รายงาน

Chatbot ให้ความรู้เกี่ยวกับกระท่อม

จัดทำโดย

นาย ปิ่นปวัฒน์ ลิ้มสุวัฒน์ 6510450658

นาย พจนินท์ วชิรวิทยากุล 6510450691

นาย ศุภณัฐ สร้อยเพชร 6510450976

เสนอ

ผู้ช่วยศาสตราจารย์ ดร.อรวรรณ อิ่มสมบัติ

มหาวิทยาลัยเกษตรศาสตร์

01418363-65 Natural Language Processing

**ที่มาและความสำคัญ**

การเรียนรู้และแบ่งปันความรู้ที่สนใจในการเรียนรู้เกี่ยวกับพืชกระท่อมการสร้าง *Chatbot* ที่เชื่อมโยงไปยังข้อมูลที่เกี่ยวข้องเป็นวิธีที่ดีในการให้ข้อมูลและคำแนะนำในการประมวลผลข้อมูลสำหรับผู้ใช้ที่สนใจเพื่อประโยชน์ในสังคม การเสริมสร้างความเข้าใจเกี่ยวกับพืชกระท่อมมีประโยชน์ในด้านการเกษตรการยาและสรรพคุณทางยา ดังนั้น *Chatbot* สามารถเสริมสร้างความเข้าใจในสังคมเกี่ยวกับคุณค่าของพืช

**วัตถุประสงค์**

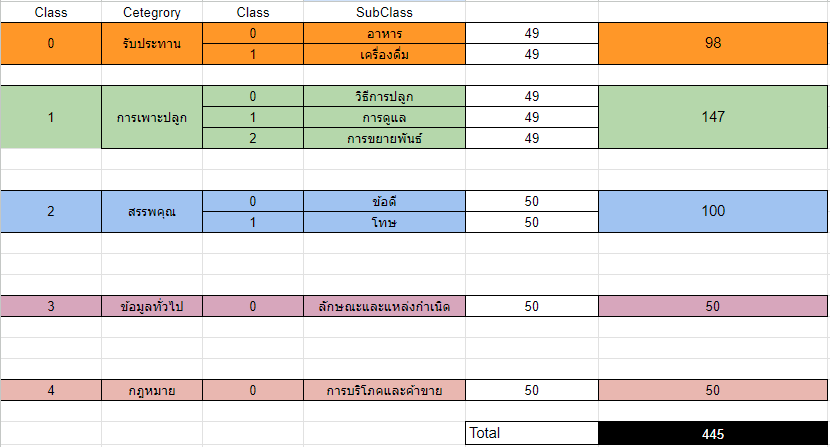
* *Chatbot* สามารถให้ข้อมูลเกี่ยวกับวิธีการปลูก การดูแลรักษาและปัญหาที่อาจเกิดขึ้นในเพาะปลูกพืชกระท่อม
* การกระจายข้อมูลการวิจัยเกี่ยวกับพืชกระท่อมที่สามารถนำเสนอผ่าน *Chatbot* เพื่อเผยแพร่ข้อมูลและความรู้ใหม่ๆที่เป็นประโยชน์ต่อสาธารณชน
* *Chatbot* สามารถส่งเสริมให้คำแนะนำเกี่ยวกับการใช้พืชกระท่อมในการทำยา เครื่องปรุง หรือการใช้งานในอุตสาหกรรม
* การให้คำแนะนำเกี่ยวกับการต่อต้านศัตรูพืช และการจัดการปัญหาที่เปิดขึ้นในการปลูกกระท่อม

**QR CODE**



รูป : QR CODE ของ dataset สำหรับการทำ *Chatbot* กระท่อม

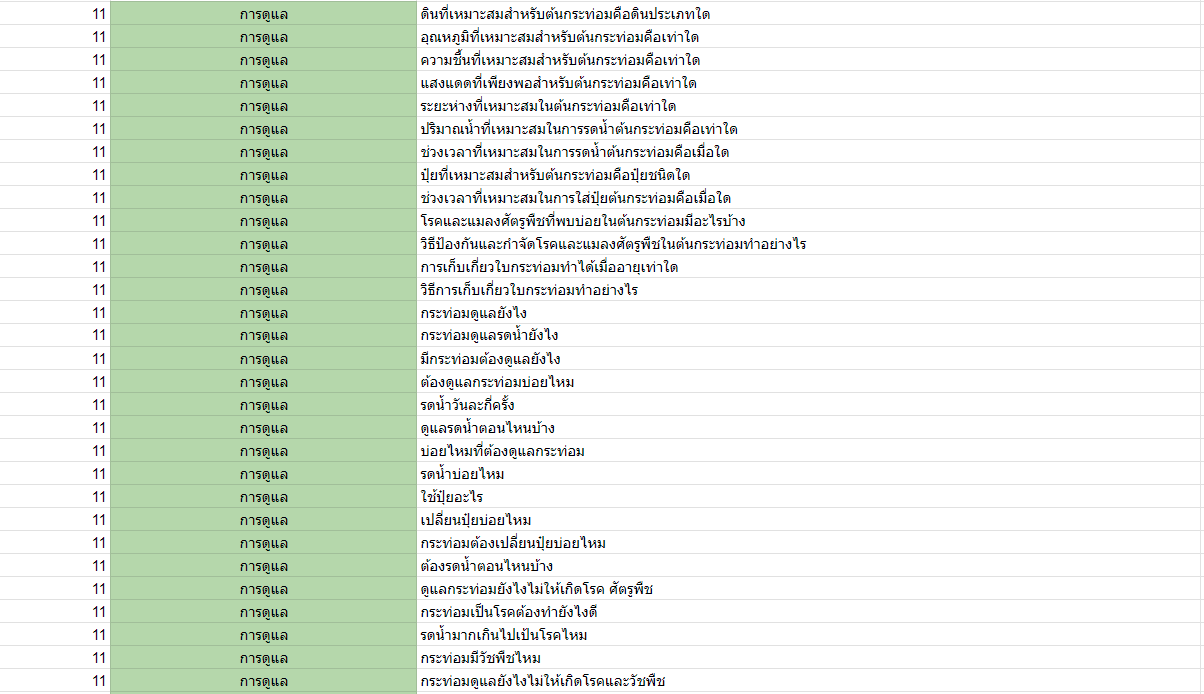
**ภาพตัวอย่าง dataset**



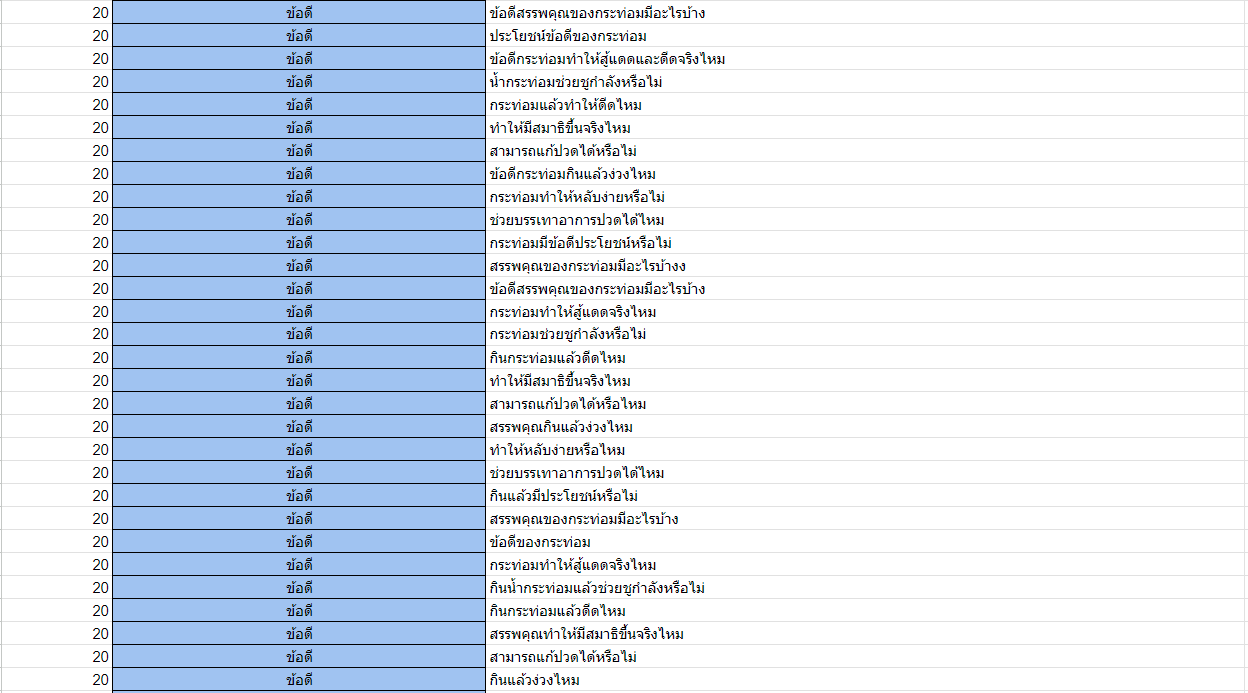
รูป : ภาพรวม ของ dataset สำหรับการทำ *Chatbot* กระท่อม

 Figure 1.1: ตัวอย่างคำถามที่ใช้ในการฝึกแชทบอทเกี่ยวกับอาหารและเครื่องดื่ม

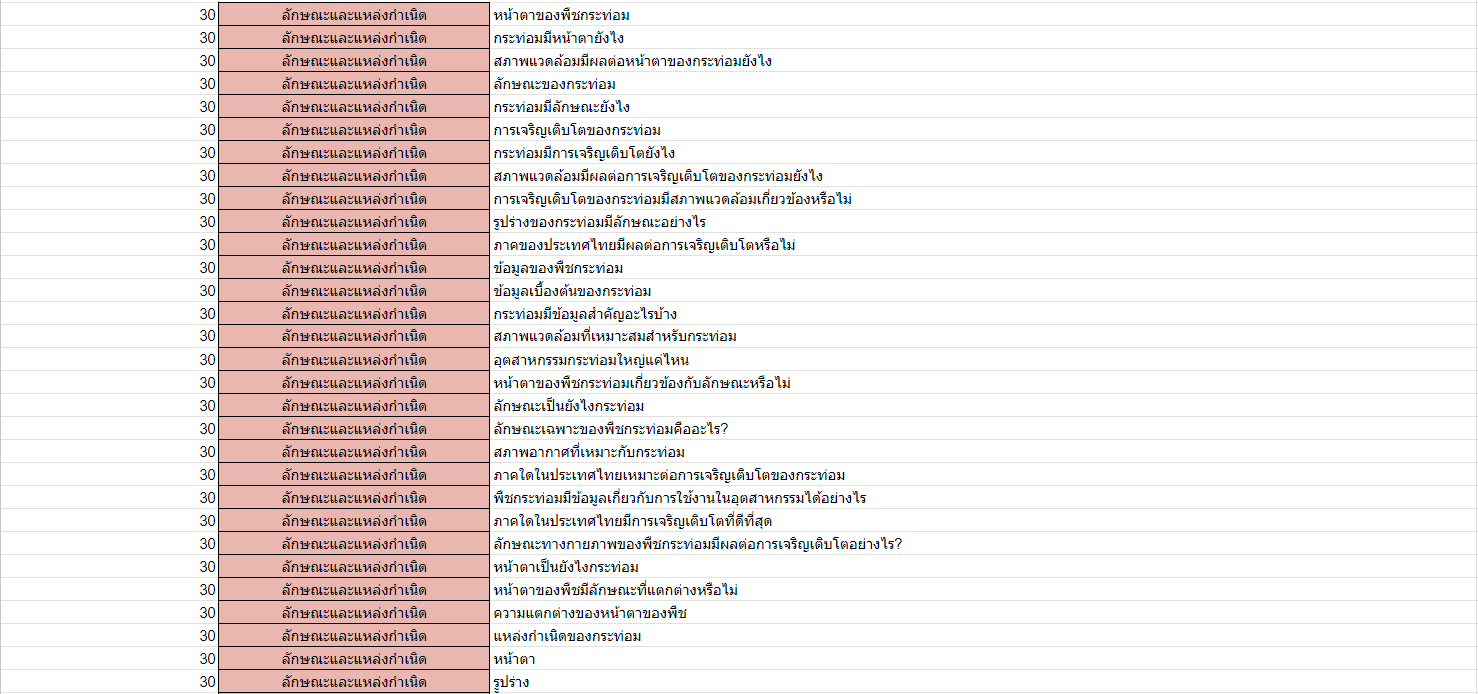

รูป : ตัวอย่าง dataset Class อาหารและเครื่องดื่ม



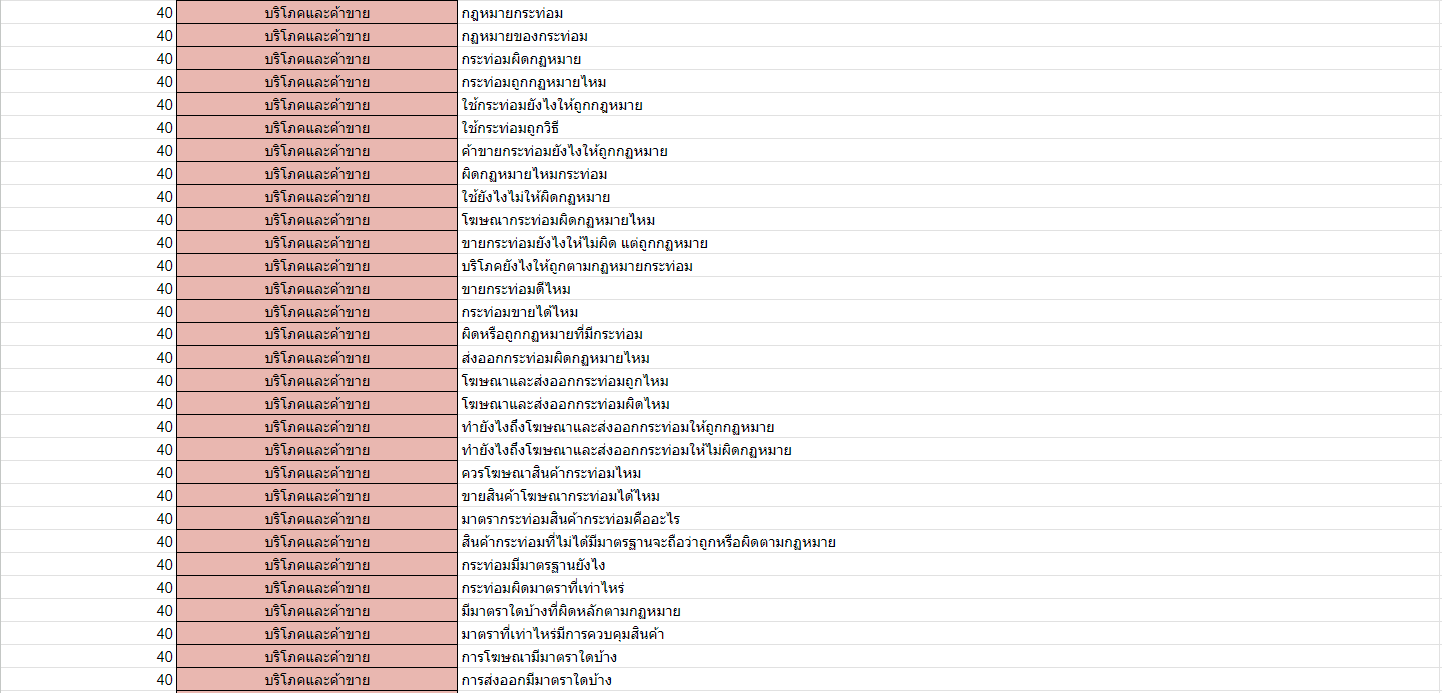
รูป : ตัวอย่าง dataset Class เกี่ยวกับการเพาะปลูก



รูป : ตัวอย่าง dataset Class สรรพคุณของกระท่อม



รูป : ตัวอย่าง dataset Class ข้อมูลทั่วไปเกี่ยวกับกระท่อม



รูป : ตัวอย่าง dataset Class การบริโภคและค้าขาย

**กระบวนการ pre-processing ต่างๆ ที่ทำ**

รูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, ตัวอักษร

คำอธิบายที่สร้างโดยอัตโนมัติ

1.Dowload Library ที่ต้องใช้

2. import Library ที่ต้องใช้   
เช่น word\_tokenize = ใช้เพื่อตัดคำภาษาไทย , thai\_stopwords = คำที่ไม่จำเป็นในภาษาไทย,

dict\_trie = Dictionary ภาษาไทย, correct = ตรวจสอบการสะกดภาษาไทย

thai\_words = คำศัพท์ในภาษาไทย

รูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, ตัวอักษร, แสดง

คำอธิบายที่สร้างโดยอัตโนมัติ

รูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, ซอฟต์แวร์, ซอฟต์แวร์มัลติมีเดีย

คำอธิบายที่สร้างโดยอัตโนมัติรูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, ตัวอักษร

คำอธิบายที่สร้างโดยอัตโนมัติ

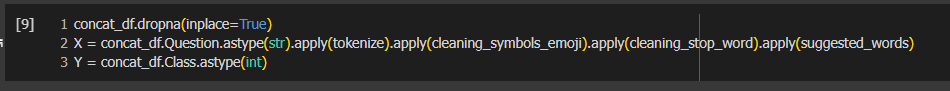
3. Download dataset ของแต่ละคลาส

4. นำข้อมูลในแต่ละคลาสมาทำเป็น DataFrame

รูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, ซอฟต์แวร์, ระบบปฏิบัติการ

คำอธิบายที่สร้างโดยอัตโนมัติ

5. สร้างฟังก์ชั่นต่างๆที่ต้องใช้ เช่น ฟังก์ชั่นแนะนำคำศัพท์ , ฟังก์ชั่นตัดคำภาษาไทย, ฟังก์ชั่นลบคำที่ไม่จำเป็น, ฟังก์ชั่นลบอิโมจิ



รูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, ซอฟต์แวร์

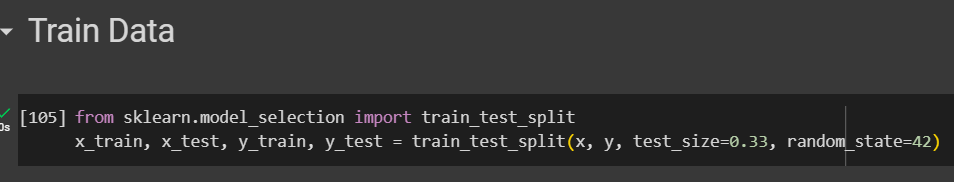
คำอธิบายที่สร้างโดยอัตโนมัติ

6. นำข้อมูลคำถามในแต่ละคลาสมาทำความสะอาดด้วยฟังก์ชั่นข้างต้น ส่วนคลาสข้อมูลคำตอบนำมาแปลงให้อยู่ในรูปแบบจำนวนเต็ม

7. นำข้อมูลที่ทำความสะอาดแล้วมาทำเป็น DataFrame และแปลงเป็นไฟล์ csv และนำคำภาษาไทยที่ถูกตัดแล้วมาทำเป็น word index

8. นำ word index มาผ่านกระบวนการ text to sequence โดย words index จะถูกแปลงเป็นเลขจำนวนเต็ม

**Training data**

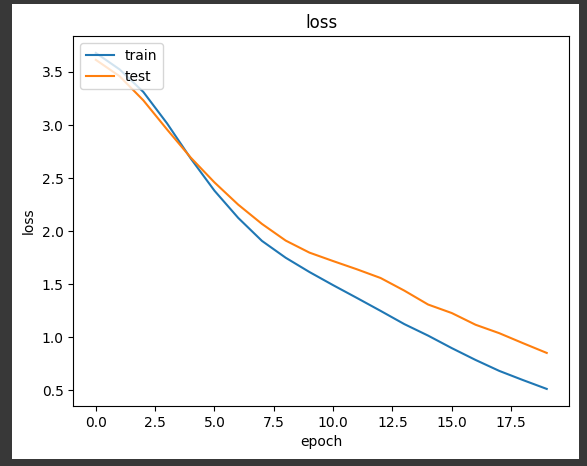
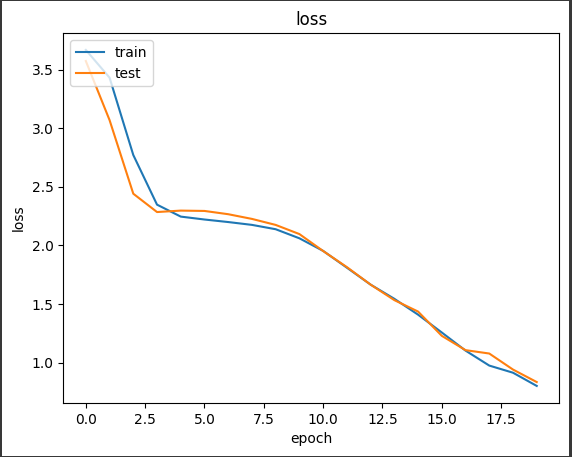


รูป : Training data

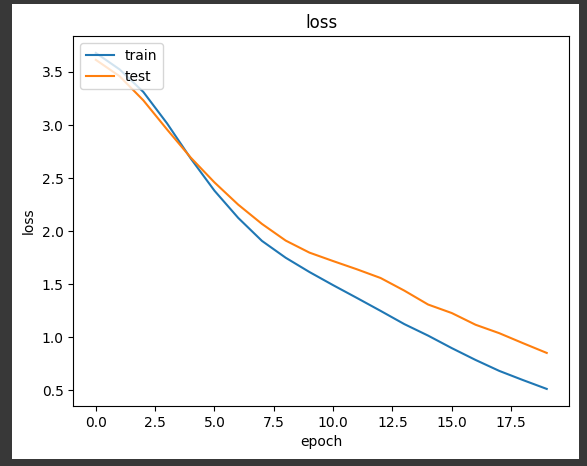
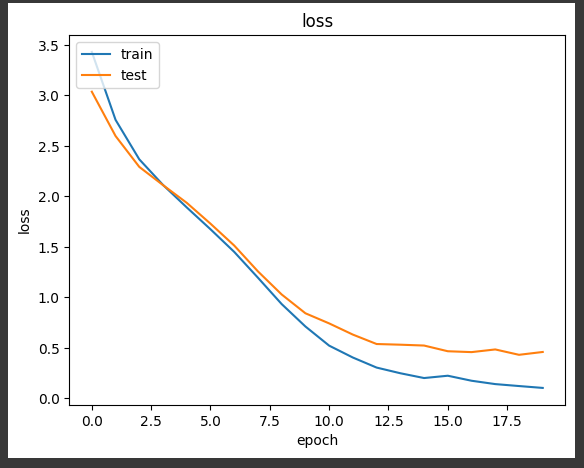
**ทดลองด้วย Model ต่างๆ**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model/Result** | **Accuracy** | **Loss** | **Val\_Accuracy** | **Val\_Loss** |
| **Fully-Connected** | **0.9698** | **0.4121** | **0.8299** | **0.7724** |
| **LSTM** | **0.7651** | **0.8004** | **0.7687** | **0.8333** |
| **Bi-LSTM** | **0.9966** | **0.0709** | **0.9456** | **0.2291** |
| **RNN** | **0.9933** | **0.0766** | **0.8367** | **0.5163** |

**กราฟ : ความสัมพันธ์ Loss กับ Epoch**

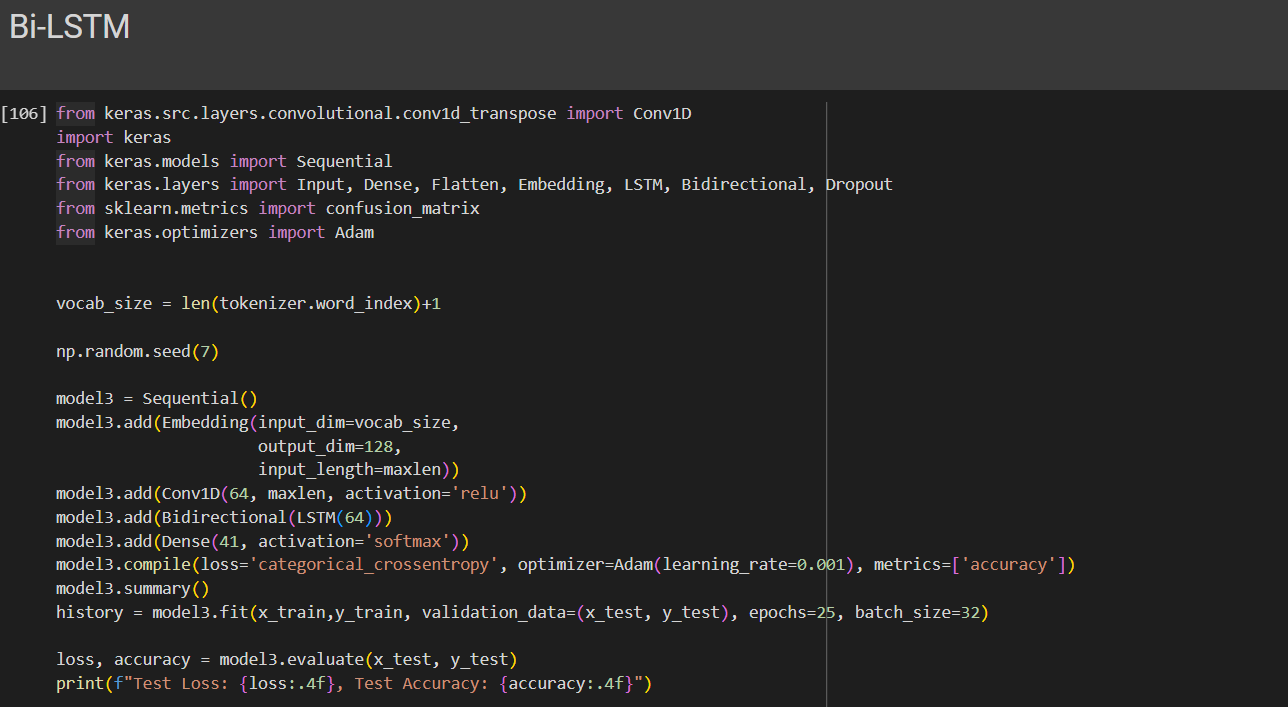
****

**Bi-LSTM LSTM**

****

**Fully-Connected RNN**

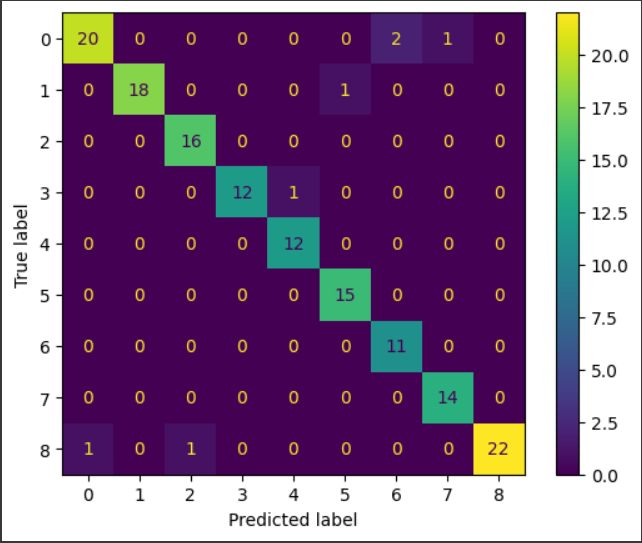
**Bi-LSTM**



**ชั้น Layer**

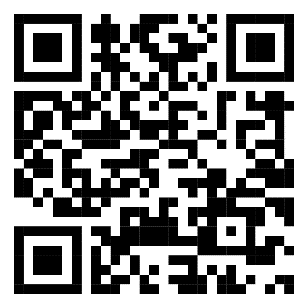
* Embredding : output\_dim = 128
* Conv1D : 64 node, activation = relu
* Bidirectional LSTM : 64 node
* Dense : 41 node, activation = softmax

**Confusion Matrix**



|  |  |  |
| --- | --- | --- |
| Class | Predict Class | Class Name |
| 00 | 0 | อาหาร |
| 01 | 1 | เครื่องดื่ม |
| 10 | 2 | วิธีการปลูก |
| 11 | 3 | การดูแล |
| 12 | 4 | การขยายพันธุ์ |
| 20 | 5 | ข้อดี |
| 21 | 6 | โทษ |
| 30 | 7 | ลักษณะและแหล่งกำเนิด |
| 40 | 8 | การบริโภคและค้าขาย |

\*\*\*ถ้าค่า predict มี prob น้อยกว่า 0.5 จะให้เข้าเงื่อนไขไม่เข้าข่ายทันที\*\*\*

**Colab’s link** <https://colab.research.google.com/drive/1Cx6pmsH2ERRYsgnWMRtWiIflMPg2VMGP#scrollTo=03_GCsd4pSLR>

Chatbot QR-Code

