

The background features abstract, translucent geometric shapes in shades of blue, purple, and pink, floating in the upper right quadrant.

Sentiment Analysis

Final Project

DADS6005 : Data Streaming and Real Time Analysis

Our Team



Chokchai Kenpho
6410422004



Noppol Anakpluek
6410422009



Supanan Sukhamta
6410422020



Natthaporn Wattanakul
6410422026



Watcharakorn Pasanta
6420422006

Contents

Q1

Project
Description

Q2

Diagram

Q3

Modeling Process

- Data Source
- Data Management
- Real Time Process
- Analytics Result

Q4

Insights and
Conclusions

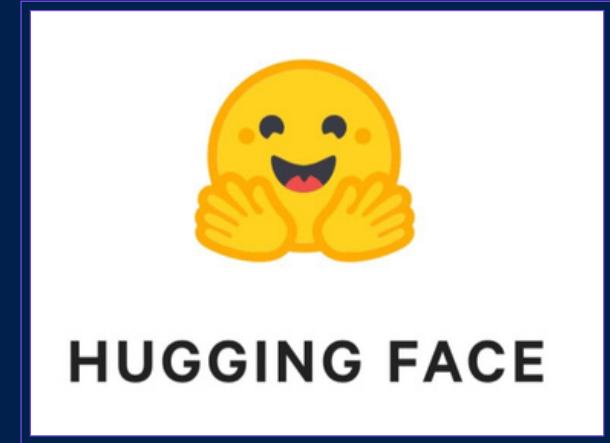
Q1 : Project Discription

This project is about analyzing social media comments on live Streaming of HEARTROCKER Channel and predicting trends with sentiment classification

Q2 : Diagram



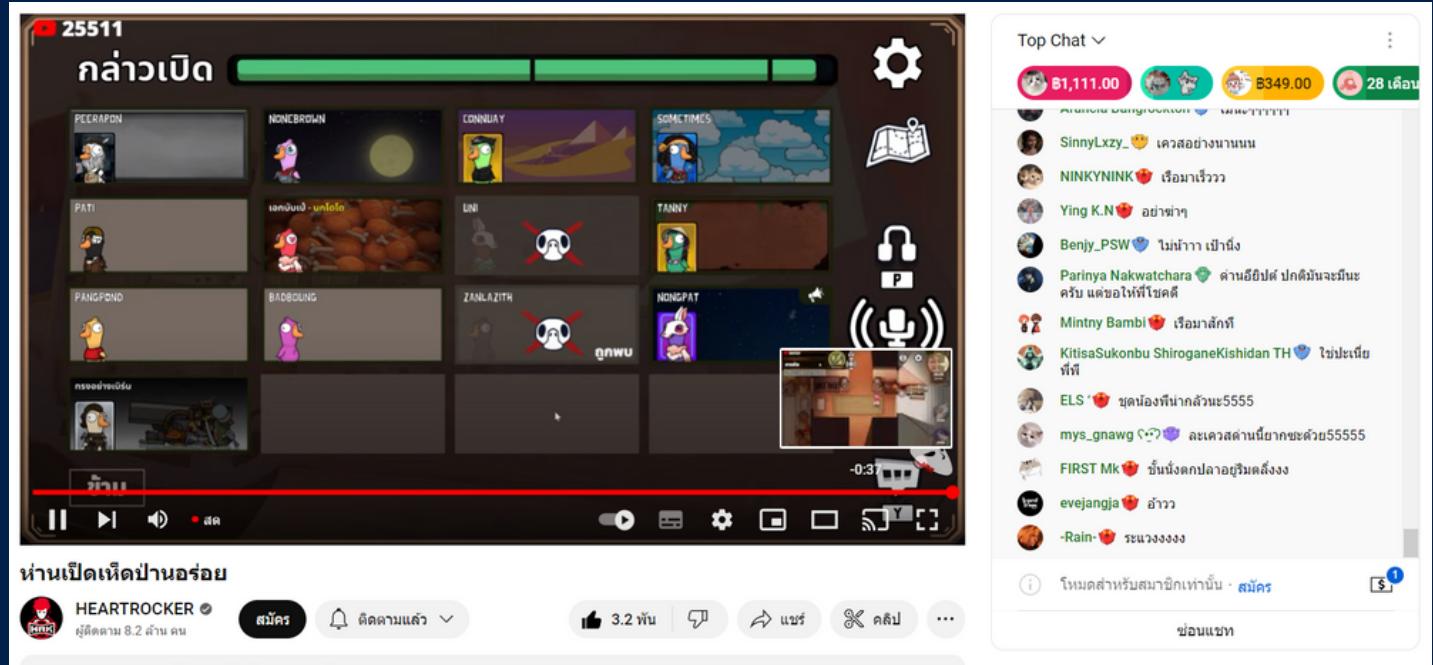
Hive MQ



Pre-Train
Sentiment model



Q3 : Data Collection and Modeling Process



<https://www.youtube.com/watch?v=EBcvHK2Moys>



Pre-Train Sentiment model



Negative



Neutral



Positive

Sentiment model

The screenshot shows the Hugging Face homepage with a sidebar on the left containing sections for Tasks, Libraries, and Datasets. The Tasks section lists Image Classification, Translation, Image Segmentation, Fill-Mask, Automatic Speech Recognition, Token Classification, Sentence Similarity, Audio Classification, Question Answering, Summarization, and Zero-Shot Classification. The Libraries section lists PyTorch, TensorFlow, JAX, and others. The Datasets section lists common datasets like common_voice, wikipedia, glue, and emotion.

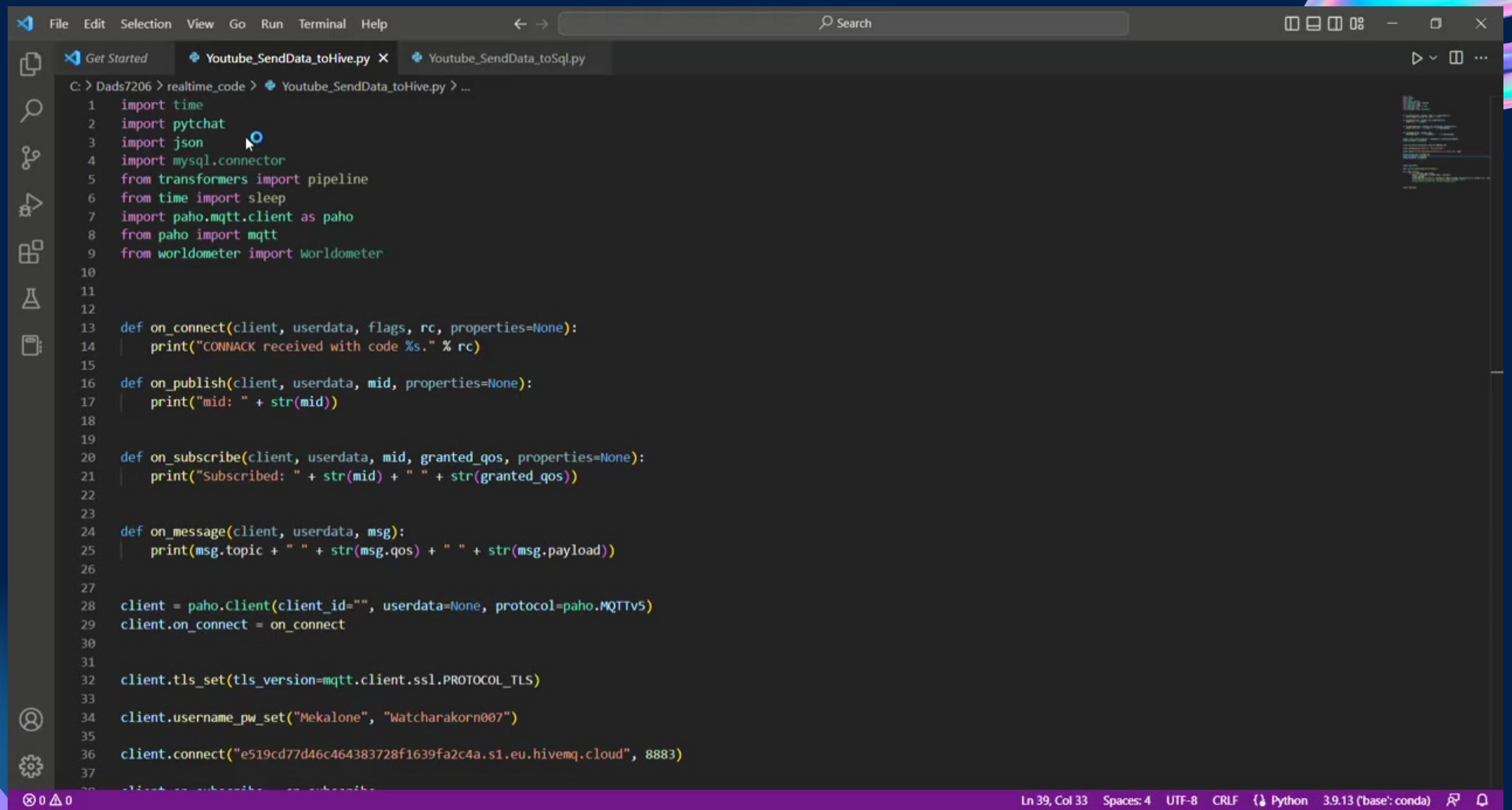
WangchanBERTa: Pre-trained Thai Language Model

โมเดลภาษาสำหรับงานประมวลผล และการเข้าใจภาษาไทย

สถาบันวิจัยปัญญาประดิษฐ์ประเทศไทย (Thailand Artificial Intelligence Research Institute) ได้ทำการเทรนโมเดลภาษา (language model) บนชุดข้อมูลในภาษาไทยที่ได้จากแหล่งต่างๆ เช่น ข่าว, วิกิพีเดีย, ข้อความในโซเชียลมีเดีย และข้อมูลที่ได้จากการ crawl เว็บไซต์ในอินเทอร์เน็ต ซึ่งมีขนาดข้อมูลรวม 78.5 GB และได้วัดประสิทธิภาพของโมเดลภาษาที่ finetune แล้ว ได้ผลคะแนน micro-averaged F1 score สูงที่สุดบน 5 ชุดข้อมูล จากห้องทดลอง 6 ชุดข้อมูล โดยเป็นชุดข้อมูลทดสอบในโจทย์การจำแนกข้อความ (text classification) และการจำแนกคำ (token classification) เมื่อเทียบกับ baseline model และโมเดลภาษาแบบหลายภาษา (multilingual language model) ที่มีอยู่ในปัจจุบัน (mBERT และ XLMR)

The screenshot shows the WangchanBERTa-finetuned-sentiment model card on the Hugging Face platform. The card includes the model name, a brief description, and various configuration options. It lists the following tags: Text Classification, PyTorch, TensorBoard, Transformers, wongnai_reviews, wisegraph_sentiment, generated_reviews_enth, Thai, camembert, sentiment-analysis, and License: apache-2.0. Below the card are navigation links for Model card, Files and versions, Training metrics, Community, and several deployment and integration buttons.

Source Code : publish



The screenshot shows a code editor window with the following details:

- Title Bar:** File Edit Selection View Go Run Terminal Help
- Search Bar:** Search
- Toolbar:** Includes icons for file operations (New, Open, Save, Close), search, and other common functions.
- Left Sidebar:** Shows project navigation with "Get Started", "Youtube_SendData_toHive.py", and "Youtube_SendData_toSql.py".
- Code Area:** Displays Python code for an MQTT publisher. The code includes imports for time, pytchat, json, mysql.connector, transformers, paho.mqtt.client, paho.mqtt, and worldometer. It defines four callback functions: on_connect, on_publish, on_subscribe, and on_message. The on_connect function prints the connection code. The on_publish function prints the message ID. The on_subscribe function prints the subscribed topic and QoS. The on_message function prints the topic, QoS, and payload. The main part of the code creates a client, sets TLS, and connects to "e519cd77d46c464383728f1639fa2c4a.s1.eu.hivemq.cloud" on port 8883, using the previously defined callbacks.
- Status Bar:** ShowsLn 39, Col 33 Spaces: 4 UTF-8 CRLF {Python 3.9.13 ('base':conda) R Q

```
C:> Dads7206 > realtime_code > Youtube_SendData_toHive.py > ...
1  import time
2  import pytchat
3  import json
4  import mysql.connector
5  from transformers import pipeline
6  from time import sleep
7  import paho.mqtt.client as paho
8  from paho import mqtt
9  from worldometer import Worldometer
10
11
12
13 def on_connect(client, userdata, flags, rc, properties=None):
14     print("CONNACK received with code %s." % rc)
15
16 def on_publish(client, userdata, mid, properties=None):
17     print("mid: " + str(mid))
18
19
20 def on_subscribe(client, userdata, mid, granted_qos, properties=None):
21     print("Subscribed: " + str(mid) + " " + str(granted_qos))
22
23
24 def on_message(client, userdata, msg):
25     print(msg.topic + " " + str(msg.qos) + " " + str(msg.payload))
26
27
28 client = paho.Client(client_id="", userdata=None, protocol=paho.MQTTv5)
29 client.on_connect = on_connect
30
31
32 client.tls_set(tls_version=paho.client.ssl.PROTOCOL_TLS)
33
34 client.username_pw_set("Mekalone", "Watcharakorn007")
35
36 client.connect("e519cd77d46c464383728f1639fa2c4a.s1.eu.hivemq.cloud", 8883)
```

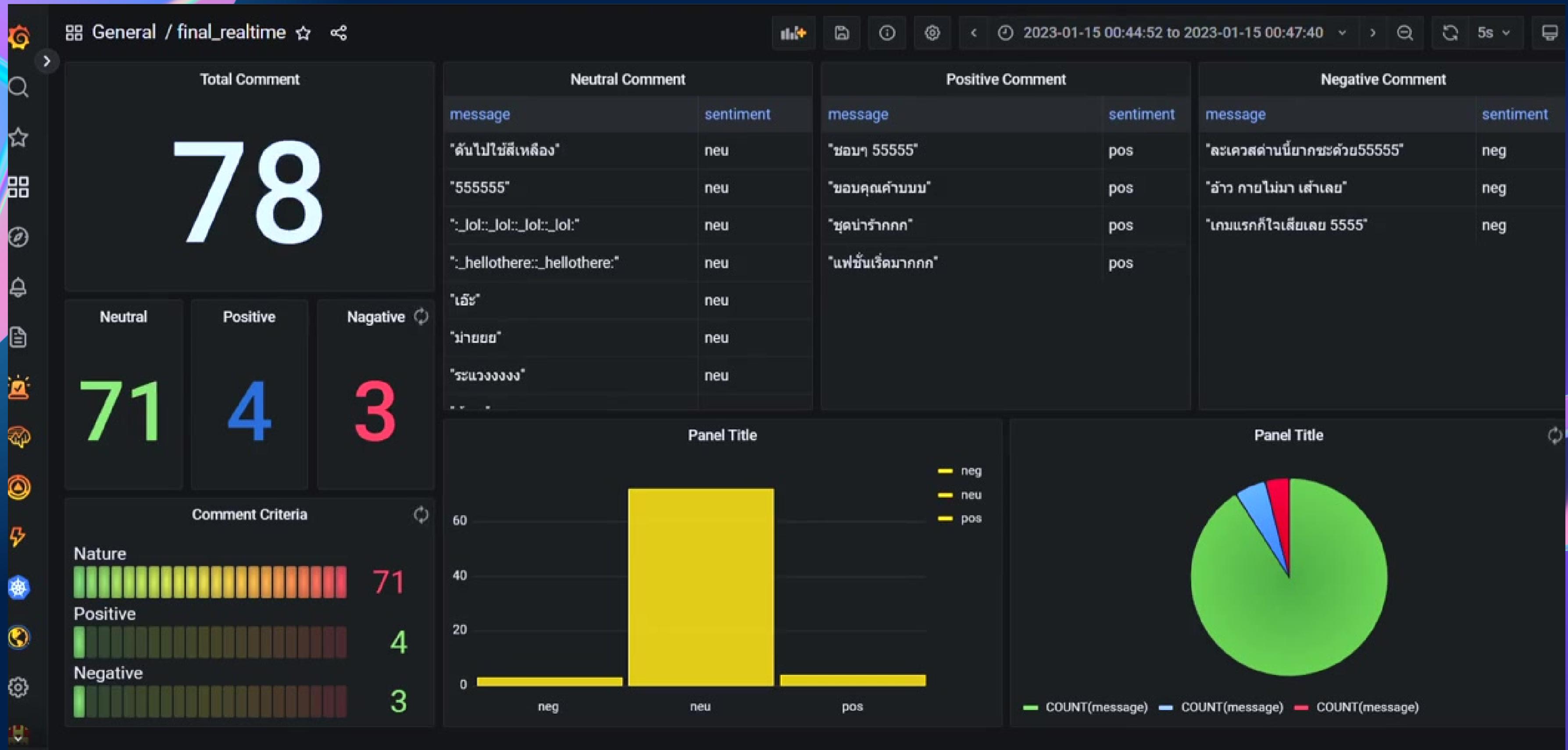
Source Code : subscribe

The screenshot shows a terminal window with the following details:

- File Bar:** File, Edit, Selection, View, Go, Run, Terminal, Help.
- Search Bar:** Search.
- Toolbar:** Includes icons for Get Started, Youtube_SendData_toHive.py, andYoutube_SendData_toSql.py (the active tab).
- Left Sidebar:** Icons for file operations like Open, Save, Find, Copy, Paste, and a gear icon.
- Code Area:** Displays Python code for an MQTT client. The code includes imports for time, paho.mqtt.client, paho, mysql.connector, transformers, and requests. It defines several callback functions: on_connect, on_publish, on_subscribe, and on_message. The on_subscribe function prints the subscribed topic and QoS level. The on_message function prints the topic, QoS, and payload, and inserts the data into a MySQL database named "youtube". A sentiment analysis function is also defined using a BERT model.
- Bottom Status Bar:** Shows line 19, column 62, spaces: 4, encoding: UTF-8, CRLF, Python 3.9.13 ('base': conda), and other status indicators.

```
C:\> Dads7206 > realtime_code > Youtube_SendData_toSql.py > on_subscribe
1  import time
2  import paho.mqtt.client as paho
3  from paho import mqtt
4  import mysql.connector
5  from transformers import pipeline
6  import requests
7
8
9
10 def on_connect(client, userdata, flags, rc, properties=None):
11     print("CONNACK received with code %s." % rc)
12
13
14 def on_publish(client, userdata, mid, properties=None):
15     print("mid: " + str(mid))
16
17
18 def on_subscribe(client, userdata, mid, granted_qos, properties=None):
19     print("Subscribed: " + str(mid) + " " + str(granted_qos))
20
21
22 def on_message(client, userdata, msg):
23     print(msg.topic + " " + str(msg.qos) + " " + str(msg.payload))
24     insertMySQL("youtube",str(msg.payload.decode()),sentiment(msg.payload.decode()))
25
26
27 def sentiment(message):
28     classifier = pipeline('sentiment-analysis',model="poom-sci/WangchanBERTa-finetuned-sentiment")
29     sentiment = classifier(message)
30     mylist = sentiment[0]
31     sentclass = mylist['label']
32
33     return sentclass
34
35
36
37
```

Analytics Result Visualization with Grafana



Q4 : Conclusions



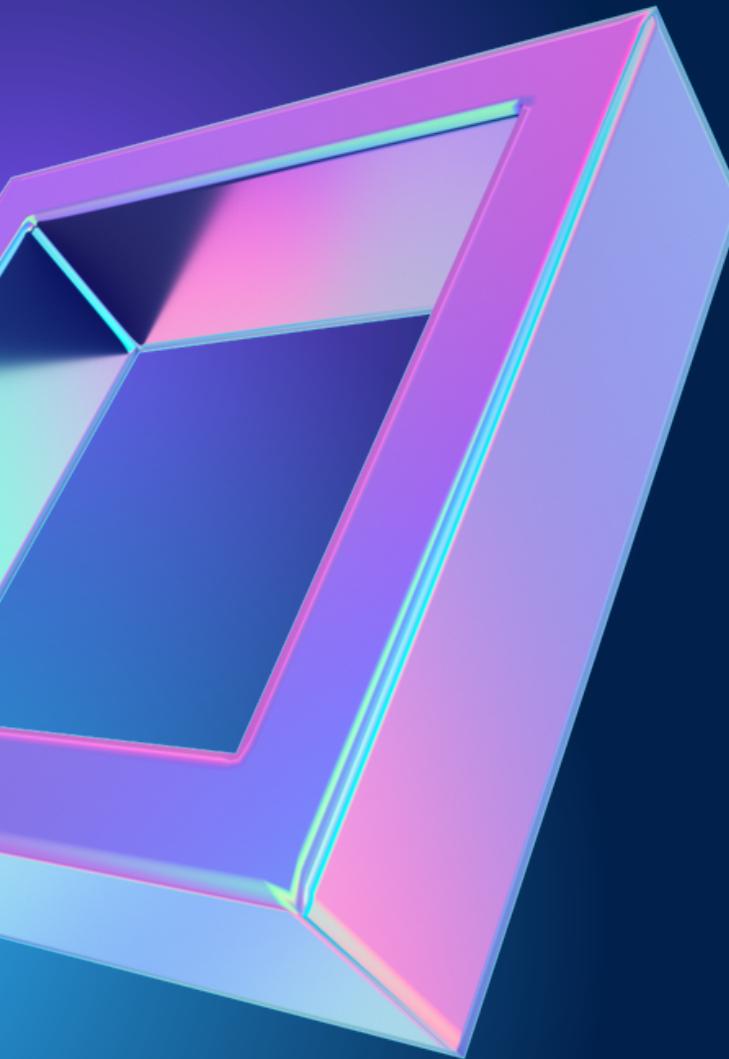
Real-Time Dashboard Show

1. Total Comment
2. Sentiment Analysis
 - o Positive Comments
 - o Neutral Comments
 - o Negative Comments
3. Comment Text Group by Sentiment Analysis
4. Sentiment Analysis Proportion



Discussion

- Before training model, the model has not high accuracy.
- customer's opinion or recommendations are a treasure trove of information that can be analyzed and processed to assess customer sentiment. So that the webpage can plan and improve for future work.



Thank You