



E-Commerce Sales Analysis and Customer Insights for Store

จัดทำโดย

นายภูวรินทร์ สังฆมาศ 6320500620

นายเฉลิมชัย กำลั้งเดช 6320502363

นายปวีณวิษณุ ทาดี 6320502461

นายธนภัทร ทูเรียนงาม 6320503017

ประกอบรายวิชา 02204390-60 การเตรียมความพร้อมสหกิจศึกษา

ประจำภาคต้น ปีการศึกษา 2566

หลักสูตรวิศวกรรมศาสตรบัณฑิตสาขาวิศวกรรมคอมพิวเตอร์ (หลักสูตรปรับปรุง พ.ศ. 2560)

คณะวิศวกรรมศาสตร์กำแพงแสน ภาควิชาวิศวกรรมคอมพิวเตอร์

มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน

สารบัญ

1. ที่มาและความสำคัญ	1
2. วัตถุประสงค์.....	1
3. ขอบเขตและข้อจำกัดของโครงการ	2
ขอบเขตของโครงการ	2
ข้อจำกัดของโครงการ.....	2
4. ทฤษฎีและหลักการที่เกี่ยวข้อง.....	3
4.1 Sales Forecasting.....	3
4.1.1 Prophet.....	3
4.2 RFM Analysis.....	4
5. แนวทางการดำเนินงาน	5
6. การออกแบบระบบ	7
7. ประโยชน์ที่คาดว่าจะได้รับ	7
8. ผลการดำเนินงาน.....	7
9. อภิปรายผล	10
10. เอกสารอ้างอิง.....	11

1. ที่มาและความสำคัญ

ธุรกิจพาณิชย์อิเล็กทรอนิกส์ หรือ E-commerce ถือเป็นหนึ่งในการค้ารูปแบบใหม่ในปัจจุบันมีอัตราการเติบโตอย่างก้าวกระโดด ด้วยความมาถึงของยูทิลิตี้เน็ตและการพัฒนาอย่างต่อเนื่องจนในปัจจุบันปฏิเสธไม่ได้เลยว่า E-commerce กลายเป็นธุรกิจที่มีความนิยมและสร้างผลกำไรมหาศาลให้กับผู้ประกอบการ แต่สิ่งหนึ่งที่เกิดขึ้นภายใต้การเติบโตที่วุ่นวายนั้นก็คือ การแข่งขันที่มากขึ้นเช่นเดียวกัน นั่นทำให้ผู้ประกอบการต้องปรับตัวและสร้างกลยุทธ์ใหม่ๆ มากขึ้นเพื่อที่จะยังคงซึ่งผลกำไรและการเติบโตของธุรกิจ และวิธีการหนึ่งที่น่าสนใจก็คือ การใช้วิธีการทาง Data Science เพื่อนำข้อมูลที่เป็นผลพลอยได้จาก E-commerce มาทำการศึกษา ดึงข้อมูลเชิงลึก ไปจนถึงนำไปผสมผสานกับการตลาดและสินค้า

วิทยาศาสตร์ข้อมูล หรือ Data Science เป็นแนวคิดที่ใช้เทคโนโลยี สถิติ และธุรกิจมาวิเคราะห์ข้อมูลเพื่อสกัดเป็นข้อมูลเชิงลึกหรือสร้างระบบช่วยสนับสนุนการตัดสินใจ (DSS) นั่นทำให้ Data Science กลายเป็นหนึ่งในศาสตร์ที่สำคัญในการช่วยให้ผู้ประกอบการนั้นสามารถสร้างความได้เปรียบและยกระดับธุรกิจของตนได้ผ่านวิธีการต่างๆ เช่น Sales Forecasting[1] , RFM Analysis [2]

จากที่เห็นได้ว่าวิทยาศาสตร์ข้อมูลนั้นสามารถที่จะช่วยวิเคราะห์และสร้างข้อมูลเชิงลึกให้กับธุรกิจได้ ทางผู้จัดทำจึงได้จัดทำโครงการนี้ขึ้น โดยเป็นการวิเคราะห์ข้อมูลผ่านการใช้ข้อมูลของ UK-based and registered non-store online retail มาวิเคราะห์ด้วยเทคนิคทาง Data Science เพื่อวิเคราะห์การขายและข้อมูลเชิงลึกของลูกค้า

2.วัตถุประสงค์

โครงการฉบับนี้ถูกสร้างขึ้นเพื่อศึกษาและวิเคราะห์แนวโน้ม, แนวทางและข้อมูลการขายที่เกิดขึ้นในแหล่งซื้อขายของ UK-based and registered non-store online retail

- เพื่อสำรวจและวิเคราะห์ข้อมูลเกี่ยวกับลูกค้าที่ใช้บริการ Store เพื่อค้นหาความสัมพันธ์ระหว่างพฤติกรรมและความพึงพอใจของลูกค้า
- ใช้ข้อมูลวิเคราะห์เพื่อทำข้อเสนอแนะที่สอดคล้องกับความต้องการของลูกค้า และพัฒนาสินค้าและบริการเพื่อสร้างความพึงพอใจและการพัฒนาของธุรกิจ
- พัฒนาและสร้างแบบจำลองทำนายเพื่อทำการคาดการณ์ยอดขายในอนาคต และเพิ่มความมั่นใจในการวางแผนสำหรับอนาคต

3.ขอบเขตและข้อจำกัดของโครงการ

ขอบเขตของโครงการ

- ข้อมูลของ UK-based and registered non-store online retail ที่มีมากกว่า 500,000 transaction สามารถแบ่งได้ดังภาพที่ 1

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
InvoiceNo	ID	Categorical		a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation		no
StockCode	ID	Categorical		a 5-digit integral number uniquely assigned to each distinct product		no
Description	Feature	Categorical		product name		no
Quantity	Feature	Integer		the quantities of each product (item) per transaction		no
InvoiceDate	Feature	Date		the day and time when each transaction was generated		no
UnitPrice	Feature	Continuous		product price per unit	sterling	no
CustomerID	Feature	Categorical		a 5-digit integral number uniquely assigned to each customer		no
Country	Feature	Categorical		the name of the country where each customer resides		no

ภาพที่ 1 | แสดงชื่อและคำจำกัดความของตัวแปร

ที่มา : [Online Retail II Data Set from ML Repository \(kaggle.com\)](https://www.kaggle.com/datasets/mlgroup/online-retail)

- การวิเคราะห์ข้อมูลด้วยเทคนิค 2 ได้แก่
 - Sales Forecasting[1] : Predict future sales based on historical data
 - RFM Analysis[2]

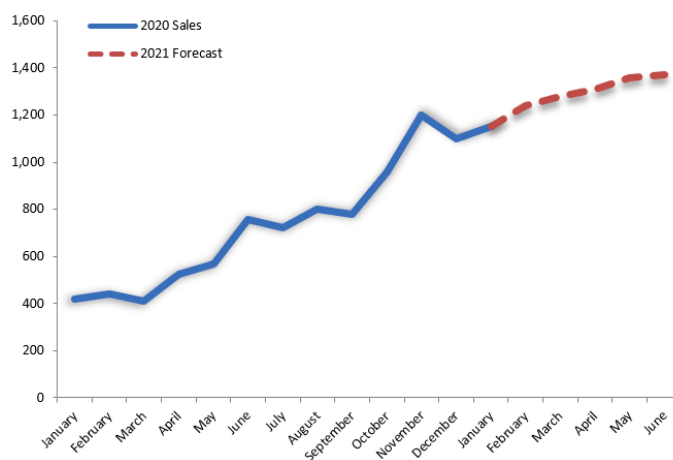
ข้อจำกัดของโครงการ

ข้อมูลในโครงการมาจากแหล่งซื้อขายออนไลน์ของ UK-based and registered non-store online retail เท่านั้น และอาจจำกัดเฉพาะช่วงเวลาหรือขอบเขตที่กำหนด ดังนั้นการวิเคราะห์ข้อมูลอาจจะไม่ครอบคลุมทุกปัจจัยที่ส่งผลต่อการขายและพฤติกรรมของลูกค้า และอาจจะไม่สามารถระบุปัจจัยที่ก่อให้เกิดการขายที่แน่นอนได้ อย่างไรก็ตาม การนำเสนอข้อมูลและข้อเสนอแนะยังคงเป็นประโยชน์ในการพัฒนารูทกิจ

4. ทฤษฎีและหลักการที่เกี่ยวข้อง

4.1 Sales Forecasting

Sales Forecasting[1] คือ กระบวนการพยากรณ์ยอดขายในอนาคตโดยอาศัยข้อมูลการขายย้อนหลัง และปัจจัยที่เกี่ยวข้องอื่น ๆ ซึ่งใช้ในการวางแผนและการตัดสินใจทางธุรกิจ การคาดการณ์การขายที่แม่นยำช่วยเพิ่มประสิทธิภาพการจัดการสินค้าคงคลัง การจัดสรรทรัพยากร และกลยุทธ์ทางการตลาด ดังภาพที่ 2



ภาพที่ 2 | Sales Forecasting

4.1.1 Prophet

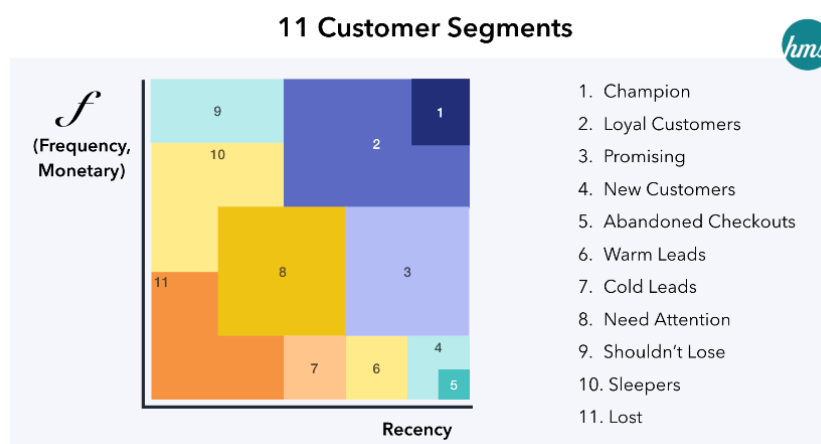
Prophet[3] คือ กระบวนการวิเคราะห์ข้อมูลอนุกรมเวลาโดยใช้โมเดลแบบเพิ่มเติมที่จะพยากรณ์แนวโน้มที่คิดตาม non-linear โดยจะครอบคลุมเวลาตามฤดูกาล, รายปี, รายสัปดาห์, และรายวัน ซึ่งโมเดลนี้จะทำงานได้ดีที่สุดกับข้อมูลชุดที่เกี่ยวข้องกับฤดูกาล การใช้ Prophet ยังสามารถใช้กับข้อมูลที่จะเกิดการขาดหายและการเปลี่ยนแปลงในแนวโน้มและจัดการกับค่าผิดปกติได้ดี

$$y_t = g(t) + s(t) + h(t) + \epsilon_t,$$

ภาพที่ 3 | สมการ non-linear regression

4.2 RFM Analysis

RFM Analysis[2] เป็นวิธีการทำ Customer Segmentation ผ่านการใช้ Transactions Data มาแบ่งกลุ่ม โดยใช้ 3 criteria หลักนั่นคือ R – Recency การซื้อล่าสุด , F – Frequency ความถี่ในการซื้อ และ M – Monetary ปริมาณการซื้อ โดยการแบ่งด้วยหลักการเหล่านี้จะสามารถทำให้สามารถมองเห็นถึง ลูกค้าในกลุ่มต่างๆ ความชอบ ไปจนถึงพฤติกรรม โดยสามารถนำไปใช้ในการพิจารณา Business Action ในอนาคตได้ ดังภาพที่ 4



ภาพที่ 4 | แสดง R(FM) segments map ของการแบ่งกลุ่มลูกค้าทั้งหมด 11 กลุ่ม ด้วย

Rule Based Segmentation

5.แนวทางการดำเนินงาน

Maket											
ทำการเอกสาร											
Project Start Date Project End		7/1/2023 (Saturday) Before Today		Display Week		Gantt Chart Template © 2006-2018 by Vertex42.com Learn about the Pro version >					
		Today Today		2		< >					
WBS TASK		START	END	DAYS	% DONE	WORK DAYS	Week 2	Week 3	Week 4	Week 5	Week 6
1											
1.1	ทำวิจัย	Mon 8/07/23	Sun 8/13/23	7	100%	7					
1.2	ทำ Dataset	Sun 8/13/23	Sat 8/19/23	7	100%	7					
1.3	ทำเอกสารอ้างอิง	Sat 8/19/23	Fri 8/25/23	7	100%	7					
2											
2.1	A	Mon 8/21/23	Mon 9/04/23	15	100%	0					
2.2	ทำ RFM analysis	Sun 8/20/23	Sat 9/02/23	14	100%	0					
3											
3.1	ทำการเอกสารเสนอโครงการ	Sun 8/20/23	Sun 9/03/23	15	100%	15					
3.2	ทำการเอกสารรายงานผล	Fri 9/01/23	Fri 9/15/23	15	100%	0					

Market

Market

กิจกรรมการตลาด

Project Start Date

Tue 2025 (Saturday)

Project Lead

ผู้รับผิดชอบ

Display Week

2

Week 7

7 Aug 2023

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

Week 8

14 Aug 2023

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

Week 9

21 Aug 2023

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

Week 10

28 Aug 2023

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

Week 11

4 Sep 2023

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

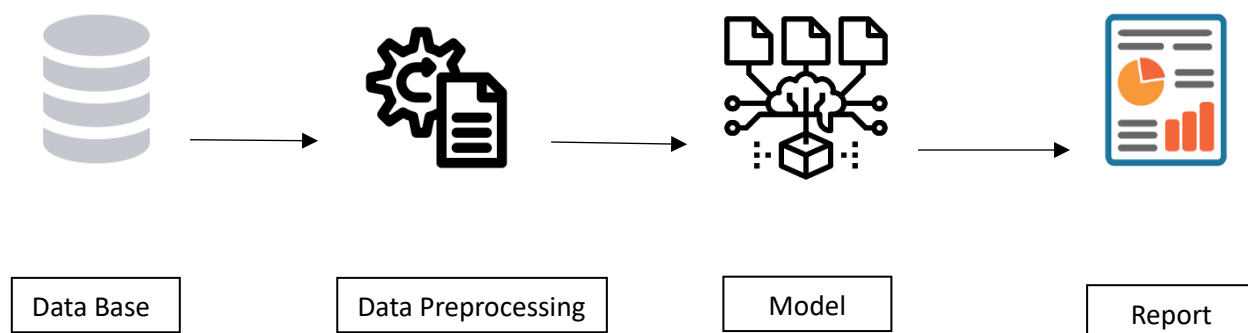
31

1

รายละเอียด

6.การออกแบบระบบ

โดยจะนำข้อมูลจาก Store มาเตรียมให้เหมาะสำหรับการนำไปเข้าโมเดลเพื่อวิเคราะห์และนำมาเสนอผลผ่านทางรายงานสรุปผล

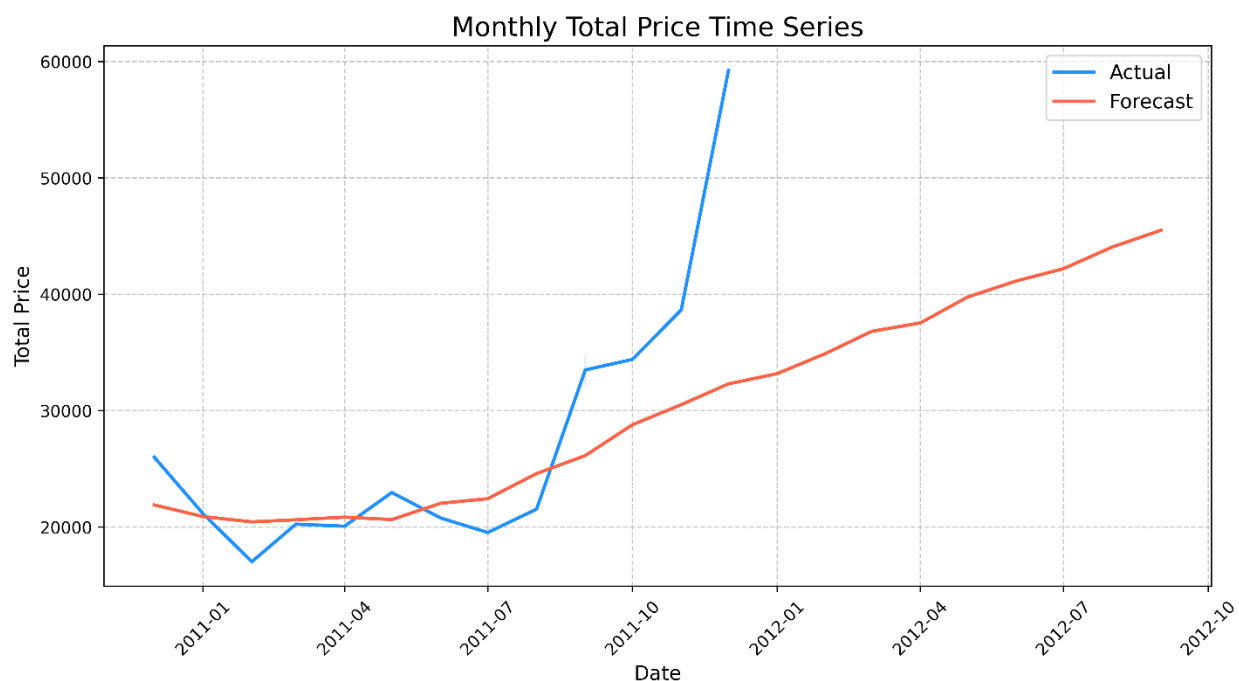


7.ประโยชน์ที่คาดว่าจะได้รับ

- ข้อมูลที่จะสามารถนำไปพัฒนาการบริการและสินค้าที่ให้บริการแก่ลูกค้าได้
- เพิ่มความเข้าใจในพฤติกรรมและความพึงพอใจของลูกค้า
- เข้าใจความสัมพันธ์ระหว่างผลิตภัณฑ์ที่ถูกซื้อพร้อมกัน

8.ผลการดำเนินงาน

- Prophet Forecasting
 - จากการดำเนินงานพบว่า มีเปอร์เซ็นต์ความผิดพลาดอยู่ที่ 28.17% โดยคาดว่าจะมีสาเหตุมาจากข้อมูลที่มีปริมาณไม่มากพอคือระยะเวลาประมาณ 1 ปี ทำให้โมเดลยังมีความผิดพลาดสูง ทั้งนี้เรายังสามารถใช้ผลการ Forecast ของโมเดลเพื่อดูแนวโน้มของยอดขายได้จากกราฟซึ่งบ่งบอกว่ายอดขายใน 1 ปีข้างหน้ามีแนวโน้มที่จะเพิ่มขึ้น



- RFM Analysis

- จากการดำเนินงานทางทีมได้แบ่งกลุ่มลูกค้าออกเป็น 11 กลุ่มด้วย criteria แบบ RFM ได้ตั้งตารางที่ 1 ตารางที่ 1 RFM pattern ในการแบ่งกลุ่มลูกค้า

Pattern	Segment	Description
5[4-5][4-5]	Champions	เพิ่งซื้อไปไม่นานมานี้ ซื้อบ่อย และ ใช้จ่ายเยอะ
[3-5][4-5][4-5]	Loyal customers	ใช้จ่ายเยอะ และซื้อบ่อย
[3-5][2-3][3-5]	Potential loyalists	เพิ่งซื้อไปไม่นาน ใช้จ่ายปานกลางถึงสูง และซื้อมากกว่า 1 ครั้ง
[3-5][3-5][3-5]	Need attention	เหนือกว่าค่าเฉลี่ยทั้งหมด
[1-5][3-5][1-5]	Normal Customer	กลุ่มลูกค้าทั่วไป
5[1-5][1-5]	New customers	เพิ่งซื้อไปไม่นาน และซื้อสินค้าไม่บ่อย
4[1-5][1-2]	Promising small money	เพิ่งซื้อไปไม่นาน และใช้จ่ายน้อย
4[1-5][3-5]	Promising big money	เพิ่งซื้อไปไม่นาน และใช้จ่ายมาก
[1-2][1-2][4-5]	Long time big buy	ใช้จ่ายเยอะ แต่นานๆครั้ง
[2-3][1-3][1-5]	At Risk	มิใช้จ่ายน้อย ซื้อไม่บ่อย และ ซื้อครั้งล่าสุดเมื่อนานมาแล้ว และมีแนวโน้มที่จะเลิกใช้บริการ
[1-2][1-2][1-3]	Hibernating	ใช้จ่ายน้อย ซื้อไม่บ่อย และ ซื้อครั้งล่าสุดเมื่อนานมาแล้ว และอาจเลิกใช้บริการ

ได้ผลลัพธ์การประมวลผลหลังจากวิเคราะห์และจำแนกกลุ่มแล้วดังตารางที่ 2
 ตารางที่ 2 ผลการแบ่งจำนวนลูกค้าด้วย RFM analysis

Segment	Customer
Normal Customer	866
At Risk	653
Loyal customers	631
Hibernating	627
Champions	562
Potential loyalists	438
Need attention	218
Promising small money	167
New customers	74
Long time big buy	73
Promising big money	16

จากข้อมูลที่ได้จากการวิเคราะห์นั้นทำให้ได้เห็นข้อมูลเชิงลึกเพิ่มเติมจากข้อมูลดังกล่าวถึงกลุ่มลูกค้าที่น่าสนใจ 3 กลุ่มหลักที่สามารถนำมาขยายผลการตลาดได้ดังนี้

1. ลูกค้าในกลุ่ม Champions และ Loyal Customers

ลูกค้ากลุ่มนี้โดยทั่วไปนั้นเป็นลูกค้าประจำและมียอดการซื้อที่สูงอยู่แล้วจึงควรที่จะจัด product ที่เขาสนใจเพื่อเพิ่มยอดขายและทำการ cross-sales เพิ่มเติมให้กับลูกค้ากลุ่มนี้

2. ลูกค้าในกลุ่ม Potential loyalists และ Need attention

ลูกค้ากลุ่มนี้โดยทั่วไปนั้นเป็นกลุ่มลูกค้าที่เริ่มจะขยับขึ้นเป็นลูกค้าขาประจำมากขึ้นจึงควรวางแผนการตลาดสร้างแรงจูงใจในการซื้อ เช่น ข้อเสนอหรือส่วนลดพิเศษ

3. ลูกค้าในกลุ่ม At Risk

จากข้อมูลทำให้เห็นว่าลูกค้าในกลุ่มนี้มียอดสูงถึง 653 คน หรือคิดเป็น 15.1% ของจำนวนลูกค้าทั้งหมดแสดงให้เห็นว่าจำนวนลูกค้าที่เข้าใกล้ความเสี่ยงที่จะเลิกซื้อกับทางเรามีจำนวนสูงมาก จึงควรที่จะศึกษาและตรวจสอบถึง pain point นั้นเพิ่มเติม

9.อภิปรายผล

จากผลการศึกษาการวิเคราะห์แนวโน้ม แนวทางและข้อมูลการขายที่เกิดขึ้นในแหล่งซื้อขายของ UK-based and registered non-store online retail ด้วยการทำการวิเคราะห์ผ่านทาง Data Science 2 วิธีได้แก่ Prophet Forecasting และ RFM analysis มีประเด็นที่น่าสนใจที่จะนำมาอภิปรายผลดังนี้

1. ข้อมูลที่ใช้ในการ forecast ควรจะมีระยะ 5-10 ปี เนื่องจากทางทีมได้ทดสอบตัวโมเดลกับข้อมูลอื่นที่มีระยะเวลา 5 ปี พบว่ามีประสิทธิภาพดีขึ้น และข้อมูลควรที่จะใกล้เคียงกับปัจจุบันมากที่สุด
2. ข้อมูลที่นำมาใช้ควรมีความต่อเนื่อง เพราะจะสามารถขยายผลไปใช้โมเดลการ forecast อื่นๆได้อีก
3. การตั้ง score สำหรับ RFM นั้นจะมีลักษณะเปลี่ยนแปลงตามข้อมูลด้วยจึงทำให้ทางทีมต้องปรับเปลี่ยนตัวเลขบ้างเล็กน้อยอาจไม่เป็นไปตามอุดมคติ เนื่องจากพฤติกรรมและยอดต่างๆของลูกค้าที่ไม่เหมือนกัน

10.เอกสารอ้างอิง

- [1] C. H. Wang and Y. W. Gu, "Sales Forecasting, Market Analysis, and Performance Assessment for US Retail Firms: A Business Analytics Perspective," *Applied Sciences (Switzerland)*, vol. 12, no. 17, Sep. 2022, doi: 10.3390/app12178480.
- [2] J.-T. Wei, S.-Y. Lin, and H.-H. Wu, "A review of the application of RFM model," *African Journal of Business Management*, vol. 4, no. 19, pp. 4199–4206, 2010, [Online]. Available: <http://www.academicjournals.org/AJBM>
- [3] S. J. Taylor and B. Letham, "Forecasting at Scale", doi: 10.7287/peerj.preprints.3190v2.