

การจำแนกผู้ที่สนใจการเปิดบัญชีเงินฝากผ่านทางโทรศัพท์

นาวิชัย ทุมวงศ์

เทคโนโลยีสารสนเทศ

มหาวิทยาลัยศรีนครินทรวิโรฒ

กรุงเทพมหานคร

napawit.toomwong@g.swu.ac.th

ปัจจุบันการเพิ่มฐานลูกค้าของธนาคารนั้นไม่ได้มาจากการที่ลูกค้าต้องเดินทางไปยังสาขาอีกต่อไป การติดต่อลูกค้าผ่านช่องทางโทรศัพท์ถือเป็นทางเลือกหนึ่งที่ธนาคารใช้ในการลดต้นทุนค่าใช้จ่าย และประหยัดทรัพยากรบุคคลในสาขาที่ต้องใช้เวลานานโดยเฉพาะการที่ต้องอธิบายเงื่อนไขต่างๆให้ลูกค้าเข้าใจ ซึ่งส่งผลกระทบต่อผู้ทำธุรกรรมอื่นๆ

I. INTRODUCTION

ปัจจุบันการเพิ่มฐานลูกค้าของธนาคารนั้นไม่ได้มาจากการที่ลูกค้าต้องเดินทางไปยังสาขาอีกต่อไป การติดต่อลูกค้าผ่านช่องทางโทรศัพท์ถือเป็นทางเลือกหนึ่งที่ธนาคารใช้ในการลดต้นทุนค่าใช้จ่าย และประหยัดทรัพยากรบุคคลในสาขาที่ต้องใช้เวลานาน โดยเฉพาะการที่ต้องอธิบายเงื่อนไขต่างๆให้ลูกค้าเข้าใจ ซึ่งส่งผลกระทบต่อผู้ทำธุรกรรมอื่นๆ

การนำข้อมูลเบื้องต้นจากการทำธุรกรรมอื่นๆของลูกค้านั้น สามารถนำมาวิเคราะห์และทำนายความเป็นไปได้ในการให้ความสนใจกับผลิตภัณฑ์อื่นๆ ของธนาคารได้ ซึ่งสามารถมุ่งเป้าไปยังกลุ่มลูกค้าที่มีความเป็นไปได้ว่าจะสนใจ ซึ่งโอกาสประสบความสำเร็จในการขายอาจจะมากกว่าการโทรสุ่ม ซึ่งสามารถประหยัดงบประมาณและทรัพยากรบุคคลได้

โดยจุดประสงค์ของงานนี้คือการเพิ่มรายได้และช่องทางในการเสนอขาย ในเงื่อนไขที่มีทรัพยากรบุคคลและทุนที่จำกัด เพิ่มความสะดวกสบายให้แก่ลูกค้าที่สนใจในผลิตภัณฑ์นั้นๆ อย่างแท้จริง

II. EXPERIMENT

โดยการนำวิเคราะห์ข้อมูลครั้งนี้ใช้แบ่งออกเป็น เทคโนโลยีที่เกี่ยวข้อง วิธีการเก็บข้อมูล แบบจำลองที่ใช้ เทคนิคที่ใช้และวิธีการทำนาย ดังต่อไปนี้

ก. เทคโนโลยีที่เกี่ยวข้อง

- Python ภาษาโปรแกรมมิ่งที่ใช้ในการเรียกใช้ไลบรารีต่างๆ ซึ่งคุณสมบัติสำคัญของภาษานี้คือการอ่านข้อมูลและเขียนข้อมูลที่สามารถทำได้อย่างรวดเร็ว
- Jupyter notebook เครื่องมือที่ใช้อำนวยความสะดวกในการเขียนภาษา Python

ข. วิธีการเก็บข้อมูล

ข้อมูลที่ทำนายได้มาจาก Open source ในอินเทอร์เน็ต

โดยเป็นข้อมูลเกี่ยวกับประวัติการทำธุรกรรมของลูกค้าธนาคารแห่งหนึ่งในปี 2014

ค. แบบจำลองที่ใช้

- Logistic regression
- Random forest
- Decision tree

ง. เทคนิคที่ใช้

- Cross validation

การทำ Cross validation เป็นเทคนิคในการสร้างแบบจำลองที่เหมาะสมกับข้อมูล โดยการนำข้อมูลทดสอบมาทดสอบกับแบบจำลองต่างๆ

- Grid search

การหาตัวแปรที่ดีที่สุดที่ใช้กับแบบจำลอง

- ROC curve (Receiver operating characteristic curve)

เป็นตัวบอกว่าการทดลองนั้นให้ผลที่แม่นยำหรือไม่

จ. วิธีการทำนาย

- การนำข้อมูลมาวิเคราะห์เพื่อให้เข้าใจในพฤติกรรมของตัวแปรแต่ละประเภท
- การเตรียมข้อมูลโดยการนำข้อมูลที่ไม่มีค่า และข้อมูลที่มีคุณสมบัติไม่อยู่ในกลุ่มเป้าหมายออกไปจากการทำนาย
- การทำ Cross validation เป็นเทคนิคในการสร้างแบบจำลองที่เหมาะสมกับข้อมูล โดยการนำข้อมูลทดสอบมาทดสอบกับแบบจำลองต่างๆ
- ทำนายเพื่อหา Baseline หลังจากการทำ Cross validation จะได้แบบจำลองที่เหมาะสมกับข้อมูลชุดนั้นและทำการทำนายเพื่อหา Baseline ก่อนจะทำการปรับปรุงชุดข้อมูล
- การหาตัวแปรที่เหมาะสมกับแบบจำลองที่เลือกมา (Grid search)

6. ทำนายอีกครั้งหลังจากการใช้ตัวแปรที่ได้จากการทำ Grid search
7. ทำการเปรียบเทียบผลการทำนายที่ได้จากการปรับปรุงและเปรียบเทียบกับ Baseline ด้วย ROC Curve

IV.RESULT AND ANALYSIS

จากผลการทดลองที่ได้หลักจากการทำการปรับปรุงตัวแปรแล้ว นั้นพบว่า Logistic regression ให้ค่า AUC [1] มากที่สุดเมื่อทำการทดสอบเปรียบเทียบกับ Decision tree และ Random forest โดยเรียงลำดับดังนี้

Logistic regression : 0.7939

Decision tree : 0.6260

Random forest : 0.7443

ดังนั้นจึงนำแบบจำลองที่ให้ค่า AUC มากที่สุดคือ Logistic regression มาทำนายเพื่อหา Baseline โดยใช้ข้อมูลชุดเดียวกัน ซึ่งได้ผลดังต่อไปนี้

Accuracy : 0.9004

Precision : 0.6856

Recall : 0.2199

F1-score : 0.33

ขั้นตอนต่อไปคือการทำการปรับปรุงแบบจำลองโดยการหาตัวแปรที่เหมาะสมของแบบจำลอง Logistic regression ซึ่งได้ตัวแปรที่ดีที่สุดดังนี้

$C = 1$, $class_weight = balanced$, $penalty = 11$

เมื่อได้ตัวแปรที่เหมาะสมที่สุดกับแบบจำลองแล้วจึงทำการทำนายผลอีกครั้ง ซึ่งได้ผลดังนี้

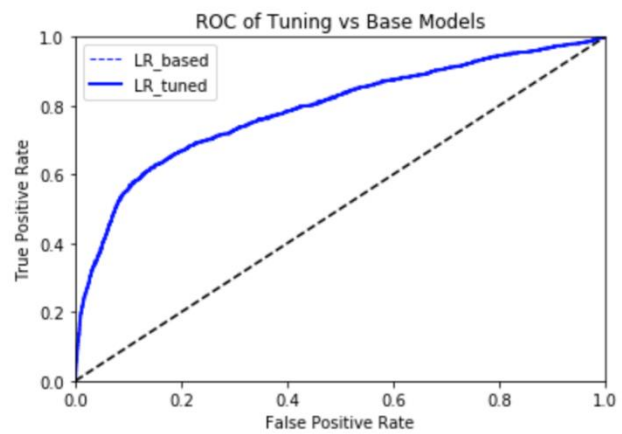
Accuracy : 0.8281

Precision : 0.3522

Recall : 0.6200

F1-score : 0.4492

ซึ่งทำการวาดกราฟได้รูปดังต่อไปนี้



ภาพที่ 1 ผลการทำนายด้วย Logistic regression

ซึ่งจากผลที่ได้สามารถสรุปได้ว่าการทำ Grid search กับ Logistic regression นั้นไม่ได้ส่งผลให้การทำนายแม่นยำขึ้น ซึ่งอาจเป็นผลมาจากการนำข้อมูลทั้งหมดมาทำการทดสอบ และข้อมูลที่ใช้ทดสอบนั้นมีการกระจายตัวที่ไม่เพียงพอ

ทั้งนี้จากกราฟผลที่ได้จากการทำนายนั้นถือว่าค่อนข้างแม่นยำเพราะได้ค่า True positive rate สูงกว่า False positive rate

REFERENCES

- [1] Towards Data Science. (2019). *Understanding AUC - ROC Curve*. [online] Available at: <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5> [Accessed 5 May 2019].

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.