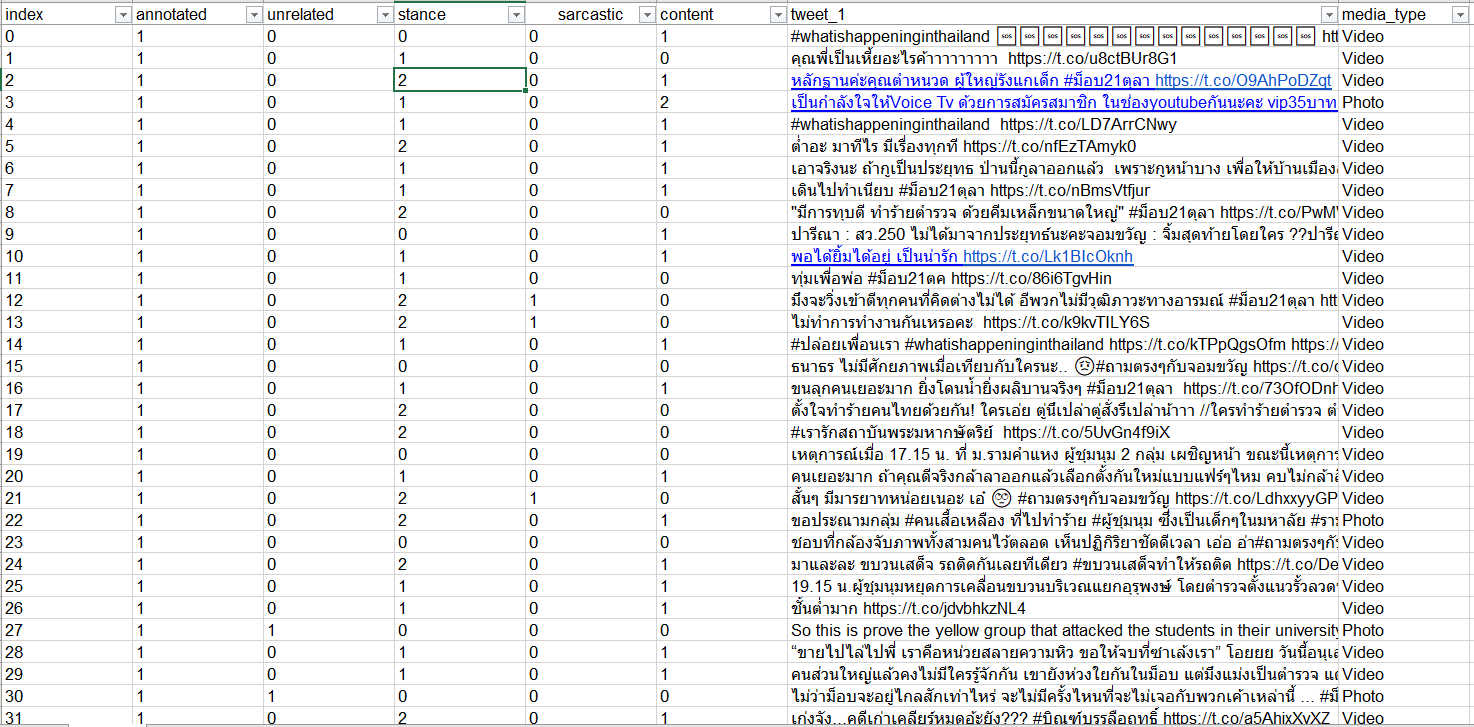
**Auto-Labeling Report**

**08.01.2021**



Data Set ที่ใช้ในการ Auto-Labeling คือ Data ใน Task2 ที่มีการ Label Data แล้วว่าเกี่ยวข้องกับการเมืองหรือไม่ ฯลฯ มีขั้นตอนในการทำดังนี้

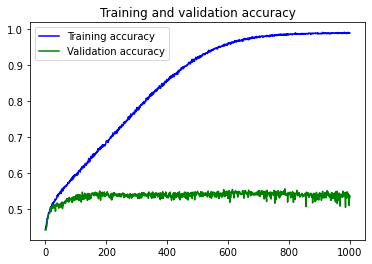
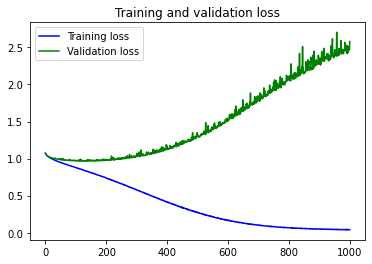
1. แบ่งข้อมูลออกเป็น 2 ส่วน คือ Training Set และ Validation Set เพื่อใช้ในการทดสอบประสิทธิภาพของ Model (ในเคสนี้แบ่งเป็น 80% Training Set : 20% Validation Set)

2. Image Processing: Extract รูปภาพให้เป็นข้อมูลที่สามารถนำมาวิเคราะห์ได้ (X: pixel of images, y: labeled ซึ่งจะลอง stance ก่อนเพื่อเป็น case study)

3. สร้างโมเดลเพื่อทำ Pattern Recognition (Convolutional Neural Network)

4. นำโมเดลมาประเมินผล (Model Evaluation)

ผลที่ได้ (Prototype)

ผลที่ได้คือ โมเดลมีประสิทธิภาพใน Training Set (High Accuracy, Low Loss) แต่ยังไม่มีประสิทธิภาพใน Validation Set (Steadily Accuracy, High Loss) แปลว่าโมเดลยังเกิด Overfitting

สิ่งที่จะทำต่อไป

* แก้ไขปัญหา Overfitting
* เพิ่มประสิทธิภาพของโมเดล (Optimization)
* ทดลองกับ Data จริง เมื่อมีข้อมูลมากพอ