# **Traditional Programming VS Machine Learning**

Graphical user interface, application

Description automatically generated with medium confidenceChart

Description automatically generated with low confidence

จากภาพจะเห็นได้ว่า Traditional Programming เราจำเป็นต้องคิด rules เพื่อให้ได้คำตอบ แต่ Machine Learning นั้นเรานั้นต้องการสร้าง Rules ที่ใช้หาคำตอบ

# **Machine Learning Problems**

Machine Learning Problems สามารถแบ่งปัญหาออกเป็น 2 กลุ่มหลักๆ ได้แก่

1. ปัญหาแบบ Supervised ซึ่งเป็นปัญหาที่รู้คำตอบอยู่แล้ว ได้แก่
   1. Classification เป็นปัญหาในด้านการเลือกอย่างใดอย่างหนึ่งจาก N กลุ่ม เช่น การทำนายว่าเป็น หมา แมว หรือ หมี
   2. Regression เป็นปัญหาในด้านการทำนายตัวเลยที่เป็น continuous เช่น การทำนายราคาหุ้น
   3. Ranking เป็นปัญหาในการหาความเชื่อมโยงของข้อมูลเพื่อจัดอันดับ เช่น การค้นหาข้อมูลบน search engine
2. ปัญหาแบบ Unsupervised ซึ่งเป็นปัญหาที่ไม่มีคนสอน (ไม่รู้คำตอบ) ได้แก่
   1. Clustering เป็นปัญหาในการจัดกลุ่มตัวอย่างที่คล้ายกัน
   2. Association rule learning เป็นปัญหาในด้านการหา Pattern ในข้อมูล

# **Reinforcement Learning**

reinforcement learning เป็นการเรียนแบบพฤติกรรมของมนุษย์ ผ่านการลองผิดลองถูก ทำให้มีการเรียนรู้การกระทำไหนดีหรือไม่ดี โดย Agent จะกระทำอะไรบ้างอย่าง ที่ส่งผลกระทบต่อ Environment และจะมี Policy ที่เอาไว้ใช้ตัดสิน reward เพื่อให้ agent รู้ว่าการกระทำนั้นเป็นการกระทำที่ดีหรือไม่ดี

# **Overfit**

คือ การที่โมเดลเรานั้นทำนายตามที่เทรนได้อย่างถูกต้อง จนเมื่อเจอข้อมูลที่ไม่รู้จัก ทำให้ได้ผลลัพธ์ผิดพลาด

# **Tuning**

คือ การที่เราพยายามหาจุดที่ทำให้ model มีประสิทธิภาพดียิ่งขึ้นผ่านการปรับประสิทธิภาพใน Hyper-Parameter

# **Validation**

คือการทดสอบประสิทธิภาพของโมเดล โดยสามารถแบ่งได้ 3 ประเภท

1. Consistency คือการนำ data ที่เอาไปเทรนมาทำการ validate ซึ่งไม่หมดสมในการวัดประสิทธิภาพ เพราะเป็นข้อมูลที่โมเดลเคยเจออยู่แล้ว
2. Splitting คือการแบ่งชุดข้อมูลออกเป็น 2 ชุด ซึ่งในบ้างครั้งถ้าเราสุ่มข้อมูลที่มีความคล้ายกับข้อมูลที่เทรนมากๆ ทำให้วัดผลได้ประสิทธิภาพดี แต่ถ้าสุ่มได้แตกต่างกัน ก็จะทำให้ผลลัพธ์ออกมาแย่
3. Cross-validation คือการแบ่งข้อมูลออกเป็นหลายๆส่วน ซึ่งแต่ละส่วนมีจำนวนข้อมูลที่เท่ากัน และข้อมูลหนึ่งส่วนจะใช้ในการทดสอบประสิทธิภาพของโมเดล และทำซ้ำจนครบจำนวนข้อมูลที่แบ่งไว้

# **Data Science in Business**

ในการทำโปรเจคด้านนี้เราควรคำนึงถึง Impact ที่เกิดขึ้นของโปรเจคว่ามีประโยชน์นระดับไหน รวมทั้งประเภทโมเดล และจำนวนคนที่ต้องใช้งาน พร้อมทั้งฝ่ายอื่นๆที่เกี่ยวข้องกับโปรเจคนี้

Table

Description automatically generated with medium confidence

# **POC vs MVP**

Prototype คือการคิดให้มากขึ้นที่พัฒนามาจาก PoC และเมื่อสรุปงานเสร็จแล้วก็ว่าเป็นผลิตภัณฑ์ที่ใช้งานได้ก็จะกลายเป็น MVP

A screenshot of a computer

Description automatically generated with medium confidence

Text

Description automatically generated

Graphical user interface, text

Description automatically generated with medium confidence

นอกเหนือจาก MVP ยังมี MMP และ MLP ซึ่งคือสามารถเข้าสู่ตลาดได้แต่คนอาจจะไม่นิยมเท่าไหร่ และผลิตภัณฑ์ที่คำนึงถึงความรู้สึกของผู้ใช้ หรือเกิดอารมณ์ว่าใช้แล้วอยากบอกต่อ

# **Confusion Matrix**

Chart, treemap chart

Description automatically generated

ซึ่งในทาง Business ในการที่ model error นั้นเรามี cost ที่เสียหายไปไม่เท่ากัน ดังนั้นเราจริงต้องประเมิน cost ที่เสียไป เช่นภาพด้านล่าง

Table

Description automatically generated

เหตุผลที่ FP เสีย cost น้อยกว่า FN เพราะว่าการที่เราทำนายว่า Churn ทั้งๆที่ไม่ Churn เราอาจจะเสียค่าการตลาดเล็กน้อยเพื่อรักษาลูกค้า แต่ถ้าทำนายตรงข้ามกัน เราต้องเสียลูกค้าไป พร้อมกับจำเป็นต้องมี cost ในการดึงลูกค้ากลับ ซึ่งมากกว่า