ที่มาและความสำคัญ

ในปัจจุบัน มีผู้ที่สนใจในการลงทุนมากยิ่งขึ้น โดยเฉพาะวัยในช่วงมหาลัยไปจนถึงวัยทำงาน และตัวผมเอง เนื่องจากตลาดforex เป็นพื้นที่ที่เข้าถึงง่าย และ สามารถสร้างกำไรได้อย่างรวดเร็ว แต่จะตามมาด้วยความเสี่ยงที่สูง จึงทำให้เกิด Ai ในท้องตลาด แต่Ai ตัวนี้ผมจะขอเรียกว่า EA ในส่วนของ EA จะสามารถเทรดให้เราได้เองโดยอัตโนมัตซึ่งเราก็ต้องไปเซ๊ทค่าต่างๆใน EA แต่ก็มีข้อเสียที่มีความเสี่ยงอยู่ เนื่องจาก EA จะเทรดได้แค่ตามที่เราเซ๊ทค่า หรือตามทฤษกราฟเท่านั้น แต่ปัจจัยที่ทำให้บ่งบอกถึงขาขึ้นหรือขา

ลงของตลาดForex ก็จะมีข่าวสารข้อมูลต่างๆ จากกระทรวงต่างๆอีกด้วย เพื่อให้การเทรดอัตรโนมัต แม่นยำยิ่งขึ้นทางผู้จัดทำโครงงานจึงเสนอไอเดีย Ai วิเคราะห์ข่าวสาร

AI วิเคราะห์ข่าวหุ้นได้มีการนำ Machine Learning NLP มาวิเคราะห์ ซึ่งเป็นตัวช่วยทำให้คอมพิวเตอร์วิเคราะห์ข่าวสารหรือข้อความได้อย่างง่ายดาย NLP เป็นสาขาหนึ่งในการเรียนรู้ของเครื่องด้วยความสามารถของคอมพิวเตอร์ในการทำความเข้าใจ วิเคราะห์ จัดการ และสร้างภาษามนุษย์ได้

การสรุปข้อความเป็นกระบวนการสร้างข้อความที่สั้นลงโดยไม่ต้องลบโครงสร้างความหมายของข้อความ

**การแยกข้อความ**สิ่งนี้เกี่ยวข้องกับการสรุปข้อความโดยอัตโนมัติและค้นหาข้อมูลที่สำคัญ

ตัวอย่างเช่น เครื่องมืออาจดึงคำที่ใช้บ่อยที่สุดในข้อความออกมา อีกตัวอย่างหนึ่งคือชื่อการรู้จำเอนทิตี ซึ่งแยกชื่อบุคคล สถานที่ และเอนทิตีอื่นๆ ออกจากข้อความ

**การจำแนกข้อความ**สิ่งนี้เกี่ยวข้องกับการกำหนดแท็กให้กับข้อความเพื่อจัดหมวดหมู่ สิ่งนี้มีประโยชน์สำหรับการวิเคราะห์ความรู้สึกซึ่งช่วยคาดเดาสิ่งที่ผู้พูดหรือผู้เขียนอาจทำโดยอิงจากข้อความที่พวกเขากำลังสร้าง

**เครื่องแปลภาษา.**นี่เป็นกระบวนการที่คอมพิวเตอร์แปลข้อความจากภาษาหนึ่ง เช่น ภาษาอังกฤษ เป็นอีกภาษาหนึ่ง

Sentences \_Tokenizing – การแยกประโยคและคำออกจากเนื้อความของข้อความ

Words\_ Tokenizing เป็นการแยกคำออกมาจากประโยค

* + - 1. Stop words with NLTK

เป็นการใช้ คำหยุด โดยอ้างอิงค์จากคลังข้อมูลของภาษาอังกฤษ คำหยุด คือ คำที่เป็นคำวลีหรือเป็นคำที่เสริมทำให้ประโยคดีขึ้น แต่ในเชิงคอมพิวเตอร์ คำหยุด คือคำที่ไม่มีความหมาย เหตุผลที่ต้องลบคำหยุดเนื่องจากประโยคที่ยาวขึ้นจะกินทรัพยากรและเวลาในการประมวลผล

* + - 1. Stemming words with NLTK

Stemming หรือที่เรียกว่า suffix stripping เป็นเทคนิคที่ใช้ในการลดขนาดข้อความ **ต้นกำเนิดยังเป็นประเภทของข้อความที่ทำให้เป็นมาตรฐานที่ช่วยให้สร้างมาตรฐานของคำบางคำให้เป็นนิพจน์เฉพาะ ข้อเสีย** ในไพพ์ไลน์การทำเหมืองข้อความหลายๆ แบบ มีตัวเลือกมากมายที่เกี่ยวข้องซึ่งอาจทำให้ข้อมูลสูญหายได้

* + - 1. Part of Speech Tagging with NLTK

1. **POS Tagging** (ส่วนหนึ่งของการแท็กคำพูด) เป็นกระบวนการในการทำเครื่องหมายคำในรูปแบบข้อความสำหรับส่วนใดส่วนหนึ่งของคำพูดตามคำจำกัดความและบริบท มีหน้าที่รับผิดชอบในการอ่านข้อความในภาษาและกำหนดโทเค็นเฉพาะ (Parts of Speech) ให้กับแต่ละคำ เรียกอีกอย่างว่าการติดแท็กทางไวยากรณ์
   * + 1. **Lemmatizing with NLTK**
2. Lemmatization เป็นกระบวนการของการจัดกลุ่มคำในรูปแบบผันแปรต่างๆ เพื่อให้สามารถวิเคราะห์เป็นรายการเดียวได้ Lemmatization คล้ายกับการกำเนิด แต่นำบริบทมาสู่คำ ดังนั้นจึงเชื่อมโยงคำที่มีความหมายคล้ายกันเป็นคำเดียว
   * + 1. Wordnet with NLTK
3. [**WordNet**](https://wordnet.princeton.edu/)เป็นฐานข้อมูลคำศัพท์สำหรับภาษาอังกฤษ ซึ่งสร้างโดย Princeton และเป็นส่วนหนึ่งของคลังข้อมูล NLTK

ในสคริปต์ นำเข้าไลบรารีที่สำคัญที่จำเป็นสำหรับการดึงข้อมูลจากเว็บก่อน

ใช้BeautifulSoupอ็อบเจ็กต์และส่งผ่านออบเจ็กต์ข้อมูลที่คัดลอกมา เช่นarticleและlxmlตัวแยกวิเคราะห์

article\_textg เป็นการลบวงเล๊บออก แต่จะไม่ลบพวกตัวเลขและอักษรพิเศษนะครับ

ในการหาความถี่ของการเกิดขึ้นของแต่ละคำ ใช้formatted\_article\_textตัวแปร ขั้นตอนแรกจะเป็นการลลบคำหยุดโดยการใช้ Stopword จากนั้นเราจะตรวจสอบคำที่เหลือ ว่ามีคำหยุดอีกไหม โดยใช้ word\_frequency ถ้าไม่มีคำหยุดแล้วจะสร้างคำใหม่ และจะเก็บว่าคำแต่ละคำเจอกี่ครั้ง และแปลงเป็นความถี่ถ่วงน้ำหนักจะเก็บที่ sentence\_scores

เพื่อสรุปบทความสามารถใช้ประโยค N อันดับแรกที่มีคะแนนสูงสุด สคริปต์ต่อไปนี้ดึงประโยค 7 อันดับแรก

* + - 1. **Sentiment Analasis**

อาต์พุตของ TextBlob สำหรับงานวิเคราะห์การวิเคราะห์จะภายในช่วง[-1.0, 1.0]ที่-1.0เป็นขั้วลบและ1.0เป็นบวก คะแนนนี้ยังสามารถเท่ากับ0ซึ่งหมายถึงการประเมินที่เป็นกลางของคำสั่ง