

LAPORAN AKHIR PROYEK PENGOLAHAN DATA BESAR

SENTIMENT ANALYSIS PADA

STREAMING DATA TWITTER UNTUK GAME

MOBILE LEGEND



DISUSUN OLEH :

12S18001 Cindy A. Siregar

12S18014 Giovanni Situmorang

12S18021 Lastris S. N. Marbun

12S18061 Angela F. Simamora

PROGRAM STUDI SARJANA SISTEM INFORMASI

FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO

INSTITUT TEKNOLOGI DEL

2022

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

I. PENDAHULUAN

Pada bab ini akan membahas mengenai latar belakang, tujuan serta ruang lingkup dari pengerjaan proyek yang akan dilakukan.

1.1 Latar Belakang

Analisis Sentimen adalah proses untuk mengidentifikasi dan menganalisis sebuah opini sehingga pengguna lebih mudah memahami maksud dari suatu data dengan menggunakan teknik pemrosesan data. Dengan menggunakan bantuan teknologi ini kita bisa mengidentifikasi lebih banyak jenis sentiment seperti tweet, review hotel, customer review dan masih banyak lagi yang nantinya dapat dimanfaatkan. Dalam analisis sentimen, tujuannya yaitu untuk menemukan pendapat orang yang diungkapkan dalam bahasa tertulis (teks). Sentimen dalam istilah berarti "apa yang dirasakan seseorang tentang sesuatu", "pengalaman pribadi, perasaan sendiri", "sikap terhadap sesuatu" atau "pendapat".

Twitter merupakan salah satu media sosial yang memiliki konsep sederhana dan cepat. Karena merupakan pesan singkat, berita atau informasi di Twitter bisa lebih mudah dicerna. Media sosial ini juga banyak digunakan sebagai objek para peneliti atau industri untuk melakukan analisis sentimen di bidang sosial, ekonomi, politik, atau bidang lainnya. Opinion mining atau juga biasa disebut dengan analisis sentimen adalah proses menganalisis teks untuk mendapatkan informasi tertentu dalam sebuah kalimat berupa opini. Penelitian ini membangun sebuah sistem untuk menganalisis sentimen dari twitter secara real-time, otomatis, dan berkelanjutan. Hasil uji coba sistem berhasil mengambil data, melakukan analisis sentimen, dan menampilkannya dalam grafik dan website real-time, dan diperbarui secara otomatis. Selanjutnya, penelitian ini akan dikembangkan dengan fokus pada keakuratan algoritma yang digunakan dalam melakukan proses analisis sentimen.

Permasalahan proyek Pengolahan Data Besar yang akan diselesaikan adalah melakukan klasifikasi analisis sentimen pada *streaming* data twitter. Sumber data yang akan digunakan merupakan data mentah yang diperoleh dari Twitter dengan memanfaatkan layanan API Stream yang disediakan oleh Twitter. Data mentah diproses sampai analisis sentimen yang diharapkan dihasilkan. Sebelum dilakukan analisis sentimen terhadap data yang diperoleh dari Twitter, dilakukan pre-processing. Pada tahap ini, pembersihan data tweet dilakukan dengan langkah-langkah sebagai berikut: *folding case*, penghapusan simbol, tokenisasi, konversi slang word, dan penghapusan stopword.

1.2 Rumusan Masalah

Bagaimana suatu sistem dapat mengklasifikasikan suatu *game mobile legend* berdasarkan *review* yang diberikan *gamers* sebagai ulasan yang bernilai positif atau ulasan yang negatif.

1.3 Tujuan

Adapun tujuan dari pengerjaan proyek ialah untuk melakukan analisis sentimen terhadap dataset *live streaming twitter* dengan menggunakan *keyword game mobile legend*.

1.4 Scope

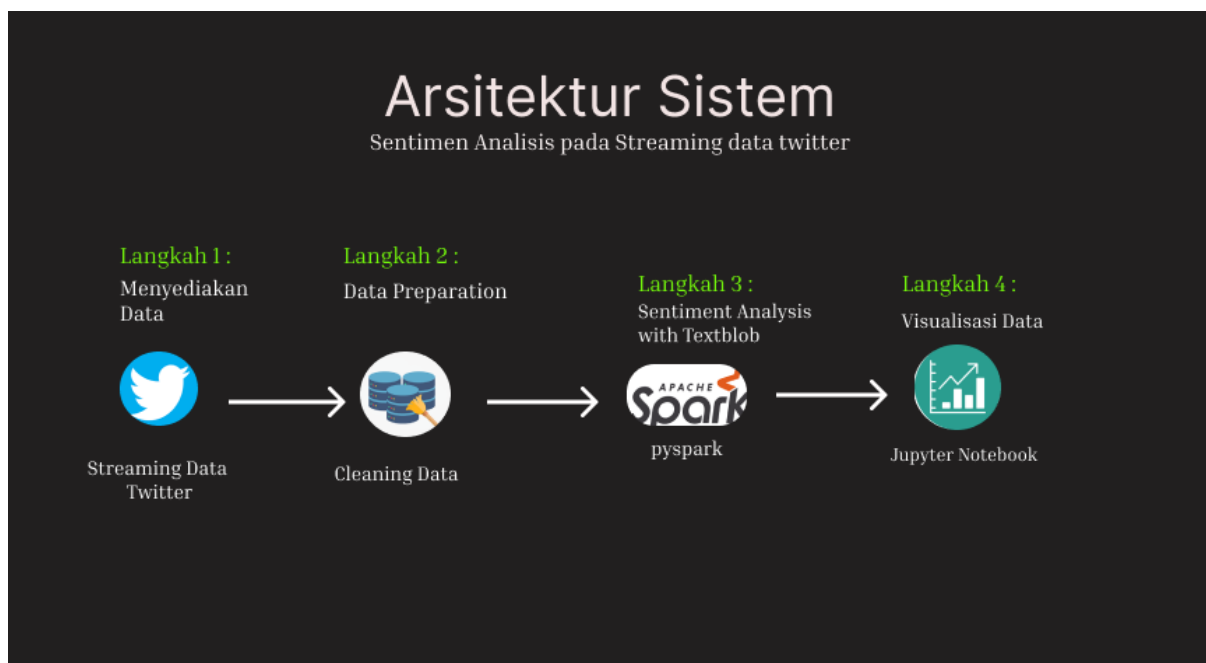
Adapun yang menjadi batasan dari pengerjaan proyek sentimen analisis pada dataset *live streaming twitter* sebagai berikut.

1. Algoritma yang digunakan dalam melakukan analisis sentimen terhadap *Games Mobile Legend Streaming Data Twitter* adalah Textblob
2. Data yang digunakan adalah data *games mobile legend* yang diperoleh dari Twitter dengan memanfaatkan layanan API Stream yang disediakan oleh Twitter

II. ARSITEKTUR SISTEM

Pada bab ini diberikan sebuah arsitektur sistem yang dapat menjadi solusi yang digunakan dalam masalah pengklasifikasian sentiment analysis *Live Streaming Data Twitter*. Saat ini *games mobile legend* merupakan sebuah produk aplikasi yang sangat populer. Maraknya perkembangan *game mobile legend* juga menimbulkan tantangan seperti banyaknya pendapat yang diberikan oleh pengguna *mobile legend*. Analisis sentimen dapat digunakan untuk memberikan informasi sentimen yang terjadi pada komentar-komentar mengenai *games mobile legend* dari streaming data twitter. Informasi yang diberikan yaitu berupa klasifikasi sentimen positif dan negatif serta metode yang digunakan yaitu *TextBlob Algorithm*.

Arsitektur big data merupakan suatu struktur keseluruhan yang mempresentasikan dari logical dan physical sistem dari big data yang akan disimpan, diakses serta diolah.



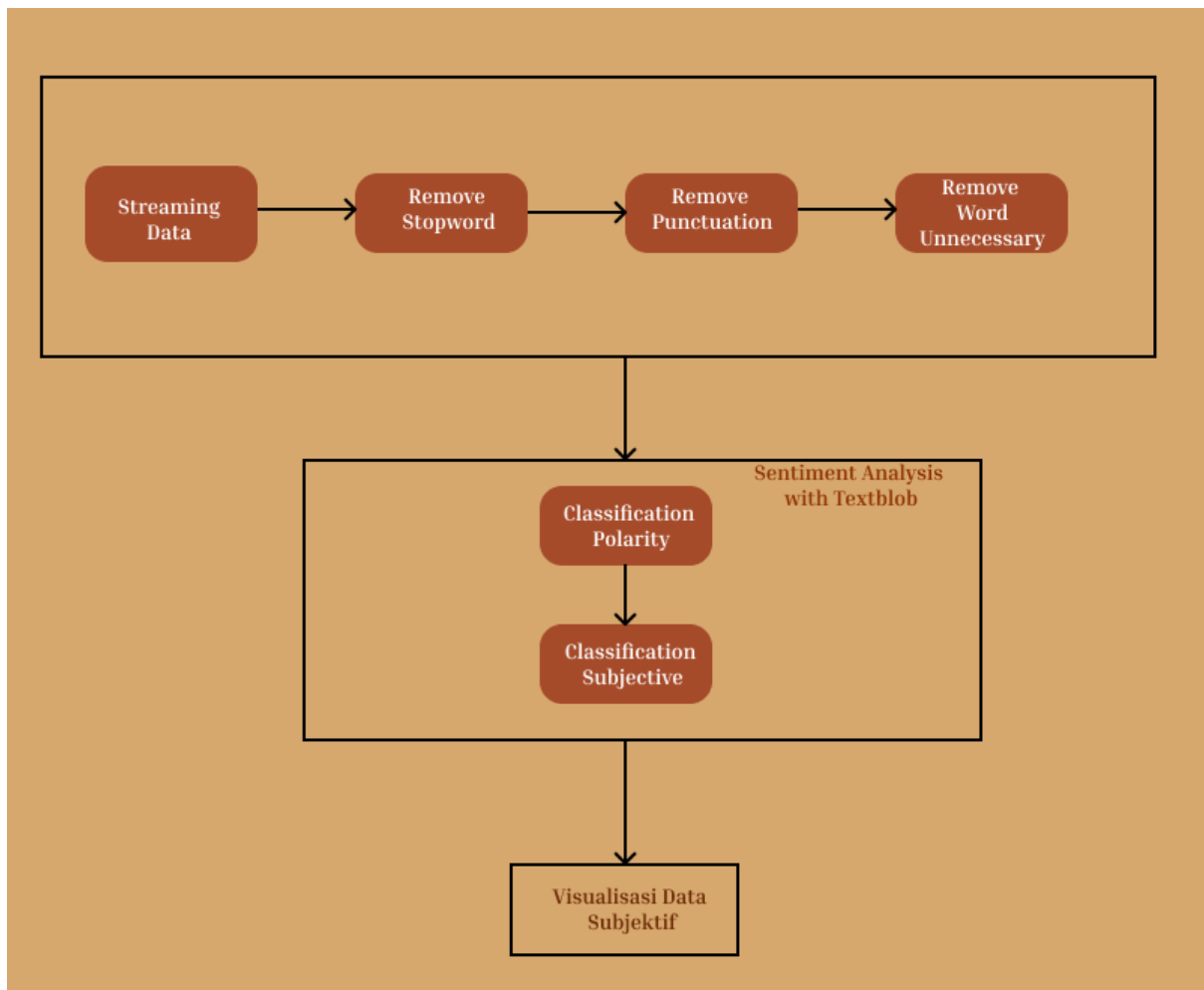
Gambar 1. Arsitektur Sistem

Pada gambar diatas dijelaskan alur arsitektur sistem yang akan digunakan dalam melakukan klasifikasi *sentiment analysis* pada *Live Streaming Twitter*. Arsitektur sistem dibentuk agar dapat mengolah data yang terstruktur dan tidak terstruktur. Data diambil menggunakan tweepy yang akan di filter menggunakan fungsi tweepy. Data yang disimpan akan dilakukan pemrosesan data dimana berguna untuk menghilangkan data yang noise, kemudian data yang telah dilakukan preprocessing, akan dipakai kembali untuk melakukan klasifikasi analisis sentimen pada Live Streaming Data Twitter. Selanjutnya dilakukan implementasi sentimen

analisis menggunakan *TextBlob*. *TextBlob* adalah pustaka sederhana yang mendukung analisis dan operasi kompleks pada data tekstual. *TextBlob* mengembalikan polaritas dan subjektivitas sebuah kalimat. Polaritas terletak antara $[-1,1]$, -1 mendefinisikan sentimen negatif dan 1 mendefinisikan sentimen positif. Kata-kata negasi membalikkan polaritas. Untuk tahapan terakhir, hasil klasifikasi sentimen analisis akan diolah kembali untuk divisualisasikan berdasarkan hasil klasifikasi sentimen analisis untuk mempermudah dalam memahami hasil pengolahan data.

III. MACHINE LEARNING PIPELINE

Pada bab ini akan dijelaskan Machine Learning Pipeline yang digunakan dalam kasus *Sentiment Analysis - Game Mobile Legend Streaming Data Twitter*.



Gambar 2. Machine Learning Pipeline

Dalam memproses data didapatkan dengan melakukan streaming data. Pada data dilakukan dengan menghapus kata-kata seperti stop words untuk proses selanjutnya. Lalu dilakukan juga penghapusan simbol dan tanda baca yang tidak penting serta penghapusan kata-kata yang tidak relevan. Setelah itu dilakukan sentimen analisis dengan textblob, secara langsung digunakan untuk mendeteksi sentimen. TextBlob melakukan klasifikasi pada polarity dan subjectivity. Rentang dari "subjectivity" itu 0.0 sampai 1.0 jika mendekati 0.0 berarti sentimen tersebut sangat objektif begitu pula sebaliknya. Rentang "polarity" itu dari -1.0 sampai 1.0 jika diterjemahkan polarity = polaritas, artinya hal yang memiliki dua sifat yang berlawanan. Jadi kalau polarity-nya 0.0 berarti sentimen tersebut bersifat netral. Setelah data

selesai diklasifikasi dilanjutkan dengan tahap akhir yaitu menggambarkannya pada bentuk grafik.

IV. IMPLEMENTASI

Pada bab ini akan dijelaskan mengenai lingkungan implementasi sentiment analysis pada streaming data twitter untuk game mobile legend.

4.1 Stream Data

Jika belum mempunyai akun Twitter Developers, akan diminta untuk mengisi detail seperti identitas lanjutan, alasan pengajuan API, digunakan untuk apa dan seterusnya. Setelah disetujui akun developernya selanjutnya Masuk di Tab Keys and Access Tokens dan simpan *consumerKey*, *consumerSecret*, *accessTokenKey* dan *accessTokenSecret*. Dengan melakukan import tweepy dan setup autentikasi Twitter API sebagai berikut.

```
In [3]: # Authentication
consumerKey = "VCXi57kKSV9XknYF6URbKsZLX"
consumerSecret = "RLRuh0KUVgGjGjGczmRIC66HbyUPrM6ltM9pVGXigI0Ub15qyd"
accessToken = "722625417081802756-Ma1ZvIpwrIZQK4K83umzgMSHQhg5bVr"
accessTokenSecret = "tjuHWYt23cLjpbqJI03ewTMKV9zmcPiKFGXatmSdd5ksB0"

auth = tweepy.OAuthHandler(consumerKey, consumerSecret)
auth.set_access_token(accessToken, accessTokenSecret)
api = tweepy.API(auth)
```

Selanjutnya dilakukan import library yang akan digunakan dalam memproses data pada code sebagai berikut.

```
In [1]: #Import Libraries

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import tweepy
from tweepy import OAuthHandler
```

4.2 Remove Stopword, Punctuation, Word Unnecessary

Setelah melakukan import library, dilanjutkan dengan menghapus stopwords, tanda baca yang tidak sesuai serta kata-kata yang tidak relevan dengan code sebagai berikut.

```
In [7]: # Fungsi untuk membersihkan data tweet

import re
def clean_tweet(tweet):
    return ' '.join(re.sub('([A-Za-z0-9+]|([^0-9A-Za-z \t])|(\w+:\w+\w\S+)|<[RT]>)', ' ', str(tweet).lower()).split())

In [8]: #Menghapus RT, Punctuation etc
remove_rt = lambda x: re.sub('RT \w+: ', " ", x)
rt = lambda x: re.sub("(@[A-Za-z0-9+]|([^0-9A-Za-z \t])|(\w+:\w+\w\S+)", " ", x)
df['Tweet'] = df.Tweet.map(remove_rt).map(rt)
df['Tweet'] = df.Tweet.str.lower()

In [9]: #Menghapus STOPWORDS
from nltk.corpus import stopwords
stop_words=stopwords.words('english')
df['Tweet'] = df['Tweet'].apply(lambda x:" ".join([i for i in str(x).split(" ") if i not in stop_words]))
```

4.3 Sentiment Analysis with Textblob

Sentimen analisis adalah proses menganalisis data dan mengklasifikasikannya berdasarkan kebutuhan proyek penelitian. Pada proyek ini, sentimen analisis menggunakan TextBlob dengan code sebagai berikut.

```
: # Fungsi untuk melakukan analisis sentiment

from textblob import TextBlob
def analyze_sentiment(tweet):
    analysis = TextBlob(tweet)
    if analysis.sentiment.polarity > 0:
        return 'Positive'
    elif analysis.sentiment.polarity == 0:
        return 'Neutral'
    else:
        return 'Negative'

: def prepCloud(Topic_text,Topic):
    Topic = str(Topic).lower()
    Topic=' '.join(re.sub('([^0-9A-Za-z \t])', ' ', Topic).split())
    Topic = re.split("\s+",str(Topic))
    stopwords = set(STOPWORDS)
    stopwords.update(Topic)

    text_new = " ".join([txt for txt in Topic_text.split() if txt not in stopwords])
    return text_new

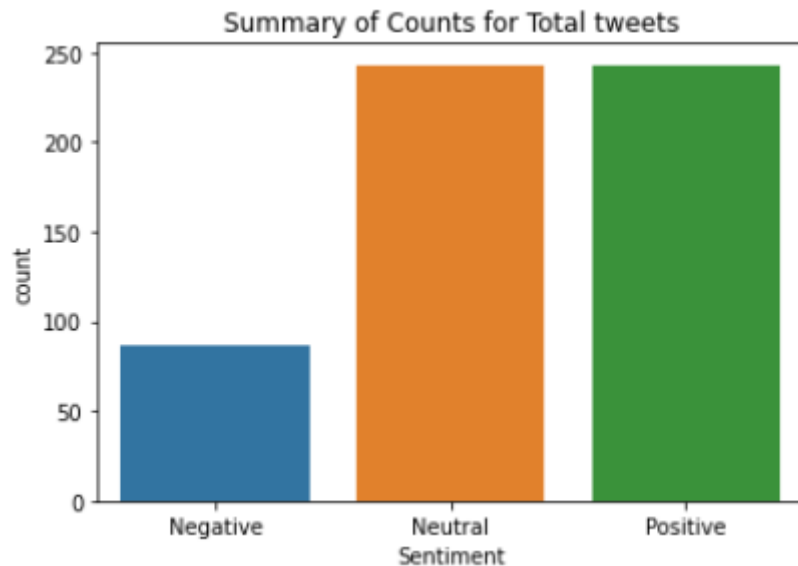
: df['clean_tweet'] = df['Tweet'].apply(lambda x : clean_tweet(x))
df.head(5)
```

4.4 Visualisasi Data

Setelah data tweet sudah dilakukan proses sentiment analysis, maka tahap selanjutnya adalah membuat hasil sentiment analysis ke dalam bentuk visualisasi. Berikut visualisasi yang telah dilakukan.

```
sns.countplot(df["Sentiment"])
plt.title("Summary of Counts for Total tweets")
```

Visualisasi yang digunakan untuk memvisualisasikan data polarity hasil dari sentiment analysisnya yaitu bar chart.



Berdasarkan chart diatas, dapat dilihat banyak data pada setiap label. Untuk data tweet yang berlabel positive lebih banyak dibandingkan dengan label yang lain. Oleh karena data yang digunakan merupakan data yang bersifat positive dan neutral.



V. HASIL DAN PEMBAHASAN

4.1 Hasil Prediksi

Pada sentiment analisis, kami menggunakan textblob dan telah dilakukan prediksi terhadap data tweet pada twitter mengenai polarity dan subjectivity.

Pada perlakuan prediksi polarity, terdapat 3 tabel yaitu positive, negative dan neutral.

Dibawah ini adalah jumlah data tweet sesuai label prediksi :

Tabel Jumlah data tweet sesuai label prediksi

| No | Label | Jumlah Data |
|----|----------|-------------|
| 1 | Positive | 42.41% |
| 2 | Neutral | 42.41% |
| 3 | Negative | 15.18% |

VI. KESIMPULAN DAN SARAN

Kesimpulan yang dapat diambil yaitu algoritma Textblob dapat memberikan solusi sentiment analisis dalam memprediksi terhadap data tweet pada twitter mengenai polarity dan subjectivity. Dimana jumlah data tweet positif dan neutral adalah 42.41% dan negative adalah 15.18%. Sehingga yang digunakan adalah data yang bersifat positive dan neutral.