

Progress Note - [ML Project Name]

[AUTHOR]

Week 1 – Problem Statement

ในสัปดาห์ที่ 1 พยายามหาปัญหาที่สมควรค่าแก่การแก้ด้วย machine learning; ปัญหานี้ในอนาคตจะไม่จำเป็นต้องมีมูลค่าทางธุรกิจหรืองานวิจัยก็ได้ แต่ต้องเป็นปัญหาที่สำคัญสำหรับตัวเราเอง; หาเหตุผลว่าทำไมเราจึงจำเป็นต้องแก้ปัญหานี้ด้วย machine learning

ทำอะไรไปบ้างในสัปดาห์ที่แล้ว

ปัญหาที่พบในสัปดาห์ที่แล้ว

ปัญหาควรมีความชัดเจนว่าต้องการความช่วยเหลืออย่างไร

ตัวอย่าง:

- ไม่เข้าใจบทเรียน 4. Lists and For Loops ในปรับพื้นฐาน #1

วางแผนจะทำอะไรในสัปดาห์นี้

สิ่งที่วางแผนจะทำควรมีผลลัพธ์เป็นรูปธรรมและสามารถทำได้ใน 1 สัปดาห์

ตัวอย่าง:

- คิดไอเดียทำโครงการเกี่ยวกับการจำแนกรูปภาพผลไม้ตามความสุกเพื่อการเก็บเกี่ยว; เขียนรวบรวมเป็นข้อเสนอโครงการ

Week 2 – Metrics and Baselines

ในสัปดาห์ที่ 2 เมื่อเราได้ปัญหาที่ต้องการแก้แล้ว ตรวจสอบว่าปัจจุบันปัญหานี้ถูกแก้ (หรือไม่ถูกแก้) อย่างไรบ้าง สิ่งนี้คือ baseline ของเรา หาตัวเลขที่ใช้บ่งบอกประสิทธิภาพของการแก้ไขปัญหา เราจะใช้ตัวเลขนี้ในการเปรียบเทียบว่าโมเดล machine learning ของเราแก้ปัญหาได้ดีกว่าทางแก้ปัจจุบันอย่างไร

ทำอะไรไปบ้างในสัปดาห์ที่แล้ว

เรียนเรื่อง AI Machine Learning and Deep Learning

ปัญหาที่พบในสัปดาห์ที่แล้ว

การทำความเข้าใจคำถามและตอบคำถามในแบบทดสอบ

วางแผนจะทำอะไรในสัปดาห์นี้

วางแผนว่าจะทำ project อะไร

Week 3 - Data Collection

สัปดาห์แห่งการเริ่มหาแหล่งข้อมูลสำหรับเทรนโมเดลและวัดผล

เริ่มจากการหาข้อมูลจากแหล่งต่างๆมาให้มากที่สุดเท่าที่จะทำได้เพื่อทำการเปรียบเทียบทั้งเชิงปริมาณและคุณภาพ เพื่อเลือกใช้ในขั้นตอนต่อไป

ทำอะไรไปบ้างในสัปดาห์ที่แล้ว

ได้คิดว่าจะทำ **project** อะไรแล้วตั้งเป็น **AimBot**

ปัญหาที่พบในสัปดาห์ที่แล้ว

คิดหัวข้อ**project** ไม่ออก

วางแผนจะทำอะไรในสัปดาห์นี้

คิดหัวข้อใหม่ และหา **dataset** ของปลาน้ำจืดในประเทศไทย

Week 4 – Data Cleaning

โครงการส่วนใหญ่จะเริ่มอยู่ในขั้นตอนของการทำความสะอาดข้อมูลเพื่อทำโมเดล

ไม่ต้องกังวลหากคุณใช้เวลาทำความสะอาดข้อมูลมากกว่า 2-3 สัปดาห์ ขั้นตอนนี้เป็นขั้นตอนที่ใช้เวลามากที่สุดในการทำ **machine learning**;

คุณอาจจะเทรนโมเดลไปแล้วพบว่าจริงๆแล้วข้อมูลคุณยังไม่สะอาดพอจนต้องกลับมาทำความสะอาดใหม่ก็ได้

ทำอะไรไปบ้างในสัปดาห์ที่แล้ว

/แยกกลุ่มออกมาแล้วมาทำโปรเจกเกมแทน

ปัญหาที่พบในสัปดาห์ที่แล้ว

/ต้องมาหาข้อมูลใหม่

วางแผนจะทำอะไรในสัปดาห์นี้

/หาdata setให้ได้

Week 5 – Exploratory Data Analysis

ก่อนที่จะเริ่มทำโมเดล ระหว่างที่ทำความสะอาดข้อมูล ลองพลอตกราฟจากข้อมูลออกมาเพื่อทำความเข้าใจก่อน เช่น

หากคุณทำข้อมูลรูปภาพ ลองดูว่ารูปภาพของคุณมีกี่รูป ขนาดเท่ากันไหม ถ้ารูปมุมตรงกันหรือเปล่า,

หากคุณทำข้อมูลข้อความ ลองนับคำดูว่าข้อมูลของคุณมีคำหรือวลีไหนบ่อยที่สุด แต่ละ **label**

มีคำหรือวลีไหนเด่นเป็นพิเศษ, หากคุณทำข้อมูลตาราง ลองดูว่าตัวเลขแต่ละคอลัมน์กระจายตัวเป็นอย่างไร

มีข้อมูลหายไปเยอะไหม มีค่าไหนมากเกินไปหรือน้อยไปหรือเปล่า เป็นต้น ไม่มีกฎตายตัวของการทำความเข้าใจข้อมูล

มันเหมือนกับการพูดคุยทำความเข้าใจชุดข้อมูลของคุณเอง ยิ่งคุณพูดคุยกับมันมากเท่าไร
คุณก็ยิ่งทำโมเดลได้ง่ายขึ้นเท่านั้น

ทำอะไรไปบ้างในสัปดาห์ที่แล้ว

/หาและของรึนdataset

ปัญหาที่พบในสัปดาห์ที่แล้ว

/ข้อมูลยังไม่หลากหลาย

วางแผนจะทำอะไรในสัปดาห์นี้

/หาdata setเพิ่ม

Week 6 – Your First Model

หากคุณโชคดีทำความสะอาดข้อมูลเรียบร้อยแล้ว สัปดาห์นี้จะเป็นสัปดาห์ที่คุณเริ่มเทรนโมเดลแรก

ถ้าโชคดีคุณอาจจะได้โมเดลที่ทำได้ดีตามที่ต้องการทันที

แต่โดยทั่วไปแล้วไม่แปลกที่โมเดลแรกของคุณจะออกมาแย่มาก เราจะเริ่มทำการปรับปรุงโมเดลในสัปดาห์ต่อไป

ทำอะไรไปบ้างในสัปดาห์ที่แล้ว

/หาdata set

ปัญหาที่พบในสัปดาห์ที่แล้ว

/ความแม่นยำตกหวน

วางแผนจะทำอะไรในสัปดาห์นี้

/พยายามทำให้ความแม่นยำเพิ่มขึ้น+data setเพิ่ม

Week 7 – Error Analysis

หากโมเดลได้ผลไม่ดีอย่างที่คุณคิด มีสิ่งเดียวที่คุณต้องทำคือ ทดลอง ทดลอง ทดลอง;

ลองตั้งสมมุติฐานว่าทำไมผลถึงไม่ดีและทดลองว่าจริงหรือไม่ เช่น หาก你认为มันไม่ดีเพราะข้อมูลคุณไม่สะอาด

ลองทำความสะอาดใหม่แล้วดูว่าได้ผลใหม่, หาก你认为ว่าปริมาณข้อมูลไม่พอ

ลองเพิ่มหรือลดปริมาณข้อมูลดูว่าได้ผลต่างกันไหม เป็นต้น

ทำอะไรไปบ้างในสัปดาห์ที่แล้ว

/เปลี่ยนdata setเพิ่ม

ปัญหาที่พบในสัปดาห์ที่แล้ว

/ยังไม่ค่อยดีขึ้น

วางแผนจะทำอะไรในสัปดาห์นี้

/ทำแผนผังข้อมูล

Week 8 – Prototype Deployment

หากคุณฝ่าฟันมาจนได้โมเดลที่พอใจแล้ว สัปดาห์นี้จะเป็นการนำโมเดลของคุณออกสู่สายตาชาวโลก

ไม่ว่าจะเป็นในรูปแบบของ **streamlit app** ตามบทเรียน, **chatbot** หรือรูปแบบอื่นๆ

ทำอะไรไปบ้างในสัปดาห์ที่แล้ว

/ลองรันเล่นๆ

ปัญหาที่พบในสัปดาห์ที่แล้ว

/ไม่ค่อยมีเวลา

วางแผนจะทำอะไรในสัปดาห์นี้

/เขียนNotebooks และ scripts ใน github บล็อกใน Medium.com อธิบายโครงการ