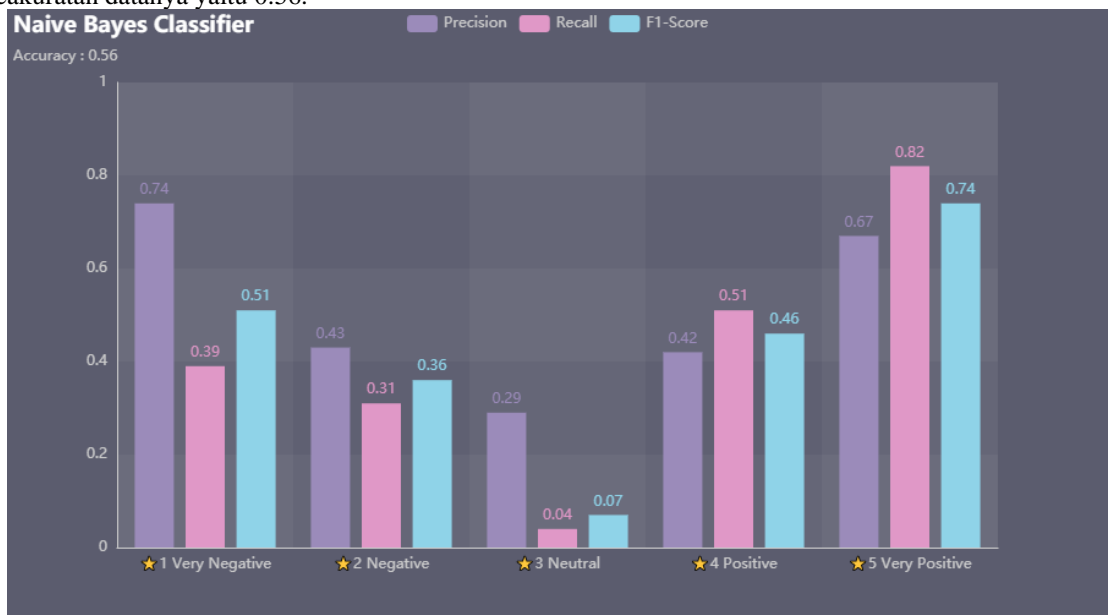
Pada Gambar 10 ini merupakan hasil dari KNearest Neighbors Classifier. Dengan melakukan metode ini, menghasilkan tingkat keakuratan datanya yaitu 0.46.



Gambar 10 KNearest Neighbors Classifier

5) *Naïve Bayes Classifier*

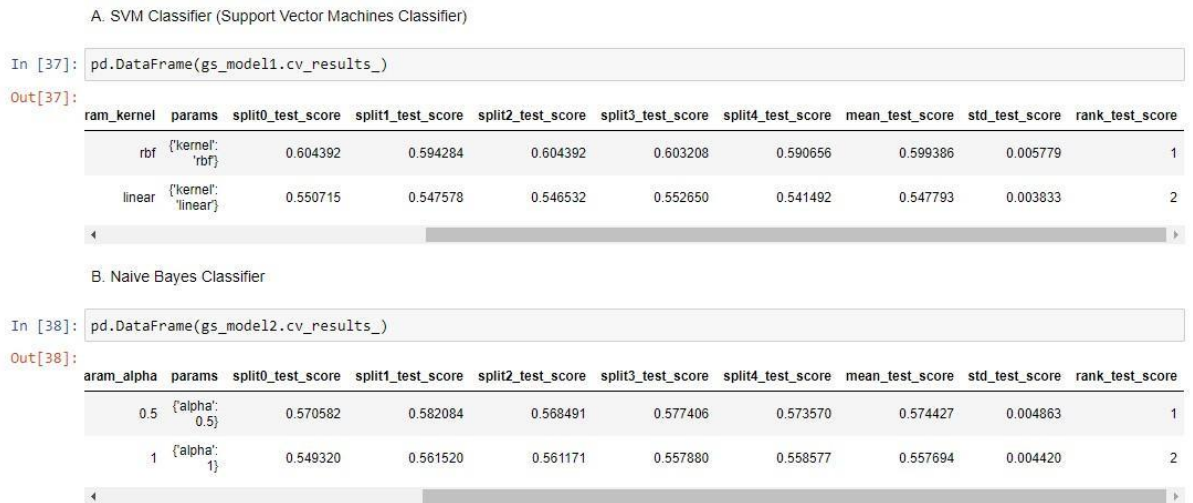
Pada Gambar 11 ini merupakan hasil dari *Naïve Bayes Classifier*. Dengan melakukan metode ini, menghasilkan tingkat keakuratan datanya yaitu 0.56.



Gambar 11 Naïve Bayes Classifier

6) *Implementing Hyperparameter Tunning menggunakan Grid Search*

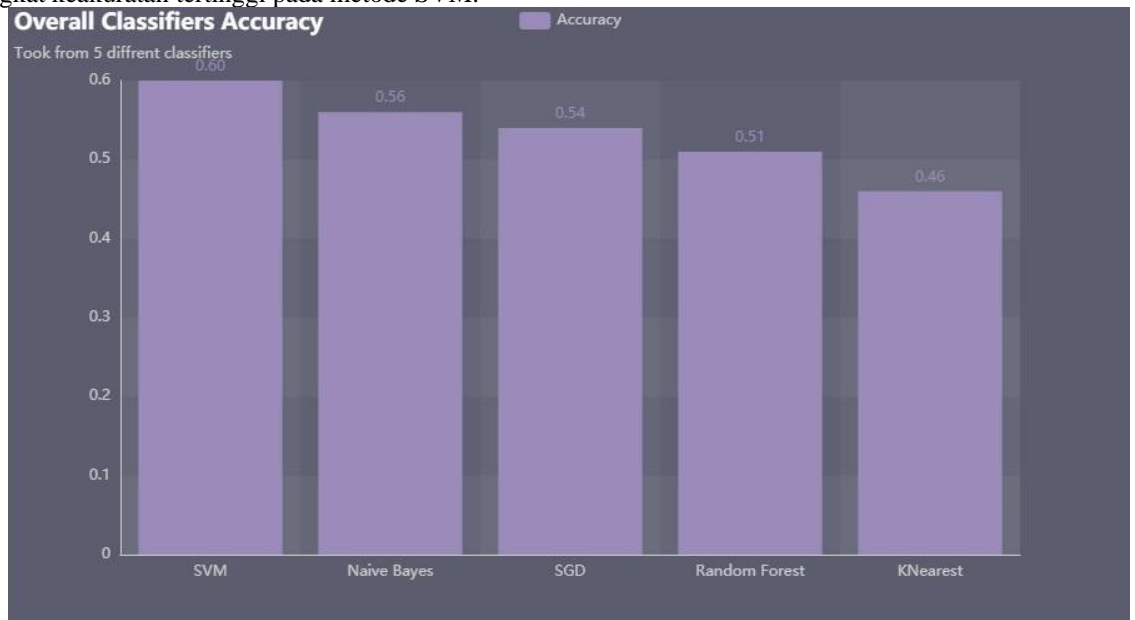
Pada Gambar 12 merupakan hasil dari implementasi *Grid Search*, top 2 model berdasarkan accuracy tertinggi yaitu *SVM Classifier* dan *Naïve Bayes Classifier*.



Gambar 12 Implementing Hypertunning menggunakan Grid Search

7) Overall Classifier Accuracy

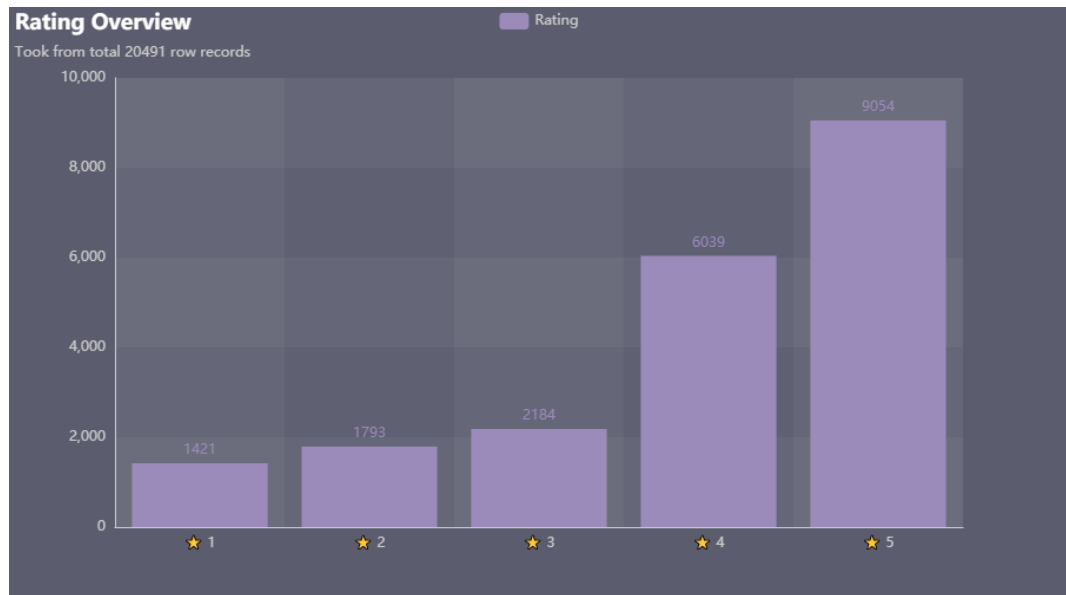
Pada Gambar 13 ini merupakan hasil dari Overall Classifier Accuracy. Dapat dibandingkan dari keempat metode bahwa tingkat keakuratan tertinggi pada metode SVM.



Gambar 13 Overall Classifier Accuracy

F. Data Visualization

- 1) Rating Overview : Pada Gambar 14 berikut merupakan Rating Overview pada Dataset Tripadvisor berbentuk bar chart. Terdapat 44.19% yang memilih rating bintang 5, 19.47% yang memilih rating bintang 4, 10.66% yang memilih rating bintang 3, 8,75% yang memilih rating bintang 2, dan terdapat 6,93% yang memilih bintang 1. Rating ini berdasarkan dari keseluruhan rating yang terdapat pada dataset Tripadvisor dengan total 20.491 records.

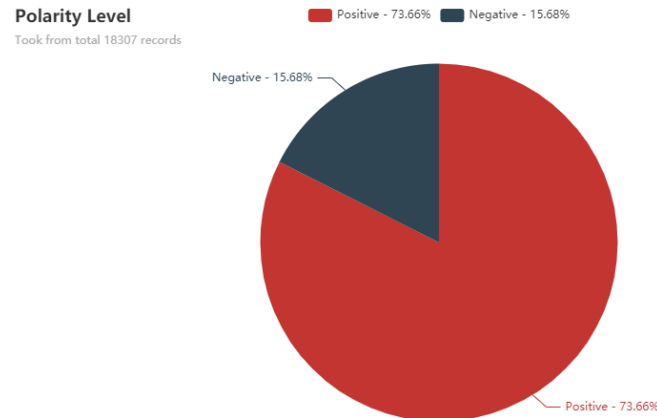


Gambar 14 Rating Overview

G. Sentiment Analysis

Pada tahapan *sentiment analysis* ini menghasilkan analisis berupa *polarity level* dan hasil *positif* dan *negatif* dari *polarity level* berbentuk *word cloud*.

- 1) *Tingkat Polaritas* : *Polarity Level* berfungsi untuk mengukur tingkat polaritas. Pada Gambar 15 merupakan *sentiment analysis* berdasarkan *Polarity Level*. Hasil polaritas ini mendapatkan 73.66% untuk *sentiment positive*, 15.68% untuk *sentiment negative*. Untuk *sentiment netral* tidak dipakai, karena data netral tidak mempunyai *trend positive / negative* maka data netral tidak digunakan pada saat kalkulasi tingkat polaritas. tingkat polaritas ini ditujukan untuk mengambil sisi *positif* atau *negatif* agar dapat melihat trend yang sedang kuat di sisi *positif* atau *negatif* nya.



Gambar 15 Polarity Level

- 2) *Positive Result by Polarity Level* : Pada Gambar 16 , hasil gabungan *rating* dari bintang 5 dan bintang 4 yang divisualisasikan menggunakan *word cloud*. Semakin besar kata semakin sering kata tersebut keluar dalam sebuah review. Kata yang sering keluar yaitu seperti “*hotel*”, “*room*”, “*stay*”, “*staff*”, “*great*”. Yang dapat disimpulkan bahwa pelayanan menginap di hotel sangat memuaskan. Ini dapat digunakan untuk mempertahankan kualitas hotel tersebut.

Positive

Took from total 15093 records



Gambar 16 Positive Result

- 3) *Negative Result by Polarity Level* : Pada Gambar 17, hasil gabungan rating dari bintang 2 dan bintang 1 yang divisualisasikan menggunakan word cloud. Semakin besar kata semakin sering kata tersebut keluar dalam sebuah review. Kata yang sering keluar yaitu seperti “hotel”, “room”, “stay”, “service”, “bed”. Yang dapat disimpulkan bahwa menginap di hotel tidak nyaman karena pelayanan yang buruk dan fasilitas kamar yang buruk. Ini dapat digunakan untuk evaluasi kepada pihak hotel untuk meningkatkan kualitas hotel tersebut.

Negative

Took from total 3214 records



Gambar 17 Negative Result

V. KESIMPULAN

Berdasarkan penelitian yang berjudul “Analisis Dataset menggunakan Sentiment Analysis (Studi Kasus Pada Tripadvisor)”, dilakukan data pre-processing lalu menerapkan sentiment analysis. Hasil dari penelitian ini didapatkan bahwa hasil polaritas mendapatkan 73.66% untuk sentiment positive, 15.68% untuk sentiment negative. Pada hasil sentiment positive, saat di custom dengan menghapus beberapa kata menghasilkan kata yang sering keluar seperti “staff”, “breakfast”, “location”, “night”, “clean”, “restaurant”. Yang dapat disimpulkan bahwa hotelnya bersih, pelayanan nya memuaskan, breakfast di hotel sangat enak. Ini dapat digunakan untuk mempertahankan kualitas hotel tersebut. Pada hasil sentiment

negative, saat di custom dengan menghapus beberapa kata tersebut menghasilkan kata yang sering keluar seperti “service”, “people”, “night”, “time”. Yang dapat disimpulkan bahwa pelayanan di hotel sangat buruk seperti orang-orang disekitar hotel mengganggu, tidak sigap atau lambat dalam melayani. Ini dapat digunakan untuk evaluasi kepada pihak hotel untuk meningkatkan kualitas hotel tersebut. Kata-kata yang sering keluar tersebut bisa digunakan untuk mengevaluasi kepada pihak hotel untuk meningkatkan kualitas hotel tersebut.

UCAPAN TERIMA KASIH

Segala puji dan syukur saya ucapkan kepada Tuhan Yang Maha Esa, karena saya masih diberikan kesempatan untuk dapat menyelesaikan Laporan Kerja Praktek yang berjudul “Analisis Dataset menggunakan Sentiment Analysis (Studi Kasus Pada Tripadvisor)” dengan lancar. Penyusunan Laporan ini sebagai syarat kelulusan mata kuliah Tugas Akhir. Dalam proses pembuatan laporan tak lepas dari bantuan, arahan dan masukan dari berbagai pihak terkhususnya saya mendapatkan banyak dukungan dari dosen Ibu Diana Trivena Yulianti, S. Kom., M.T., yang sangat membantu karena telah meluangkan waktu dan pikiran untuk saran dan bimbingan kepada penulis. Untuk itu saya ucapkan banyak terima kasih atas segala partisipasinya dalam menyelesaikan laporan ini.

Ucapan terima kasih selanjutnya penulis sampaikan kepada:

1. Bapak Ir. Teddy Markus Z, M.T. selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Maranatha.
2. Bapak Sedy Ferdian Sujadi, S.Kom., M.T. selaku Kaprodi S1 Sistem Informasi.
3. Adelia, S.Kom., MT. selaku Dosen koordinator Mata Kuliah Tugas Akhir.
4. Seluruh keluarga besar Sistem Informasi angkatan 2017, khususnya teman-teman yang juga mengambil Mata Kuliah Tugas Akhir.

Semoga dengan adanya laporan ini, semua pihak yang membaca mendapatkan sesuatu yang berguna dan bermanfaat. Tuhan Memberkati.

DAFTAR PUSTAKA

- [1] B. A. M. Zhaparov dan R. Suliyev, “Python to learn programming.,” Journal of Physics: Conference Series, vol. 423, 2013.
- [2] D. M D, S. Ca dan A. Ganesha, “Sentiment analysis: a comparative study on different approaches,” Procedia Computer Science, pp. 44-49, 2016.
- [3] F. A. Pozzi, E. Fersini, E. Messina dan B. Liu, Sentiment Analysis in Social Networks, Morgan Kaufmann, 2016.
- [4] B. Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool, 2012.
- [5] J. Thanaki, Python Natural Language Processing, Packt Publishing Ltd, 2017.
- [6] F. Millstein, Natural Language Processing With Python: Natural Language Processing Using NLTK, Frank Millstein, 2020.
- [7] N. Hardeniya, J. Perkins, D. Chopra, N. Joshi dan I. Mathur, Natural Language Processing: Python and NLTK, Packt Publishing Ltd, 2016.
- [8] S. Raj, Building Chatbots with Python: Using Natural Language Processing and Machine Learning, Apress, 2018.
- [9] J. A. G. Kulkarni dan V. Shah, “Natural language processing,” International Journal of Computer Sciences and Engineering, vol. 6, no. 1, pp. 161-167, 2018.
- [10] D. Chopra, N. Josh dan I. Mathur, Mastering Natural Language Processing with Python, Packt Publishing Ltd, 2016.
- [11] B. Lubanovic, Introducing Python: Modern Computing in Simple Packages, O'Reilly Media, Inc., 2019.
- [12] Python, “About Python,” Python Software Foundation, [Online]. Available: <https://www.python.org/about/>. [Diakses 24 February 2021].
- [13] S. K. R, “Python—the fastest growing programming language.,” International Research Journal of Engineering and Technology, vol. 04, no. 12, pp. 354-357, 2017.
- [14] Kaggle, “Kaggle : Your Machine Learning and Data Science Community,” Kaggle Inc, [Online]. Available: <https://www.kaggle.com/>. [Diakses 24 February 2021].
- [15] Tripadvisor, “About Tripadvisor,” Tripadvisor, 2017. [Online]. Available: <https://tripadvisor.mediaroom.com/us-about-us>. [Diakses 25 February 2021].