

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

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



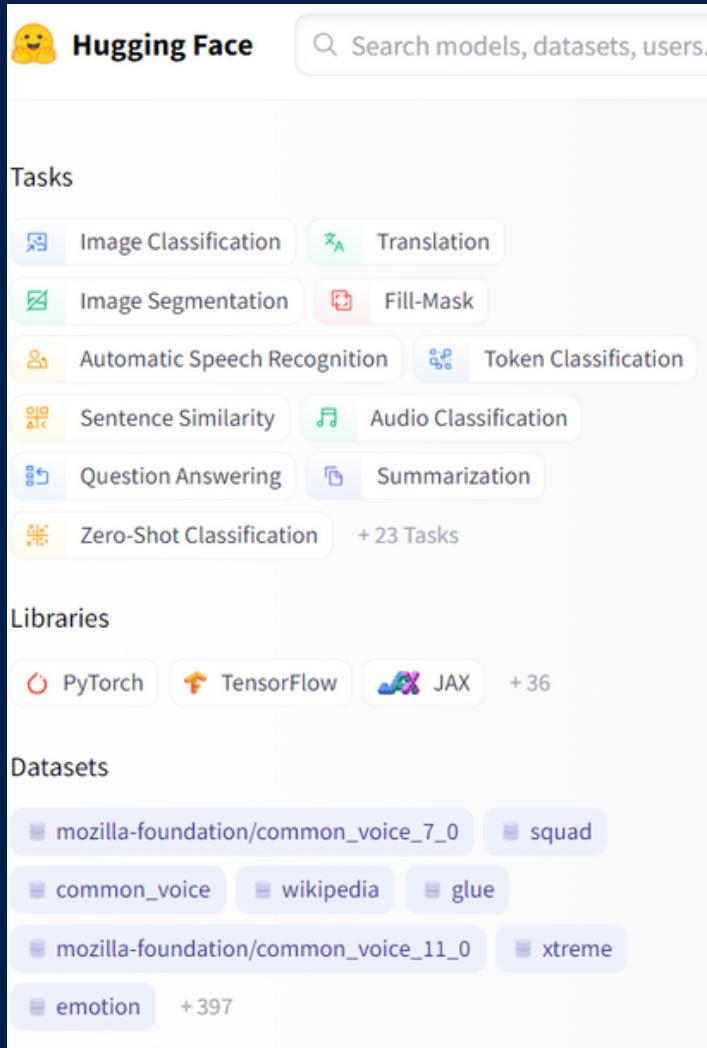
Hive MQ



Pre-Train
Sentiment model



Sentiment model



The sidebar of the Hugging Face homepage, featuring a search bar at the top, followed by sections for Tasks, Libraries, and Datasets.

- Tasks:** Image Classification, Translation, Image Segmentation, Fill-Mask, Automatic Speech Recognition, Token Classification, Sentence Similarity, Audio Classification, Question Answering, Summarization, Zero-Shot Classification, + 23 Tasks.
- Libraries:** PyTorch, TensorFlow, JAX, + 36.
- Datasets:** mozilla-foundation/common_voice_7_0, squad, common_voice, wikipedia, glue, mozilla-foundation/common_voice_11_0, xtreme, emotion, + 397.

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 จากทั้งหมด 6 ชุดข้อมูล โดยเป็นชุดข้อมูลทดสอบในโจทย์การจำแนกข้อความ (text classification) และการจำแนกคำ (token classification) เมื่อเทียบกับ baseline model และโมเดลภาษาแบบหลายภาษา (multilingual language model) ที่มีอยู่ในปัจจุบัน (mBERT และ XLMR)



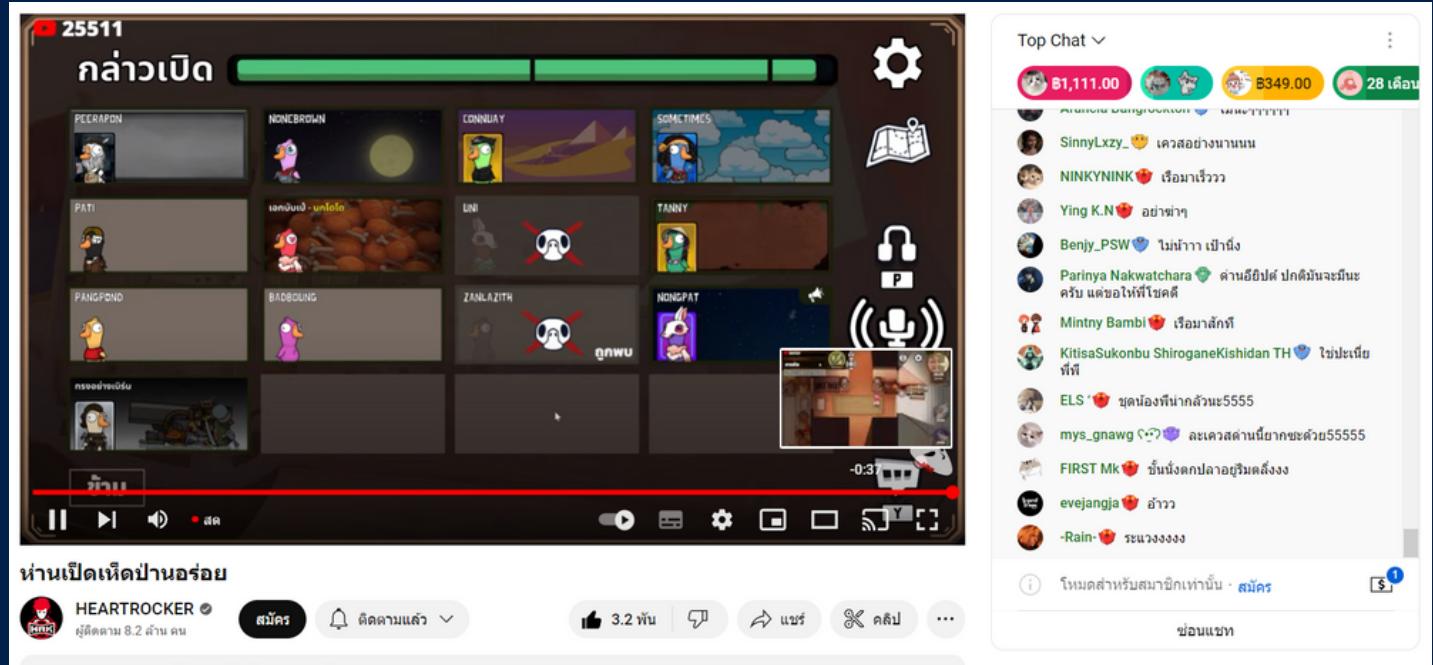
The model card for the WangchanBERTa-finetuned-sentiment model, showing details such as the owner (poom-sci), name, likes, and various tags.

poom-sci/WangchanBERTa-finetuned-sentiment like 3

Tags: Text Classification, PyTorch, TensorBoard, Transformers, wongnai_reviews, wisegraph_sentiment, generated_reviews_enth, Thai, camembert, sentiment-analysis, License: apache-2.0

Model card, Files and versions, Training metrics, Community, Train, Deploy, Use in Transformers

Q3 : Data Collection and Modeling Process



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

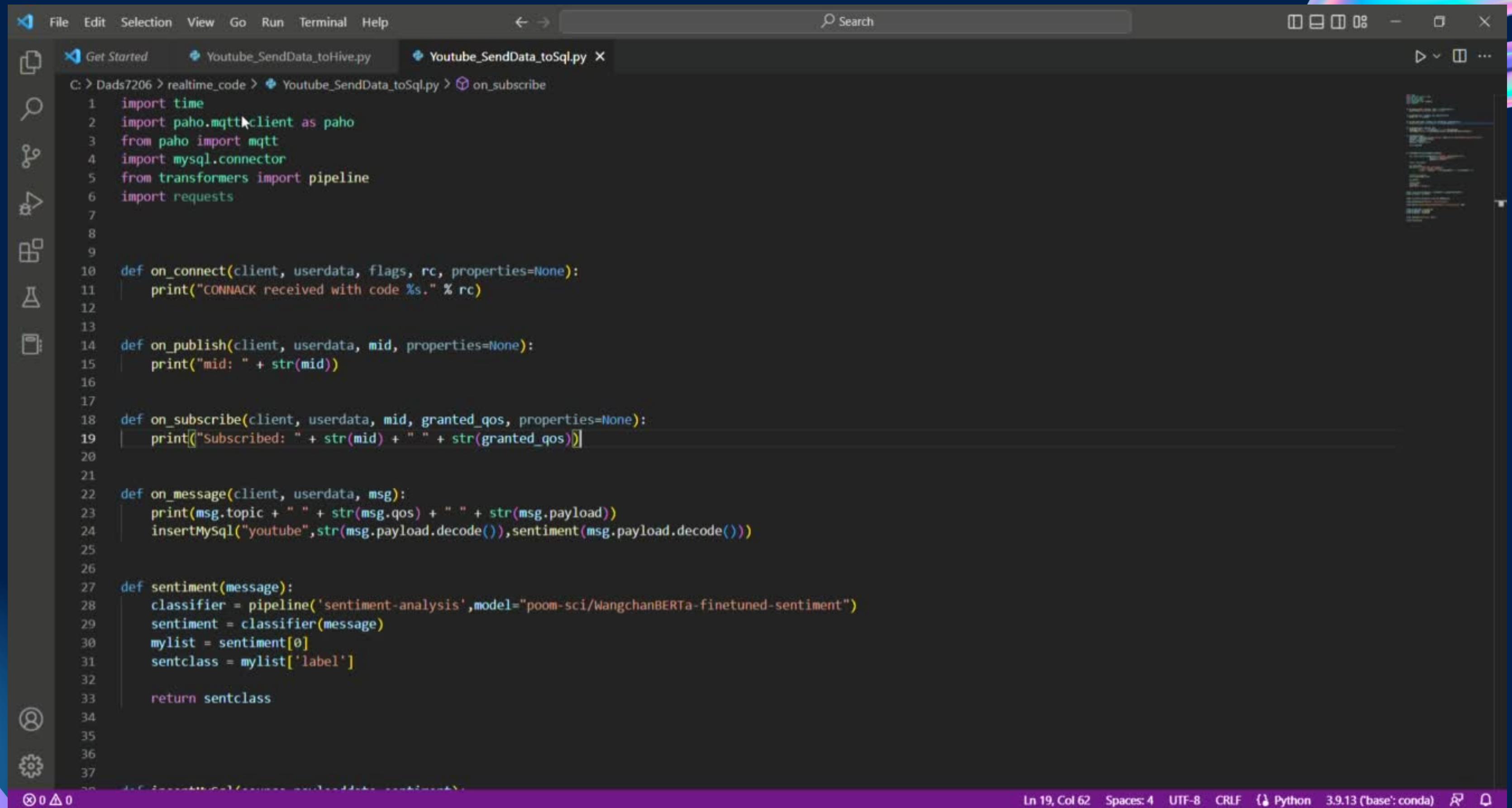


Pre-Train Sentiment model



Source Code : publish

Source Code : subscribe

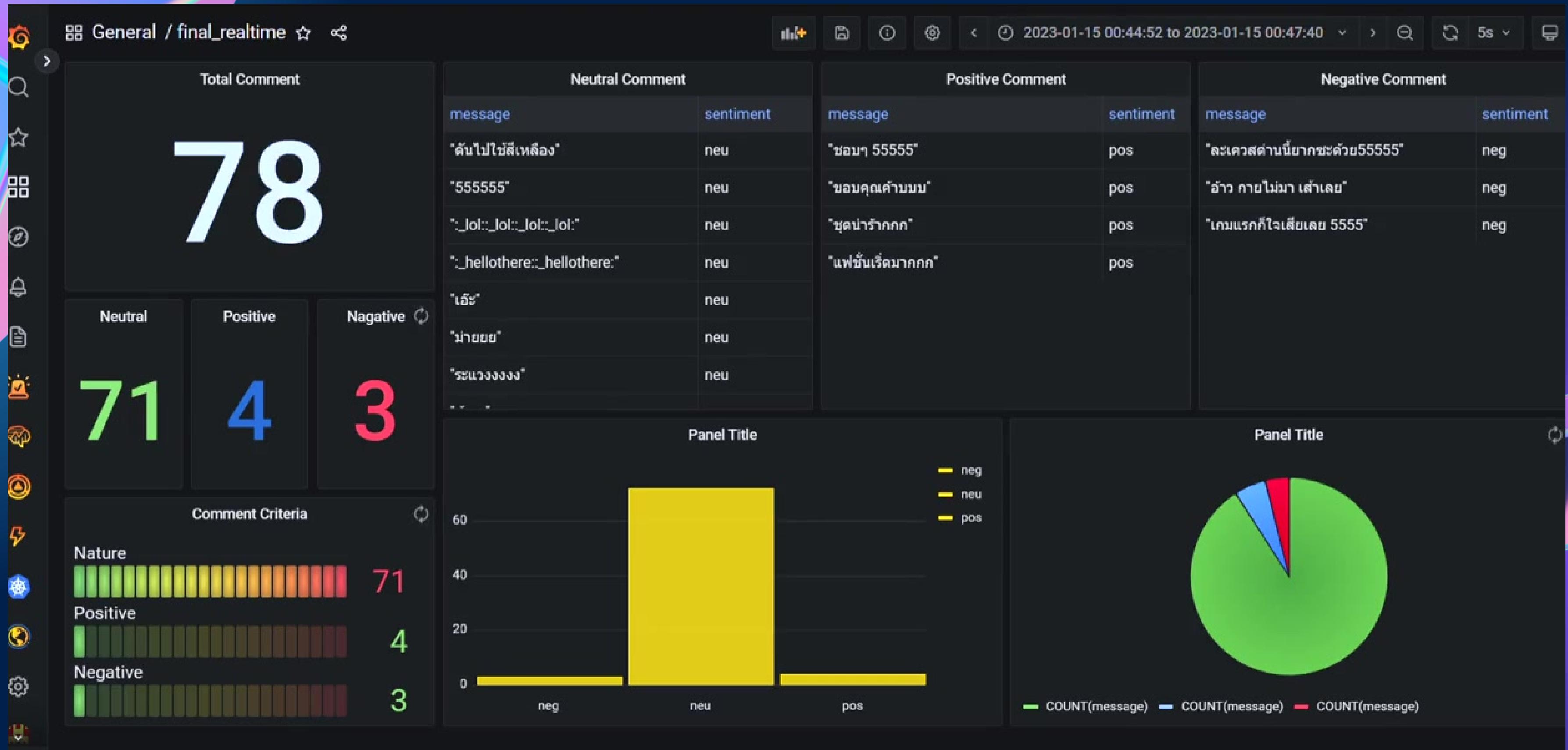


The screenshot shows a code editor window with the following details:

- File Menu:** 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:** Contains icons for file operations (New, Open, Save, Find, Copy, Paste, Delete), a search icon, a refresh icon, a project icon, a file icon, a person icon, 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.pipeline, and requests. It defines several callback functions: on_connect, on_publish, on_subscribe, and on_message. The on_subscribe function prints "Subscribed: " followed by the message ID and granted QoS. The on_message function prints the topic, QoS, and payload, then inserts the data into a MySQL database using insertMySQL. A sentiment analysis function is also defined using a BERT model from the transformers library.
- 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.

Our Team



Chokchai Kenpho
6410422004



Noppol Anakpluek
6410422009



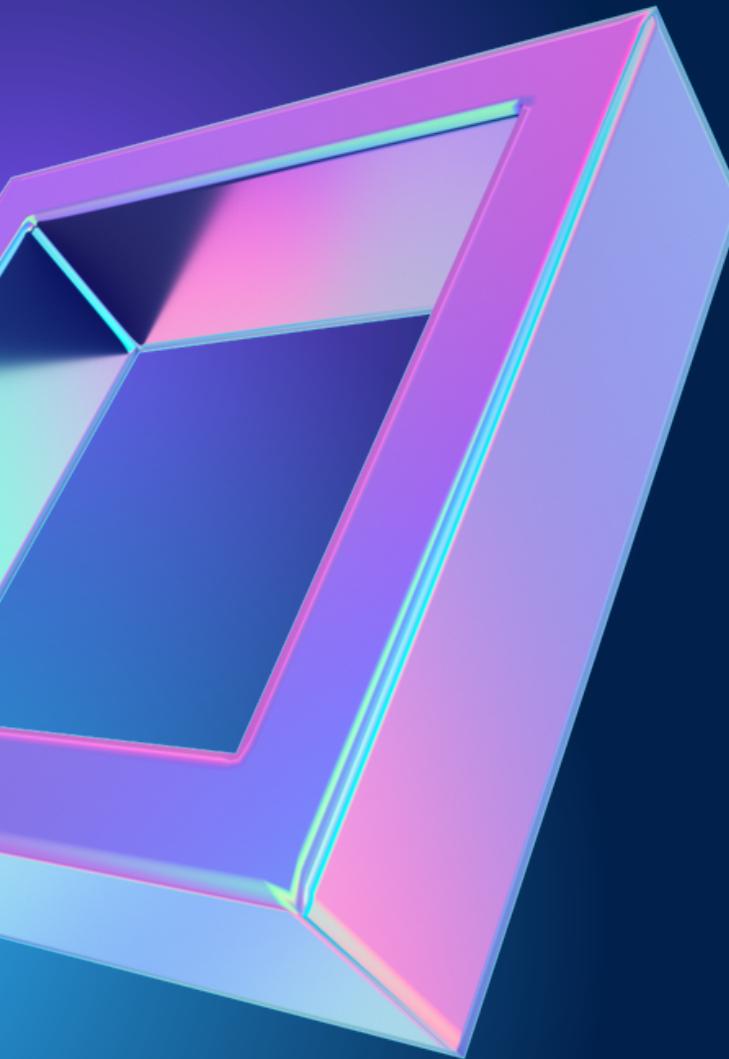
Supanan Sukhamta
6410422020



Natthaporn Wattanakul
6410422026



Watcharakorn Pasanta
6420422006



Thank You