



การวิเคราะห์ความรู้สึกในโพสต์ Twitter Sentimental Analysing using Twitter data

ชิง แซ่เสี่ย, กฤษณะ นีโรลา และ อ.ดร.ศิริสรพ เหล่าหะเกียรติ
หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

Abstract

Nowadays, many people post things on various social networks like Facebook, Twitter and Youtube. In this study, we perform sentiment analysis over posts in Twitter to determine how users think about a topic in the social media. In this study, we adopt long short-term memory (LSTM) as a classifier used in the prediction. Posts are mapped using Word2Vec method to map text to vectors for analyzing further. We evaluate the performance of the proposed system in comparison with other algorithms to determine the suitability of the proposed system. The experimental results show that the proposed system slightly outperform the compared algorithms.

ภาพรวมการดำเนินงาน

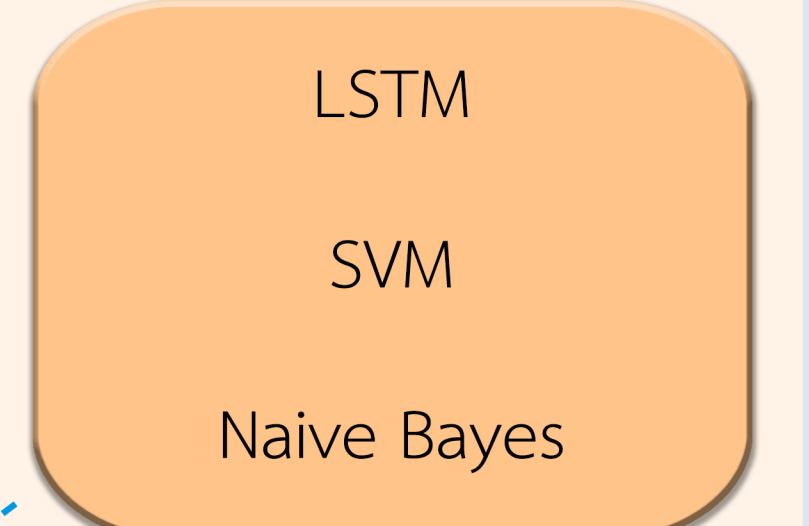
Training step



ดึงโพสต์จาก Twitter



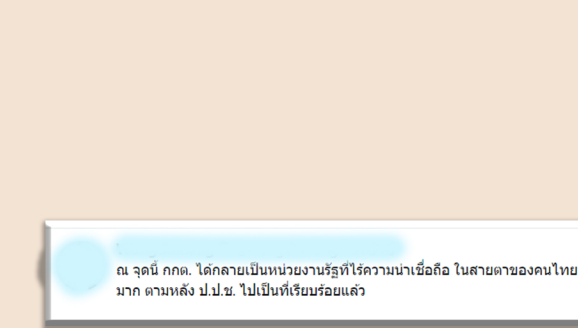
แยกแยะอารมณ์ของโพสต์



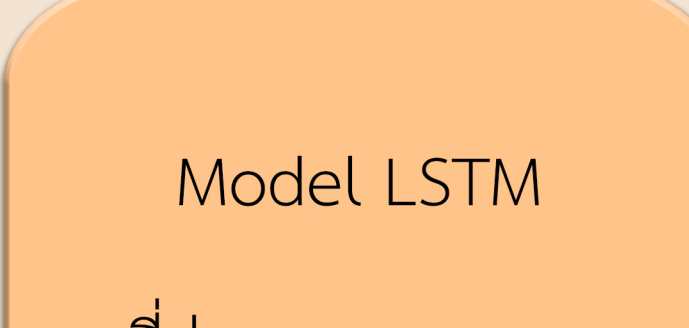
LSTM
SVM
Naive Bayes

Training Model

Pretiction step



ดึงโพสต์จาก Twitter



แยกแยะอารมณ์ของโพสต์



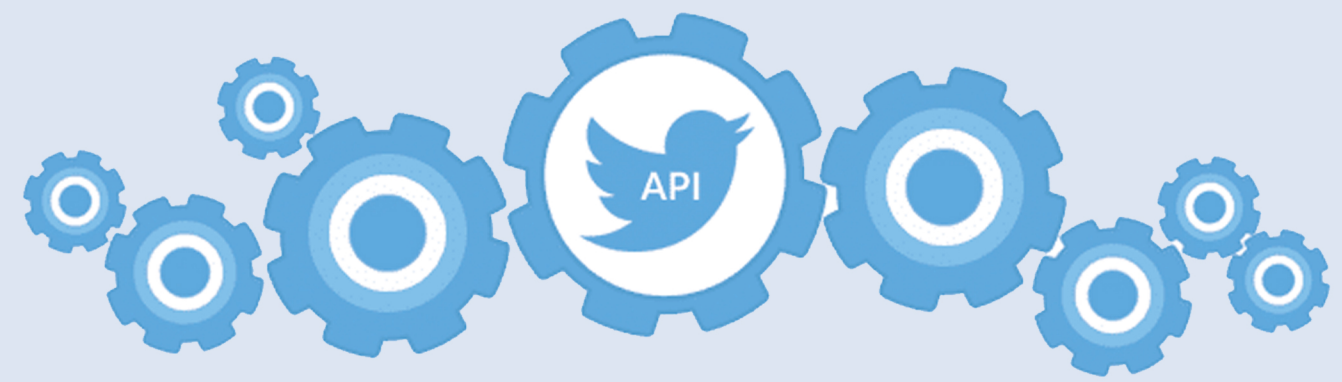
LSTM
SVM
Naive Bayes

Training Model

องค์ความรู้ที่เกี่ยวข้อง

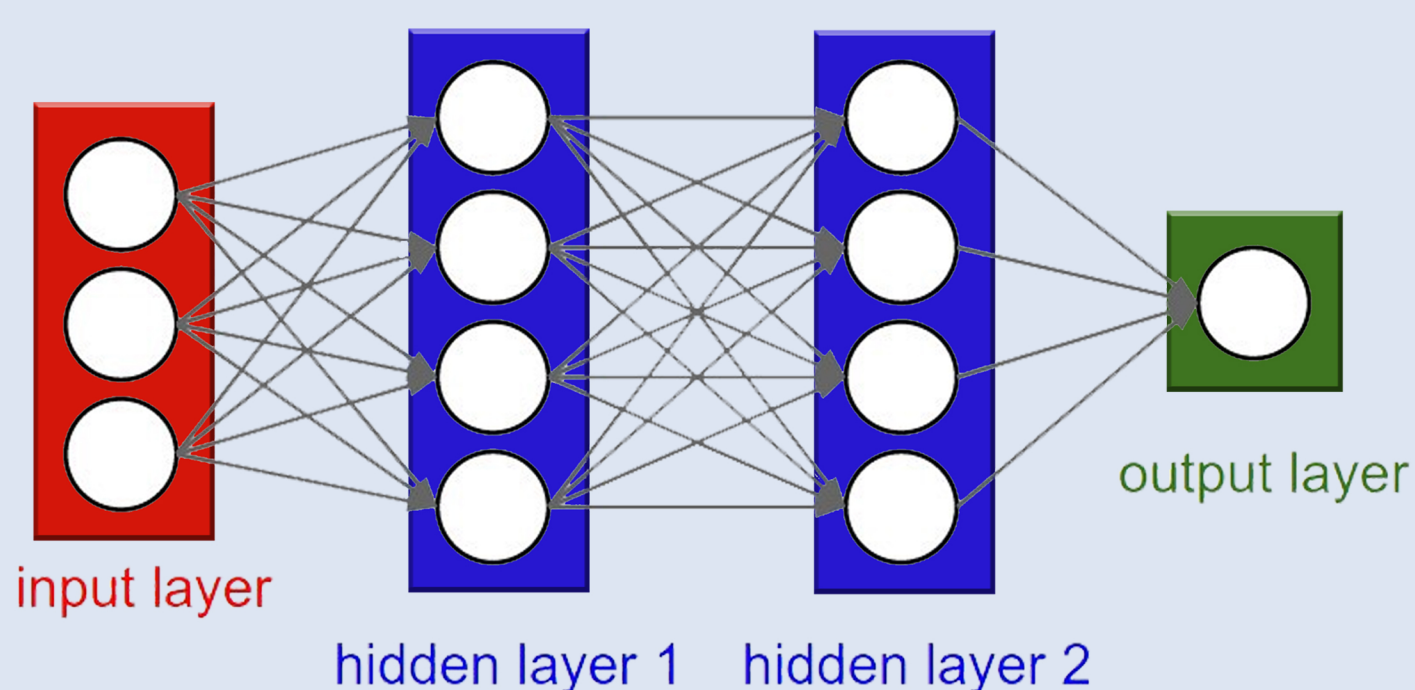
Twitter API^[1]:

- ถูกพัฒนาโดย Twitter เพื่อให้ผู้ใช้สามารถดึงข้อมูลจาก Twitter มาใช้ได้



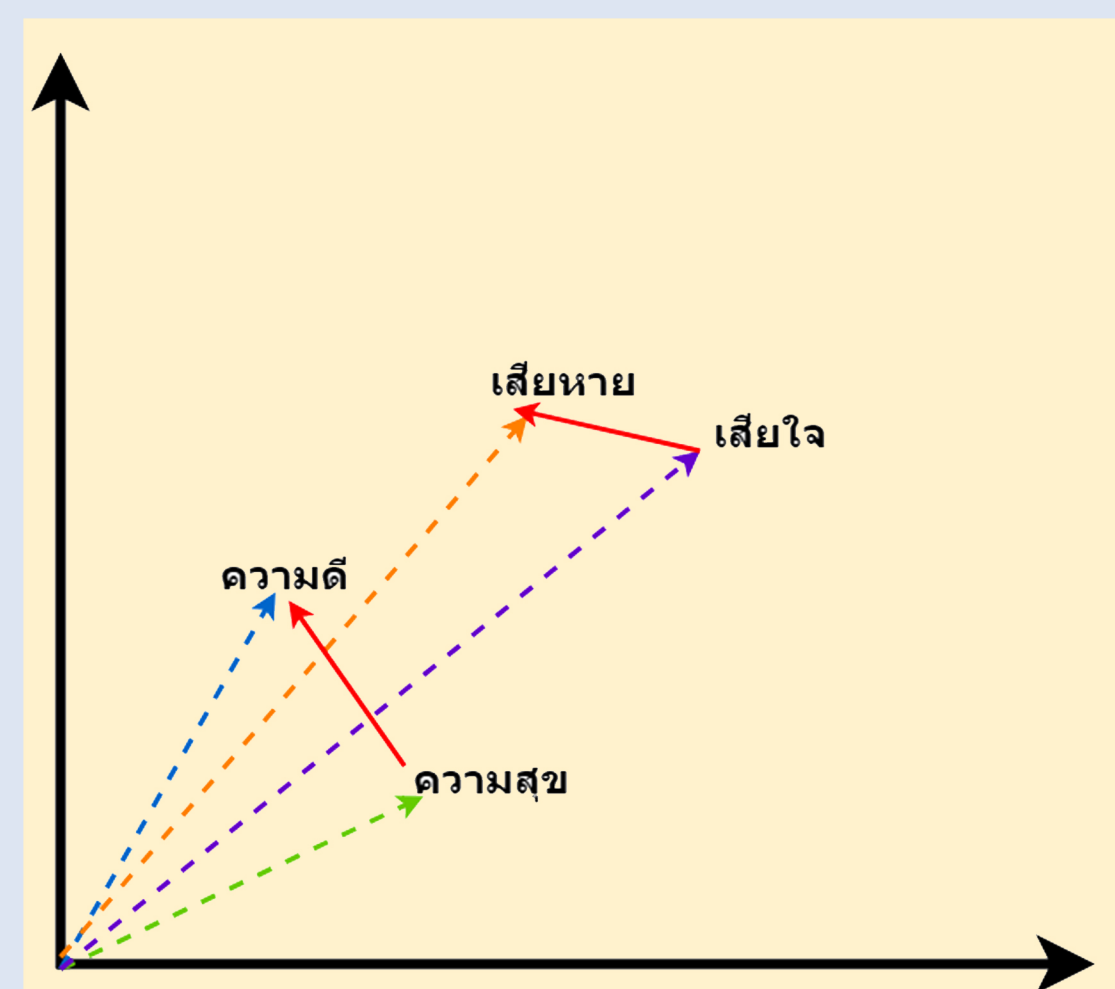
เทคนิคการเรียนรู้เชิงลึก (Deep Learning)^[3]:

- เป็นส่วนหนึ่งของกระบวนการเรียนรู้ของเครื่อง (Machine Learning)
- ใช้เครือข่ายประสาทเทียม (Neural Network) โดยมี Hidden Layers ลึกหลายชั้น



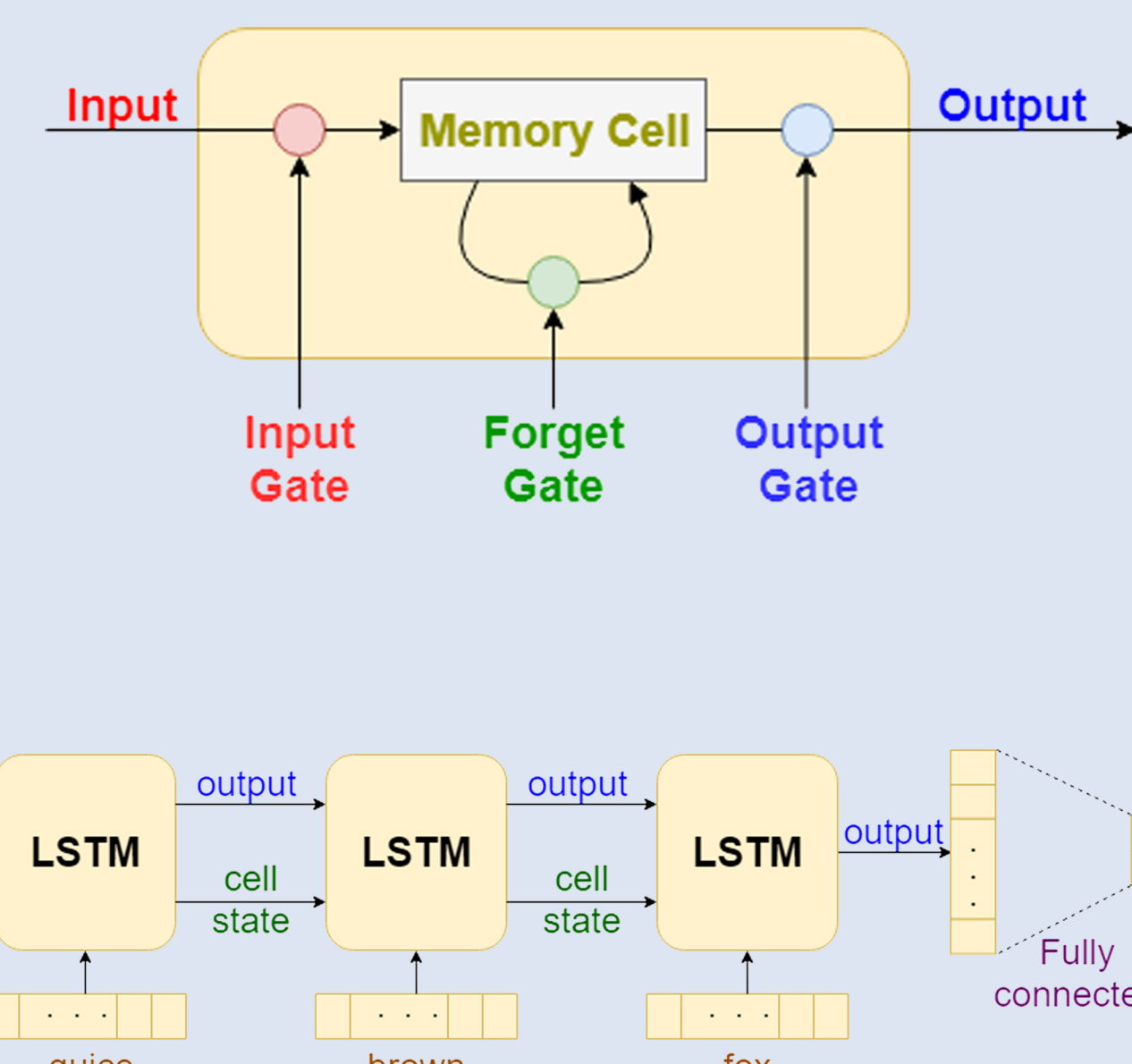
Word2Vec^[2]:

- การแปลง "คำ" ให้อยู่ในรูปของ "vector" โดยใช้วิธีการคำนวณตัวเลขของคำนั้น ๆ จาก context รอบๆ คำนั้น
- สามารถหาความคล้ายของคำต่างๆ จากการคำนวณ scalar product ระหว่าง vectors ของคำต่างๆ
- คำที่มีความหมายคล้ายกัน จะมี Word Vector คล้ายกัน



Long short-term memory (LSTM)^[4]:

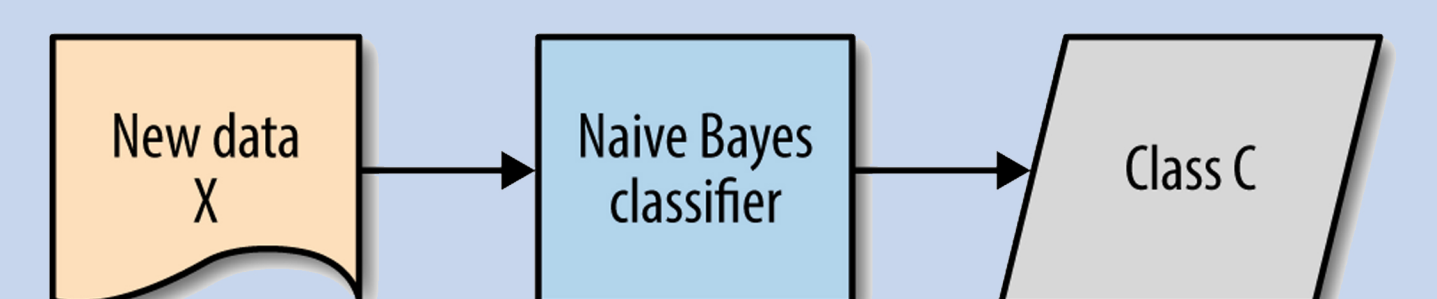
- เป็น Recurrent Neural Network (RNN) ที่เพิ่มหน่วยความจำเข้ามาภายใน Cell ซึ่งเป็นเทคนิคหนึ่งใน Deep Learning
- ประกอบด้วย 3 Gates: Input Gate, Forget Gate และ Output Gate
- แก้ปัญหาของ RNN ที่มีต่อ Sequence ยาวๆ ของข้อมูล



โมเดลที่ใช้เปรียบเทียบ

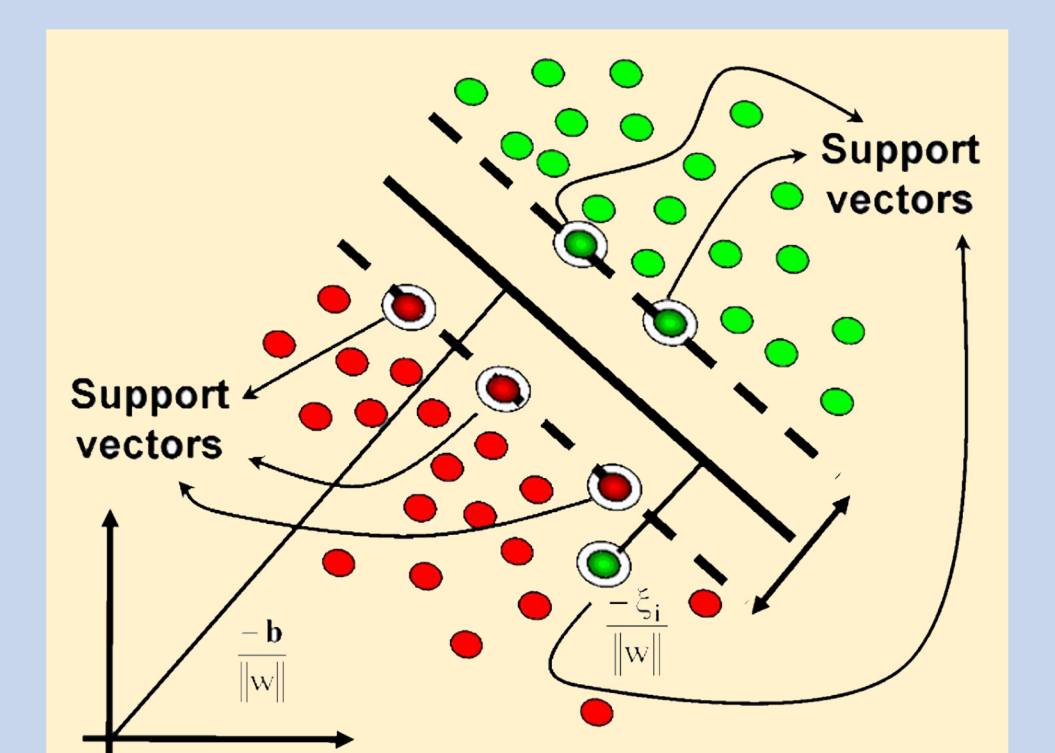
Naive Bayes^[5]:

- เป็นเทคนิคการจำแนกแยกแยะข้อมูล โดยใช้หลักการความน่าจะเป็น Bayes theory เพื่อวิเคราะห์ความรู้สึกของโพสต์



Support Vector Machine (SVM)^[6]:

- เป็นเทคนิคการจำแนกแยกแยะข้อมูลที่ได้รับ ความนิยม ใช้ Support Vectors เป็นตัวกำหนดจำแนกขอบเขตของข้อมูล



ผลการดำเนินงาน

Model	Accuracy
Long Short-Term Memory	63.12%
Support Vector Machine	60.04%
Naïve Bayes	57.64%

ตารางแสดงความแม่นยำของแต่ละ Model

จากตารางเราจะพบว่า LSTM หรือ long short-term memory มีประสิทธิภาพมากกว่า SVM และ Naive Bayes เล็กน้อย ทำให้เห็นว่าการนำ Deep Learning ในการจำแนกความรู้สึก จะช่วยให้ปรับปรุงความแม่นยำของ Model ได้ดียิ่งขึ้น

ข้อความของโพสต์ : " ไม่ทำมาไว้อะไรกับเขาแล้ว ก็บอกเขาว่าตัวเองไม่เป็น "	ผลการทำนาย : " แย่เลย :(" "
ข้อความของโพสต์ : " ก็ดีนะการันตีเนี่ยทำให้เห็นมุมมองของแต่ละพรรคได้ดีมาก "	ผลการทำนาย : " แย่เลย :(" "
ข้อความของโพสต์ : " อย่าลืมไปซื้อลิขัยเลือกตั้ง 24 กันละครับ "	ผลการทำนาย : " แย่เลย :(" "
ข้อความของโพสต์ : " เลือกใครก็ได้ที่ไม่ใช่ลุงจะไม่มีวันประเทศล่มแน่ "	ผลการทำนาย : " แย่เลย :(" "
ข้อความของโพสต์ : " เป็นกำลังใจให้ลุงนะ "	ผลการทำนาย : " แย่เลย :(" "
ข้อความของโพสต์ : " อย่าให้พรรคอนาคตใหม่ชนะเลือกตั้งละ พวกเขาน่าจะพาประเทศเราไปได้อีก "	ผลการทำนาย : " แย่เลย :(" "

จากรูปข้างล่าง ได้มาจากการนำ Model LSTM มาทำการทำนาย โพสต์ที่เราดึงมาจาก Twitter ว่าแต่ละโพสต์อยู่ในแง่ใดบ้าง โดยที่เราจะนำมาแต่เพียง ข้อความที่โพสต์มาทำนายในรูปของ ข้อความ หลังทำนายเราจะได้ผลลัพธ์ออกมา ตามภาพจะมีบอกว่า แต่ละข้อความของโพสต์นั้นเป็นแง่ใด

สรุปผล

จากการทดลองพบว่าโมเดล LSTM สามารถทำงานได้แม่นยำที่สุดเมื่อเทียบกับ SVM และ Naive Bayes ซึ่งแสดงให้เห็นว่า เราสามารถนำเทคนิคการเรียนรู้เชิงลึกแบบ LSTM มาใช้ในการวิเคราะห์ความรู้สึกได้อย่างมีประสิทธิภาพ ขั้นตอนในการทำนั้นอาจมี บางข้อความที่ไม่สามารถทำนายได้แม่นยำ เพราะตัว Model มีความแม่นยำไม่สูงมากนัก และปัญหาก็อย่างคือ คำพูดจากโพสต์ที่ ดึงจาก Twitter ที่เป็นคำพูดกำกวม ไม่ชัดเจน ทำให้การ Training Model ไม่ชัดเจนตาม ส่งผลทำให้การทำนายได้ยากขึ้น แต่โดยรวมแล้ว โมเดล LSTM สามารถนำไปพัฒนาต่อให้มีความแม่นยำมากขึ้น โดยมีโครงงานนี้เป็นกรณีตัวอย่างการศึกษาต่อไป

แหล่งอ้างอิง

- [1]About Twitter's APIs.(2019)[Online]Available : help.twitter.com/en/rules-and-policies/twitter-api
- [2]Word2Vec ด้วย Gensim ใน Python(2560).Word2Vec คืออะไร ?[Online]Available : <https://python3.wannaphong.com/2017/01/word2vec-gensim-python.html>
- [3]What Is Deep Learning?.How Deep Learning Works(2018)[Online]Available : au.mathworks.com/discovery/deep-learning.html..html
- [4]Long Short-Term Memory (LSTM)(2017)[Online]Available : medium.com/@sinart.t/long-short-term-memory-lstm-e6cb23b494c6
- [5]Applying Gaussian Naïve Bayes Classifier in Python.(2017)[Online]Available : https://medium.com/@gp_pulipaka/applying-gaussian-na%C3%AFve-bayes-classifier-in-python-part-one-9f82aa8d9ec4
- [6]อัลกอริทึม Support Vector Machine(2016)[Online]Available : <http://kokzard.blogspot.com/2011/10/jfkdshfkjsldf.html>