# รายงานความก้าวหน้าวิชา Computer Engineering Project Preparation

**ครั้งที่ 1**

1. ชื่อโครงงาน (อังกฤษ) Classifying Emotion of Thai‐Human Voice by Using Deep Learning

## การดำเนินงานมีความก้าวหน้า 24 %

## ความก้าวหน้าระหว่างวันที่ 01 ก.พ. 65 ถึงวันที่ 04 มี.ค. 65

## รายละเอียดความก้าวหน้า

* ทำการนัดประชุมกับอาจารย์ที่ปรึกษาเพื่อหาแนวทางการดำเนินการทำโปรเจค
* ทำการหาข้อมูลเรื่องคลื่นเสียงว่า เสียงคืออะไรสามารถแสดงเป็นตัวเลข (numeric) ได้อย่างไร

มีวิธีการ preprocess data และ ทำ Features extraction ที่เป็นเสียงได้อย่างไร

* ทำการหาข้อมูลว่าโปรเจคในลักษณคล้ายกันก่อนหน้า มีวิธีการทำอย่างไร ในการแก้ปัญหา

ใช้ model อะไร มี accuracy ในการ Classification เท่าไหร่ ใช้ matrix อื่นๆไหมในการ evaluation model

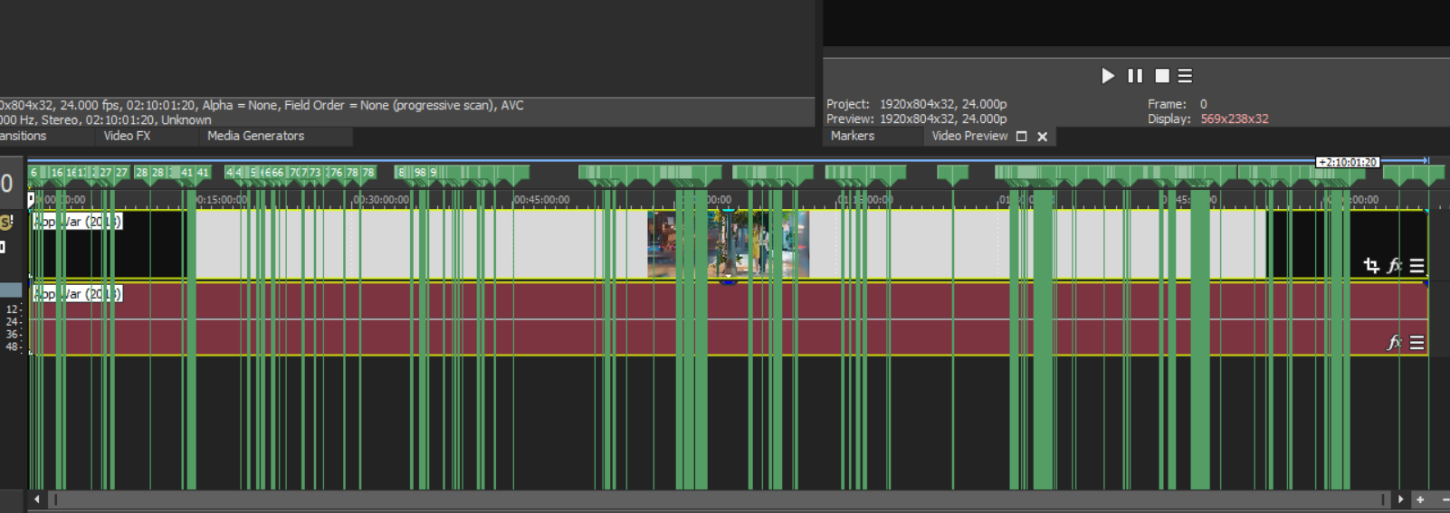
ทำการหาว่าการทำเป็น speech to text ก่อน แล้วค่อยนำผลลัพธ์จาก speech to text มาประกอบ Emotional Classification และการใช้หลายภาษาในเป็น Input ให้ผลอย่างไร ดีกว่าการใช้ input ภาษาเดียวหรือไม่

* ทำการทดลองทำการ data sourcing ว่าได้ผลตามคาดหวังหรือไม่
* ในการหาข้อมูลหนังและข้อมูลของหนัง เราจำเป็นต้องหาหนังที่มี เสียง 5.1 เพื่อให้เรามาแยกเอฟเฟคเสียงต่างๆในโปรแกรมเพื่อที่จะได้แค่เสียงพูดเปล่าๆเท่านั้น และ เมื่อแยกเสียงเสร็จแล้ว ก็ต้องไปตัดส่วนที่ไม่มีเสียงออกเพื่อให้ได้นำไป train ต่อได้

รูปภาพประกอบด้วย โต๊ะ

คำอธิบายที่สร้างโดยอัตโนมัติ

รูปที่ 1.1 List รายชื่อหนัง



รูปที่ 1.2 รูปการตัดต่อวิดีโอ

รูปภาพประกอบด้วย โต๊ะ

คำอธิบายที่สร้างโดยอัตโนมัติ

รูปที่ 1.3 Gantt Chart

## ปัญหาที่เกิดขึ้นและแนวทางการแก้ไข

**Problem No. 1 ไม่มีตัวอย่างเสียง (Input)ที่แสดงอารมณ์แท้จริงได้**

**พบปัญหาในการรายงานความก้าวหน้า ครั้งที่ 1**

**สถานะ**   กำลังดำเนินการ  แก้ไขสำเร็จ

**รายละเอียดปัญหา**

การ Train model มีหลักการคือการนำ data ที่เป็น input เข้าไปใน Train model แต่ถ้า input data ที่เข้าไปนั้นไม่มีความสมบูรณ์ หรือ มี Bias ก็อาจจะทำให้ Model ทำงานได้ไม่ดี แม้เราจะทำการ Parameter Tuning ดีแค่ไหนก็ตาม ( Garbage in Garbage out )

ปัญหาที่เราพบคือ เราไม่สามารถเก็บตัวอย่างเสียงที่แสดงอารมณ์แบบแท้จริงได้ ยกตัวอย่างเช่น เราต้องการเสียงของคนที่กำลังโกรธ ซึ่งแน่นอนเราไม่สามารถขอบันทึกเสียงบุคคลนั้น ขณะเขากำลังโกรธได้ และ ถ้าเราจะขอให้อาสาสมัครแกล้งทำเป็นโกรธก็อาจจะทำให้เกิดการ Bias ได้

**แนวทางแก้ไข/การแก้ไข**

เราจะนำเสียงมาจากภาพยนต์ที่เกิดจากนักแสดงมืออาชีพวิธีการนี้ ก็ยังมี Bias แต่คาดว่าน้อยกว่า การที่เราขอให้อาสาสมัครแสดงอารมณ์ และ เรายังสามารถหาตัวอย่างได้มากกว่าในระยะเวลาเท่ากัน ซึ่งจะทำให้ Model มีประสิทธิภาพในการ Classification มากกว่า ขณะนี้กำลังทดลองว่าตัวอย่างนั้นใช้ได้หรือไม่ ต้องนำตัวอย่างมา Train model แล้วใส่ input เข้าไป Test ก่อน

**Problem No. 2 การแยก**เสียงพูดและเอฟเฟคออกจากกัน

**พบปัญหาในการรายงานความก้าวหน้า ครั้งที่ 1**

**สถานะ**   กำลังดำเนินการ  แก้ไขสำเร็จ

**รายละเอียดปัญหา**

ไม่สามารถหาข้อมูลหนังที่มีเสียง 5.1 ได้ยาก กับ วิธีการแยกไฟล์เสียงพูดและเอฟเฟคออกจากกัน

**แนวทางแก้ไข/การแก้ไข**

**เราจะนำเข้าโปรแกรมแยกเสียงกับ**เอฟเฟคออกก่อน หลัง จากนั้นทำการตัดส่วนที่ไม่ต้องการออก จะได้ไฟล์วิดีโอที่มีภาพและเสียงพูดเท่านั้น

## สิ่งที่จะดำเนินการต่อไป

เราจะฝึก Technical Skills ที่ต้องใช้ในการทำ Project ต่อไปเป็นระยะเวลา 3 weeks ได้แก่

* Machine Learning | Deep Learning
* Data preparation สำหรับเสียงเพิ่มเติม
* Statistic สำหรับเสียง
* Django and Deploying Model to Application
* Flutter
* Framework for application