

ระบบแช่ตบอดและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ



สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศทางธุรกิจ ไม่สังกัดภาควิชา/เทียบเท่า

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2562

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

CHATBOT AND ADVANCED DATA ANALYTICS SYSTEM OF INTERNATIONAL SHIPPING
SERVICE BUSINESS



An Independent Study Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Information Technology in Business

Common Course

FACULTY OF COMMERCE AND ACCOUNTANCY

Chulalongkorn University

Academic Year 2019

Copyright of Chulalongkorn University

หัวข้อสารนิพนธ์	ระบบแช่ตบอดและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจ
	บริการจัดส่งพัสดุไปต่างประเทศ
โดย	น.ส.ทัณฑิกา นันยากรสกุล
สาขาวิชา	เทคโนโลยีสารสนเทศทางธุรกิจ
อาจารย์ที่ปรึกษาหลัก	อาจารย์ ดร.วิชรา จันทาทับ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับสารนิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

คณะกรรมการสอบสารนิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.จันทร์เจ้า มงคลนาวิน)

..... อาจารย์ที่ปรึกษาหลัก
(อาจารย์ ดร.วิชรา จันทาทับ)

..... กรรมการ
(รองศาสตราจารย์ ดร.ธีรยุทธ วัฒนาศุภโชค)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ทัศนिका นันยากรสกุล : ระบบแช็ตบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการ
จัดส่งพัสดุไปต่างประเทศ. (

CHATBOT AND ADVANCED DATA ANALYTICS SYSTEM OF INTERNATIONAL
SHIPPING SERVICE BUSINESS) อ.ที่ปรึกษาหลัก : อ. ดร.วัชรวิทย์ จันทาทับ

ทุกวันนี้การก้าวสู่ยุคดิจิทัลส่งผลให้ธุรกิจต้องปรับตัวรองรับกระแสการเปลี่ยนแปลงที่เกิดขึ้นอย่างรวดเร็ว โดยเฉพาะธุรกิจพาณิชย์อิเล็กทรอนิกส์หรืออี-คอมเมิร์ซ ที่เชื่อมโยงผู้ซื้อและผู้ขายเข้าด้วยกันผ่านระบบอินเทอร์เน็ต และเนื่องจากธุรกิจโลจิสติกส์มีรายละเอียดค่อนข้างมาก ตั้งแต่ต้นทางจนถึงปลายทาง ทั้งยังมีปัจจัยเพิ่มเติมในเรื่องของเงื่อนไขเวลา เทคโนโลยีแช็ตบอตจึงนับเป็นอีกเทคโนโลยีหนึ่งที่จะสามารถช่วยให้ธุรกิจโลจิสติกส์สามารถให้บริการแก่ลูกค้าได้แทบทุกที่ทุกเวลาและอย่างชาญฉลาด รวมทั้งยังสามารถเก็บข้อมูลจากลูกค้าเพื่อที่จะนำมาวิเคราะห์ต่อยอดด้วยเทคโนโลยีการวิเคราะห์ข้อมูลขั้นสูงได้อีกด้วย

โครงการ “ระบบแช็ตบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” ที่พัฒนาขึ้นนี้ประกอบด้วย 5 ระบบ ได้แก่ (1) ระบบแช็ตบอต (2) ระบบวิเคราะห์ข้อมูลจากการใช้งานแช็ตบอต (3) ระบบทำนายการสมัครใช้บริการเว็บไซต์ (4) ระบบจัดกลุ่มผู้ใช้งานเว็บไซต์ และ (5) ระบบทำนายการเลิกใช้บริการของลูกค้า โดยระบบแช็ตบอตได้ถูกพัฒนาขึ้นด้วยโปรแกรม DialogFlow และ LINE Messaging API และระบบการวิเคราะห์ข้อมูลขั้นสูงทั้งหมดได้ถูกพัฒนาขึ้นด้วยโปรแกรม RapidMiner ร่วมกับระบบจัดการฐานข้อมูล Microsoft SQL Server 2019 Express

ระบบสารสนเทศจากโครงการพิเศษนี้จะช่วยให้ผู้บริหารและผู้ใช้สามารถเพิ่มศักยภาพในการให้บริการลูกค้าของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ รวมถึงเพิ่มขีดความสามารถในการเก็บข้อมูล การนำข้อมูลมาใช้ให้เกิดประโยชน์สูงสุด และเพิ่มความได้เปรียบในการแข่งขันให้กับองค์กรได้ในระยะยาว

สาขาวิชา เทคโนโลยีสารสนเทศทางธุรกิจ ลายมือชื่อนิสิต

ปีการศึกษา 2562

ลายมือชื่อ อ.ที่ปรึกษาหลัก

6181511426 : MAJOR INFORMATION TECHNOLOGY IN BUSINESS

KEYWORD: Logistics, Advanced Data Analytics System, Chatbot, Data Mining,
International Shipping Service Business

Tantika Nanyakornsakun :
CHATBOT AND ADVANCED DATA ANALYTICS SYSTEM OF INTERNATIONAL
SHIPPING SERVICE BUSINESS. Advisor: WACHARA CHANTATUB, Ph.D.

At present, the advancement of digital era caused businesses to adapt swiftly, especially in electronic commerce or E-Commerce. Hence, chatbot is one of the technologies that can help logistics business intelligently provide services to customers almost anywhere and anytime. Also, it enables the organization to collect data from customers for further analysis and can be used perfectly with advanced data analytics technology.

The "Chatbot and Advanced Data Analytics System of International Shipping Service Business" project consists of five systems, namely (1) Chatbot System, (2) Data Analysis System from Chatbot Usage, (3) Website Subscription Prediction System, (4) Website Users Clustering System, and (5) Customer Churn Prediction System. The Chatbot System was developed using DialogFlow and LINE Messaging API. All advanced data analytics systems were developed using RapidMiner and Microsoft SQL Server 2019 Express database management system.

The information systems from this special project will enable users to increase customer service capabilities, enhance the ability of store data to maximize benefits, and expand the competitive advantage for the organization in the long run.

Field of Study: Information Technology in Student's Signature
Business

Academic Year: 2019 Advisor's Signature

กิตติกรรมประกาศ

โครงการพิเศษเรื่อง “ระบบแช็ตบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” ที่ได้พัฒนาขึ้นนี้สามารถสำเร็จลุล่วงได้เป็นอย่างดีด้วยความช่วยเหลือ การให้คำแนะนำ และการสนับสนุนจากหลายฝ่าย โดยเฉพาะอย่างยิ่ง อาจารย์ ดร. รัชรา จันทาทับ อาจารย์ที่ปรึกษาโครงการ ที่ได้สละเวลาอันมีค่าในการให้คำปรึกษา แนะนำ ตรวจสอบ และแก้ไขจุดบกพร่องต่าง ๆ เสมอมาจนโครงการนี้เสร็จสิ้นโดยสมบูรณ์

ขอขอบพระคุณคณาจารย์ทุก ๆ ท่าน ที่ได้ประสิทธิ์ประสาทวิชาความรู้ ตลอดระยะเวลาในการศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต ทำให้สามารถนำความรู้ต่าง ๆ เหล่านั้น มาประยุกต์ใช้ประกอบการจัดทำโครงการพิเศษนี้

ขอขอบคุณบริษัทที่นำมาเป็นต้นแบบในการพัฒนาโครงการพิเศษ ที่ให้ความรู้ทางด้านธุรกิจ เป็นแหล่งให้ข้อมูล สนับสนุนการวิเคราะห์ข้อมูล ซึ่งเป็นส่วนสำคัญที่ช่วยให้การดำเนินโครงการพิเศษนี้สำเร็จได้ด้วยดี

ขอขอบคุณครอบครัวอันเป็นที่รักของข้าพเจ้า พ่อ แม่ และสมาชิกในครอบครัวทุกท่านรวมไปถึงเพื่อน ๆ ทุกคน สำหรับกำลังใจ และการสนับสนุนข้าพเจ้าอย่างเต็มที่มาโดยตลอด

สุดท้ายนี้หากโครงการพิเศษนี้มีข้อผิดพลาดหรือมีข้อบกพร่องประการใด ทางผู้จัดทำต้องขออภัยมา ณ ที่นี้ และหวังว่าโครงการพิเศษนี้จะเป็นประโยชน์แก่ผู้ที่สนใจต่อไป

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ค
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
บทที่ 1 บทนำ	1
1.1 ความสำคัญและที่มาของโครงการ.....	1
1.2 วัตถุประสงค์ของโครงการ	3
1.3 ขอบเขตของโครงการ.....	3
1.4 วิธีการดำเนินโครงการ.....	5
1.5 เทคโนโลยีที่ใช้ในการพัฒนาระบบ.....	8
1.6 ประโยชน์ที่คาดว่าจะได้รับ.....	8
บทที่ 2 เหตุผลและแนวคิด.....	10
2.1 แนวคิดเกี่ยวกับธุรกิจโลจิสติกส์ (Kiglogistics).....	10
2.2 แนวคิดเกี่ยวกับแช็ตบอต	16
2.3 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูลขั้นสูง (Advanced Data Analytics) หรือการทำเหมืองข้อมูล (Data Mining).....	20
บทที่ 3 โครงสร้างองค์กรและการดำเนินงาน	40
3.1 ประวัติองค์กร	40
3.2 โครงสร้างองค์กร.....	40
3.3 ลักษณะและการดำเนินงานขององค์กร	42
3.4 ปัญหาในการดำเนินงาน.....	43

บทที่ 4 การพัฒนาระบบงาน.....	44
4.3.1 ระบบแชทบอต (Chatbot System)	45
4.3.2 ระบบวิเคราะห์ข้อมูลจากการใช้งานแชทบอต (Data Analysis System from Chatbot Usage).....	64
4.3.3 ระบบทำนายการสมัครใช้บริการเว็บไซต์ (Website Subscription Prediction System).....	66
4.3.4 ระบบจัดกลุ่มผู้ใช้งานเว็บไซต์ (Website Users Clustering System)	82
4.3.5 ระบบทำนายการเลิกใช้บริการของลูกค้า (Customer Churn Prediction System) 94	
บทที่ 5 บทสรุป ปัญหา และข้อเสนอแนะ.....	109
5.1 บทสรุป	109
5.2 ปัญหา.....	111
ภาคผนวก ก การติดตั้งซอฟต์แวร์ที่ใช้งานในระบบ.....	116
ภาคผนวก ข เมนูการทำงานของระบบ.....	123
ภาคผนวก ค ตัวอย่างการใช้งานระบบ	130
บรรณานุกรม.....	135
ประวัติผู้เขียน.....	140

สารบัญตาราง

ตารางที่ 1-1: เทคโนโลยีที่ใช้ในการพัฒนาระบบ	8
ตารางที่ 4-1: การแก้ปัญหาของบริษัทด้วยการใช้งานแชทบอท	47
ตารางที่ 4-2: การตั้งค่าการทำงานหลักของ DialogFlow	52
ตารางที่ 4-3: ข้อมูลนำเข้าสำหรับการทำนายการสมัครใช้บริการ	67
ตารางที่ 4-4: ข้อมูลนำเข้าสำหรับการจัดกลุ่มผู้ใช้งานเว็บไซต์	83
ตารางที่ 4-5: ข้อมูลนำเข้าสำหรับการทำนายการเลิกใช้บริการของลูกค้า.....	94



สารบัญรูปภาพ

รูปที่ 1-1: มูลค่าตลาดขนส่งพัสดุของไทย	2
รูปที่ 2-1: กิจกรรมโลจิสติกส์.....	11
รูปที่ 2-2: ประโยชน์ของแช็ตบอต	20
รูปที่ 2-3: ขั้นตอนการทำ Data Mining.....	23
รูปที่ 2-4: รากฐานของการทำเหมืองข้อมูล.....	25
รูปที่ 2-5: เทคนิคที่พบได้บ่อยในการทำเหมืองข้อมูล.....	26
รูปที่ 2-6: กฎความสัมพันธ์ (Association Rule) (Budkod, 2561).....	27
รูปที่ 2-7: การจำแนกกลุ่ม (Clustering) (Budkod, 2561).....	27
รูปที่ 2-8: การจำแนกกลุ่ม (Clustering) (Budkod, 2561).....	28
รูปที่ 2-9: ตาราง Confusion Matrix (Gatchalee, 2562)	38
รูปที่ 3-1: โครงสร้างองค์กรของบริษัท Pigeon Air Freight.....	41
รูปที่ 4-1: หลักการทำงานของแช็ตบอตที่พัฒนาด้วย DialogFlow (Programmerbonn, 2562)...	51
รูปที่ 4-2: กระบวนการทำงานของ DialogFlow (Petch Kruapanich, 2561)	55
รูปที่ 4-3: คนไทยกับการใช้งาน LINE.....	56
รูปที่ 4-4: กระบวนการทำงานของ Messaging API.....	58
รูปที่ 4-5: หมวดหมู่ Intent ของระบบแช็ตบอตที่พัฒนา	60
รูปที่ 4-6: การ Code ในรูปแบบของ JSON format	61
รูปที่ 4-7: การแสดงผลแบบปุ่มกดด้วย Flex Message ใน LINE Application	61
รูปที่ 4-8: ตัวอย่างการแสดงผลแบบรูปภาพใน LINE Application.....	62
รูปที่ 4-9: LINE ID และ QR Code สำหรับการเพิ่มแช็ตบอตเป็นเพื่อน	63
รูปที่ 4-10: ข้อมูลที่ได้จากการทดลองใช้งานโดยลูกค้า	65
รูปที่ 4-11: จำนวนการใช้งานแต่ละ Intent	65
รูปที่ 4-12: Session Flow แสดงภาพรวมการใช้งานตลอด Session ของลูกค้า.....	66

รูปที่ 4-13: Process การทำนายการสมัครใช้บริการเว็บไซต์ด้วยโมเดล Decision Tree	73
รูปที่ 4-14: Sub-Process ภายในโอเพอร์เรเตอร์ Validation	74
รูปที่ 4-15: ผลลัพธ์โมเดล Decision Tree ที่ได้สร้างขึ้นจากข้อมูล Training Set.....	75
รูปที่ 4-16: Process การทำนายการสมัครใช้บริการเว็บไซต์ด้วยโมเดล Neural Network.....	79
รูปที่ 4-17: Sub-Process ภายในโอเพอร์เรเตอร์ Validation	80
รูปที่ 4-18: โมเดล Neural Network ที่ได้สร้างขึ้นจากข้อมูล Training Set.....	81
รูปที่ 4-19: ค่าความแม่นยำของโมเดล Decision Tree ในการทำนายการสมัครใช้บริการเว็บไซต์..	81
รูปที่ 4-20: ค่าความแม่นยำของโมเดล Neural Network ในการทำนายการสมัครใช้บริการเว็บไซต์	82
รูปที่ 4- 21: Process การจัดกลุ่มข้อมูลผู้ใช้งานเว็บไซต์ของบริษัทด้วยการจัดกลุ่มแบบ K-Medoids Clustering.....	87
รูปที่ 4-22: Visualization เบื้องต้นของผลลัพธ์การจัดกลุ่มข้อมูลในโปรแกรม RapidMiner	88
รูปที่ 4-23: หน้าจอการทำ Visualization ผลลัพธ์เพิ่มเติมด้วยโปรแกรม Tableau.....	88
รูปที่ 4-24: Process การหาค่า K ที่เหมาะสมสำหรับการจัดกลุ่มข้อมูลผู้ใช้เว็บไซต์.....	89
รูปที่ 4-25: Sub-Process ภายในโอเพอร์เรเตอร์ Loop Parameter	90
รูปที่ 4-26: ผลลัพธ์จากการวน Loop เพื่อหาค่า K ที่เหมาะสมต่อการจัดกลุ่มข้อมูล	91
รูปที่ 4-27: Process การทำนายการเลิกใช้บริการของลูกค้าด้วยโมเดล Decision Tree	97
รูปที่ 4-28: Sub-Process ภายในโอเพอร์เรเตอร์ Validation	98
รูปที่ 4-29: ผลลัพธ์โมเดล Decision Tree ที่ได้สร้างขึ้นจากข้อมูล Training Set.....	99
รูปที่ 4-30: Process การทำนายการเลิกใช้บริการของลูกค้าด้วยโมเดล Random Forest.....	105
รูปที่ 4-31: Sub-Process ภายในโอเพอร์เรเตอร์ Validation	106
รูปที่ 4-32: ค่าความแม่นยำของโมเดล Decision Tree ในการทำนายการเลิกใช้บริการของลูกค้า	107
รูปที่ 4-33: ค่าความแม่นยำของโมเดล Random Forest ในการทำนายการเลิกใช้บริการของลูกค้า	107
รูปที่ 4-34: ค่าน้ำหนักซึ่งแสดงถึงความสำคัญของตัวแปรที่นำเข้าสู่โมเดล	108

รูปที่ ก-1: หน้าจอแรกในการ Sign-up เพื่อสร้าง Account ก่อนเริ่มต้นใช้งาน	116
รูปที่ ก-2: หน้าจอการ Sign-in เข้าสู่ระบบด้วยอีเมลเพื่อเริ่มพัฒนาระบบแชทบอต	116
รูปที่ ก-3: หน้าจอเมนูการสร้างฐานข้อมูลใหม่.....	117
รูปที่ ก-4: หน้าจอการสร้างฐานข้อมูลใหม่.....	118
รูปที่ ก-5: หน้าจอเมนูการนำเข้าข้อมูล.....	118
รูปที่ ก-6: หน้าจอการเลือก Data Source.....	119
รูปที่ ก-7: หน้าจอการเลือก Destination	120
รูปที่ ก-8: หน้าจอการเลือก Copy หรือ Query.....	120
รูปที่ ก-9: หน้าจอการเลือกตารางข้อมูล.....	121
รูปที่ ก-10: หน้าจอการ Save และ Run Package.....	121
รูปที่ ก-11: หน้าจอเสร็จสิ้นการนำเข้าข้อมูล	122
รูปที่ ก-12: หน้าจอแสดงผลพร้อมในการนำเข้าข้อมูลสำเร็จ.....	122
รูปที่ ข-1: หน้าจอหลักทั้ง 3 ส่วนเมื่อเข้าสู่ระบบของ DialogFlow	123
รูปที่ ข-2: หน้าจอการปรับแต่งแชทบอตภายใต้เมนู WELCOME Intent.....	124
รูปที่ ข-3: หน้าจอการเชื่อมต่อแชทบอตเข้ากับ Platform ต่าง ๆ ด้วยเมนู Integrations.....	124
รูปที่ ข-4: หน้าจอแสดงข้อมูลประวัติการใช้งานแชทบอต.....	125
รูปที่ ข-5: หน้าต่างเมนูเริ่มต้นของโปรแกรม RapidMiner Studio 9.5.1	125
รูปที่ ข-6: หน้าจอหลักในการสร้าง Process ของโปรแกรม RapidMiner Studio.....	126
รูปที่ ข-7: ฟังก์ชันในส่วนของ Repository.....	127
รูปที่ ข-8: องค์ประกอบของโอเปอเรเตอร์ (Pacharawongsakda, 2560)	128
รูปที่ ข-9: องค์ประกอบหลักในส่วนของ Process (Pacharawongsakda, 2560).....	128
รูปที่ ข-10: ลักษณะของโอเปอเรเตอร์และพารามิเตอร์ใน Process (Pacharawongsakda, 2560)	129
รูปที่ ข-11: ส่วนช่วยเหลือซึ่งแสดงข้อมูลเกี่ยวกับโอเปอเรเตอร์ชื่อ Linear Regression.....	129

รูปที่ ค-1: หน้าจอการปรับแต่ง Training Phrases ภายใต้เมนู Intents	130
รูปที่ ค-2: การปรับแต่ง Responses ภายใต้เมนู Intents.....	131
รูปที่ ค-3: การเชื่อมต่อโปรแกรม RapidMiner Studio เข้ากับฐานข้อมูลที่ต้องการ	131
รูปที่ ค-4: หน้าต่างสำหรับตั้งชื่อช่องทางในการเชื่อมต่อ.....	132
รูปที่ ค-5: หน้าต่างสำหรับกรอกข้อมูลเพื่อเชื่อมต่อกับฐานข้อมูลที่ได้สร้างไว้	132
รูปที่ ค-6: หน้าจอการนำเข้าข้อมูลผ่านช่องทางการเชื่อมต่อที่ได้สร้างขึ้น.....	133
รูปที่ ค-7: หน้าจอการปรับเปลี่ยน Parameter ของโมเดล Decision Tree.....	134



บทที่ 1

บทนำ

ในบทนี้จะกล่าวถึงความสำคัญและที่มาของโครงการ วัตถุประสงค์ของโครงการ ขอบเขตของโครงการ วิธีการดำเนินงานโครงการ เทคโนโลยีที่ใช้ในการพัฒนาระบบและการดำเนินงาน ตลอดจนประโยชน์ที่คาดว่าจะได้รับจากโครงการนี้

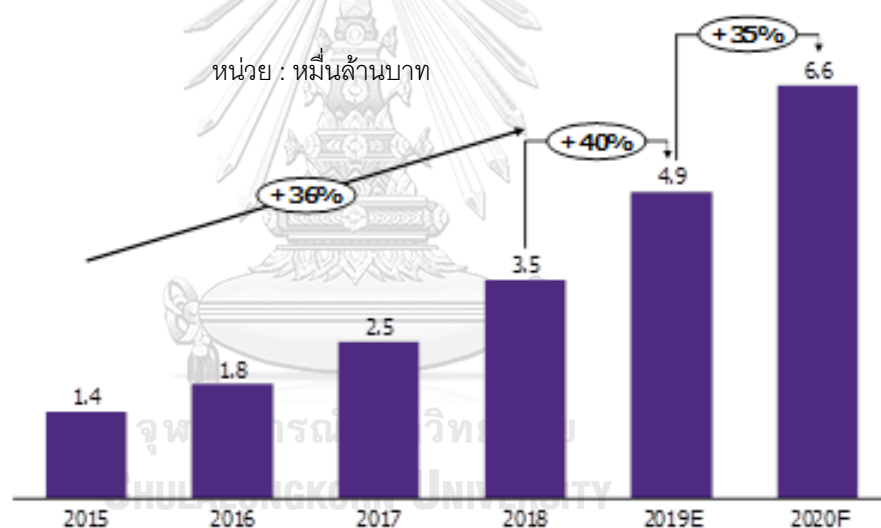
1.1 ความสำคัญและที่มาของโครงการ

ปัจจุบันธุรกิจบริการในด้านโลจิสติกส์ (Kiglogistics) มีการเติบโตอย่างต่อเนื่อง แต่ในขณะเดียวกันก็มีการแข่งขันที่รุนแรงมากขึ้น โดยบริษัทยักษ์ใหญ่ต่าง ๆ ไม่ว่าจะเป็น DHL, FedEx รวมถึงไปรษณีย์ไทยก็เริ่มหันมาให้ความสนใจกับตลาดลูกค้ารายย่อยหรือบุคคลทั่วไปที่เป็นเจ้าของธุรกิจ SME มากขึ้นกว่าแต่ก่อนซึ่งมักจะเน้นลูกค้าที่เป็นองค์กรใหญ่ (WP, 2562) ดังนั้นกุญแจสำคัญอย่างหนึ่งที่จะสามารถสร้างความได้เปรียบในการแข่งขันได้ก็คือการนำเทคโนโลยีสารสนเทศมาใช้เพื่อเพิ่มความแข็งแกร่งให้กับธุรกิจ เทคโนโลยีจะกลายเป็นปัจจัยที่สนับสนุนให้ธุรกิจโลจิสติกส์สามารถให้บริการลูกค้าโดยตอบสนองความต้องการได้ดีขึ้น รวมถึงสามารถคาดการณ์อนาคตได้อย่างแม่นยำมากยิ่งขึ้นในแบบที่ไม่เคยสามารถทำได้ในอดีต

การก้าวสู่ยุคดิจิทัลส่งผลให้ธุรกิจต้องปรับตัวรองรับกระแสการเปลี่ยนแปลงที่เกิดขึ้นอย่างรวดเร็ว โดยเฉพาะ ธุรกิจพาณิชย์อิเล็กทรอนิกส์หรืออี-คอมเมิร์ซ (E-Commerce) ที่เชื่อมโยงผู้ซื้อและผู้ขายเข้าด้วยกันผ่านระบบอินเทอร์เน็ต ซึ่งแนวโน้มการเติบโตของตลาด “อี-คอมเมิร์ซ” ก็มีตัวเลขที่น่าสนใจ สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (องค์การมหาชน) ประเมินว่า มูลค่าการซื้อขายออนไลน์ในปี พ.ศ. 2561 อยู่ที่ 3.05 ล้านล้านบาท เพิ่มขึ้นจากปีก่อน 9-10 เปอร์เซ็นต์ และคาดการณ์ว่าจากปีพ.ศ. 2562 นี้ไปจนถึงปี พ.ศ. 2565 ตลาดอี-คอมเมิร์ซไทยน่าจะเติบโตขึ้นเฉลี่ย 22 เปอร์เซ็นต์ และน่าเชื่อได้ว่าตลาดนี้จะมีโอกาสเติบโตได้อีกมาก ซึ่งเป็นผลมาจากพฤติกรรมการใช้ชีวิตของผู้คนในยุค 4.0 ที่เริ่มเปลี่ยนไปจากอดีตโดยสิ้นเชิง และความเปลี่ยนแปลงตลาดออนไลน์ที่เติบโตขึ้นอย่างมากย่อมเป็นผลดีต่อ “บริการขนส่ง” ด้วยเช่นกัน เนื่องจากกลุ่มธุรกิจอี-คอมเมิร์ซจำเป็นต้องอาศัยการบริการขนส่งเป็นองค์ประกอบหลัก ไม่ว่าจะเป็นธุรกิจขนาดใหญ่และขนาดเล็ก (Nalisa, 2562) ด้วยเหตุนี้ธุรกิจกลุ่มโลจิสติกส์จึงมีการเติบโตอย่างรวดเร็วและต่อเนื่อง โดยคาดการณ์ว่าในปัจจุบันมีมูลค่าไม่ต่ำกว่า 28,000 ล้านบาท และเติบโตโดยเฉลี่ย 10-20 เปอร์เซ็นต์ต่อปี การแข่งขัน

ในธุรกิจนี้จึงเข้มข้น ไม่เฉพาะแต่ผู้ประกอบการในประเทศเท่านั้น แต่ยังมี การเข้ามาลงทุนของบริษัท ขนส่งทั้งในระดับภูมิภาคและระดับโลกอีกด้วย

ขณะที่เม็ดเงินใน “ธุรกิจขนส่ง” หรือ “โลจิสติกส์” นั้น มีมูลค่ามากกว่า 200,000 ล้านบาท มีอัตราเติบโตปีละ 15-20 เปอร์เซ็นต์ ถือเป็นอีกธุรกิจดาวเด่นในโลกดิจิทัลยุค 4.0 จึงไม่แปลกใจที่จะ มีผู้เล่นหน้าใหม่กระโจนเข้าสู่ตลาดนี้กันไม่ขาดสายดังจะเห็นได้จากมูลค่าที่เติบโตดังรูปที่ 1-1 และ เนื่องจากธุรกิจโลจิสติกส์มีรายละเอียดค่อนข้างมากตั้งแต่ต้นทางจนถึงปลายทาง อีกทั้งยังมีปัจจัย เพิ่มเติมในเรื่องของเงื่อนไขเวลาอีกด้วย เทคโนโลยีแช็ตบอตจึงนับเป็นอีกเทคโนโลยีหนึ่งที่จะสามารถ ช่วยให้ผู้ธุรกิจโลจิสติกส์สามารถให้บริการแก่ลูกค้าได้แทบทุกที่ทุกเวลาและได้อย่างชาญฉลาด รวมทั้ง ยังสามารถเก็บข้อมูลจากลูกค้าเพื่อนำมาวิเคราะห์ต่อ และสามารถนำมาใช้งานร่วมกับเทคโนโลยีการ วิเคราะห์ข้อมูลขั้นสูง (Advanced Data Analytics) ได้เช่นเดียวกัน



รูปที่ 1-1: มูลค่าตลาดขนส่งพัสดุของไทย

ที่มา: “EIC” ซึ่งธุรกิจขนส่งพัสดุโต35%แตะ6.6หมื่นล้าน, โดย ประชาชาติธุรกิจ, 2563,

<https://www.prachachat.net/finance/news-412411>

หมายเหตุ: คำนวณจากบริษัทที่ขนส่งพัสดุในไทยรายใหญ่ประมาณ 22 ราย

จากเหตุผลข้างต้นจึงเป็นที่มาของการพัฒนาโครงการ “ระบบแช็ตบอตและการวิเคราะห์ ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” เพื่อให้องค์กรมีแช็ตบอตเข้ามาช่วยให้บริการ ลูกค้า ซึ่งเหมาะกับลักษณะของธุรกิจและสามารถตอบโจทย์ความต้องการของลูกค้าในยุคสมัย

ปัจจุบัน และผู้บริหารได้มีข้อมูลเชิงลึก (Insights) ที่ได้จากการนำเอาข้อมูลเทคโนโลยีสารสนเทศ ทั้งหมดนี้เข้ามาใช้ ทั้งหมดนี้จะช่วยเพิ่มความได้เปรียบให้กับบริษัทและส่งเสริมการขับเคลื่อนธุรกิจให้เติบโตก้าวหน้าได้ดียิ่งขึ้น

1.2 วัตถุประสงค์ของโครงการ

โครงการ “ระบบแชทบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” จัดทำขึ้นโดยมีวัตถุประสงค์ดังนี้

- 1) เพื่อพัฒนาแชทบอตสำหรับให้ข้อมูลลูกค้าและรับเรื่องเกี่ยวกับปัญหาต่าง ๆ ไว้ในเบื้องต้น ซึ่งจะช่วยเพิ่มประสบการณ์ของลูกค้า (Customer Experience) รวมถึงช่วยแบ่งเบาภาระงานของพนักงานฝ่ายขายและฝ่ายบริการลูกค้า ให้สามารถทำงานในส่วนอื่นที่มีความสำคัญหรือซับซ้อนกว่าได้มากขึ้น
- 2) เพื่อนำข้อมูลที่ได้จากแชทบอตมาวิเคราะห์เพิ่มเติมและเป็นอีกหนึ่งช่องทางในการรวบรวมอีเมลเพื่อใช้ในการติดต่อกลุ่มลูกค้าใหม่ ๆ เนื่องจากในปัจจุบันการให้พนักงานโทรศัพท์คุยกับลูกค้าเพียงอย่างเดียวนั้นส่งผลให้การเก็บข้อมูลค่อนข้างทำได้ยาก
- 3) เพื่อจัดทำระบบทำนายข้อมูลการเลิกใช้บริการของลูกค้าเดิม และการสมัครใช้บริการของกลุ่มลูกค้าใหม่ จากนั้นจึงนำมาใช้ปรับปรุงต่อยอดการให้บริการและขยายฐานลูกค้าใหม่ โดยที่ยังคงรักษากลุ่มลูกค้าเดิมเอาไว้
- 4) เพื่อจัดกลุ่มผู้ที่เข้ามาใช้บริการเว็บไซต์ของบริษัทตามปัจจัยทางด้านอายุ เพศ ระดับการศึกษา ลักษณะการใช้งาน ฯลฯ ผู้บริหารและทีมงานของบริษัทจึงจะสามารถวางแผนและตัดสินใจดำเนินการทางการตลาดได้อย่างครอบคลุมและเหมาะสมในระยะยาว

1.3 ขอบเขตของโครงการ

โครงการ “ระบบแชทบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” แบ่งออกเป็น 5 ระบบ โดยมีขอบเขตของแต่ละระบบ ดังนี้

1) ระบบแชทบอต (Chatbot System)

ระบบนี้จะครอบคลุมการให้บริการในด้านการตอบคำถามแก่ลูกค้า กล่าวคือให้ข้อมูลที่ลูกค้าต้องการโดยเป็นคำถามที่มักพบบ่อยตั้งแต่ต้นจนจบกระบวนการส่งจนถึง

การชำระเงิน ไปจนถึงการ Tracking พัสดุ เพื่อช่วยลดภาระงานของพนักงานและเพิ่มความพึงพอใจของลูกค้า (Customer Satisfaction) ให้ลูกค้าเปรียบเทียบราคาได้ง่าย ให้ข้อมูลที่สะดวก รวดเร็ว และถูกต้องแม่นยำ โต้ตอบได้แบบ Real Time และใช้งานง่ายบนมือถือ เป็นอีกหนึ่งช่องทางในการติดต่อรวบรวมข้อมูลของลูกค้า ไม่ว่าจะเป็นความคิดเห็น อีเมลล์ ฯลฯ ซึ่งจะลดโอกาสการได้ข้อมูลที่ผิดพลาดคลาดเคลื่อนจากการที่พนักงานขอข้อมูลจากลูกค้าผ่านทางโทรศัพท์ นอกจากนี้การให้ลูกค้าทำแบบสอบถามที่ดูจริงจังเป็นทางการในเว็บไซต์อาจก่อให้เกิดความรู้สึกเบื่อได้ แต่การใช้แชทบอทซึ่งมีลักษณะเป็น Persona โต้ตอบได้คล้ายมนุษย์นั้นอาจส่งผลให้การพูดคุยเพื่อขอความคิดเห็นหรือข้อติชมต่าง ๆ เป็นไปได้ง่ายมากยิ่งขึ้น รวมถึงเป็นอีกหนึ่งช่องทางในการทำการตลาด โดยใช้ส่ง Message ข้อมูลข่าวสารของบริษัท รวมถึงโฆษณาและโปรโมชั่นต่าง ๆ ไปให้ลูกค้าเพื่อกระตุ้นให้มาใช้บริการได้อีกทางหนึ่ง

2) ระบบวิเคราะห์ข้อมูลจากการใช้งานแชทบอท (Data Analysis System from Chatbot Usage)

ระบบนี้จะครอบคลุมการนำเอาข้อมูลจากระบบแชทบอทมาทำการวิเคราะห์ว่าจากการใช้งานของลูกค้านั้นมีคำถามในประเด็นใดเกิดขึ้นมากเป็นพิเศษหรือไม่ และยังมีข้อบกพร่องใดที่ยังต้องปรับปรุงเพื่อเพิ่มประสิทธิภาพการทำงานของระบบแชทบอทให้ดียิ่งขึ้น เช่น ลูกค้าส่วนใหญ่กดยอกจากการใช้งานเมื่อใด มีคำถามอะไรบ้างที่แชทบอทยังตอบไม่ได้ ลูกค้าสนใจในการใช้งานแชทบอทมากเพียงใด และมีการแลกเปลี่ยนข้อมูลหรือ Engagement ระหว่างลูกค้ากับแชทบอทมากน้อยแค่ไหน ฯลฯ

3) ระบบทำนายการสมัครใช้บริการเว็บไซต์ (Website Subscription Prediction System)

ระบบนี้จะครอบคลุมการทำนายว่าจากผู้ที่เข้าใช้งานเว็บไซต์ของบริษัทนั้น ผู้ใช้ที่มีลักษณะและการใช้งานแบบใดที่มีแนวโน้มจะกดสมัครใช้บริการกับบริษัทในท้ายที่สุด เพื่อให้สามารถมองเห็นได้ว่าปัจจัยใดที่น่าจะส่งผลต่อการสมัครใช้บริการของลูกค้าได้บ้าง และดำเนินการเพื่อให้อัตราการสมัครใช้งานจากผู้ที่เข้ามายังเว็บไซต์นั้นเพิ่มสูงขึ้น

4) ระบบจัดกลุ่มผู้ใช้งานเว็บไซต์ (Website Users Clustering System)

ระบบนี้จะครอบคลุมการวิเคราะห์เพื่อให้สามารถมองเห็นถึงภาพรวมของกลุ่มผู้ที่เข้ามาใช้เว็บไซต์ทั้งหมดได้ว่ามีลักษณะอย่างไร ซึ่งจะช่วยให้สามารถนำไปวิเคราะห์ต่อยอดถึงแผนทางการตลาดให้เหมาะกับลูกค้าแต่ละกลุ่มได้มากที่สุด เพื่อเพิ่มโอกาสในการเข้าถึงลูกค้ากลุ่มใหม่ ๆ ได้มากขึ้นอย่างที่ไม่เคยทำได้ในอดีต

5) ระบบทำนายการเลิกใช้บริการของลูกค้า (Customer Churn Prediction System)

ระบบนี้จะครอบคลุมการทำนายว่าจากลูกค้าของบริษัทนั้น ลูกค้ารายใดที่มีแนวโน้มจะเลิกใช้บริการของบริษัท (Churn) และลูกค้ารายใดที่จะยังคงใช้บริการกับทางบริษัทต่อไป (Loyal) เพื่อให้บริษัทสามารถแก้ปัญหาและลดอัตราการเลิกใช้บริการ (Churn Rate) ให้ลดลงเหลือน้อยที่สุด

1.4 วิธีการดำเนินโครงการ

โครงการ “ระบบแชทบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” มีขั้นตอนในการจัดทำ ดังนี้

1) การเตรียมโครงการ (Project Preparation)

- ศึกษาข้อมูลที่จะนำมาเป็นกรณีศึกษา โดยศึกษาขั้นตอน ลักษณะการดำเนินงาน และความต้องการทางธุรกิจของธุรกิจการจัดส่งพัสดุ และข้อมูลเกี่ยวกับการนำเอาแชทบอตมาใช้งานในธุรกิจต่าง ๆ
- ศึกษาความเป็นไปได้ในการพัฒนาระบบ
- ปรึกษาอาจารย์ที่ปรึกษาเกี่ยวกับความเป็นไปได้ในการจัดทำโครงการ “ระบบแชทบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ”
- วิเคราะห์และสรุปขอบเขตของระบบงานที่จะพัฒนา
- จัดทำแผนการดำเนินการและกำหนดระยะเวลาการดำเนินโครงการ
- ศึกษาความต้องการทางด้านฮาร์ดแวร์และซอฟต์แวร์ที่ต้องใช้ในการพัฒนาโครงการ

- ศึกษาเทคโนโลยีที่จะนำมาใช้ในการพัฒนาระบบ ทั้งในการพัฒนาระบบแช็ตบอต ออกแบบโมเดลเพื่อการทำนาย รวมถึงการจัดกลุ่มข้อมูล เพื่อให้สามารถเลือกใช้เทคโนโลยีได้อย่างเหมาะสม

2) การวิเคราะห์ระบบ (System Analysis)

- เก็บรวบรวมความต้องการและปัญหาในการทำงานปัจจุบัน เพื่อนำมาออกแบบระบบงานให้ตรงกับความต้องการ
- วิเคราะห์และสรุปความต้องการของผู้บริหารและผู้มีส่วนเกี่ยวข้อง
- กำหนดความต้องการด้านฮาร์ดแวร์และซอฟต์แวร์ที่ต้องใช้ในการพัฒนาโครงการ

3) การออกแบบระบบ (System Design)

- ออกแบบลักษณะฟังก์ชันการทำงานของระบบแช็ตบอตให้สามารถแก้ปัญหาที่พบในการให้บริการลูกค้าได้มากที่สุด
- ออกแบบโครงสร้างโมเดลการทำนายด้วย Decision Tree, Random Forest และ Neural Network Method เพื่อให้ตรงกับวัตถุประสงค์ของโครงการ รวมถึงสามารถช่วยในการวางแผนและดำเนินธุรกิจของผู้บริหาร
- ออกแบบกระบวนการจัดกลุ่มข้อมูลลูกค้า เพื่อให้สอดคล้องกับลักษณะเฉพาะของธุรกิจและได้ข้อมูลที่ถูกต้องเหมาะสมมากที่สุดในการจัดกลุ่ม

4) การพัฒนาระบบ (System Development and Implementation)

- ติดตั้ง (Installation and Deployment) และ กำหนดค่า (Configuration) ตามที่ได้ออกแบบไว้
- พัฒนารูปแบบของระบบแช็ตบอตตามที่ได้มีการออกแบบไว้
- พัฒนาระบบการประมวลผลการทำนายข้อมูลด้วยโมเดลต่าง ๆ
- พัฒนาระบบการจัดกลุ่มข้อมูลของลูกค้าตามที่ได้มีการออกแบบไว้

5) การทดสอบระบบ (System Testing)

- ทดสอบฟังก์ชันการทำงานของระบบแช็ตบอตให้มีการตอบสนองตรงตามวัตถุประสงค์ (Chatbot Test)

- ทดสอบฟังก์ชันการทำงานของแต่ละโมเดลให้ประมวลผลการทำนายออกมาอย่างแม่นยำสูงสุด (Prediction Model Test)
- ทดสอบฟังก์ชันการจัดกลุ่มข้อมูลรูปแบบต่าง ๆ ตามตัวแปรที่กำหนด (Clustering Model Test)
- ตรวจสอบความถูกต้องและความสมบูรณ์ของระบบ
- ปรับปรุงและแก้ไขข้อผิดพลาดที่เกิดขึ้นให้มีความสมบูรณ์ เพื่อให้ตรงตามความต้องการ

6) การจัดทำเอกสารประกอบการใช้งาน (User Manual)

- จัดทำคู่มือสำหรับการใช้งานระบบ (User Manual) ซึ่งเป็นเอกสารที่ระบุถึงขั้นตอนการใช้งานระบบที่ได้พัฒนาขึ้น เพื่อช่วยให้เกิดความเข้าใจในระบบและสามารถใช้งานได้อย่างถูกต้อง

1.5 เทคโนโลยีที่ใช้ในการพัฒนาระบบ

เทคโนโลยีที่ใช้สำหรับการพัฒนาโครงการ “ระบบแช็ตบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” มีรายละเอียดดังตารางที่ 1-1

ตารางที่ 1-1: เทคโนโลยีที่ใช้ในการพัฒนาระบบ

ด้าน Software:	
ระบบปฏิบัติการ (Operating System)	Microsoft Windows 10
ระบบจัดการฐานข้อมูล	Microsoft SQL Server 2019 Express
เครื่องมือที่ใช้ในการพัฒนาระบบแช็ตบอต	DialogFlow LINE Messaging API
เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลขั้นสูง	RapidMiner Studio Version 9.5.1
ด้าน Hardware:	
หน่วยประมวลผลกลาง (CPU)	2.30 GHz Intel® Core i5-3700U
หน่วยความจำ (Memory)	8.0 GB
หน่วยเก็บข้อมูล (Storage)	475 GB

1.6 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่คาดว่าจะได้รับจากโครงการ “ระบบแช็ตบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” มีดังนี้

- 1) สามารถเป็นหนึ่งในเครื่องมือที่ช่วยให้บริการลูกค้าจำนวนมากได้อย่างรวดเร็วและแม่นยำภายใต้ต้นทุนที่ไม่สูง ทั้งยังช่วยลดภาระงานของพนักงานได้
- 2) สามารถจัดเก็บและวิเคราะห์ข้อมูลเกี่ยวกับคำถามของลูกค้าได้อย่างครบถ้วนและเป็นมาตรฐาน ส่งผลให้ช่วยเพิ่มประสิทธิภาพในการทำงานในอนาคต

- 3) สามารถสร้างองค์ความรู้เกี่ยวกับการเพิ่มโอกาสผู้เข้าชมเว็บไซต์กวดสมัครใช้บริการของบริษัท เพื่อให้สามารถจัดทำโปรโมชั่นให้เหมาะกับลักษณะหรือความต้องการของลูกค้าได้
- 4) สามารถมองเห็นภาพรวม ลักษณะเชิงลึก และพฤติกรรมของกลุ่มผู้ที่เข้ามาใช้บริการเว็บไซต์ของบริษัท เพื่อใช้เป็นข้อมูลในการทำนายแนวโน้มการเพิ่มขึ้นของลูกค้าใหม่ วางแผนกลยุทธ์ รวมถึงเพิ่มสัดส่วนจำนวนลูกค้าใหม่ที่จะสมัครใช้บริการกับทางบริษัทได้
- 5) สามารถมองเห็นแนวโน้มที่ลูกค้าเดิมอาจเลิกใช้บริการ (Churn) และกำหนดแนวทางในการแก้ไขหรือป้องกัน เพื่อลดอัตราการเลิกใช้บริการของลูกค้าได้อย่างทัน่วงที่



บทที่ 2

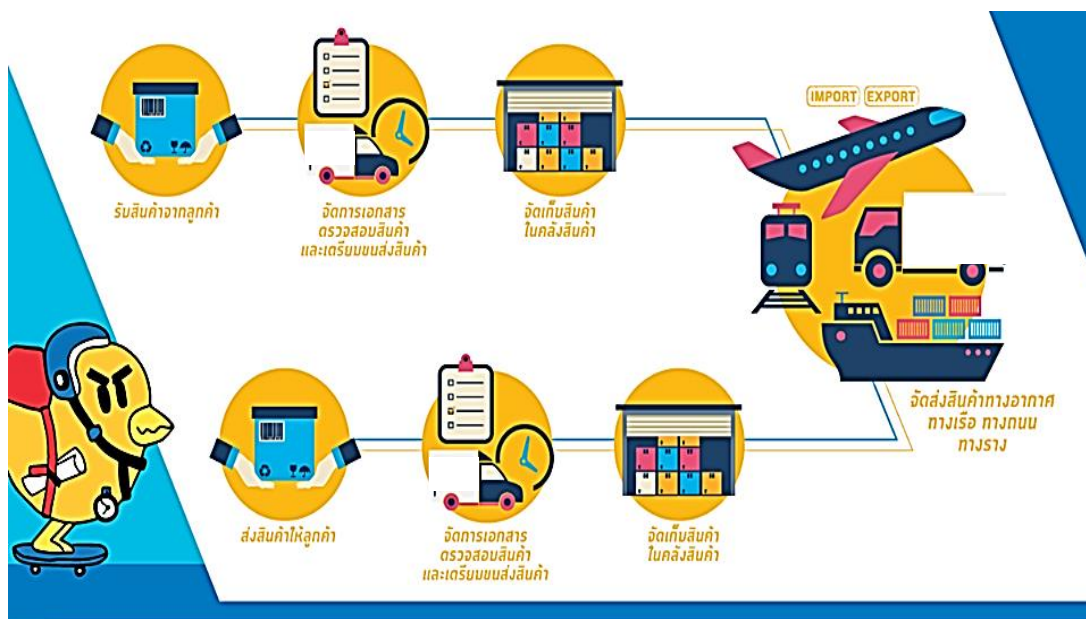
เหตุผลและแนวคิด

ในบทนี้จะกล่าวถึงแนวคิดและเทคโนโลยีสำคัญที่นำมาประยุกต์ใช้ในการพัฒนาโครงการ “ระบบแช่ตบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” ซึ่งได้แก่ แนวคิดเกี่ยวกับธุรกิจโลจิสติกส์ แนวคิดเกี่ยวกับแช่ตบอต และแนวคิดเกี่ยวกับการวิเคราะห์ข้อมูลขั้นสูง

2.1 แนวคิดเกี่ยวกับธุรกิจโลจิสติกส์ (Kiglogistics)

โลจิสติกส์หมายถึงระบบการจัดการการส่งสินค้า ข้อมูลการขนส่ง และทรัพยากรอย่างอื่นจากจุดต้นทางไปยังจุดบริโภคตามความต้องการของลูกค้า ดังนั้นโลจิสติกส์จึงมีความเกี่ยวข้องกับการผสมผสานของข้อมูล การขนส่ง การบริหารวัสดุคงคลัง การจัดการวัตถุดิบ และการบรรจุหีบห่อ หรือกล่าวได้ว่าโลจิสติกส์เป็นส่วนหนึ่งของกระบวนการจัดการห่วงโซ่อุปทาน (Supply Chain Management หรือ SCM) ที่เพิ่มมูลค่าของการใช้ประโยชน์ของเวลาและสถานที่ โดยที่มาของคำว่าโลจิสติกส์ (Kiglogistics) นั้นมาจากภาษาฝรั่งเศสคือคำว่า Logistique ที่มีรากศัพท์คำว่า โลเจอร์ (Loger) ที่หมายถึงการเก็บ โดยมีจุดเริ่มต้นมาจากการขนส่งสินค้าทางการทหาร ในการส่งกำลังบำรุง ทั้งเสบียง อาวุธ กำลังพล เพื่อสนับสนุนการรบหรือกิจกรรมที่มีการเคลื่อนย้ายและจัดเก็บจากที่หนึ่งไปยังอีกที่หนึ่ง ซึ่งอาจมีการจัดเก็บระยะเวลาสั้นหรือระยะเวลาชั่วคราว เช่น เอกสาร สินค้าสำเร็จรูป วัตถุดิบ และอื่น ๆ (Topten2U, 2556) ส่วนคำว่า Logistics ในภาษาอังกฤษก็ต้องลงท้ายด้วย “s” เสมอเพราะเป็นศาสตร์ ไม่ใช่เติม s เพราะเป็นพหูพจน์ ดังนั้นไม่ว่าจะเปิดหลักสูตรเพียงหลักสูตรเดียวหรือเขียนหนังสือเพียงเล่มเดียวเกี่ยวกับโลจิสติกส์ก็ต้องมี s เสมอ (อัศม์เดช วานิชชินชัย, 2553)

กล่าวโดยสรุปคือแม้ว่าหากพูดถึงโลจิสติกส์แล้วหลายคนอาจนึกถึง “ระบบการขนส่งทางถนน” เป็นอย่างแรก แต่ในความเป็นจริงแล้วกิจกรรมโลจิสติกส์นั้นครอบคลุมตั้งแต่การจัดการคลังสินค้า การขนส่งทางอากาศ ทางเรือ ทางถนน ทางราง กิจกรรมนำเข้าและส่งออก การเคลียร์ภาษีนำเข้า และระบบไอทีในการบริหารสต็อกสินค้า เป็นต้น ดังจะเห็นได้จากรูปที่ 2-1



รูปที่ 2-1: กิจกรรมโลจิสติกส์

ที่มา: โลจิสติกส์ คืออะไร ใครรู้บ้าง, โดย เจตต์เบ็ญญูดี อินโฟโลจิสติกส์, 2561,

<https://www.facebook.com/JWDGroup.2014/posts/1686661461417922/>

ข้อแตกต่างระหว่างบริษัท Shipping และ Freight Forwarder ในแวดวงโลจิสติกส์

Freight Forwarder คือตัวแทนหรือตัวกลางระหว่าง ผู้นำเข้า-ส่งออกสินค้าข้ามพรมแดน โดยจะทำหน้าที่ในการนำสินค้าไปดำเนินการจัดส่งระหว่างประเทศ ซึ่งในการขนส่งสินค้า สามารถขนส่งได้หลายช่องทางไม่ว่าจะเป็นทางเรือ ทางรถบรรทุก หรือทางอากาศด้วยเครื่องบิน แต่ทั้งนี้ Freight Forwarder จะไม่ใช่เจ้าของเรือหรือเครื่องบิน กล่าวคือ เมื่อสินค้ามาถึงท่าเรือขนถ่ายสินค้า ปลายทาง Freight Forwarder จะทำหน้าที่เป็นคนกลางช่วยติดต่อประสานงาน รับสินค้าไปส่งมอบให้กับผู้รับสินค้าเพื่อให้การนำเข้าส่งออกนั้นเป็นไปด้วยความราบรื่นมากที่สุด

Freight Forwarder เป็นบริการแบบ One Stop Service ซึ่งมีลักษณะการให้บริการดังต่อไปนี้ (Kiglogistics, 2561)

- 1) จัดการขนส่งสินค้าระหว่างประเทศ
- 2) ดำเนินพิธีการศุลกากรแทนผู้นำเข้าและส่งออก
- 3) ทำหน้าที่เป็นผู้จัดการจองการระวางเรือ
- 4) ส่งมอบสินค้าจากต่างประเทศให้ถึงมือผู้นำเข้า – ส่งออก

5) ให้บริการรับส่งสินค้าด้วยรถบรรทุกรวมถึงบริการเก็บรักษาสินค้า บรรจุภัณฑ์ รวบรวม และกระจายสินค้าให้แก่ลูกค้า

6) ช่วยเหลือในการแก้ปัญหาเฉพาะหน้า รวมถึงติดต่อเครือข่ายที่มีอยู่เพื่อดำเนินการจัดหา อุปกรณ์ เครื่องมือ บุคลากร และจัดการกับปัญหาในการขนส่ง

แม้ว่าหน้าที่หลักของ Freight Forwarder คือการให้บริการช่วยเหลือด้านการนำเข้า-ส่งออกสินค้า แต่ก็เป็นที่รู้จักประเภทที่สามารถช่วยจัดการทุกขั้นตอน ทุกเรื่องที่เกี่ยวข้องกับการนำเข้าและส่งออกสินค้าไปจนถึงจุดหมายปลายทางได้อย่างราบรื่น เรียบร้อย และรวดเร็ว โดยบริการของ Freight Forwarder ถือเป็นตัวช่วยที่ดียิ่งสำหรับธุรกิจส่งออกและนำเข้า เพราะช่วยลดต้นทุนในการซื้อหรือเช่าตู้คอนเทนเนอร์ของตนเอง ไม่ต้องจ้างพนักงานหลายคนเพื่อมาทำงาน ตั้งแต่การขนส่งสินค้าและการจัดการเอกสารนั่นเอง

อย่างไรก็ตาม คำว่า “Freight Forwarder” กับ “Shipping” สองธุรกิจนี้มีความคล้ายคลึงกันอยู่ แต่ก็ได้ต่างไปเสียทีเดียว โดย Shipping ก็คือผู้ทำหน้าที่ด้านพิธีการเอกสารศุลกากร หรืออาจเรียกว่าตัวแทนออกของ ซึ่งจะทำหน้าที่แทนผู้นำเข้าในการติดต่อหน่วยงานต่าง ๆ ที่เกี่ยวข้อง เช่น กรมศุลกากร ธนาคาร บริษัทขนส่งสินค้า รถขนส่งสินค้าในต่างประเทศ รถลาก โกดัง ฯลฯ เป็นตัวแทนติดต่อรับส่งสินค้าระหว่างสายการบินและการเดินเรือ หรือ “มีเรือเดินทะเลและตู้คอนเทนเนอร์เป็นของตนเอง” ทำหน้าที่จัดการเอกสารผ่านธนาคาร ใบขนส่งสินค้าให้กับด้านศุลกากร เดินเอกสารของบริษัทเรือเดินทะเล และบริษัทขนส่งทางเครื่องบิน ซึ่งลูกค้าส่วนใหญ่มักเป็นลูกค้ารายใหญ่ ในขณะที่ Freight Forwarder (ตัวแทนของผู้ส่งสินค้า) “ไม่มีเรือเดินทะเลและตู้คอนเทนเนอร์เป็นของตนเอง” แต่ไปเข้ามาเพื่อให้บริการในการนำเข้าและส่งออกสินค้า ผ่านสายการบินและการเดินเรือ ส่วนการทำหน้าที่ต่าง ๆ ก็แทบจะไม่ต่างกัน เพียงแต่ลูกค้าของ Freight Forwarder มักเป็นลูกค้ารายเล็กและมีหลาย ๆ รายมารวมกันเพื่อใช้บริการตู้คอนเทนเนอร์สำหรับหนึ่งตู้ ซึ่งทำให้มีค่าใช้จ่ายสูงกว่า Shipping

เกณฑ์ในการใช้พิจารณาบริษัทขนส่งต่างประเทศหรือบริษัทโลจิสติกส์ที่ดีที่สุด

เกณฑ์ที่สำคัญมีดังนี้ (Kongrath, 2560)

1) บริษัทต้องมีการตอบสนองอย่างรวดเร็ว

บริษัทชิปปิงต้องมีความเต็มใจที่จะช่วยเหลือและให้บริการผู้ใช้หรือลูกค้าเป็นสิ่งสำคัญ เนื่องจากการขนส่งมักจะมีปัญหาอยู่บ่อยครั้ง บริษัทต้องสามารถจัดการปัญหา

อย่างทันท่วงที อาทิเช่น พนักงานมีความพร้อมในการให้บริการ มีความรวดเร็วในการขนส่ง สามารถจัดส่งให้อยู่ในช่วงเวลาเร่งด่วนตามที่ลูกค้ากำหนดหรือหลังเวลาทำการ มีการติดต่อกลับเพื่อยืนยันการสั่งซื้อ สามารถตอบคำถามของลูกค้าได้ว่าสินค้าอยู่ที่ไหน

2) มีความเชี่ยวชาญสูงและต้องสามารถให้ความมั่นใจลูกค้าได้

สำหรับธุรกิจด้านการขนส่งนั้นนอกจากเรื่องของการตอบสนองอย่างรวดเร็วแล้ว ยังต้องมีเรื่องการให้ความมั่นใจในการให้บริการกับลูกค้าด้วย เช่น การให้บริการเกี่ยวกับการป้องกันสินค้าเพื่อไม่ให้เกิดการชำรุด โดยอาจมีอุปกรณ์ยึด รััด ห่อหุ้มสินค้า และมีการติดชื่อผู้รับสินค้า นอกจากนี้ยังมีเรื่องของการจัดเรียงสินค้าให้เป็นระเบียบเพื่อป้องกันความผิดพลาด เสียหาย และทำให้ขนส่งได้สะดวกปลอดภัย ส่วนพนักงานก็ต้องมีความรู้และมนุษยสัมพันธ์ในฐานะของผู้ให้บริการ ทำให้ผู้ใช้บริการมีความเชื่อมั่นในบริการที่ได้รับ รวมถึงความมีมารยาท ความน่าศรัทธา และความปลอดภัย ที่ผสมรวมเข้าไว้ด้วยกัน มีทักษะและปฏิบัติตามมาตรฐานการจัดส่งสินค้า เช่น วิธีการยกหรือขนถ่ายสินค้าแต่ละประเภท การดูแลสินค้าระหว่างการขนส่ง เป็นต้น และมี Certificate เพื่อรับรองมาตรฐานหรือประสิทธิภาพการทำงาน

3) อุปกรณ์และสิ่งอำนวยความสะดวกต่าง ๆ ต้องพร้อม

สิ่งที่ปรากฏให้เห็นหรือสิ่งที่จับต้องได้ เช่น วัสดุ อุปกรณ์ เครื่องมือ อาคาร สถานที่ บุคคล และสิ่งอำนวยความสะดวก ตัวอย่างเช่น พนักงานแต่งกายสุภาพ เรียบร้อยเหมาะสม มีจำนวนพนักงานที่เหมาะสมกับปริมาณสินค้าที่จัดส่ง มีอุปกรณ์และเครื่องมือที่จำเป็นพร้อมสำหรับการให้บริการขนส่งสินค้า รถส่งสินค้ามีความสะอาด เรียบร้อย อยู่ในสภาพพร้อมใช้งาน ไม่ก่อให้เกิดปัญหาระหว่างจัดส่ง และเอกสารที่ใช้ในการจัดส่งสินค้ามีความถูกต้องและครบถ้วนตามระเบียบข้อบังคับของบริษัท

4) มีความถูกต้องแม่นยำในการให้บริการ

บริษัทต้องมีความสามารถที่จะแสดงผลงานหรือความสามารถในการให้บริการได้ตามที่สัญญาไว้อย่างมีคุณภาพ ถูกต้องแม่นยำ และน่าเชื่อถือ ทำให้สินค้าอยู่ในสภาพเรียบร้อย หีบห่อหรือบรรจุภัณฑ์ไม่ชำรุด ฉีกขาด หรือเสียหาย มีการรับผิดชอบเพื่อเปลี่ยนสินค้าหรือรับสินค้าตามระยะเวลาที่กำหนด ในกรณีที่สินค้ามีปัญหาจากการจัดส่งสินค้า และตรงตามระยะเวลาที่กำหนด

5) มีการดูแลเอาใจใส่ในงานและคำนึงถึงผลประโยชน์ของลูกค้า

การดูแลเอาใจใส่และความสนใจที่ผู้ให้บริการมีต่อผู้ใช้บริการ รวมทั้งเป็นการรวมปัจจัยในการเข้าถึงบริการ การสื่อสาร และความเข้าใจเข้าด้วยกัน ได้แก่ การให้บริการต้องคำนึงถึงผลประโยชน์สูงสุดของลูกค้า โดยนำการสื่อสารทางอินเทอร์เน็ตเข้ามาเกี่ยวข้อง เพื่อให้ง่ายต่อการเข้าถึงบริการของลูกค้า

ห้าแนวโน้มเทคโนโลยีธุรกิจขนส่งและโลจิสติกส์ ในปี พ.ศ. 2563

อุตสาหกรรมขนส่งและโลจิสติกส์เป็นอุตสาหกรรมที่มีการแข่งขันของผู้ประกอบการอย่างเข้มข้น และยังมีการนำเทคโนโลยี เช่น หุ่นยนต์ ระบบอัตโนมัติ ปัญญาประดิษฐ์ และโมบายล์แอปพลิเคชันแบบเรียลไทม์ มาเป็นตัวช่วยบริหารจัดการเพิ่มประสิทธิภาพงานขนส่งและโลจิสติกส์ให้สอดคล้องกับกระแสการปรับเปลี่ยนสู่สังคมเศรษฐกิจดิจิทัล สิ่งสำคัญที่จำเป็นต้องตระหนักเพื่อรับความเปลี่ยนแปลงหรือเรียกได้ว่าเป็นหัวใจสำคัญของธุรกิจขนส่งและโลจิสติกส์ นั่นคือ “บริหารต้นทุนให้ต่ำ ทำให้เร็วกว่าเดิม และเพิ่มการลงทุนในเทคโนโลยี”

โดยสิ่งที่ผู้ประกอบการต้องคำนึงถึงคือความเร็วในการจัดส่งสินค้า การส่งมอบสินค้าตรงเวลา สินค้าอยู่ในสภาพสมบูรณ์ และพนักงานให้บริการอย่างมีอาชีพ เพื่อให้ผู้รับสินค้ามีความพึงพอใจมากที่สุด ขณะที่ธุรกิจขนส่งและโลจิสติกส์ไทยที่กำลังเติบโตท่ามกลางการแข่งขันที่ดุเดือดได้เร่งปรับตัวนำเทคโนโลยีเข้ามาบริหารงานมากขึ้น การแข่งขันจึงไม่ใช่แค่การใช้เทคโนโลยี แต่คือการลงทุนเทคโนโลยีด้วยการอ่านเกมให้ขาด ลงทุนให้ถูกจุด เพื่อครองชัยชนะในสมรภูมิยุค Digital Disruption

ผู้อำนวยการส่วนผลิตภัณฑ์เชิงพาณิชย์ บริษัท จีไอเอส จำกัด ในเครือ CDG Group ผู้ให้บริการด้านระบบเทคโนโลยีสารสนเทศแบบครบวงจรแก่องค์กรภาครัฐ รัฐวิสาหกิจ และเอกชน ดำเนินธุรกิจมายาวนานกว่า 50 ปี ได้เปิดเผยถึง 5 แนวโน้มความท้าทายของโลกขนส่งและโลจิสติกส์ที่ต้องเผชิญในปี 2563 นี้ ประกอบด้วย (Pp, 2563)

1) Digital Logistics

การบริหารงานโลจิสติกส์ด้วยข้อมูลดิจิทัล โดยมีข้อมูลเป็นตัวแปรสำคัญที่ทำให้ผู้ประกอบการสามารถเอาชนะคู่แข่งได้ ดังนั้นการปรับเปลี่ยนกระบวนการทำงานเข้าสู่ระบบดิจิทัลเป็นเรื่องที่เลี่ยงไม่ได้อีกต่อไป เพราะดิจิทัลมีผลต่อการทำงานแบบ Day-to-

Day ในธุรกิจ ข้อมูลในระบบดิจิทัลช่วยสนับสนุนการตัดสินใจได้ดีและเร็วขึ้น ทำให้มองเห็นซัพพลายเชนทั้งระบบ และสามารถควบคุมการเปลี่ยนแปลงที่รวดเร็วในทุกวันนี้ได้ง่ายขึ้น

2) Real Time Supply Chain Visibility

ในการจัดการห่วงโซ่อุปทานแบบทันทีที่ชัดเจนได้นั้น เจ้าของธุรกิจจำเป็นต้องรู้ข้อมูลแบบเรียลไทม์ ว่าสินค้าจะอยู่ที่ไหน จำนวนเท่าไร เคลื่อนย้ายเมื่อใด ตลอดจนการใช้ IoT Sensor เพื่อการติดตามการขนส่ง เช่น เส้นทาง จุดส่งสินค้า เงื่อนไขพิเศษต่าง ๆ เพื่อตอบสนองความต้องการของลูกค้าให้เร็วที่สุดและควบคุมค่าใช้จ่ายภายในธุรกิจ การเชื่อมโยงข้อมูลเหล่านี้ในกระบวนการธุรกิจจะช่วยสนับสนุนการตัดสินใจและการคาดการณ์อนาคต เพื่อปรับเปลี่ยนการจัดการคำสั่งซื้อและห่วงโซ่อุปทานให้เกิดประสิทธิภาพสูงสุด

3) Consolidation of Goods

เป็นการรวบรวมสินค้าจากหลายบริษัทในการบรรทุกจัดส่งไม่ว่าทางบก ทางน้ำ หรือทางอากาศ หรือเรียกได้ว่าเป็นการจัดการขนส่งที่จะเกิดขึ้น เพื่อให้ประหยัดค่าใช้จ่ายในการขนส่งและช่วยให้กระบวนการโลจิสติกส์มีเสถียรภาพอย่างยั่งยืน ปัจจัยหลักคือขนาดของตลาดขนส่งที่เปลี่ยนแปลงอย่างรวดเร็วเช่นเดียวกับขนาดของการบรรทุกจัดส่ง เมื่อทุกสิ่งสามารถสั่งซื้อได้แบบออนไลน์ ส่งผลให้เกิดการเพิ่มขึ้นของจำนวนสินค้าที่จะถูกจัดส่งจำนวนมาก ขนาดของการไหลดบรรทุกจัดส่งของบริษัทจึงเล็กลง เพื่อให้สามารถส่งออกสินค้าได้รวดเร็วมากขึ้น

4) Artificial and Augmented Intelligence

AI หรือ ปัญญาประดิษฐ์ กำลังจะมาแทนที่งานบางส่วนของมนุษย์ที่ทำซ้ำ ๆ ซึ่งสามารถถูกแทนที่ด้วยระบบ เพราะเป็นระเบียบและวัดผลได้มากขึ้น แต่งานบางประเภทเป็นงานที่จำเป็นต้องใช้ทักษะของมนุษย์ จึงเกิดเทรนด์ใหม่ที่เรียกว่า “Augmented Intelligence” หรือ “ปัญญาเสริม” ซึ่งเป็นแนวคิดที่ว่า “มนุษย์” กับ “AI” สามารถทำงานร่วมกันได้ คือการเลียนแบบความฉลาดของมนุษย์และขยายขีดความสามารถทางปัญญาของมนุษย์ลงในซอฟต์แวร์ เช่น หน่วยความจำ การจัดลำดับ การรับรู้ การคาดการณ์ การแก้ปัญหา ไปจนถึงการตัดสินใจ เพื่อให้สามารถเข้าใจความต้องการของ

มนุษย์ได้ง่ายขึ้นและเพิ่มความสามารถในการเชื่อมโยงข้อมูลต่าง ๆ เข้าด้วยกัน ตัวอย่างเช่น การวางแผนงานโลจิสติกส์ สามารถใส่ข้อมูลที่มีมนุษย์เป็นผู้สร้าง เช่น แผนงาน ความรับผิดชอบของส่วนงานการบริการลูกค้า ความยืดหยุ่นในการทำงาน ซึ่งต้องใช้ประสบการณ์ ความรู้สึกหรือสามัญสำนึก และอื่น ๆ ผสานเข้ากับการวิเคราะห์ ประเมินผลโดย AI ที่สามารถทำนายและทำความเข้าใจสิ่งที่จะเกิดขึ้นในอนาคต ช่วยวิเคราะห์ผลลัพธ์ของการตัดสินใจจากการเรียนรู้ข้อมูลในอดีต อะไรจะเกิดขึ้นหากเลือกเดินในแต่ละเส้นทาง ทำให้มีวิธีรับมือกับเหตุการณ์เหล่านั้นได้

5) Data Standardization and Predictive Analytics Platform

เทคโนโลยีแพลตฟอร์มสำหรับการวิเคราะห์ข้อมูลความเสี่ยงและจัดการข้อมูลให้ เป็นหนึ่งเดียว คือ เครื่องมือที่จะช่วยรวบรวมข้อมูลโลจิสติกส์ขนาดใหญ่ในโลกดิจิทัล สิ่ง ที่ธุรกิจโลจิสติกส์ควรทำคือ การศึกษากระบวนการทำงานของโซลูชันแพลตฟอร์ม และ นำมาใช้กับธุรกิจเพื่อเพิ่มประสิทธิภาพการจัดการงานขนส่งได้ดีขึ้น เพิ่มสมรรถนะการใช้ รถขนส่งได้มากขึ้น และจัดการการใช้เชื้อเพลิงการขนส่งได้น้อยลง ข้อมูลและการ วิเคราะห์พยากรณ์ด้วยแพลตฟอร์มจะทำให้เห็นรูปแบบของความเสี่ยง รวมถึงโอกาสที่ อาจเกิดขึ้นด้วยข้อมูลที่มีในระบบ และอาจใส่เงื่อนไขเฉพาะลงไป ทำให้ได้ข้อมูล พยากรณ์ล่วงหน้า เช่น อุบัติเหตุ การใช้น้ำมัน การซ่อมบำรุงรถ การใช้จ่าย เส้นทาง ขนส่งและจุดจอดรถที่เหมาะสม

2.2 แนวคิดเกี่ยวกับแชทบอต

แชทบอตคือโปรแกรมคอมพิวเตอร์ชนิดหนึ่ง ซึ่งถูกพัฒนาขึ้นมาให้มีบทบาทในการตอบกลับ การสนทนาด้วยตัวอักษรแบบอัตโนมัติผ่าน Messaging Application เสมือนการโต้ตอบของคนจริง ๆ หรืออาจเรียกง่าย ๆ ว่า โปรแกรมตอบกลับอัตโนมัติ ซึ่งเวลานี้กลายเป็นสุดยอดผู้ช่วยอัจฉริยะที่ทุก บริษัทต้องการนำมาใช้กับธุรกิจออนไลน์ในการสื่อสารกับกลุ่มลูกค้าแบบเรียลไทม์ (Krungsri Plearn Plearn, 2560)

แชทบอต หรือ Chatbot นั้นย่อมาจาก Chat Robot ซึ่งเป็นโปรแกรมตอบกลับการสนทนา อัตโนมัติ ภายใต้การใช้ระบบ Database บันทึกคำถามและคำตอบ ตรวจสอบ Keyword จากคำถาม เพื่อประมวลคำตอบส่งกลับไปหาลูกค้า (Rule-Based Chatbot) ซึ่งถูกนำมาประยุกต์ใช้ในด้านการ สอบถามข้อมูลหรือร้องเรียนปัญหา แต่ยังคงมีความเป็นหุ่นยนต์ให้ผู้บริโภคจับสังเกตได้

ตัวอย่างของแชทบอตที่มักพบได้บ่อย

สำหรับแชทบอตรูปแบบหลัก ๆ ที่พบได้บ่อย 2 รูปแบบ ได้แก่ แบบที่ถูกกำหนดด้วยกฎต่าง ๆ (Based on Rules) กับแบบที่มีปัญญาประดิษฐ์ AI (Artificial Intelligence) คือใช้ Machine learning เข้ามาช่วย (Coraline Team, 2562a)

1) Rule-Based Bot

เป็นระบบตอบรับอัตโนมัติที่มีการกำหนดการถามตอบชัดเจน เช่น พิมพ์ 1 เพื่อเลือกหัวข้อการซื้อผลิตภัณฑ์ พิมพ์ 2 เพื่อเลือกหัวข้อการบริการหลังการขาย นอกจากนี้เป็นตัวเลข ยังสามารถใช้คำหรือ ประโยคเป็นตัว Input ได้ แต่ต้องพิมพ์ให้ถูกต้องเท่านั้น แชทบอตแบบ Rule-Based จึงไม่ใช่ AI แต่เป็นแชทบอตที่ถูกพัฒนาขึ้นโดยใช้กฎต่าง ๆ เป็นตัวตั้งต้นคำสั่งในการสื่อสาร โดยการสร้าง กฎ หรือคีย์เวิร์ด ลงในระบบ และคำตอบที่ตรงกับคีย์เวิร์ดนั้น ๆ ถ้าหากคำถามที่ถูกคำถามตรงกับคีย์เวิร์ดใด ระบบก็จะตอบคำถามตามที่ได้ถูกกำหนดไว้นั่นเอง ซึ่งระบบแชทบอตแบบนี้จะมีข้อเสียตรงที่ผู้สร้างจะต้องทำคีย์เวิร์ดและการตอบคำถามไว้หลาย ๆ กรณี หากผู้ใช้ตอบกลับแล้วไม่ตรงกับคำสั่งที่กำหนดไว้ แชทบอตก็อาจไม่เข้าใจว่าผู้ใช้ต้องการอะไร และไม่สามารถตอบคำถามผู้ใช้ได้อย่างถูกต้องนั่นเอง ดังนั้นหากต้องการให้แชทบอตแบบที่ใช้กฎ (Based on Rules) มีความเก่ง จึงอาจจะต้องสร้างกฎไว้ให้หลาย ๆ ข้อ เพื่อให้ครอบคลุมหลาย ๆ กรณีและตรงตามเป้าหมายที่ต้องการให้แชทบอตเป็น เพราะแชทบอตแบบนี้จะสามารถโต้ตอบได้เฉพาะคำสั่งที่ได้สร้างขึ้นไว้ตามกฎ หากผู้ใช้ตอบกลับในบางคำสั่งที่ไม่ได้เตรียมไว้แชทบอตเองก็อาจไม่เข้าใจว่าผู้ใช้ต้องการอะไร

2) AI Bot

เป็นการนำการเรียนรู้ของเครื่อง หรือ ML (Machine Learning) เข้ามาพัฒนาให้แชทบอตมีความฉลาดเหมือนมนุษย์ สามารถโต้ตอบและสื่อสารกับผู้ใช้งานได้อย่างเข้าใจมากยิ่งขึ้น ซึ่งก็มีความยากในการพัฒนามากกว่า แต่ก็ก็เป็นแบบที่นิยมกันมากในบริษัทใหญ่ ๆ ที่พัฒนาระบบแชทบอตอย่าง IBM, Microsoft, และ Google เป็นต้น ซึ่งแบบที่ใช้ AI นั้นจะมีความยากในการพัฒนามากกว่า เพราะอาจจะต้องมีการนำการประมวลผลภาษาธรรมชาติ หรือ NLP (Natural Language Processing) และการทำความเข้าใจภาษาธรรมชาติ หรือ NLU (Natural Language Understanding) มาใช้

เพื่อช่วยให้ Chatbot สามารถเข้าใจภาษามนุษย์ รูปประโยค และความหมายที่มนุษย์ต้องการสื่อได้ดีขึ้น ซึ่งตอนนี้ก็มีบริษัทใหญ่ ๆ หลายราย ที่ได้พัฒนาเทคโนโลยี NLP-NLU ไม่ว่าจะเป็น IBM, Microsoft, Google หรือแม้กระทั่ง Facebook เอง

— NLP-Based Chat-Bot

NLP ย่อมาจาก Natural Language Processing ถือเป็น AI ประเภทหนึ่ง ที่เรียนรู้การใช้ภาษา และมีการตอบสนองจากการใส่ Input ประเภทข้อความ ทำให้ระบบการเรียนรู้สามารถเข้าใจประโยคได้ จากการให้คะแนน "คำ" หรือ "ตัวอักษร" ต่าง ๆ ทำให้เมื่อนำคำนั้น ๆ มาเรียงกันเป็นประโยค ก็สามารถทำความเข้าใจประโยคนั้นได้ การทำงานของแชตบอตคือ ทำความเข้าใจ "ประโยค" ที่พิมพ์เข้าไป โดยไม่จำเป็นต้องมีการจดจำประโยคดังกล่าวเอาไว้ก่อน ต่อมาแชตบอตจะส่งคำตอบที่มีการกำหนดเอาไว้ เช่น ถ้าคำถามเกี่ยวข้องกับสภาพอากาศ แชตบอตจะทำการดึงข้อมูลสภาพอากาศมาแสดงผล เป็นต้น อย่างไรก็ตามข้อจำกัดของการเรียนรู้คือจะต้องมีการใส่ Input เป็น "คำศัพท์ต่าง ๆ" เพื่อเป็นตัวตั้งต้นให้แชตบอตเรียนรู้ และกำหนดผลลัพธ์ที่เป็นคำตอบเอาไว้ เนื่องจากแชตบอตจะตอบคำถามเท่าที่ได้กำหนดกรอบเอาไว้เท่านั้น

— NLP แบบ Dynamic Learning

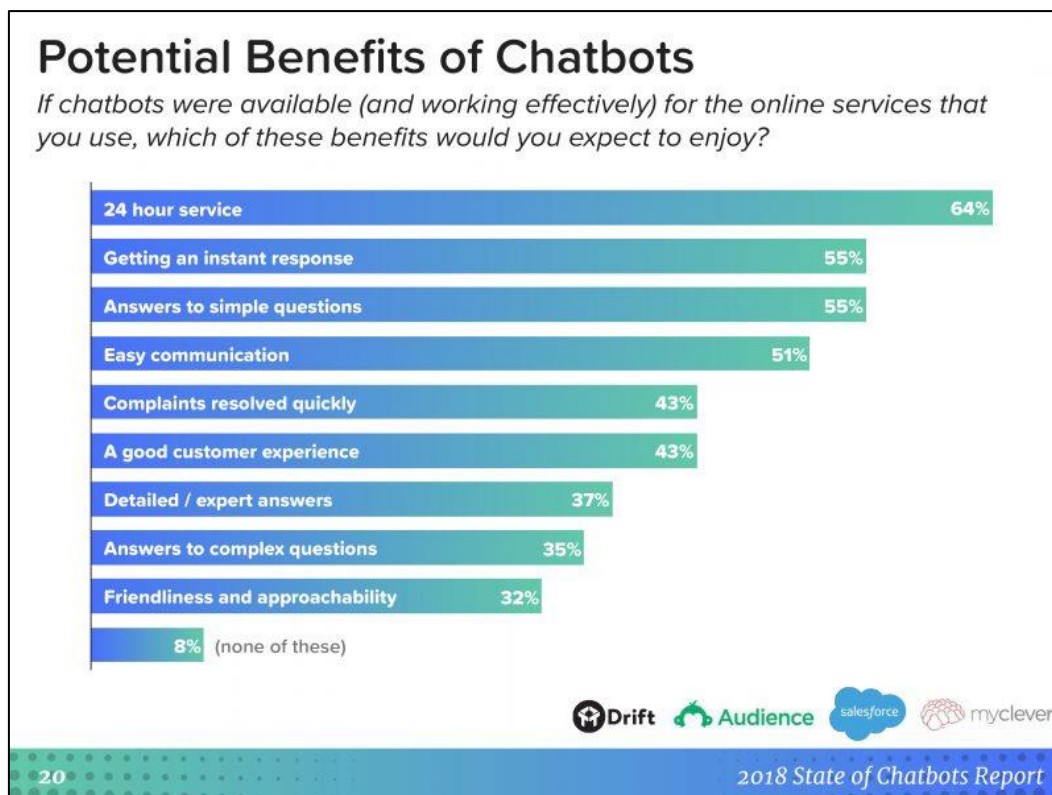
เป็นการต่อยอดจาก NLP แบบเดิมตรงที่ตัว AI จะเก็บข้อมูลถามตอบที่เกิดขึ้นใหม่ไปเป็นต้นแบบในการเรียนรู้ต่อไป ดังนั้นในกรณีที่ผู้ใช้เกิดการพิมพ์ผิดหรือพิมพ์ตกหล่น AI จะสามารถเข้าใจรายละเอียดได้ และการเก็บรายละเอียดของ AI ทำได้โดยตั้งคำถามเพื่อให้ผู้ตอบยืนยันว่า AI เข้าใจถูก จากนั้นตัว AI จะค่อย ๆ เรียนรู้ประโยคใหม่ ๆ ทั้งนี้สิ่งสำคัญในการพัฒนาแชตบอตคือข้อมูลที่เป็นตัวตั้งต้น และคำตอบสำหรับแต่ละโจทย์ ดังนั้นแชตบอตของแต่ละอุตสาหกรรมจะมีตัวตั้งต้นที่ไม่เหมือนกัน ซึ่งการจะเลือกให้แชตบอตเป็น AI หรือไม่ขึ้นอยู่กับแนวทางการใช้งาน หากต้องการใช้เพื่อตอบคำถามที่เรียบง่ายสามารถใช้แบบ Rule-Based ได้ แต่หากต้องการแชตบอตขั้นสูงก็อาจต้องพัฒนาเป็น Dynamic AI และควรมีผู้เชี่ยวชาญทำหน้าที่เป็นผู้ดูแลรักษาระบบ

แช็ตบอตคืออะไร

ด้วยความที่แช็ตบอตนั้นสามารถประยุกต์ให้เข้ากับหลากหลายแพลตฟอร์มและหลากหลายธุรกิจได้ จึงเปรียบเสมือนตัวช่วยที่ดีมาก ๆ ตัวหนึ่งที่จะคอยช่วยเหลือลูกค้าหรือผู้ใช้งานแพลตฟอร์มและธุรกิจ (Anak Mirasing, 2560) ตัวอย่างเช่น

- คอยตอบปัญหาเกี่ยวกับการใช้งาน
- นำเสนอข่าวสาร โปรโมชัน จากร้านค้าหรือสำนักข่าวต่าง ๆ
- ช่วยในการปิดการขายสินค้า คอยแนะนำลูกค้า สร้างบิล และหน้าจ่ายเงิน
- ให้คำปรึกษาในด้านต่าง ๆ (เช่น ในประเทศอังกฤษซึ่งมีแช็ตบอตที่ให้คำปรึกษาทางกฎหมายเกี่ยวกับที่จอดรถและที่อยู่อาศัย)
- ช่วยดูแลลูกค้า เช่น ในธุรกิจการบิน ซึ่งลูกค้าสามารถดูตาราง เช็คอิน เที่ยวบินและข้อมูลอื่น ๆ ได้
- ช่วยปรับปรุงสินค้า เช่น แช็ตบอตที่ช่วยคุยกับลูกค้าเพื่อหา Feedback ว่าเบียร์ที่ลูกค้าดื่มมีรสชาติดีหรือไม่ดี กลิ่นเป็นอย่างไร ราคาเป็นอย่างไร ซึ่งเจ้าของกิจการก็สามารถเก็บเอาข้อมูลส่วนนี้มาพัฒนารสชาติและราคาสินค้าได้ในทันที
- ผู้ช่วยด้านการเงิน อย่างของบริษัท American Express ซึ่งได้ปล่อย Amex Bot ที่คอยช่วยเหลือเรื่องของสิทธิประโยชน์ แจ้งการทำธุรกรรมหรือแม้แต่อัปเดตทางการเงิน ทำให้ผู้ใช้บัตรเครดิตได้รับความสะดวกเพิ่มขึ้นอีกขั้น

นอกจากนี้แช็ตบอตยังมีความสามารถและประโยชน์อีกมากมายดังจะเห็นได้จากรูปที่ 2-2



รูปที่ 2-2: ประโยชน์ของแชทบอต

ที่มา: Chatbot พนักงานคนใหม่ เปลี่ยนโลกการสื่อสารกับผู้บริโภค, โดย OPEN-TEC, 2561,

<https://www.open-tec.com/th/chatbot->

%E0%B8%9E%E0%B8%99%E0%B8%B1%E0%B8%81%E0%B8%87%E0%B8%B2%E0%B8%99%E0%B8%84%E0%B8%99%E0%B9%83%E0%B8%AB%E0%B8%A1%E0%B9%88/

2.3 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูลขั้นสูง (Advanced Data Analytics) หรือการทำเหมืองข้อมูล (Data Mining)

การวิเคราะห์ข้อมูลขั้นสูง (Advanced Data Analytics) หรือการทำเหมืองข้อมูล (Data Mining) เป็นกระบวนการในการค้นหาความผิดปกติ รูปแบบ และความสัมพันธ์บางอย่างที่ซ่อนอยู่ภายในชุดข้อมูลขนาดใหญ่ (Big Data) ซึ่งไม่สามารถสังเกตได้จากการดูข้อมูลเพียงอย่างเดียว เนื่องจากข้อมูลมีปริมาณมาก เช่น การค้นหาความสัมพันธ์ (Association Rules) ของสินค้าในห้างสรรพสินค้า อาจทำให้พบว่าลูกค้าร้อยละ 90 ที่ซื้อเปียร์ จะซื้อผ้าอ้อมเด็กด้วย ซึ่งเป็นข้อมูลที่จะช่วยให้ทางห้างสามารถวางแผนรายการส่งเสริมการขายใหม่ ๆ ได้ เป็นต้น

การคาดการณ์ผลลัพธ์โดยใช้เทคนิคที่หลากหลาย ไม่ว่าจะเป็นการจำแนกประเภท รูปแบบ เชื่อมโยงข้อมูลที่มีความสัมพันธ์กัน และหาความน่าจะเป็นที่จะเกิดขึ้น จะช่วยให้ผู้ใช้สามารถนำข้อมูล นี้ไปเพิ่มรายได้ ลดต้นทุน พัฒนาความสัมพันธ์กับลูกค้า ลดความเสี่ยง และอื่น ๆ อีกมากมาย อีกทั้ง องค์ความรู้ใหม่ที่เกิดขึ้นยังสามารถนำไปใช้ประกอบการตัดสินใจในด้านต่าง ๆ เช่น ตลาดหลักทรัพย์ กลยุทธ์ทางธุรกิจ ทางด้านการแพทย์ ยุทธศาสตร์ทางการทหาร ฯลฯ

การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้ จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล โดยสามารถแยกเป็นหัวข้อต่าง ๆ ได้ดังนี้ (SAS Institute Inc., 2560)

- กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง
- การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงิน หรือการนำข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัย มาใช้งานในด้านวิทยาศาสตร์
- การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล
- การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ในเชิงสถิติและตรรกะของข้อมูลขนาดใหญ่ เป็นการมองหารูปแบบที่สามารถช่วยในการตัดสินใจได้

กระบวนการทำงาน

กระบวนการทำงานทางด้านวิทยาศาสตร์ข้อมูลเป็นการนำขั้นตอน วิธีการ (Processes & Methods) มาใช้เพื่อหาคูณค่าของข้อมูลซึ่งไม่สามารถทำได้โดยผ่านการอ่าน การมองเห็น หรือการคิดวิเคราะห์ตามตรรกะพื้นฐาน หรืออาจเรียกบริการนี้ว่าเป็นงานบริการวิเคราะห์ข้อมูลขั้นสูง (Advanced Analytics) หรือการทำเหมืองข้อมูล (Data Mining) โดยมีขั้นตอนการทำงานอย่างเป็นระบบดังนี้ (G-able, 2561)

- 1) การหาเป้าหมายหรือจุดประสงค์ของการวิเคราะห์
- 2) การรวบรวมข้อมูลที่เกี่ยวข้องกับการวิเคราะห์รวมถึงการปรับปรุงข้อมูลให้อยู่ในรูปแบบที่เหมาะสม
- 3) การตั้งสมมติฐานผ่านวิธีการทางสถิติ (Statistical Models)

- 4) การพิสูจน์ ทดสอบผลลัพธ์จนกว่าจะผ่านระดับความน่าเชื่อถือที่กำหนด และในความเป็นจริงแล้วมีการทำงานในขั้นตอนที่ 2-4 ซ้ำอยู่หลายครั้ง จนกว่าจะได้วิธีการทางสถิติที่เหมาะสมสำหรับนำไปใช้วิเคราะห์ในระบบงานจริง

เพราะเหตุใดจึงต้องมีการทำเหมืองข้อมูล

ข้อมูลภายในฐานข้อมูลต่าง ๆ หากเก็บไว้โดยไม่นำมาใช้ก็จะไม่เกิดประโยชน์ ดังนั้นจึงต้องมีการสกัดสารสนเทศหรือการคัดเลือกข้อมูลออกมาใช้งานในส่วนที่ต้องการ ซึ่งในอดีตมนุษย์นั้นใช้แรงงานคนเป็นผู้สืบค้นข้อมูลต่าง ๆ ในฐานข้อมูล ซึ่งผู้สืบค้นจะทำการสร้างเงื่อนไขขึ้นมาตามภูมิปัญญาของผู้สืบค้น ทว่าในปัจจุบันการวิเคราะห์ข้อมูลจากฐานข้อมูลเดียวอาจไม่ให้ความรู้เพียงพอ และลึกซึ้งสำหรับการดำเนินงานภายใต้ภาวะที่มีการแข่งขันสูงและมีการเปลี่ยนแปลงที่รวดเร็ว จึงเกิดความจำเป็นที่จะต้องรวบรวมฐานข้อมูลหลาย ๆ ฐานข้อมูลเข้าด้วยกัน เรียกว่า “คลังข้อมูล” (Data Warehouse) ดังนั้นจึงมีความจำเป็นที่จะต้องใช้การทำเหมืองข้อมูล หรือ Data Mining ในการดึงข้อมูลจากฐานข้อมูลที่มีขนาดใหญ่ เพื่อนำข้อมูลนั้นมาใช้งานให้เกิดประโยชน์สูงสุดด้วยเช่นกัน (ศวี วานิช, 2558)

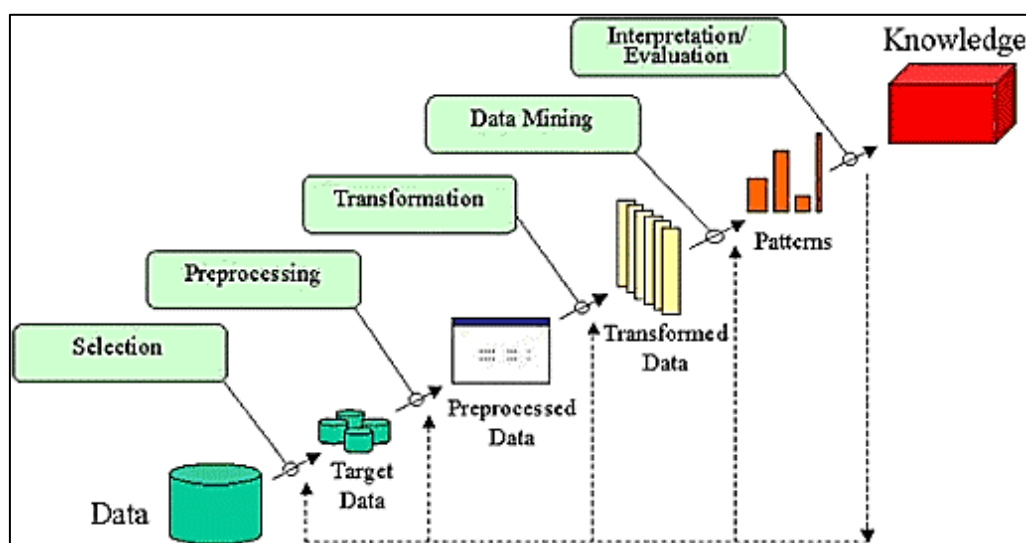
วัตถุประสงค์ของการทำเหมืองข้อมูล

- 1) เพื่อค้นพบองค์ความรู้ใหม่ในฐานข้อมูล (Knowledge discovery in databases)
- 2) เพื่อสกัดองค์ความรู้ที่ซ่อนเร้นอยู่ (Knowledge extraction)
- 3) เพื่อจัดการกับข้อมูลในอดีต (Data archeology)
- 4) เพื่อสำรวจข้อมูล (Data exploration)
- 5) เพื่อค้นหารูปแบบหรือ Pattern ที่ซ่อนอยู่ของข้อมูล (Data pattern processing)
- 6) เพื่อใช้ขุดเจาะข้อมูล (Data dredging)
- 7) เพื่อเก็บเกี่ยวผลประโยชน์ให้ได้มาซึ่งสารสนเทศที่มีประโยชน์

ขั้นตอนการทำเหมืองข้อมูล

- 1) Data Cleaning เป็นขั้นตอนสำหรับการคัดกรองข้อมูลที่ไม่เกี่ยวข้องออกไป
- 2) Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้กลายเป็นข้อมูลชุดเดียวกัน
- 3) Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้

- 4) Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน
- 5) Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่
- 6) Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล
- 7) Knowledge Representation เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบ โดยใช้เทคนิคในการนำเสนอเพื่อให้เกิดความเข้าใจ



รูปที่ 2-3: ขั้นตอนการทำ Data Mining

ที่มา: Data Mining (เหมืองข้อมูล), โดย ศจี วานิช, 2558,

<http://sajeegm301.blogspot.com/2015/11/data-mining.html>

ส่วนประกอบของการทำเหมืองข้อมูล

- 1) Database, Data Warehouse, World Wide Web และ Other Info Repositories เป็นแหล่งข้อมูลสำหรับการทำเหมืองข้อมูล
- 2) Database หรือ Data Warehouse Server ทำหน้าที่นำเข้าข้อมูลตามคำขอของผู้ใช้
- 3) Knowledge Base ได้แก่ ความรู้เฉพาะด้านในงานที่ทำจะเป็นประโยชน์ต่อการสืบค้นหรือประเมินความน่าสนใจของรูปแบบผลลัพธ์ที่ได้
- 4) Data Mining Engine เป็นส่วนประกอบหลักประกอบด้วยโมดูลที่รับผิดชอบงานทำเหมืองข้อมูลประเภทต่าง ๆ ได้แก่ การหากฎความสัมพันธ์ การจำแนกประเภท หรือการจัดกลุ่ม

- 5) Pattern Evaluation Module ทำงานร่วมกับ Data Mining Engine โดยใช้มาตรฐานความน่าสนใจในการกลั่นกรองรูปแบบผลลัพธ์ที่ได้ เพื่อให้การค้นหามุ่งเน้นเฉพาะรูปแบบที่น่าสนใจ
- 6) User Interface ส่วนติดต่อประสานระหว่างผู้ใช้กับระบบการทำเหมืองข้อมูล ช่วยให้ผู้ใช้สามารถระบุงาน ทำเหมืองข้อมูลตามที่ต้องการ และดูข้อมูลหรือโครงสร้างการจัดเก็บข้อมูล รวมถึงประเมินผลลัพธ์ที่ได้

ประวัติความเป็นมาและความก้าวหน้าของการทำเหมืองข้อมูลในปัจจุบัน

กระบวนการขุดผ่านข้อมูลเพื่อค้นหาความเชื่อมโยงที่ซ่อนอยู่และคาดการณ์แนวโน้มที่จะเกิดขึ้นในอนาคตนั้นมีประวัติศาสตร์อันยาวนาน บางครั้งได้ถูกเรียกว่า "การค้นพบองค์ความรู้ในฐานข้อมูล" ส่วนคำว่า "การทำเหมืองข้อมูล" นั้นยังไม่ได้มีการใช้งานจนถึงช่วงทศวรรษที่ 1990 ทั้งนี้รากฐานของสิ่งนี้ประกอบขึ้นจาก 3 ศาสตร์ ดังรูปที่ 2-4 ได้แก่ สถิติ (การศึกษาเชิงตัวเลขของความสัมพันธ์ของข้อมูล) ปัญญาประดิษฐ์ (ปัญญาเสมือนมนุษย์ที่เกิดขึ้นจากซอฟต์แวร์และ/หรือเครื่องจักร) และการเรียนรู้ของเครื่อง (อัลกอริทึมที่สามารถเรียนรู้จากข้อมูลเพื่อสร้างการคาดการณ์)

กว่าทศวรรษที่ผ่านมาความก้าวหน้าด้านขีดความสามารถและความเร็วในการประมวลผลของคอมพิวเตอร์ทำให้มนุษย์สามารถก้าวข้ามการปฏิบัติในแบบทำด้วยมือ (Manual) ที่น่าเบื่อและใช้เวลานาน ไปสู่การวิเคราะห์ข้อมูลที่รวดเร็ว ง่ายตาย และเป็นไปแบบอัตโนมัติ โดยยิ่งชุดข้อมูลที่รวบรวมมีความซับซ้อนมากเท่าใดก็ยังมีโอกาสค้นพบข้อมูลเชิงลึกที่เกี่ยวข้องมากขึ้นเท่านั้น ร้านค้าปลีก ธนาคาร ผู้ผลิต ผู้ให้บริการโทรคมนาคมและบริษัทประกัน และธุรกิจอื่น ๆ จึงได้มีการใช้การวิเคราะห์เหมืองข้อมูลเพื่อค้นหาความสัมพันธ์ระหว่างทุกอย่าง นับตั้งแต่การเพิ่มประสิทธิภาพด้านราคา โปรโมชัน และข้อมูลประชากร ไปจนถึงเศรษฐกิจ ความเสี่ยง การแข่งขัน และสื่อสังคม ซึ่งมีผลกระทบต่อรูปแบบการทำธุรกิจ รายได้ การดำเนินงาน และความสัมพันธ์กับลูกค้า (SAS Institute Inc., 2560)



รูปที่ 2-4: รากฐานของการทำเหมืองข้อมูล

ที่มา: การวิเคราะห์เหมืองข้อมูลคืออะไร และสำคัญอย่างไร, โดย SAS Institute Inc., 2560,
https://www.sas.com/th_th/insights/analytics/data-mining.html

เพราะเหตุใดการทำเหมืองข้อมูลจึงมีความสำคัญ

การทำเหมืองข้อมูลนั้นมีความสำคัญเนื่องจากช่วยให้ผู้ใช้สามารถ... (SAS Institute Inc., 2560)

- กลับกรองข้อมูลที่ยุ่งเหยิงและมีความซ้ำซ้อน
- ทำความเข้าใจว่าสิ่งใดที่เกี่ยวข้องและใช้ประโยชน์จากข้อมูลนั้นเพื่อประเมินผลลัพธ์ที่เป็นไปได้
- เร่งความเร็วให้การตัดสินใจที่ชาญฉลาด
- เรียนรู้เพิ่มเติมเกี่ยวกับเทคนิคการทำเหมืองข้อมูลในเอกสารการทำเหมืองข้อมูลตั้งแต่เริ่มต้น ซึ่งเป็นเอกสารที่แสดงวิธีการต่าง ๆ ในการใช้การวิเคราะห์เชิงคาดการณ์และการทำเหมืองข้อมูลที่องค์กรสามารถใช้งานเพื่อเปิดเผยข้อมูลเชิงลึกใหม่ ๆ จากข้อมูลที่มีอยู่

เทคนิคที่พบได้บ่อยในการทำเหมืองข้อมูล

สำหรับเทคนิคที่พบได้บ่อยในการทำเหมืองข้อมูลนั้นมีดังรูปที่ 2-5 (Mr. Automated, 2559)

Algorithm of Data mining

Descriptive Modeling : Unsupervised Learning

1. Association Algorithm
2. Clustering Algorithm
3. Time Series Algorithm

Predictive Modeling : Supervised Learning

Classification

4. Decision Trees Algorithm
5. Naive Bayes Algorithm
6. Neural Network Algorithm

Regression

7. Linear Regression Algorithm
8. Logistic Regression Algorithm



รูปที่ 2-5: เทคนิคที่พบได้บ่อยในการทำเหมืองข้อมูล

ที่มา: เหมืองข้อมูล (data mining), โดย Mr. Automated, 2559,

[http://www.autosoft.in.th/data-warehouse/%E0%B9B9%89%E0%](http://www.autosoft.in.th/data-warehouse/%E0%B9B9%89%E0%89%B8%B9%E0%B8%A5-data-mining/)

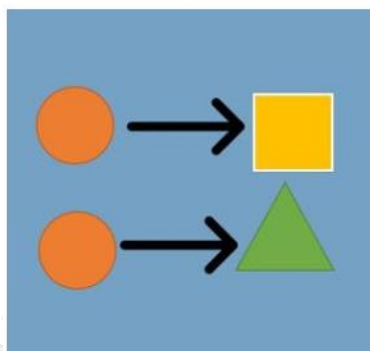
[%E0%B8%B9%E0%B8%A5-data-mining/](http://www.autosoft.in.th/data-warehouse/%E0%B8%B9%E0%B8%A5-data-mining/)

โมเดลเชิงอธิบาย (Descriptive Modeling): การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

ความสัมพันธ์ (Association)

Association Algorithm เป็นอัลกอริทึมการค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ (Big Data) เพื่อนำไปใช้ในการวิเคราะห์ หรือทำนายปรากฏการณ์ต่าง ๆ เช่น การวิเคราะห์การซื้อสินค้าของลูกค้าที่เรียกว่า “Market Basket Analysis” โดยนำ Transaction การซื้อสินค้ามาทำการค้นหาวิเคราะห์ว่าลูกค้าซื้อสินค้าอะไรคู่กับสินค้าอะไรบ่อย ๆ ทำให้สามารถออกโปรโมชั่นแคมเปญคู่กัน ทำให้มีราคาถูกลง เพื่อเพิ่มมูลค่าให้กับสินค้า โดยผลการวิเคราะห์ที่ได้จะเป็นคำตอบของปัญหา และการวิเคราะห์ลักษณะนี้เป็นการใช้ “กฎความสัมพันธ์ (Association Rule)” เพื่อหาความสัมพันธ์ของข้อมูลหรือแสดงความสัมพันธ์ของเหตุการณ์หรือวัตถุที่เกิดขึ้นพร้อมกัน (Mr. Automated, 2559) ดังรูปที่ 2-6 และตัวอย่างของการประยุกต์ใช้กฎเชื่อมโยงคือ การวิเคราะห์ข้อมูลการขายสินค้า โดยเก็บข้อมูลจากระบบ ณ จุดขาย (POS – Point of Sale) หรือร้านค้าออนไลน์ แล้วพิจารณา

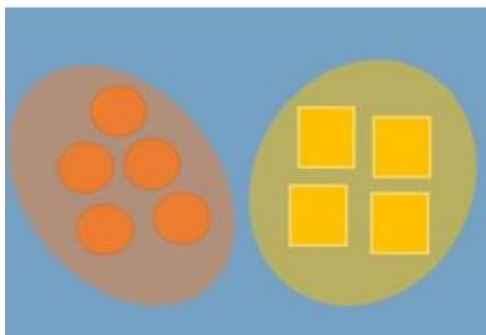
สินค้าที่ผู้ซื้อมักจะซื้อพร้อมกัน เช่น ถ้าพบว่าผู้ที่ซื้อเทปวิดีโอมักจะซื้อเทพกาวด้วย ร้านค้าก็อาจจัดร้านให้สินค้าสองอย่างวางอยู่ใกล้กันเพื่อเพิ่มยอดขาย หรืออาจพบว่าหลังจากที่ลูกค้าซื้อหนังสือ ก แล้ว มักจะซื้อหนังสือ ข ด้วย ก็สามารถนำความรู้นี้ไปใช้แนะนำผู้ที่กำลังจะซื้อหนังสือ ก ได้



รูปที่ 2-6: กฎความสัมพันธ์ (Association Rule) (Budkod, 2561)

การจำแนกกลุ่ม (Clustering)

Clustering Algorithm เป็นเทคนิคที่ใช้ในการจำแนกกลุ่มข้อมูลใหม่ที่มีลักษณะคล้ายกันไว้ในกลุ่มเดียวกัน ขั้นตอนวิธีที่ใช้ในการแบ่งกลุ่มจะอาศัยความเหมือน (Similarity) หรือ ความใกล้ชิด (Proximity) โดยคำนวณจากการวัดระยะระหว่างเวกเตอร์ของข้อมูลนำเข้า ดังรูปที่ 2-7 ตัวอย่างเช่น บริษัทจำหน่ายรถยนต์ได้แยกกลุ่มลูกค้าออกเป็น 3 กลุ่ม คือ กลุ่มผู้มีรายได้สูง (มากกว่า 80,000 บาท) กลุ่มผู้มีรายได้ปานกลาง (25,000 ถึง 80,000 บาท) และกลุ่มผู้มีรายได้ต่ำ (น้อยกว่า 25,000 บาท) หรือจำแนกกลุ่มลูกค้าช่วงตามช่วงอายุ และเพศ รวมถึงการวิเคราะห์ปัจจัยเสี่ยงเกี่ยวกับโอกาสในการเกิดโรคต่าง ๆ เพื่อจัดแคมเปญเสนอขายประกันคุ้มครองชีวิตให้ตรงกลุ่มเป้าหมาย เป็นต้น



รูปที่ 2-7: การจำแนกกลุ่ม (Clustering) (Budkod, 2561)

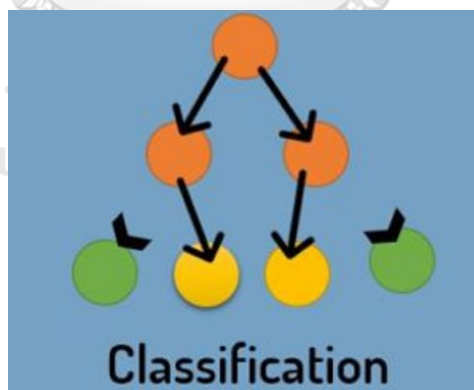
อนุกรมเวลา (Time Series)

Time Series Algorithm เป็นวิธีการพยากรณ์แบบข้อมูลอนุกรมเวลา เช่น การพยากรณ์การขาย (Sales Forecasting) คือการประมาณหรือการคาดคะเนว่าอะไรจะเกิดขึ้นในอนาคต เช่น การพยากรณ์ยอดขายของ 3 ปีข้างหน้า ซึ่งการพยากรณ์นั้นมีบทบาทสำคัญกับทุกภาคส่วน ทั้งหน่วยงานของรัฐบาลและเอกชน เนื่องจากรัฐบาลต้องมีการประมาณหรือพยากรณ์รายได้และรายจ่ายของหนึ่งปีข้างหน้าเพื่อนำมาวางแผน ส่วนด้านเอกชนก็ต้องมีการพยากรณ์ยอดขายเพื่อวางแผนการผลิตสินค้าคงคลัง และแรงงาน เป็นต้น

โมเดลเชิงทำนาย (Predictive Modeling): การเรียนรู้แบบมีผู้สอน (Supervised Learning)

การจำแนกประเภท (Classification)

Classification เป็นการหากฎเพื่อระบุประเภทของวัตถุจากคุณสมบัติของวัตถุ ดังรูปที่ 2-8 เช่น หาความสัมพันธ์ระหว่างผลการตรวจร่างกายต่าง ๆ กับการเกิดโรค โดยใช้ข้อมูลผู้ป่วยและการวินิจฉัยของแพทย์ที่เก็บไว้ เพื่อนำมาช่วยวินิจฉัยโรคของผู้ป่วยหรือการวิจัยทางการแพทย์ ในทางธุรกิจจะใช้เพื่อดูคุณสมบัติของผู้ที่จะก่อกำเนิดหรือหนี้เสีย เพื่อประกอบการพิจารณาการอนุมัติเงินกู้



รูปที่ 2-8: การจำแนกกลุ่ม (Clustering) (Budkod, 2561)

ต้นไม้การตัดสินใจ (Decision Tree)

Decision Trees Algorithm เป็นการแยกข้อมูล (Classification) ออกเป็นกลุ่มโดยใช้คุณสมบัติของข้อมูล (Attribute) เป็นตัวกำหนด ซึ่งประกอบไปด้วย โหนดภายใน (Internal Node), กิ่ง (Link) และโหนดใบ (Leaf Node) วิธีการวิเคราะห์แบบต้นไม้ตัดสินใจ

เป็นการค้นหาจากบนลงล่าง (Top-down) โดยเริ่มจากการเลือกคุณสมบัติที่ดีที่สุดมาเป็น โหนดราก (Root Node) และวนสร้างโหนดลูกและเส้นเชื่อมไปเรื่อย ๆ จนกว่าข้อมูลที่ได้จะถูกจัดไว้เป็นกลุ่มเดียวกันจึงจะหยุดการสร้างต้นไม้

Naïve Bayesian Classification

Naive Bayes Algorithm คือ การทำเหมืองข้อมูลในแบบ Classifier ที่ถูกสร้างขึ้น โดยหลักความน่าจะเป็น Naïve Bayesian Classification จะใช้วิเคราะห์หาความน่าจะเป็นของสิ่งที่ยังไม่เคยเกิดขึ้น โดยการคาดเดาจากสิ่งที่เคยเกิดขึ้นมาก่อน

โครงข่ายประสาท (Neural Network)

Neural Network Algorithm เป็นแนวคิดที่ได้มาจากการจำลองการทำงานของ เซลล์สมองของมนุษย์ ซึ่งมีโครงสร้างประกอบด้วย Input Layer, Hidden Layer และ Output Layer มีหน่วยย่อยเรียกว่า Perceptron ซึ่งเทียบเท่าได้กับเซลล์สมองของมนุษย์ หนึ่ง Neuron โดยหลักการของ Neural Network จะมีการกำหนดค่าน้ำหนัก (Weight) และ Threshold ให้แก่ Input แต่ละตัวโดยใช้ Back-Propagation Algorithm ในการ คำนวณ และการสร้างโมเดล Neural Network สามารถทำได้ทั้งหมด 2 วิธีคือ Supervised Learning และ Unsupervised Learning ซึ่งสามารถนำมาประยุกต์ใช้กับงานด้านต่าง ๆ เช่น การพยากรณ์ การจดจำใบหน้า เรียนรู้จำลายมือ ลายเซ็นต์ หรือใช้ในทางการแพทย์ ฯลฯ

การวิเคราะห์ความถดถอย (Regression Analysis)

Linear Regression Algorithm คือการใช้วิธีการทางสถิติในการหาความสัมพันธ์ ระหว่างตัวแปรต้น (Independent Variable) และ ตัวแปรตาม (Dependent Variable) ส่วน Logistic Regression Algorithm ถูกนำมาใช้เพื่อทำนายว่าจะเกิดเหตุการณ์หนึ่งขึ้น หรือไม่มีหรือมีโอกาสเกิดขึ้นมากน้อยเพียงใด โดยมีการกำหนดค่าตัวแปรตัวหนึ่งหรือหลายตัว ที่คาดว่าจะส่งผลต่อการเกิดเหตุการณ์นั้น ๆ และในที่สุดก็จะช่วยให้ผู้ใช้สามารถเข้าใจสาเหตุ การเกิดเหตุการณ์นั้น ๆ ได้ (Lowphansirikul, 2561)

ทั้งนี้ Classification และ Regression ต่างก็เป็น Model ประเภท Supervised Model เหมือนกัน ซึ่ง Model ประเภทนี้จำเป็นต้องมี Target หรือ ตัวแปรที่ต้องการศึกษา เป็นตัวต้นแบบ ยกตัวอย่างเช่น ต้องการศึกษากลไกในการอนุมัติสินเชื่อ ซึ่งตัวแปรที่ต้องการศึกษาคือ การอนุมัติสินเชื่อ (อนุมัติ หรือ ไม่อนุมัติ) โดยใช้ข้อมูลเก่าในอดีตเป็นตัวต้นแบบ ความต่างของ Classification กับ Regression คือ ลักษณะของ Target ที่ต่างกัน โดย Target ของ Classification คือ ข้อมูลที่เป็นลักษณะกลุ่ม หรือข้อมูลที่ไม่มีความต่อเนื่อง เช่น ใช่/ไม่ใช่ หรือ 1, 2, 3 เป็นต้น แต่ Target ของ Regression คือ ข้อมูลที่มีความต่อเนื่อง ไม่ได้แบ่งเป็นกลุ่ม เช่น 0-100

ถึงแม้หลายคนอาจมองว่าในเชิงการพัฒนาโปรแกรม Classification และ Regression Model นั้นเหมือนกัน เนื่องจากเป็นการมองหาความสัมพันธ์ของปัจจัยต่าง ๆ เพื่อให้ได้ผลลัพธ์ที่เป็น Target แต่รายละเอียดในการใช้ Classification และ Regression Model นั้นมีความแตกต่างกันในส่วนต่าง ๆ ตัวอย่างเช่น (Coraline Team, 2562)

1) การประเมินผลลัพธ์

Classification ประเภท Binary หรือ ตัวแปรที่มีเพียง 2 กลุ่ม จะใช้ Confusion Matrix เป็นตัววัดค่าความแม่นยำ ในขณะที่ Regression จะใช้วิธีการวัดผลที่หลากหลายกว่า เช่น ค่า R-squared ค่า Mean Absolute Percentage Error (MAPE) หรือ RMSE เป็นตัววัดผล

และข้อสังเกตคือ การวัดผลของ Supervised Model จะสามารถวัดค่าความแม่นยำ หรือ Accuracy ได้ เนื่องจากมี Target เป็นตัวตั้งต้น ทำให้สามารถนำค่า Predict ที่ได้ มาวัดค่าความคาดเคลื่อนจากตัวตั้งต้นได้นั่นเอง ดังนั้นในกรณีของทั้ง Classification และ Regression จึงสามารถประเมินผลลัพธ์ของโมเดลได้จากค่า Accuracy

2) ความเข้าใจในเชิงการทำงาน หรือบุคคลที่ไม่ใช่ Technician

โมเดลทั้ง 2 ประเภทนี้ มีวัตถุประสงค์ในการใช้งานต่างกัน โดย Classification มักจะใช้เพื่อการจัดประเภทและการตัดสินใจ เช่น การวิเคราะห์สินเชื่อ ที่มีผลประเมินเป็น อนุมัติ/ไม่อนุมัติ, Fraud Detection หรือ การวิเคราะห์กลไก ที่มีผลประเมินเป็น ความระดับความเสี่ยง 1/2/3/4 ในขณะที่ Regression มักจะใช้เพื่อวิเคราะห์ข้อมูลที่มี

Range ข้อมูลค่อนข้างกว้าง เช่น Weather Forecast หรือ การพยากรณ์สภาพอากาศ, การประเมินปัจจัยที่ทำให้เกิดมลพิษ ที่มีผลลัพธ์เป็นตัวเลขในช่วง 0-100 เป็นต้น

ข้อสังเกตคือโจทย์บางข้อจะมีแนวความคิดการใช้โมเดลที่ชัดเจน เช่น Forecasting จะใช้ Time Series Model หรือ Regression Model ส่วนการประเมินความเสี่ยงที่มีระดับความเสี่ยงเป็นกลุ่ม จะใช้ Classification Model เป็นต้น ซึ่ง Data Scientist จะค่อนข้างมีความเคยชิน และจะสามารถบอกประเภทของโจทย์ได้จากชื่อ

3) Algorithm ที่ใช้

ทั้ง 2 โมเดลเป็น Supervised Model เหมือนกัน และมีลักษณะการใช้งานเพื่อศึกษาปัจจัยที่ให้ผลลัพธ์เป็น Target เหมือนกัน ดังนั้นจึงมี Algorithm บางตัวที่สามารถใช้งานได้เหมือนกัน ขึ้นอยู่กับการ Set up หรือ ตั้งค่าโมเดล เช่น Neural Network, Random forests เป็นต้น แต่ก็มี Algorithm บางประเภทที่ใช้งานกับโมเดลได้เพียงอย่างเดียวหนึ่ง เช่น Logistic Regression เป็นโมเดลที่ใช้สำหรับการวิเคราะห์โอกาสของการเป็น Binary ซึ่งการเป็น Binary คือ การเป็น Classification

ตัวอย่างการใช้งานเหมืองข้อมูล

ปัจจุบันได้มีการนำเหมืองข้อมูลไปใช้งานในด้านต่าง ๆ ดังนี้ (Budkod, 2561)

ด้านการตลาด

- การทำนายผลการตอบสนองกับการเปิดตัวสินค้าใหม่
- การทำนายยอดขายเมื่อมีการลดราคาสินค้า
- การทำนายกลุ่มลูกค้าที่น่าจะใช้สินค้าของบริษัท

การเงินการธนาคาร

- การคาดการณ์ถึงโอกาสในการชำระหนี้ของลูกค้าว่าสูงเท่าใด
- ค้นหาลูกค้าขาดคุณภาพ เพื่อหลีกเลี่ยงความเสี่ยงในการปล่อยกู้
- ค้นหาลูกค้าชั้นดี เพื่อเสนอการปล่อยกู้
- ทำนายแนวโน้มของพฤติกรรมการใช้บัตรเครดิต

สถานีโทรทัศน์หรือวิทยุ

- ค้นหารายการที่ดีและเหมาะสมต่อช่วงเวลาที่ดู เพื่อวางแผนรายการในแต่ละเดือน

ฮาร์ดแวร์และซอฟต์แวร์คอมพิวเตอร์

- ค้นหาช่วงเวลาที่เหมาะสมกับการผลิตชิพคอมพิวเตอร์ตัวใหม่เพื่อป้อนสู่ตลาด
- การทำนายอายุการใช้งานของ Disk Drive หรือ อุปกรณ์ต่าง ๆ

รายละเอียดของโมเดลที่ใช้ทำนายข้อมูลของโครงการ “ระบบแช่ตบอดและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ”

1) ระบบการทำนายด้วย Decision Tree Model

Decision Tree เป็นแบบจำลองทางคณิตศาสตร์เหมือนกราฟหรือแผนผัง มีลักษณะเป็นต้นไม้กลับหัวที่มีเป้าหมายเพื่อหาทางเลือกที่ดีที่สุด โดยนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปของโครงสร้างต้นไม้ ซึ่งเป็นการเรียนรู้ข้อมูลแบบมีผู้สอน (Supervised Learning) และมีองค์ประกอบคือ

- (1) ลำต้นหรือราก (Root node): แต่ละโหนดแทนตัวแปรนำเข้า (Input Attribute) ต่าง ๆ ในชุดข้อมูล โดยโมเดล Decision tree จะคัดเลือกเอาทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของต้นไม้ และเป็นตัวเริ่มต้นที่ใช้บอกว่าการตัดสินใจอะไรที่สำคัญที่สุด รวมถึงเป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทางใด
- (2) กิ่ง (Branch): แทนค่าของตัวแปร (Numerical Attributes) เหมือนแขนขาของโหนดที่แยกออกมา ใช้ขยายโหนดออกไปเรื่อย ๆ โดย Branch จะเป็นตัวระบุเงื่อนไขในการตัดสินใจ ถ้าผ่านเงื่อนไขในนี้ก็จะย้ายไปยังการตัดสินใจที่สำคัญรองลงมา
- (3) ใบ (Leaf): โหนดลำดับสุดท้ายในซึ่งอยู่ในส่วนล่างสุดของต้นไม้ หลังจากโหนดนี้จะไม่มีการตัดสินใจอีกต่อไป ดังนั้น Leaf จึงเป็นเป้าหมายที่ผู้สร้างโมเดลต้องการจะได้หรือเป็นกลุ่มของผลลัพธ์ในการแยกแยะข้อมูลนั่นเอง

จุดเด่นของโมเดล Decision Tree

โมเดล Decision Tree มีจุดเด่น ดังนี้ (Satangmongkol, 2561)

- มีความยืดหยุ่นในการรับตัวแปรได้ทุกประเภท และสามารถรับข้อมูลที่มี Missing หรือ Noise ได้
- สามารถเลือกตัวแปรที่มีความสำคัญที่ช่วยแบ่งแยกข้อมูลออกมาได้เอง
- Train ได้อย่างรวดเร็ว
- เป็นโมเดลที่ชัดเจน เข้าใจง่าย สามารถแปลความจากโมเดลได้เลย จึงทำให้สะดวกต่อการแปลผล

จุดด้อยของโมเดล Decision Tree

โมเดล Decision Tree มีจุดด้อย ดังนี้ (Satangmongkol, 2561)

- ค่อนข้าง Overfit หากไม่มีการตัดแต่งกิ่ง (Prune) ก่อน
- Accuracy อยู่ในระดับปานกลาง

หลักการพื้นฐานของการสร้าง Decision Tree

เป็นการสร้างในลักษณะจากบนลงล่าง (Top-Down) คือเริ่มจากการสร้างรากของต้นไม้แล้วจึงแตกกิ่งไปจนถึงใบ โดยมีขั้นตอนดังนี้ (รุจิรา ธรรมสมบัติ, 2554)

- ต้นไม้เริ่มต้นโดยมีโหนดเพียงโหนดเดียวแสดงถึงชุดข้อมูลฝึก (Training Set)
- หากข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกันแล้ว โหนดนั้นจะกลายเป็นใบและตั้งชื่อแยกตามกลุ่มของข้อมูล
- หากในโหนดมีข้อมูลหลายกลุ่มปะปนอยู่ จะต้องวัดค่าเกน (Gain) ของแต่ละแอททริบิวต์เพื่อใช้เป็นเกณฑ์ในการคัดเลือกแอททริบิวต์ที่มีความสามารถในการแบ่งแยกข้อมูลออกเป็นกลุ่มต่าง ๆ ได้ดีที่สุด โดยแอททริบิวต์ที่มีค่าเกนมากที่สุดจะถูกเลือกให้เป็นตัวทดสอบหรือเป็นแอททริบิวต์ที่ใช้ในการตัดสินใจโดยอยู่ในรูปของโหนดบนต้นไม้
- กิ่งของต้นไม้ถูกสร้างขึ้นจากค่าต่าง ๆ ที่เป็นไปได้ของโหนดทดสอบ และข้อมูลจะถูกแบ่งไปตามกิ่งต่าง ๆ ที่สร้างขึ้น

- ทำการวนซ้ำเพื่อหาแอททริบิวต์ที่มีค่าเกินมากที่สุดสำหรับข้อมูลที่ถูกรวบรวมแยกออกมาในแต่ละกิ่งเพื่อนำแอททริบิวต์นั้นมาสร้างเป็นโหนดตัดสินใจต่อไป โดยที่แอททริบิวต์ที่ถูกเลือกมาเป็นโหนดแล้วจะไม่ถูกเลือกมาอีกสำหรับโหนดในระดับต่อไป
- แบ่งข้อมูลและแตกกิ่งของต้นไม้ไปเรื่อย ๆ โดยวนซ้ำจนกว่าจะได้ตามเงื่อนไขที่กำหนด เช่น ความลึกสูงสุดของต้นไม้ (Max Dept) ไม่เกิน 10 ชั้น หรือ จำนวนข้อมูลในแต่ละกลุ่มที่แบ่งออกมา (Leaf Node) มีจำนวนขั้นต่ำ 5 Observation (Min Sample)

2) ระบบการทำนายด้วย Neural Network Model

Neural Network Model เป็นโมเดลทางคณิตศาสตร์ที่เลียนแบบการทำงานของสมองมนุษย์ โดยลักษณะการทำงานโดยส่วนใหญ่จะมี 3 ชั้น ได้แก่ Input Node , Hidden Node และ Output Node ซึ่งหากมีการซ่อน Hidden Node เข้าไปเป็นจำนวนมากอาจเสี่ยงต่อการเกิด Overfitting ได้ และค่าน้ำหนักบนเส้น Network แต่ละเส้นก็มีความสำคัญกับคำตอบเช่นกัน เนื่องจากเป็นส่วนที่ใช้หาน้ำหนักของความสัมพันธ์ระหว่างข้อมูลนำเข้า ว่าข้อมูลนำเข้าใดมีความสัมพันธ์กับข้อมูลนำเข้าอื่นในระดับใด ซึ่งจะทำให้สามารถเชื่อมโยงไปหาข้อสรุปได้ด้วยการลองผิดลองถูกในความสัมพันธ์แต่ละแบบ และเก็บไว้เป็นแบบแผนหรือรูปแบบ (Pattern) ของประสบการณ์เพื่อการเรียนรู้ของโครงข่าย (นงคราญ คำวิชัย, 2559)

จุดเด่นของโมเดล Neural Network

โมเดล Neural Network มีจุดเด่น ดังนี้ (ศราวุธ เลิศพลังสันติ, 2553)

- เกิดข้อผิดพลาดได้ยาก (Fault Tolerance) เนื่องจากหากระบบโครงข่ายใดประสาทย่อยประกอบไปด้วยโครงข่ายที่ใช้ในการประมวลผลมากมายหลายโครงข่าย ดังนั้นความผิดพลาดที่เกิดขึ้นจากเพียงหนึ่งหรือสองโครงข่ายจะไม่ทำให้ทั้งระบบเกิดข้อผิดพลาดได้
- ความสามารถในการหาเหตุผล (Generalization) เมื่อระบบโครงข่ายใดประสาทย่อยเหมือนได้รับข้อมูลนำเข้าที่ไม่ครบถ้วนหรือไม่เพียงพอต่อการหาข้อสรุป หรือได้รับ

ข้อเท็จจริงที่ไม่เคยได้รับมาก่อน ระบบจะสามารถลำดับการเชื่อมโยงข้อเท็จจริงจนสามารถให้ข้อสรุปและเหตุผลได้ ส่งผลให้สามารถทำงานกับข้อมูลที่มีความซับซ้อนได้ดีกว่า

- ความสามารถในการปรับเปลี่ยน (Adaptability) โครงข่ายใยประสาทเสมือนสามารถเรียนรู้สภาพแวดล้อมใหม่ได้ ดังนั้นเมื่อมีเหตุการณ์ใหม่ ๆ เข้าสู่ระบบก็จะสามารถปรับเปลี่ยนหรือปรับปรุงองค์ความรู้ให้ทันสมัยตามเหตุการณ์ใหม่นั้นได้ สามารถ Fit ตัวเข้าหาข้อมูลได้ดี
- ความสามารถในการพยากรณ์ (Forecasting Capability) มีความสามารถในการจำชุดของคู่อินพุตและเอาต์พุตที่มีความซับซ้อนมากจนไม่สามารถจำลองแบบในเชิงความน่าจะเป็นได้ ดังนั้น Neural Network จึงสามารถนำข้อมูลทางสถิติเดิมที่มีอยู่ในระบบ มาใช้คาดการณ์หรือพยากรณ์ข้อมูลในอนาคตได้และมีความแม่นยำในการทำนายสูง

จุดด้อยของโมเดล Neural Network

- รับเฉพาะตัวแปรที่เป็นตัวเลข (Numeric) เท่านั้น
- ผู้ใช้ต้องเลือกตัวแปรที่จะนำไปใช้ด้วยตนเอง
- อาจต้องใช้เวลาในการสอนหรือเรียนรู้าน
- เกิดการ Overfit ได้ง่าย

Neural Network จะมีลักษณะการทำงานแบบ Black Box มากกว่า จึงเป็นเทคนิคที่ไม่เหมาะกับการคาดการณ์ที่ต้องนำไปอธิบายต่อ เนื่องจากผลลัพธ์หรือค่าน้ำหนักของตัวแปรฟังก์ชันที่ได้จากการเรียนรู้ค่อนข้างเข้าใจยากส่งผลให้อาจไม่สามารถอธิบายย้อนกลับได้

3) ระบบการทำนายด้วย Random Forest Model

Random Forest เป็น Model ประเภทหนึ่งของ Machine Learning ซึ่งถูกพัฒนาขึ้นจาก Decision Tree แต่ต่างกันว่า Random Forest นั้นเป็นการเพิ่มจำนวนต้นไม้เป็นหลาย ๆ ต้น ทำให้ประสิทธิภาพในการทำงานสูงขึ้นและแม่นยำมากขึ้น ซึ่ง

โมเดล Random Forest เป็นโมเดลที่ได้รับความนิยมอย่างมากในการใช้ Machine Learning นอกจากนี้ Random Forest ยังเข้ามาช่วยแก้ปัญหาการทำ Bagging ซึ่งจะ มีประเด็นเรื่องความไม่เป็นอิสระของข้อมูล เนื่องจากถึงแม้จะมีการแยกต้นไม้ออกเป็น หลายต้น แต่ก็ยังคงเป็นข้อมูลเดียวกัน ดังนั้นจึงเกิดการทำให้ Random Sample Feature คือนอกจากจะแบ่งต้นไม้ออกเป็นหลายต้นแล้ว ยังแบ่ง Feature ของต้นไม้ แต่ละต้นให้มี Feature ที่ไม่เหมือนกันทั้งหมด เพื่อให้ต้นไม้แต่ละต้นมีความ หลากหลายและมีความเป็นอิสระจากกันมากขึ้นอีกด้วย จึงเป็นการช่วยลดโอกาสเกิด ปัญหาเรื่อง Overfit ได้อย่างมาก (Daroontham, 2561)

จุดเด่นของโมเดล Random Forest

- Random Forest ใช้ได้ทั้งกับปัญหา Classification และ Regression
- Random Forest ใช้ได้ทั้งกับข้อมูล Structured (ข้อมูลลักษณะเป็น Column/Table) และ Unstructured (เช่น รูปภาพ, Text)
- สามารถปรับค่าพารามิเตอร์ให้ Random Forest ไม่ Overfit ได้ง่าย
- Random Forest ไม่ตั้ง Assumption กับ Feature ว่าจะต้องกระจายข้อมูลแบบ Normal Distribution หรือสัมพันธ์กับ Target แบบ Linear และ ไม่ต้องสร้างความสัมพันธ์ระหว่าง Feature เพิ่มเติม
- จากข้อ 4 จึงส่งผลให้ประหยัดแรงทำ Feature Engineering เช่น ไม่จำเป็นต้อง ทำ Log Transform หรือสร้าง Interaction จาก Feature
- ความแม่นยำ (Accuracy) สูง

จุดด้อยของโมเดล Random Forest

- Train โมเดลได้ช้า
- อธิบายผลลัพธ์ยาก เนื่องจากกระบวนการสร้างป่าซึ่งประกอบด้วยต้นไม้จำนวนมากนั้นเกิดขึ้นแบบสุ่ม

หลักการเบื้องต้นในการสร้าง Random Forest Model

การสร้าง Random Forest Model มีหลักการเบื้องต้น ดังนี้ (Daroontham, 2561)

- Sample ข้อมูล (Bootstrapping) จาก Data Set ทั้งหมด ให้ได้ข้อมูลออกมา n ชุด ที่ไม่เหมือนกัน ตามจำนวน Decision Tree ใน Random Forest เช่น Data Set ตั้งต้นมีอยู่ 10 feature (X_1, X_2, \dots, X_{10})
- สร้างโมเดล Decision Tree สำหรับแต่ละชุดข้อมูล คือ สร้างโมเดลจาก Decision Tree หลายโมเดลย่อย (ตั้งแต่ 10 โมเดลถึงมากกว่า 1000 โมเดล) โดยแต่ละโมเดลจะได้รับชุดข้อมูล (Data Set) ที่ไม่เหมือนกัน ซึ่งเป็น Subset ของ Data Set ทั้งหมด
- ทำ Prediction โดยให้แต่ละ Decision Tree ทำ Prediction ของตนเองและทำ Aggregation คำนวณผลลัพธ์ จากแต่ละโมเดล (Bagging) ด้วยการ Vote Output ที่ถูกเลือกโดย Decision Tree มากที่สุด (กรณี Classification) หรือหาค่า Mean จาก Output ของแต่ละ Decision Tree (กรณี Regression)
- Decision Tree แต่ละโมเดลใน Random Forest นั้นถือว่าเป็น Weak Learner หรือเป็นโมเดลที่ไม่เก่งนัก แต่เมื่อนำเอาแต่ละ Decision Tree มาทำ Prediction ร่วมกัน จะได้โมเดลรวมที่มีความเก่ง และแม่นยำมากกว่า Decision Tree ที่ทำ Prediction แบบเดี่ยว ๆ

การประเมินประสิทธิภาพในการทำนายของโมเดล (Predictive Model Evaluation)

การประเมินประสิทธิภาพของ Predictive Model สามารถทำได้โดยพิจารณาจากค่าความถูกต้อง (Accuracy) ของทั้งสองโมเดลที่ถูกสร้างขึ้น เพื่อให้สามารถเลือกใช้ได้อย่างเหมาะสม โดยอาจมีการปรับปรุงโมเดลเพิ่มเติมอย่างต่อเนื่อง เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด โดยการ Evaluate โมเดลที่สร้างขึ้นนั้นสามารถพิจารณาได้จาก Confusion Matrix ซึ่งถือเป็นเครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนาย หรือ Prediction จากโมเดล โดยมีแนวคิดในการวัดว่าสิ่งที่โมเดลทำนายกับสิ่งที่เกิดขึ้นจริงนั้นมีสัดส่วนเป็นอย่างไร (Gatchalee, 2562)

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

รูปที่ 2-9: ตาราง Confusion Matrix (Gatchalee, 2562)

True Positive (TP) = สิ่งที่ทำนายตรงกับสิ่งที่เกิดขึ้นจริง คือโมเดลทำนายว่า “จริง” และสิ่งที่เกิดขึ้นก็คือ “จริง”

True Negative (TN) = สิ่งที่ทำนายตรงกับสิ่งที่เกิดขึ้น คือโมเดลทำนายว่า “ไม่จริง” และสิ่งที่เกิดขึ้นก็คือ “ไม่จริง”

False Positive (FP) = สิ่งที่ทำนายไม่ตรงกับสิ่งที่เกิดขึ้น คือโมเดลทำนายว่า “จริง” แต่สิ่งที่เกิดขึ้นคือ “ไม่จริง”

False Negative (FN) = สิ่งที่ทำนายไม่ตรงกับที่ที่เกิดขึ้นจริง คือโมเดลทำนายว่า “ไม่จริง” แต่สิ่งที่เกิดขึ้นคือ “จริง”

โดย TP, TN, FP, FN ในตารางจะแทนด้วยค่าความถี่ และสามารถนำ Confusion Matrix มาคำนวณการประเมินประสิทธิภาพของการทำนายด้วยโมเดลในรูปแบบค่าต่าง ๆ ได้หลายค่า ได้แก่

1) **Accuracy**: ภาพรวมความถูกต้องที่โมเดลทายได้ตรงกับสิ่งที่เกิดขึ้นจริง

$$\text{Accuracy (ความถูกต้อง)} = (TPs + TNs) / (TPs + TNs + FPs + FNs)$$

หรือกล่าวได้ว่า Accuracy = ผลรวมของตัวเลขบนเส้นทแยงมุมในตาราง Confusion Matrix / จำนวน Observations ทั้งหมด และในความเป็นจริงแล้ว Confusion matrix ไม่จำเป็นต้องเป็นแบบ 2x2 หรือมีผลลัพธ์แค่ 2 แบบเสมอไป แต่อาจเป็น 3x3, 4x4, nxn ก็ได้ โดยวิธีการหา

Accuracy ก็ใช้แบบเดิมคือผลรวมของตัวเลขบนเส้นทแยงมุมในตาราง Confusion Matrix / จำนวน Observations ทั้งหมด

2) Precision: ค่าความแม่นยำ ซึ่งบอกว่าที่โมเดลทำนายว่าจริงนั้นถูกต้องเท่าใด

เป็นการเปรียบเทียบจำนวนครั้งที่โมเดลทำนายว่า Positive แล้วถูก (TP) กับจำนวนครั้งที่โมเดลทำนายว่า Positive ทั้งหมด (TPs + FPs)

$$\text{Precision} = \text{TPs} / (\text{TPs} + \text{FPs})$$

3) Recall: ค่าที่บอกว่าโมเดลทำนายได้ว่าจริง คิดเป็นอัตราส่วนเท่าใดของจริงทั้งหมด

เป็นการเปรียบเทียบจำนวนครั้งที่โมเดลทำนายว่า Positive แล้วถูก (TP) กับจำนวน Positive ทั้งหมดในข้อมูล (ทั้งที่โมเดลทำนายและเกิดขึ้นจริง)

$$\text{Recall} = \text{TPs} / (\text{TPs} + \text{FNs})$$

บทที่ 3

โครงสร้างองค์กรและการดำเนินงาน

ในบทนี้จะกล่าวถึงประวัติองค์กร โครงสร้างองค์กร การดำเนินงานขององค์กร ไปจนถึงปัญหาในปัจจุบันที่เกิดขึ้นกับองค์กรที่ใช้เป็นกรณีศึกษาสำหรับการพัฒนาโครงการ “ระบบเช็คบอต และการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ”

3.1 ประวัติองค์กร

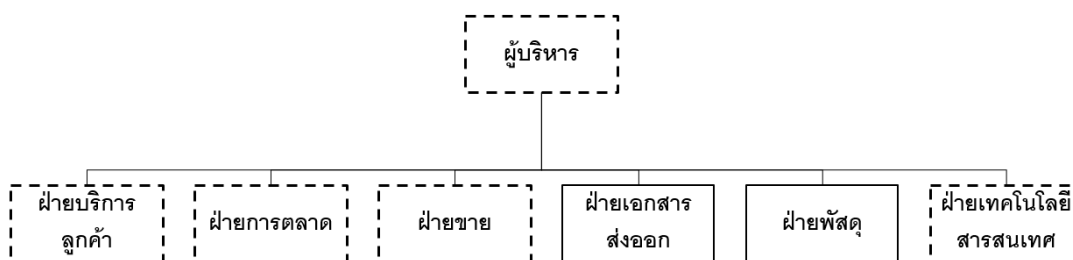
บริษัท Pigeon Air Freight (นามสมมติ) จัดตั้งขึ้นเพื่อเป็นช่องทางให้ลูกค้าสามารถส่งพัสดุไปต่างประเทศได้ในราคาประหยัด โดยบริษัทเป็นพาร์ตเนอร์กับหนึ่งในองค์กรผู้นำด้านการจัดส่งพัสดุระดับโลกคือ DHL ซึ่งครอบคลุมบริการจัดส่งพัสดุในกว่า 220 ประเทศทั่วโลก ให้บริการแบบถึงมือผู้รับ (Door-to-Door) และครบวงจร (One Stop Service) ตั้งแต่เรื่องของการจัดทำเอกสารส่งออก การบรรจุพัสดุลงกล่อง การเข้ารับพัสดุถึงที่ ไปจนถึงการจัดส่งให้ถึงมือผู้รับปลายทาง

บริษัทเปิดดำเนินการในปี พ.ศ. 2550 โดยสำนักงานใหญ่ตั้งอยู่ในกรุงเทพมหานครและมีสาขาย่อย ๆ ในต่างจังหวัด เป้าหมายหลักคือมุ่งเน้นกลุ่มลูกค้าทั่วไปซึ่งไม่จำเป็นต้องเป็นองค์กรใหญ่ ให้สามารถส่งพัสดุไปต่างประเทศได้อย่างรวดเร็วในราคาประหยัด โดยเฉพาะกลุ่มลูกค้าที่ทำธุรกิจ SME ซึ่งต้องส่งสินค้าไปยังต่างประเทศ การที่สามารถลดต้นทุนค่าส่งสินค้าให้ถูกลงได้จะช่วยเพิ่มขีดความสามารถในการแข่งขัน รวมถึงช่วยส่งเสริมด้านการเติบโตของธุรกิจในอนาคตได้อีกด้วย

ปัจจุบันธุรกิจยังมีข้อจำกัดในการขยายฐานลูกค้า เนื่องจากพนักงานที่มีไม่มากพอ ทั้งยังต้องแบ่งเวลาในการให้ข้อมูลและตอบคำถามต่าง ๆ จากลูกค้า นอกจากนี้บริษัทยังไม่เคยทำการวิเคราะห์ข้อมูลเกี่ยวกับลูกค้ามาก่อน ส่งผลให้ยังขาดข้อมูลเชิงลึกเกี่ยวกับพฤติกรรมการใช้งานและลักษณะของลูกค้ากลุ่มต่าง ๆ ซึ่งถือเป็นข้อมูลสำคัญอย่างมากสำหรับการทำความเข้าใจลูกค้าเพื่อให้สามารถตอบโจทย์และวางแผนในการดำเนินธุรกิจต่อไปในอนาคต

3.2 โครงสร้างองค์กร

โครงสร้างองค์กรของบริษัท Pigeon Air Freight แสดงดังรูปที่ 3-1



หมายถึง หน่วยงานที่เกี่ยวข้องกับโครงการพิเศษนี้

รูปที่ 3-1: โครงสร้างองค์กรของบริษัท Pigeon Air Freight

การแบ่งส่วนงานและหน้าที่ความรับผิดชอบ

- 1) ผู้บริหาร มีหน้าที่ในการบริหารดูแลทุกส่วนของกิจการ และมอบหมายความรับผิดชอบให้กับผู้จัดการของแต่ละฝ่าย รวมถึงติดตามผลการดำเนินงานให้เป็นไปตามนโยบายที่กำหนดไว้
- 2) ฝ่ายบริการลูกค้า มีหน้าที่ให้คำแนะนำ ตอบข้อสงสัย และช่วยประสานงานแก้ปัญหาให้กับลูกค้าในทุกช่องทางไม่ว่าจะเป็นโทรศัพท์หรืออีเมล เพื่อให้ลูกค้าได้รับความพึงพอใจสูงสุด
- 3) ฝ่ายการตลาด มีหน้าที่รับผิดชอบดูแลด้านการวางแผนการตลาด กำหนดแนวทางการประชาสัมพันธ์ และกลยุทธ์การขาย ศึกษาศึกษาทางการตลาด โปรโมชัน รวมถึงพฤติกรรมผู้บริโภค
- 4) ฝ่ายขาย มีหน้าที่เป็นตัวแทนในการให้ข้อมูลนำเสนอบริการของบริษัท ภูมิใจที่สนใจใช้บริการ ทำยอดขายให้ได้ตามเป้าหมายที่ตั้งไว้ ดูแลรับผิดชอบฐานลูกค้าเดิมและขยายฐานลูกค้าใหม่
- 5) ฝ่ายเอกสารส่งออก มีหน้าที่ดูแลรับผิดชอบจัดทำเอกสารส่งออกพัสดุ เช่น บัญชีราคาสินค้า (Commercial Invoice), บัญชีรายละเอียดบรรจุหีบห่อ (Packing List), ใบตราส่งสินค้า (Bill of Lading) ฯลฯ ให้มีความถูกต้องแม่นยำ เพื่อป้องกันความผิดพลาดที่จะส่งผลให้พัสดุอาจติดค้างหรือถูกตีกลับ
- 6) ฝ่ายพัสดุ มีหน้าที่ดูแลรับผิดชอบในส่วนของการพัสดุทั้งหมดที่บริษัทได้รับมา ตั้งแต่การรับพัสดุจากลูกค้าจนถึงการส่งต่อไปยังคลังเก็บพัสดุของ DHL โดยพนักงานตรวจนับและชั่ง

นำนักพัสดุเป็นผู้ทำหน้าที่ตรวจสอบรายละเอียดทั้งหมดของพัสดุที่ได้รับจากลูกค้า ทั้งในเรื่องของจำนวน น้ำหนัก ประเภท ฯลฯ ว่าเป็นไปตามข้อมูลที่ลูกค้าได้แจ้งหรือทำการตกลงกับพนักงานฝ่ายขายเอาไว้ เพื่อทำการจัดส่งต่อไปยังคลังพัสดุของ DHL ต่อไป และพนักงานขับรถขนส่งพัสดุ ทำหน้าที่ให้บริการลูกค้าในการเข้ารับพัสดุตามความสะดวกของลูกค้า เช่น อาจเป็นที่บ้าน อาคารสำนักงาน ฯลฯ สำหรับกรณีที่ลูกค้าไม่สะดวกนำพัสดุเข้ามาที่บริษัทด้วยตนเอง

- 7) ฝ่ายเทคโนโลยีสารสนเทศ (Information Technology) มีหน้าที่บริหารและจัดการระบบคอมพิวเตอร์ในองค์กร รวมถึงดูแลด้านโครงสร้างพื้นฐานของเทคโนโลยีสารสนเทศ และระบบสารสนเทศให้สามารถใช้งานได้อย่างมีประสิทธิภาพอยู่เสมอ

3.3 ลักษณะและการดำเนินงานขององค์กร

บริษัท Pigeon Air Freight ดำเนินธุรกิจโดยเป็นผู้ให้บริการจัดส่งพัสดุไปต่างประเทศ ให้บริการจัดส่งทั้งหมด 2 แบบ ได้แก่ การส่งพัสดุด่วน 3-7 วัน และการส่งพัสดุด่วนพิเศษ 1-5 วัน โดยมีขั้นตอนการดำเนินธุรกิจดังนี้

- 1) พนักงานฝ่ายขายและฝ่ายบริการลูกค้า ให้ข้อมูลกับลูกค้า รวมถึงติดตามความคืบหน้ากรณีที่ลูกค้าสนใจใช้บริการเพื่อให้สามารถปิดการขายกับลูกค้าได้สำเร็จ
- 2) นัดวันเวลาในการเข้าไปรับพัสดุจากลูกค้า หลังจากทีลูกค้าตกลงสมัครใช้บริการ ชำระเงิน และมีการให้ข้อมูลเบื้องต้นเกี่ยวกับพัสดุและประเทศปลายทางแล้วผ่านทางเว็บไซต์ อีเมลล์ หรือโทรศัพท์
- 3) นำพัสดุเข้ามาที่บริษัทเพื่อชั่งน้ำหนักและจัดทำเอกสาร ตรวจสอบความเรียบร้อยของข้อมูลในเอกสารให้ตรงกับความเป็นจริงเพื่อลดความผิดพลาดหรือปัญหาในการที่พัสดุอาจติดค้างหรือถูกตีกลับ
- 4) ส่งต่อพัสดุเข้าสู่คลังของ DHL เพื่อทำการจัดส่งต่อไป
- 5) พนักงานฝ่ายบริการลูกค้าให้ข้อมูลแก่ลูกค้าในกรณีที่ลูกค้าต้องการติดตามสถานะพัสดุที่อยู่ระหว่างการจัดส่ง

บริการของบริษัทมีลักษณะเด่นคือช่วยให้ลูกค้าที่เป็นบุคคลทั่วไปหรือผู้ประกอบการธุรกิจรายย่อยอย่างธุรกิจ SME สามารถส่งพัสดุไปต่างประเทศได้ในอัตราราคาที่ต่ำกว่าการที่ลูกค้าจะไปสมัคร Account กับทาง DHL โดยตรง นอกจากนี้บริษัทยังให้บริการดูแลการจัดส่งตั้งแต่ต้นทางจนถึงปลายทาง กล่าวคือตั้งแต่เรื่องของการเอกสารส่งออก การบรรจุพัสดุดลงกล่อง การเข้ารับพัสดุ ไปจนถึงการจัดส่งให้ถึงมือผู้รับปลายทาง

3.4 ปัญหาในการดำเนินงาน

- 1) พนักงานต้องแบ่งเวลาในการตอบคำถามที่ซ้ำซ้อน ทั้งกับลูกค้าเดิมและลูกค้าใหม่ เช่น เรื่องการสอบถามอัตราราคา การตรวจสอบสถานะพัสดุที่อยู่ระหว่างการจัดส่ง การบรรจุพัสดุดลงกล่อง การเข้ารับพัสดุ ฯลฯ ส่งผลให้พนักงานต้องเสียเวลาค่อนข้างมากในแต่ละวัน
- 2) ไม่ทราบรายละเอียดว่าคำถามที่ลูกค้ามักจะถามเข้ามานั้นเป็นอย่างไรหรือเกี่ยวกับเรื่องใดมากเป็นพิเศษหรือไม่เนื่องจากยังไม่มีระบบการเก็บข้อมูลในด้านนี้อย่างจริงจัง มีเพียงพนักงานเป็นผู้ให้ข้อมูลลูกค้าเท่านั้น
- 3) พบปัญหาลูกค้ากลุ่มลูกค้าเดิมบางส่วนเลิกใช้บริการ (Churn) ในจำนวนที่มีนัยสำคัญและไม่ควรมองข้าม แต่ยังไม่ทราบสาเหตุที่แน่ชัดว่าเกิดจากปัจจัยอะไรบ้างหรือแนวโน้มในอนาคตจะเป็นอย่างไร
- 4) ผู้บริหารมีข้อมูลไม่มากพอในการกำหนดกลยุทธ์เกี่ยวกับลูกค้าใหม่ซึ่งเข้ามาใช้งานเว็บไซต์ เนื่องจากลูกค้าบางส่วนมีการกดสมัครใช้บริการ แต่บางส่วนกลับไม่สมัครใช้งานหลังจากที่ได้เข้าชมเว็บไซต์แล้ว
- 5) ผู้บริหารและฝ่ายการตลาดยังมองไม่เห็นภาพรวมที่ชัดเจนว่ากลุ่มผู้ที่เข้ามาใช้งานเว็บไซต์ของบริษัทนั้นมีลักษณะอย่างไร ส่งผลให้ไม่สามารถวางกลยุทธ์ทางการตลาดหรือส่งเสริมการใช้บริการได้มากเท่าที่ควร

บทที่ 4

การพัฒนาระบบงาน

ในบทนี้จะกล่าวถึงคุณสมบัติระบบงาน ความต้องการระบบงาน รายละเอียดระบบงาน และการออกแบบระบบงาน ของโครงการนี้

4.1 คุณสมบัติระบบงาน

ระบบงานในการพัฒนาโครงการ “ระบบแชทบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” มีคุณสมบัติดังต่อไปนี้

4.1.1. การติดต่อกับผู้ใช้แชทบอต (Chatbot User Interface)

- 1) ระบบต้องสามารถแสดงผลของการพูดคุยได้อย่างง่ายดายทั้งทางโทรศัพท์มือถือและหน้าจอ PC และรองรับการพูดคุยเป็นภาษาไทยได้
- 2) ระบบต้องรองรับการใช้งานได้พร้อมกันหลายคน

4.1.2. การติดต่อกับผู้ใช้การวิเคราะห์ข้อมูลขั้นสูง (Advanced Data Analytics System User Interface)

- 1) มีการติดต่อระหว่างผู้ใช้งานกับระบบผ่านรูปแบบ Graphic User Interface (GUI)
- 2) สามารถบันทึกไฟล์โมเดล (Export) และรองรับการใช้งานไฟล์ได้หลายประเภท เช่น ไฟล์ Excel, PNG, JPG, PDF ฯลฯ
- 3) ต้องสามารถเก็บรวบรวมข้อมูลไว้ภายใต้ฐานข้อมูลเดียวกัน โดยระบบที่พัฒนาขึ้นมีการรวบรวมข้อมูลจากฝ่ายต่าง ๆ ที่เกี่ยวข้อง โดยข้อมูลที่ได้มาจะถูกจัดเก็บให้เป็นมาตรฐานเดียวกันและจัดเก็บไว้ในฐานข้อมูลเดียวกัน เพื่อความถูกต้องและลดความซ้ำซ้อนของข้อมูล ซึ่งทำให้การวิเคราะห์ข้อมูลเป็นไปอย่างมีประสิทธิภาพ
- 4) ระบบต้องให้ความสะดวกต่อผู้ใช้งาน มีการแสดงผลได้หลากหลายรูปแบบ เพื่อให้ผู้ใช้งานสามารถวิเคราะห์ข้อมูลได้หลากหลายมุมมอง
- 5) มีวิธีการเตรียมข้อมูล (Pre-process) และการวิเคราะห์ได้หลากหลายรูปแบบ

4.2 ความต้องการของระบบงาน

จากการศึกษาวิเคราะห์ระบบงานของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ รวมถึงความต้องการของผู้ใช้ระบบงาน สามารถสรุปความต้องการโดยรวมของระบบได้ดังนี้

- 1) สามารถให้บริการลูกค้าจำนวนมากได้อย่างรวดเร็วและแม่นยำภายใต้ต้นทุนที่ไม่สูง ทั้งยังช่วยลดภาระงานของพนักงานได้
- 2) สามารถจัดเก็บและวิเคราะห์ข้อมูลเกี่ยวกับคำถามของลูกค้าได้อย่างครบถ้วนและเป็นมาตรฐาน ส่งผลให้ช่วยเพิ่มประสิทธิภาพในการทำงานในอนาคต
- 3) สามารถเพิ่มโอกาสที่ผู้เข้าชมเว็บไซต์จะกดสมัครใช้บริการของบริษัทได้อย่างรวดเร็ว และสามารถจัดทำโปรโมชั่นให้เหมาะกับลักษณะหรือความต้องการของลูกค้าได้
- 4) สามารถมองเห็นภาพรวม ลักษณะเชิงลึก และพฤติกรรมของกลุ่มผู้ที่เข้ามาใช้บริการเว็บไซต์ของบริษัท เพื่อใช้เป็นข้อมูลในการทำนายแนวโน้มการเพิ่มขึ้นของลูกค้าใหม่ วางแผนกลยุทธ์ รวมถึงเพิ่มสัดส่วนจำนวนลูกค้าใหม่ที่จะสมัครใช้บริการกับทางบริษัทได้
- 5) สามารถมองเห็นแนวโน้มที่ลูกค้าเดิมอาจเลิกใช้บริการ (Churn) และกำหนดแนวทางในการแก้ไขหรือป้องกัน เพื่อลดอัตราการเลิกใช้บริการของลูกค้าได้อย่างทัน่วงที

4.3 รายละเอียดระบบงาน

การพัฒนาระบบของโครงการ “ระบบแชทบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” ประกอบด้วย 5 ระบบ และมีรายละเอียดของแต่ละระบบดังนี้

4.3.1 ระบบแชทบอต (Chatbot System)

ภาพรวมของระบบ (System Overview)

ระบบนี้จะครอบคลุมการให้บริการในด้านการตอบคำถามแก่ลูกค้า กล่าวคือให้ข้อมูลที่ลูกค้าต้องการโดยเป็นคำถามที่มักพบบ่อยตั้งแต่ต้นจนจบกระบวนการส่งจนถึงการชำระเงิน ไปจนถึงการ Tracking พัสดุ เพื่อช่วยลดภาระงานของพนักงานและเพิ่มความพึงพอใจของลูกค้า (Customer Satisfaction) ให้ลูกค้าเปรียบเทียบราคาได้ง่าย ให้ข้อมูลที่สะดวก รวดเร็ว และถูกต้องแม่นยำโต้ตอบได้แบบ Real Time และใช้งานง่ายบนมือถือ เป็นอีกหนึ่งช่องทางในการติดต่อรวบรวมข้อมูลของลูกค้า ไม่ว่าจะเป็นความคิดเห็น อีเมล ฯลฯ ซึ่งจะลดโอกาสการได้ข้อมูลที่ผิดพลาดคลาดเคลื่อนจากการที่พนักงานขอข้อมูลจากลูกค้าผ่านทางโทรศัพท์ นอกจากนี้ การให้ลูกค้าทำแบบสอบถามที่ดู

จริงจังเป็นทางการในเว็บไซต์อาจก่อให้เกิดความรู้สึกเบื่อได้ แต่การใช้แชทบอตซึ่งมีลักษณะเป็น Persona โต้ตอบได้คล้ายมนุษย์นั้นอาจส่งผลให้การพูดคุยเพื่อขอความคิดเห็นหรือข้อติชมต่าง ๆ เป็นไปได้ง่ายมากยิ่งขึ้น รวมถึงเป็นอีกหนึ่งช่องทางในการทำการตลาด โดยใช้ส่ง Message ข้อมูลข่าวสารของบริษัท รวมถึงโฆษณาและโปรโมชั่นต่าง ๆ ไปให้ลูกค้าเพื่อกระตุ้นให้มาใช้บริการได้อีกทางหนึ่ง

แชทบอตและคุณค่าของแชทบอต (Chatbot and Its Values)

ในยุคที่ธุรกิจ E-Commerce กำลังเฟื่องฟู การสื่อสารระหว่างผู้ให้บริการและผู้บริโภคได้ถูกย้ายมาอยู่บนแพลตฟอร์มดิจิทัลกันมากขึ้น โดยเฉพาะการโต้ตอบผ่าน Messaging Application ที่เข้ามาเป็นส่วนหนึ่งของชีวิตมนุษย์เรา แต่เพื่อให้ธุรกิจบนโลกโซเซียลมีเดียสามารถเข้าถึงกลุ่มลูกค้าได้ในวงกว้าง ง่ายตาย และรวดเร็วขึ้น การใช้มนุษย์เป็นตัวกลางสื่อสารเพียงอย่างเดียวอาจไม่ทันการณ์ ดังนั้นแชทบอตที่เป็นนวัตกรรมปัญญาประดิษฐ์จึงเข้ามาช่วยรับหน้าที่ดังกล่าว และเริ่มเข้ามามีบทบาทต่อธุรกิจในยุคนี้อย่างหลีกเลี่ยงไม่ได้

จากผลสำรวจพบว่าผู้บริโภคนิยมใช้โปรแกรมแชทคุยธุรกิจ 65 เปอร์เซ็นต์ ใช้ Messaging Application สั่งซื้อสินค้าออนไลน์ 50 เปอร์เซ็นต์ และอีก 50 เปอร์เซ็นต์ ต้องการให้ธุรกิจเปิดทำการ 24 ชั่วโมง ดังนั้นการใช้งานบนแพลตฟอร์มออนไลน์จึงเข้ามามีบทบาทสำคัญต่อชีวิตคนในยุคนี้อย่างมาก และสำหรับธุรกิจออนไลน์ที่มีการแข่งขันสูงขึ้นเรื่อย ๆ จะหยุดยั้งในตลาดได้ ก็ต่อเมื่อสร้างประสบการณ์ใหม่ ๆ ให้แก่ลูกค้า รวมถึงสร้างความประทับใจที่ดึงดูดลูกค้าให้กลับมาใช้บริการซ้ำ ซึ่งสิ่งเหล่านี้ขึ้นอยู่กับเงื่อนไขเรื่องเวลาที่ทุกธุรกิจต้องสามารถตอบสนองความต้องการของลูกค้าได้อย่างรวดเร็ว ฉับไว พร้อมให้บริการตลอด 24 ชั่วโมง และนี่คือจุดเปลี่ยนสำคัญในการดึงแชทบอตเข้ามาใช้งานแทน Agent หรือมนุษย์

นอกจากนี้แชทบอตยังช่วยสร้าง Customer Experience ใหม่ ๆ เพื่อตอบสนองกลุ่มคนรุ่นใหม่ที่เป็นการใช้โทรไปยัง Call Center ทั้งยังสามารถให้บริการได้ตลอดเวลาไม่มีวันหยุด และกรณีที่มีการติดต่อเข้ามามาก็ยังสามารถเพิ่มการบริการได้ทันที ซึ่งถ้าเทียบกับ Agent แล้ว ต้องมีเวลาหาคนและฝึกอบรมเพื่อให้บริการ ซึ่งต้องใช้เวลา 1-2 เดือน จึงจะเริ่มทำงานได้ ดังนั้นจึงทำให้การใช้แชทบอตกลายเป็นช่องทางสำคัญในการบริการลูกค้าในปัจจุบัน และที่น่าสนใจคือทุกวันนี้มีการนำแชทบอตมาใช้งานในธุรกิจต่าง ๆ เพิ่มขึ้นเรื่อย ๆ ไม่ว่าจะเป็นในธุรกิจด้านอาหาร (Taco Bell)

อุตสาหกรรมสื่อ (CNN) การค้าปลีกและอีคอมเมิร์ซ (Sephora และ H&M) FinTech (SCB Chatbot) ฯลฯ (Affinity Solution, 2561) แต่สำหรับธุรกิจด้านการจัดส่งพัสดุยังไม่พบว่ามีการนำแชทบอทมาใช้อย่างแพร่หลายมากนัก มีเพียงผู้ให้บริการรายใหญ่อย่าง UPS เท่านั้นที่เปิดให้ใช้งานแชทบอทของบริษัทได้ผ่าน Amazon Alexa และ Skype แต่ยังไม่รองรับภาษาไทย ดังนั้นการเริ่มต้นพัฒนาแชทบอทเพื่อนำมาช่วยในด้านการให้บริการลูกค้าสำหรับธุรกิจนี้จึงเป็นแนวคิดที่น่าสนใจ รวมถึงจะสามารถสร้างประโยชน์ให้แก่ธุรกิจได้อย่างมาก

การแก้ปัญหาของบริษัทด้วยการใช้งานแชทบอท (The Usefulness of Chatbot)

จาก Function ความสามารถต่าง ๆ ของแชทบอท เมื่อมีการนำมาใช้งานจะสามารถช่วยลดปัญหาที่พบเกี่ยวกับการให้บริการลูกค้าของบริษัท Pigeon Air Freight ได้ ดังตารางที่ 4-1

ตารางที่ 4-1: การแก้ปัญหาของบริษัทด้วยการใช้งานแชทบอท

เหตุการณ์ที่พบ	การให้บริการแบบเดิม	เพิ่มการให้บริการด้วยแชทบอท
1. กรณีที่ลูกค้าต้องการสอบถามข้อมูลต่าง ๆ		
- ประเทศปลายทาง, อัตราค่าส่ง	ลูกค้าต้องเข้าไปหาข้อมูลผ่านทางเว็บไซต์ของบริษัท หรือติดต่อพนักงานเพื่อขอข้อมูลทาง Email, โทรศัพท์ ซึ่งจะส่งผลให้ลูกค้าได้ข้อมูลทีค่อนข้างจำกัดเนื่องจากในแต่ละครั้งลูกค้าต้องแจ้งข้อมูลและตัวเลขต่าง ๆ ให้พนักงานหรือเข้าไปกรอกเลขหน้าหนึ่งเพื่อให้ระบบคำนวณใหม่ในแต่ละครั้ง หากต้องการเปรียบเทียบราคา	ลูกค้าจะได้รับตารางภาพรวมซึ่งได้ข้อมูลเยอะกว่าทั้งในด้านของประเทศปลายทางที่บริษัทรับจัดส่งและอัตราค่าส่ง ซึ่งสำหรับลูกค้าใหม่อาจได้ใช้บริการในส่วนนี้มาก ขณะที่ลูกค้าเดิมหรือลูกค้าประจำอาจได้ใช้ส่วนที่เป็นการ Tracking พัสดุมากกว่า

เหตุการณ์ที่พบ	การให้บริการแบบเดิม	เพิ่มการให้บริการด้วยแชทบอต
- พัสดุที่ไม่รับส่ง, บริการ Packing, ประกันภัย, รูปแบบการชำระเงิน, การติดต่อบริษัท, การเข้าไปรับพัสดุ รวมถึงโปรโมชั่นต่าง ๆ	พนักงานต้องเสียเวลาตอบคำถามซ้ำซ้อน หรือทำกระบวนการซ้ำ ๆ อย่างการช่วยตรวจสอบข้อมูลให้ ทั้งที่ในความเป็นจริง บางอย่างก็เป็นสิ่งที่ลูกค้าสามารถทำได้ หรือหากทำเองอาจสะดวกกว่า การโทรถามพนักงานเสียด้วยซ้ำ	ลูกค้าสามารถพิมพ์คำถามได้โดยไม่ต้องกดเข้าไปหาข้อมูลเองในเว็บไซต์หรือโทรศัพท์ถามพนักงาน จึงถือเป็นการประหยัดเวลาของลูกค้า สำหรับกรณีที่ต้องการข้อมูลเบื้องต้นเพื่อประกอบการตัดสินใจ
- การติดตามพัสดุที่อยู่ระหว่างการจัดส่ง (Tracking)	ลูกค้าต้องโทรศัพท์สอบถามพนักงานในช่วงเวลาทำการ หรือกดเข้าไปยังเว็บไซต์ของ DHL เท่านั้น	ลูกค้าสามารถเข้ามาสอบถามข้อมูลได้ตลอดเวลาจากแชทบอตซึ่งจะช่วย Link ไปยังหน้า Tracking ในระบบของ DHL ให้ทันที
2. การขอข้อมูลจากลูกค้า		
- การขอช่องทางในการติดต่อ	พนักงานพูดคุยกับลูกค้าผ่านโทรศัพท์เพื่อขอ Email หรือช่องทางติดต่อ ซึ่งก่อให้เกิดปัญหาทั้งความล่าช้าและอาจเกิดความผิดพลาดจากการสะกดคำ	ลูกค้าสามารถพิมพ์ช่องทางติดต่ออย่าง Email ไว้ในแชตได้เลยทำให้ง่ายต่อการรวบรวมข้อมูลในการติดต่อลูกค้าไปในภายหลัง

เหตุการณ์ที่พบ	การให้บริการแบบเดิม	เพิ่มการให้บริการด้วยแชทบอต
- Comment จากลูกค้า หลังได้รับบริการจาก บริษัท	สำหรับการขอความคิดเห็นของ ลูกค้าผ่านทางเว็บไซต์, Email, หรือโทรศัพท์นั้นอาจไม่น่าสนใจ มากพอ	การได้พูดคุยกับแชทบอตซึ่งมี ลักษณะของ Persona โต้ตอบได้คล้ายมนุษย์ จึงอาจ ช่วยเพิ่มความน่าสนใจในการ สนทนาขอความคิดเห็นได้ มากขึ้น หรือเป็นการเพิ่ม Customer Engagement ได้ อีกทางหนึ่ง
3. การทำการตลาด		
การเพิ่มโอกาสทำให้ลูกค้า รู้จักบริษัท / กระตุ้นให้ ลูกค้าอยากใช้บริการ	ส่งอีเมลแจ้งข่าวสาร หรือ พนักงานโทรศัพท์สอบถาม พูดคุยรวมถึงให้รายละเอียดแก่ ลูกค้า	สามารถใช้แชทบอตส่ง ข้อความเกี่ยวกับโปรโมชั่น ต่าง ๆ Reply ให้กับลูกค้า หรืออาจมีการต่อยอดพัฒนา ต่อในอนาคตให้สามารถส่ง Push Message ให้กับลูกค้า หรือขยายไปยัง Platform อื่น ๆ อย่าง Facebook Messenger, Twitter, Skype. เป็นต้น

รายละเอียดเกี่ยวกับ DialogFlow ซึ่งใช้ในการสร้างแชทบอต (DialogFlow in Detail)

DialogFlow เป็น Platform สำหรับทำแชทบอตหรือหนึ่งใน Software-As-A-Service (SaaS) ของ Google โดยเป็น Intent-Based Chatbot คือ แชทบอตที่หาความต้องการ (Intent) ของผู้ใช้จากข้อความพิมพ์มา แล้วหาคำตอบที่เหมาะสมสำหรับ Intent นั้น ๆ ซึ่ง DialogFlow สามารถจัดกลุ่มประโยคที่มีความต้องการ (Intent) เหมือนกันให้แชทบอตเรียนรู้ และกำหนดคำตอบ

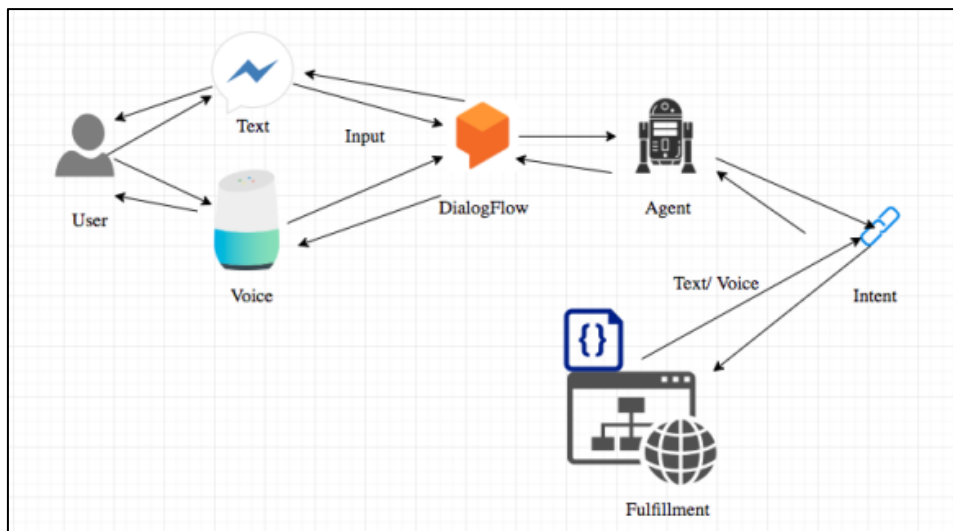
สำหรับแต่ละ Intent ได้ ถ้าคำตอบที่กำหนดไว้มีหลายแบบ DialogFlow จะสุ่มเลือกคำตอบให้แก่ผู้ใช้งานเอง และจุดเด่นของ DialogFlow มีดังนี้

- ใช้งานได้โดยไม่ต้องเขียนโปรแกรม เพียงแต่การเขียนโปรแกรมจะช่วยให้สามารถปรับแต่งการโต้ตอบของแชทบอทให้มีความน่าสนใจ ชับซ้อน หรือยืดหยุ่นมากขึ้นได้ (เรียกว่า Flex Message)
- รองรับการทำ Natural Language Processing (NLP) มากกว่า 20 ภาษารวมถึงภาษาไทย
- รองรับการเชื่อมต่อ Facebook Messenger, Twitter, LINE และเว็บไซต์ของตนเอง
- สามารถใช้งานได้โดยไม่ต้องเสียค่าใช้จ่าย

หลักการทำงานของ DialogFlow

หลักการทำงานของ DialogFlow มีดังนี้ (Programmerbonn, 2562)

- 1) ผู้ใช้พิมพ์ข้อความ Chat ผ่าน LINE, Facebook (หรือ App, Web Browser)
- 2) Text จะถูกส่งผ่าน Platform อย่าง LINE Message API หรือ Facebook Message API (ตาม Platform ที่ผู้ใช้ใช้งาน) เช่น LINE Message API Configuration
- 3) เสร็จแล้วจะถูก Hook หรือส่งต่อไปยัง DialogFlow ซึ่งเป็น Cloud Service ของ Google
- 4) DialogFlow จะทำหน้าที่สองอย่างหลัก ๆ คือตีความหมายของข้อความ (Intent) นั้นให้เข้าใจ เสร็จแล้วจะตอบกลับผลลัพธ์ที่ได้จากการ Configure หรือการเขียนโปรแกรม ซึ่งต่อกับระบบหลังบ้านกลับไป User ดังรูปที่ 4-1 และการเชื่อมต่อกับ DialogFlow นั้นไม่จำกัดที่ LINE หรือ Facebook เพียงเท่านั้น แต่ยังสามารถ Integrate กับ Custom App และช่องทางอื่น ๆ ได้อีกมากมาย



รูปที่ 4-1: หลักการทำงานของแชทบอตที่พัฒนาด้วย DialogFlow (Programmerbonn, 2562)

การส่งข้อมูลผ่าน Webhook

Webhooks เป็น HTTP POST ที่คอยแจ้งเตือนความเคลื่อนไหวทั้งหมดที่เกิดขึ้นภายในบัญชี (Trigger Event) ของผู้ใช้ไม่ว่าจะจาก API หรือบน Dashboard โดยข้อมูลที่ระบบแจ้งเตือนจะถูกส่งจากเซิร์ฟเวอร์กลับมายัง URL ที่ได้ตั้งไว้ (ซึ่งสามารถเปลี่ยน URL ได้) และส่วนมากจะนิยมนำไปใช้กับการทำงานของระบบ API บนระบบแอปพลิเคชันต่าง ๆ (Omise, 2562)

การตั้งค่าการทำงานหลักของ DialogFlow

การตั้งค่าการทำงานหลักของ DialogFlow สามารถทำได้ตามตารางที่ 4-2 และแสดงได้ดังรูปที่ 4-2

ตารางที่ 4-2: การตั้งค่าการทำงานหลักของ DialogFlow

Function	การทำงาน
Intent	<p>เป็นการตีความหมายของข้อความจากผู้ใช้ เพื่อให้แชทบอทสามารถเรียนรู้และเข้าใจความหมายของประโยคได้ เช่น เมื่อ Input เป็น "สวัสดี", "ดี ๆ", "ดีค่ะ" ระบบจะเข้าใจได้ว่าข้อความเหล่านี้เป็นข้อความทักทาย ซึ่งจะ Map กับ กลุ่มคำที่ตั้งไว้ในระบบ (Intents) และตอบกลับมาเป็นประโยคตามที่ถูกกำหนดไว้ และสามารถกำหนด Default Fallback Intent คือ เมื่อข้อความที่เข้ามาไม่ Match กับ Intent ใดเลยจะส่งต่อให้เข้ามาทำงานใน Intent นี้ได้ว่า จะให้ตอบอะไรกลับไป ซึ่งหากกำหนดให้แชทบอทขอข้อมูลจากผู้ใช้เพิ่มเติมก็จะช่วยเพิ่มโอกาสการพบคำถามที่แชทบอทจะสามารถตอบได้ โดย Intent มีส่วนประกอบสำคัญ ได้แก่</p>
	<ul style="list-style-type: none"> - Context มีลักษณะคล้ายบริบทการสนทนาในชีวิตจริงของมนุษย์ กล่าวคือ เป็นส่วนที่กำหนดว่า Input มาจาก Intent ใด และ Output ไปที่ Intent ใดต่อ (Follow-up Intent) เช่น หากลูกค้าตอบใน Intent เริ่มต้นว่า ต้องการซื้อสินค้าแล้ว ระบบก็จะส่งต่อไปยัง Intent ถัดไปเพื่อถามรายละเอียดว่าต้องการสินค้าลักษณะใดและไปต่อยัง Intent ชำระเงิน เป็นต้น ซึ่งในส่วนของ Output จะมีสิ่งที่เรียกว่า Lifespan ว่าหาก User คนนั้นตอบไม่ตรงตาม Flow ก็จะถูกหลุดออกไปจาก Flow แล้วเริ่มใหม่ โดย Default ของ Lifespan คือ 2 และสามารถแก้ไขตัวเลขได้
	<ul style="list-style-type: none"> - Event จะเป็นการผูกเหตุการณ์เข้ากับ Intent หนึ่ง ๆ นั้นหมายความว่าเมื่อ Action นี้ถูกเรียกใช้งาน Intent ที่ถูกผูกไว้ก็จะทำงานเป็นอันดับแรก แต่สำหรับกรณีที่ User ส่ง Event อื่นที่ไม่ใช่ Text เข้ามา เช่น อาจเป็น Sticker, รูปภาพ ฯลฯ จะต้องอาศัยการเขียนโปรแกรมเชื่อมต่อกับ Firebase และ Google Cloud เพิ่มเติมเพื่อแปลง Event เหล่านั้นให้

Function	การทำงาน
	<p>กลายเป็น Text Event เสียก่อน</p> <p>- Training Phrases เป็นคำหรือข้อความที่ใช้ในการสอนแชทบอตว่าข้อความใดบ้างที่จะเข้ามาในบริบทนี้ เช่น คำทักทาย มีอะไรบ้างที่ User อาจพิมพ์เข้ามา ซึ่งในส่วนนี้สามารถใส่ข้อความได้หลากหลาย และ DialogFlow จะนำข้อความเหล่านี้ไปทำ Natural Language Understanding (NLU) เพื่อหารูปแบบข้อความอื่น ๆ ที่ผู้ใช้อาจพิมพ์มาด้วย</p> <p>- Action & Parameter เป็นส่วนที่จะรับคำที่ User พิมพ์เข้ามา เพื่อเก็บไว้ในตัวแปรแล้วนำมาใช้ต่อ เช่นกรณีที่ User กรอกส่วนสูง หรือน้ำหนักเข้ามาซึ่งไม่ใช่ข้อความพูดคุยตามปกติ</p> <p>- Responses เป็นส่วนที่ใช้กำหนดว่าหากข้อความที่ User ส่งเข้ามา Match กับ Training Phrases ใน Intent นี้แล้วแชทบอตควรตอบอะไรกลับไป โดยสามารถเลือกการตอบกลับหาผู้ใช้ได้ 2 วิธี นั่นคือ</p> <ol style="list-style-type: none"> 1) Text Response: รับส่งข้อความแบบ Text เป็นแบบถามมาตอบไป โดยแชทบอตจะสามารถสุ่มเลือกคำตอบได้หากกำหนดให้มีมากกว่า 1 คำตอบ 2) Custom Payload: เป็นการตอบกลับด้วยการเขียน JSON Object ลงไป กล่าวคือต้องใส่เป็น JSON ตาม Template ของ Platform ต่าง ๆ ซึ่งการโต้ตอบอาจเป็นรูปแบบอื่น ๆ ที่ไม่ใช่ข้อความ เช่น Button Template, Flex Message หรือการส่งรูปภาพ โดยต้องมีการเขียนโค้ดและอัปโหลดรูปภาพขึ้นไปไว้บน Server ก่อน เช่นที่ https://www.picz.in.th เพื่อให้ระบบสามารถดึงข้อมูลมาแสดงผลให้แก่ User ได้

Function	การทำงาน
<p>Entities</p>	<p>Entity คือคุณลักษณะของตัวแปรแต่ละตัวที่ DialogFlow สามารถดึงออกมาได้จากประโยคที่ส่งเข้ามาในระบบ ซึ่งจะมีอยู่ด้วยกัน 3 ประเภท ได้แก่</p> <p>1) System Entities : เป็น Entity พื้นฐานที่ DialogFlow มีมาให้ โดยระบบจะพยายามแกะคุณลักษณะของตัวแปรออกมาจากประโยคให้อัตโนมัติ เช่น การดึง Date และ Time ออกมาจากประโยคได้นั่นเอง</p> <p>2) Developer Entities : เป็น Entity ที่นักพัฒนาสามารถสร้างขึ้นมาตัวเองเพื่อจัดกลุ่มคุณลักษณะของกลุ่มคำหรือประโยคต่าง ๆ เข้าด้วยกัน โดยสามารถสร้าง Entity ประเภทนี้ต่อจาก Intent ที่ได้สร้างเอาไว้แล้ว แต่ Developer Entities จะถูก Detect และถูกดึงออกมาได้ก็ต่อเมื่อประโยคที่เข้ามามีคำหรือรูปประโยคที่ตรงกันทุกอย่างกับ Synonym เท่านั้น</p> <p>3) Session Entities : Entity ประเภทนี้จะเป็น Entity ที่อยู่ใน Level ของ Session โดยจะทำหน้าที่ในการผูก Session ID เข้ากับผู้ใช้คนนั้น ๆ ซึ่งจะทำให้สามารถเก็บค่าข้อมูลต่าง ๆ และส่งต่อไปยัง Intent ต่าง ๆ ได้ (Jirawatee, 2562)</p>
<p>Fulfillment</p>	<p>เป็นการเติมเต็มด้วย Webhook โดยเป็นส่วนที่ผู้ใช้สามารถกดปิด-เปิดการใช้งานได้ ซึ่ง Webhook หรือ Fulfillment จะเป็นส่วนของ Coding javascript เพิ่มเติม เนื่องจากในบางครั้งแช็ตบอตก็ต้องนำ Input จาก User มาใช้ในการคำนวณ หรือจำเป็นต้องดึงข้อมูลบางอย่างจาก Back End มาประมวลผล จึงมีส่วนนี้เข้ามาเติมเต็มให้แช็ตบอตมีความสมบูรณ์แบบมากยิ่งขึ้น โดยสามารถใส่ได้ 2 แบบคือใส่ Webhook ลงไปกับพิมพ์ลงไป Inline Editor ซึ่งเชื่อมกันกับ Cloud Function for Firebase</p>
<p>Integrations</p>	<p>DialogFlow สามารถเชื่อมต่อ Platform ได้อย่างหลากหลาย โดยประมาณทั้งหมด 17 Platform ไม่ว่าจะเป็น Facebook Messenger, Google</p>

Function	การทำงาน
	Assistant, Twitter, LINE ฯลฯ
Training	ในส่วนนี้จะบอกว่าแชทบอตได้เคยพบคำหรือข้อความหนึ่ง ๆ มากี่รอบแล้ว ในวันที่และช่วงเวลาใด และมีรายละเอียดอะไรเกิดขึ้นบ้าง
History	เป็นส่วนที่บอกว่า User คุณอะไรกับแชทบอตบ้าง ในวันที่เท่าใด และแชทบอตตอบคำถามได้หรือไม่
Analytics	ส่วนนี้จะแสดงข้อมูลต่าง ๆ เกี่ยวกับการใช้งานแชทบอต เช่น มี User เข้ามาคุยกับแชทบอตมากน้อยเพียงใด มีการเรียกใช้งาน Intent ไหนจำนวนเท่าใด รวมถึงเรื่องที่ว่า User ออกจากการใช้งานที่ Intent ไหนเป็นจำนวนมากน้อยเท่าใด



รูปที่ 4-2: กระบวนการทำงานของ DialogFlow (Petch Kruapanich, 2561)

ขั้นตอนการสร้างแชทบอตที่ใช้งานผ่าน LINE Application (Steps of Chatbot Development)

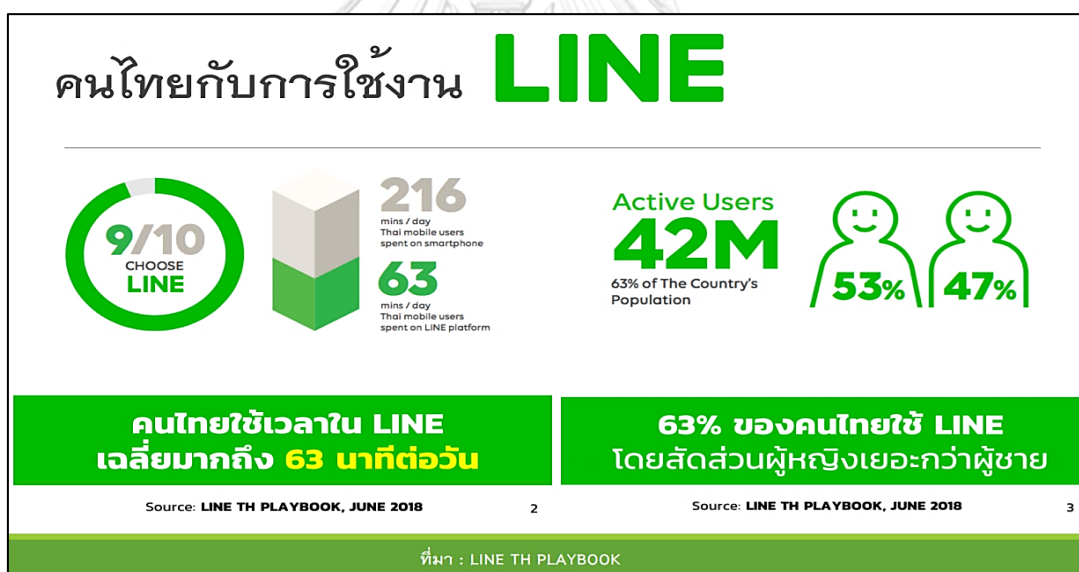
สำหรับการสร้างแชทบอตบน DialogFlow ซึ่งสามารถใช้งานผ่าน LINE Application ได้นั้น สามารถแบ่งออกเป็น 3 ขั้นตอนหลัก ดังนี้

- 1) การสร้าง Agent ที่ DialogFlow
- 2) สร้าง Intent และ Response ให้กับแชทบอท
- 3) เชื่อมต่อ Agent เข้ากับ LINE@ Account ผ่านส่วนของ Integrations

โดยสิ่งที่ต้องมีนั้น ได้แก่

- 1) LINE@ Account + Messaging API
- 2) DialogFlow Account : สามารถสมัครได้โดยไม่มีค่าใช้จ่ายที่เว็บไซต์
<https://DialogFlow.com/>

ทั้งนี้แชทบอทที่พัฒนาด้วย DialogFlow สามารถนำไปเชื่อมต่อได้กับหลากหลาย Platform ไม่ว่าจะเป็น Facebook, Twitter, Google Assistant, Skype ฯลฯ และ Platform หนึ่งซึ่งในปัจจุบันกำลังได้รับความนิยมอย่างมากก็คือ LINE Application ดังรูปที่ 4-3



รูปที่ 4-3: คนไทยกับการใช้งาน LINE

ที่มา: ข้อมูลจากโครงการสัมมนาวิชาการเรื่อง เทคโนโลยีห้องสมุดดิจิทัล 4.0, สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏพิบูลสงคราม,
<https://arit.kpru.ac.th/contents/edms2/admin-323-201903021551504265.pdf>

รายละเอียดเกี่ยวกับ LINE Messaging API (The Details of LINE Messaging API)

หลักการทำงานและประโยชน์ของของ API

API ย่อมาจาก Application Programming Interface คือช่องทางการเชื่อมต่อเพื่อแลกเปลี่ยนข้อมูลจากระบบหนึ่งไปสู่ระบบอื่น ๆ โดยผ่าน Library (Function/Module/Utility) ของผู้สร้างหรือให้บริการ API เพื่อให้ผู้พัฒนาระบบอื่น ๆ สามารถเชื่อมต่อกับ API ของผู้ที่เปิดให้บริการได้ API จึงเป็นเสมือนภาษาที่คอมพิวเตอร์ใช้คุยกับคอมพิวเตอร์เพื่อแลกเปลี่ยนข้อมูลกันระหว่างเครื่องข่าย (Server) กับผู้ใช้ (Client) หรือ Server กับ Server ก็ได้ (Click Next Company Limited., 2555) ดังรูปที่ 4-4

ประโยชน์ของ API

- สามารถรับส่งหรือเชื่อมต่อเพื่อแลกเปลี่ยนข้อมูลข้าม Server ได้
- ไม่จำเป็นต้องเข้าสู่หน้าเว็บไซต์หลัก ก็สามารถดึงข้อมูลผ่าน API ได้
- สามารถนำไปใช้งานในโปรแกรมประยุกต์ต่าง ๆ ได้
- ช่วยให้ผู้พัฒนาสามารถพัฒนาระบบ และทำงานได้รวดเร็วขึ้น

รายละเอียดของ LINE Messaging API

Line Messaging API ช่วยในการรับส่งข้อมูล อาทิ ข้อความ รูปภาพ และสติ๊กเกอร์ ระหว่าง Server ของโปรแกรมประยุกต์ที่ถูกพัฒนาขึ้นเข้ากับแพลตฟอร์มบนแอปพลิเคชันไลน์ เมื่อผู้ใช้งานส่งข้อความจะมีเหตุการณ์เกิดขึ้น (Event) และแพลตฟอร์มของไลน์จะส่งค่าข้อกลับไปยัง URL ที่ผู้ใช้งานกำหนดไว้ (Webhook) ในการใช้งานนี้จะต้องดำเนินการบน HTTPS ทั้งหมด และรูปแบบในการรับส่งระหว่าง Server ของผู้ใช้ และแพลตฟอร์มของไลน์ถูกกำหนดไว้ในรูปแบบของ JSON โดยหากต้องการใช้งานจะสามารถเพิ่มเพื่อน (Add Friend) ได้จาก Line ID หรือ QR Code ที่ถูกสร้างขึ้นภายใน Line Account (สิริวิชญ์ ธนาวิชญ์, 2561)

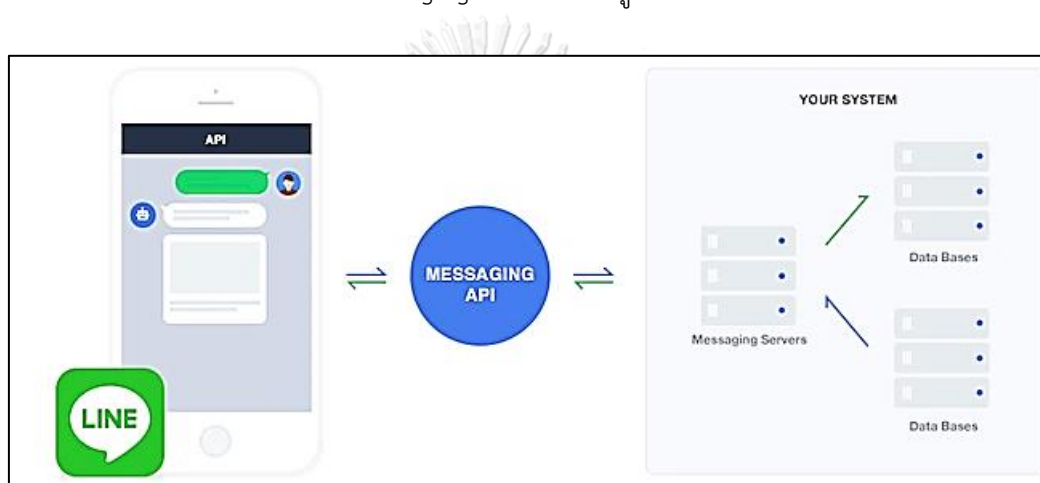
สำหรับแชทบอตบนแพลตฟอร์มของ LINE นั้นได้มีการเปิด Messaging API ให้ใช้งานได้ตั้งแต่ปีพ.ศ. 2559 และมีนักพัฒนาในไทยสนใจสร้างแชทบอตเพิ่มขึ้นเรื่อย ๆ โดย DialogFlow เป็น Platform ที่ให้บริการด้านแชทบอตโดยตรง มีความยืดหยุ่น และสามารถ Integrate เข้ากับโปรแกรม Chat ต่าง ๆ ได้หลากหลาย รวมทั้งไม่มีค่าใช้จ่าย

หลักการทำงานของ Line Messaging API

- 1) ผู้ใช้ส่งข้อความแชทผ่าน LINE BOT

- 2) ระบบส่ง Request ไปยัง URL ของ Webhook (Server) ที่กำหนดไว้
- 3) Server ทำการประมวลผล Request ที่ถูกส่งเข้ามา (รวมถึงกรณีที่มีการประมวลผลกับฐานข้อมูล)
- 4) Server ส่ง Respond ไปยัง LINE Platform เพื่อตอบสนองผู้ใช้งาน
- 5) การส่ง Request ต่าง ๆ จะถูกส่งผ่าน HTTPS ในรูปแบบของ JSON (Webhook URL จำเป็นต้องเป็น HTTPS)

กระบวนการทำงานของ Messaging API แสดงดังรูปที่ 4-4



รูปที่ 4-4: กระบวนการทำงานของ Messaging API

ที่มา: สอน+วิธีทำ Line Messaging Api, โดย Prapon Chinchon, 2560,

<https://medium.com/@prapon/%E0%B8%AA%E0%B8%AD%E0%B8%99-%E0%B8%A7%E0%B8%B4%E0%B8%98%E0%B8%B5%E0%B8%97%E0%B8%B3-line-messaging-api-ec29e8e9858e>

ความสามารถของ Messaging API

ปัจจุบัน LINE สามารถส่งข้อความหา User ได้ทั้งหมด 4 รูปแบบได้แก่ (Boontamthan, 2562)

- 1) Broadcast Messages: เป็นการส่งข้อความให้กับ Follower ทั้งหมด (All) จัดอยู่ในประเภทข้อความที่ส่งออกไปหา Follower ก่อน หรือ One-way (ทั้ง API และ CMS)

- 2) Push Messages: เป็นการส่งข้อความไปหา “ผู้รับที่เลือกไว้รายบุคคล” ด้วย API ได้ตลอดเวลาโดยไม่ต้องรอ Request จากผู้ใช้เข้ามา โดยจะต้องกำหนด User ID ของผู้รับคนนั้น ๆ เช่น การส่งหา Follower ที่ลงทะเบียนรับสิทธิพิเศษของทางร้าน หรือส่งหาลูกค้าที่เคยทำแบบสอบถาม เป็นต้น ถือเป็น การส่งแบบ One-way ไปยังผู้ใช้ โดยการใช้งานจะแตกต่างกันออกไปตามแพ็คเกจที่เลือกสมัครใช้งาน
- 3) Multicast Messages: เป็นการส่งข้อความไปหาผู้รับที่เลือกไว้ โดยต้องส่งด้วย API และต้องกำหนด User ID ของผู้รับนั้น ๆ เช่นกัน การส่งแบบนี้สามารถส่งได้สูงสุดครั้งละ 150 User ID และเป็นการส่งข้อความแบบ One-way
- 4) Reply Messages: (Text, Image, Video, Audio, Location, Sticker, Imagemap, Template, Flex Message) เป็นการตอบ Message กลับไปให้ User หลังจากที่ User ส่งข้อความมายังแชทบอตหรือเพิ่มแชทบอตเป็นเพื่อน ซึ่งผู้พัฒนาจำเป็นต้องระบุ Reply Token ไว้ใน Webhooks เพื่อที่จะส่งข้อความตอบกลับไปยัง User และการส่งข้อความแบบ Reply นี้หากนำ Reply Token ที่ได้รับจาก Webhooks มาใช้ Reply กลับจะสามารถโต้ตอบกันได้โดยไม่ต้องเสียค่าใช้จ่าย

ความสามารถของแชทบอตที่ทำการพัฒนา (Chatbot of This Project and Its Capabilities)

แชทบอตของบริษัท Pigeon Air Freight ซึ่งพัฒนาขึ้นด้วย DialogFlow เพื่อเน้นด้านการให้ข้อมูลลูกค้าเกี่ยวกับการจัดส่งพัสดุไปยังต่างประเทศจะครอบคลุมการตอบคำถามทั้งหมด 19 หมวดหมู่ตาม Intent ดังรูปที่ 4-5

<input type="checkbox"/> 01 WELCOME ▾	11 ปัญหา
02 Speedy Package	12 พื้นที่ทางไกล
03 Super Speedy Package	13 วันหยุดบริษัท
04 สอบถามอื่น ๆ	14 เอกสาร
05 บัญชีสูงสุด	15 DHL
06 การส่งสินค้า	16 สมัครใช้บริการ
07 โปรโมชั่น	17 การชำระเงิน
08 ของที่ไม่รับส่ง	18 ขอบคุณ
09 ประกันภัย	19 Comment
10 การติดต่อบริษัท	<input type="checkbox"/> 20 Default Fallback Intent

รูปที่ 4-5: หมวดหมู่ Intent ของระบบแชทบอตที่พัฒนา

โดยภายในแต่ละ Intent จะสามารถปรับแต่ง Response ให้มีความหลากหลายแตกต่างกันได้ และสำหรับคำถามที่พบบ่อยที่สุดจะมีการสร้างเป็นเมนูตั้งต้นในรูปแบบของ Flex Message ไว้ทั้งหมด 3 Intents ด้วยกัน ซึ่งจะเป็น Reply หลักที่ลูกค้าได้พบเมื่อเริ่มต้นการสนทนา (Welcome Intent) เช่น เมื่อพิมพ์คำว่า สวัสดี, ต้องการส่งของ เป็นต้น และสำหรับการส่งข้อมูลตอบกลับ (Response) ที่ไม่ใช่ Text นั้นต้องอาศัยการ Coding ในรูปแบบของ JSON Format ดังรูปที่ 4-6 ซึ่งจะทำให้เกิดการแสดงผลแบบปุ่มกดได้ด้วย Flex Message ใน LINE Application ดังรูปที่ 4-7 โดย Flex Message นั้นคือรูปแบบหนึ่งของข้อความที่ LINE Bot สามารถส่งได้ ซึ่งเพิ่งเปิดตัวไปเมื่อกลางปีพ.ศ. 2561 โดยนักพัฒนาสามารถออกแบบข้อความได้อย่างอิสระ ด้วยการสร้าง Payload ที่เป็น JSON แล้วส่งผ่าน Messaging API เพื่อมอบประสบการณ์ที่ดีกว่าในสื่อสารระหว่างผู้ใช้ และ Chatbot (Jirawatee, 2562)

```

Custom Payload
1 {
2   "line": {
3     "altText": "Flex Message",
4     "type": "flex",
5     "contents": {
6       "type": "bubble",
7       "header": {
8         "contents": [
9           {
10            "color": "#916004",
11            "type": "text",
12            "text": "สามารถกดเลือกรับข้อมูลได้เลยครับ ^ ^",
13            "align": "center",
14            "weight": "bold",
15            "size": "sm"
16          }
17        ]
18      }
19    }
20  }
21 }

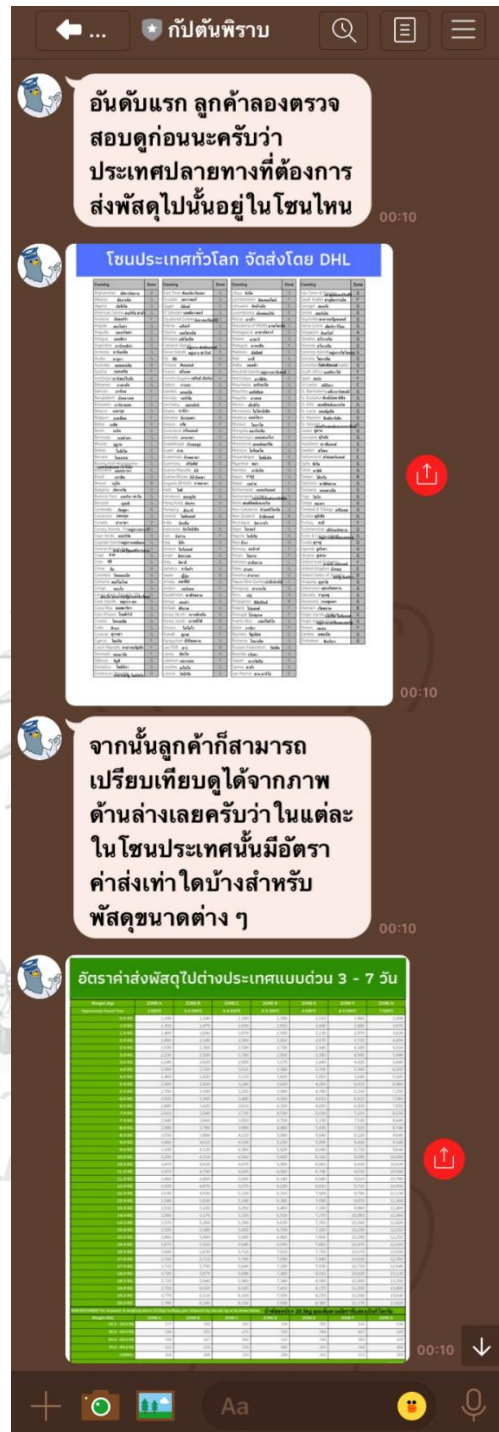
```

รูปที่ 4-6: การ Code ในรูปแบบของ JSON format



รูปที่ 4-7: การแสดงผลแบบปุ่มกดด้วย Flex Message ใน LINE Application

เมื่อกดเลือกรูปแบบของการจัดส่งแล้ว ลูกค้าน่าจะได้รับข้อมูลเกี่ยวกับโซนประเทศที่มีการจัดส่ง รวมถึงอัตราราคาค่าส่งแยกตามโซนประเทศและน้ำหนักของพัสดุ (โดยส่งเป็นรูปภาพให้แก่ลูกค้า เพื่อให้สามารถเปรียบเทียบราคาได้) ดังรูปที่ 4-8



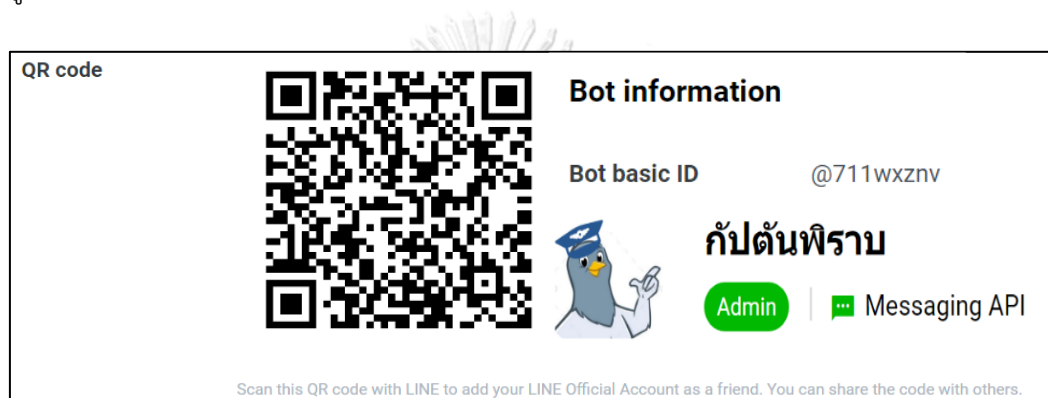
รูปที่ 4-8: ตัวอย่างการแสดงผลแบบรูปภาพใน LINE Application

และหากผู้ใช้กดเลือกตรงส่วนของ Tracking เพื่อตรวจสอบสถานะพัสดุที่อยู่ระหว่างการจัดส่ง หรือพิมพ์ข้อความสอบถามเกี่ยวพื้นที่ห่างไกล (Remote Area) ระบบก็จะทำการ Link ไปยังเว็บไซต์ของ DHL โดยตรงเพื่อให้ผู้ใช้สามารถรอกหมายเลขพัสดุหรือรายละเอียดต่าง ๆ ได้ทันที

ส่วนตัวเลือกที่เป็นคำถามเพิ่มเติมอื่น ๆ ลูกค้าสามารถพิมพ์คำถามซึ่งจะครอบคลุมทั้งหมด 20 หมวดหมู่หลัก ของ Intent เช่น เรื่องการประกันภัยพิบัติ เอกสารการส่งออก สินค้าต้องห้าม ระยะเวลาในการนำส่ง ฯลฯ โดยจะเน้นการโต้ตอบด้วยข้อความ (Text based) เป็นหลัก

ช่องทางในการใช้งานแชทบอท

สำหรับการใช้งานลูกค้าสามารถทำการเพิ่มเพื่อนเพื่อติดต่อกับแชทบอทได้ด้วย LINE ID หรือ การสแกน QR Code ซึ่งจะปรากฏให้เห็นได้ในเว็บไซต์ รวมถึงในอีเมลต่าง ๆ ที่บริษัทมีการส่งไปให้ ดังรูปที่ 4-9



รูปที่ 4-9: LINE ID และ QR Code สำหรับการเพิ่มแชทบอทเป็นเพื่อน

หลังจากนั้นลูกค้าจะสามารถพูดคุยกับแชทบอทได้ผ่าน LINE Application ทั้งจากในมือถือระบบ Android และ iOS รวมถึงคอมพิวเตอร์ และ Notebook

การพัฒนาปรับปรุงแชทบอทเพิ่มเติม

เนื่องจาก DialogFlow เป็นช่องทางหนึ่งที่เปิดโอกาสให้ผู้ใช้สามารถสร้างแชทบอทได้โดยไม่ต้องเขียนโค้ด แต่หากต้องการพัฒนาความสามารถของแชทบอทให้มีความซับซ้อนมากขึ้น อาจต้องอาศัยการ Coding หรือนำ Tools อื่น ๆ มาใช้งานเพิ่มเติม ไม่ว่าจะเป็น Google Cloud หรือ IBM Watson ซึ่งจะช่วยเสริมความสามารถต่าง ๆ ของแชทบอทให้มีความน่าสนใจมากขึ้นได้ ยกตัวอย่างเช่น

- 1) กรณีที่ต้องการเก็บข้อมูลของลูกค้าโดยละเอียดหรือดึงข้อมูลจากที่อื่นมาให้บอทตอบ เนื่องจากระบบของ DialogFlow จะเก็บเฉพาะข้อความที่ผู้ใช้พิมพ์มาเท่านั้น แต่ไม่ได้เก็บว่าใครเป็นผู้พิมพ์เข้ามา

- 2) เพิ่มเติมความสามารถของแชทบอทในการรับข้อความอื่น ๆ ที่ไม่ใช่ Text ผ่าน LINE Application เช่น ข้อความที่เป็นเสียง, Sticker, วิดีโอ ฯลฯ
- 3) การทำ NLP จากลูกค้าที่ได้เคยใช้งานแชทบอทไปสักระยะหนึ่ง

4.3.2 ระบบวิเคราะห์ข้อมูลจากการใช้งานแชทบอท (Data Analysis System from Chatbot Usage)

ภาพรวมของระบบ (System Overview)

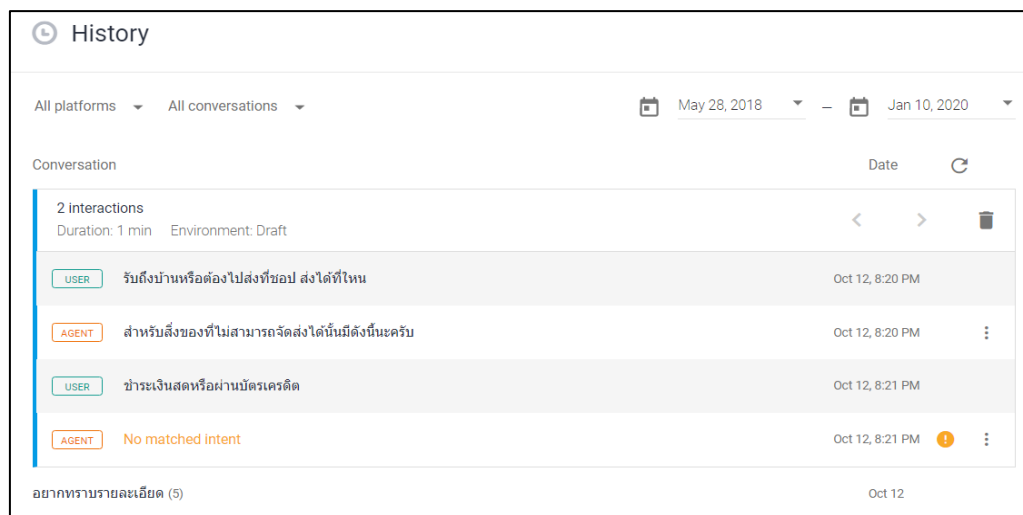
ระบบนี้จะเป็นการนำข้อมูลจากระบบแชทบอทมาทำการวิเคราะห์ว่าจากการใช้งานของลูกค้านั้นมีคำถามในประเด็นใดเกิดขึ้นมากเป็นพิเศษหรือไม่ และยังมีข้อบกพร่องใดที่ยังต้องปรับปรุงเพื่อเพิ่มประสิทธิภาพการทำงานของระบบแชทบอทให้ดียิ่งขึ้น เช่น ลูกค้าส่วนใหญ่ออกจากการใช้งานเมื่อใด มีคำถามอะไรบ้างที่แชทบอทยังตอบไม่ได้ ลูกค้าสนใจในการใช้งานแชทบอทมากเพียงใด และมีการแลกเปลี่ยนข้อมูลหรือ Engagement เกิดขึ้นระหว่างลูกค้ากับแชทบอทมากน้อยแค่ไหน ฯลฯ

ผู้ใช้ (Users)

- 1) ผู้บริหาร
- 2) ฝ่ายการตลาด
- 3) ฝ่ายบริการลูกค้า
- 4) ฝ่ายเทคโนโลยีสารสนเทศ

ข้อมูลนำเข้า (Input Data)

สำหรับข้อมูลนำเข้าของระบบนี้จะป็นข้อมูลที่ได้จากการทดลองใช้งานโดยลูกค้า โดยสามารถตรวจสอบได้จากในส่วนของ History ซึ่งจะเห็นได้ว่ามีข้อความบทสนทนาโต้ตอบเกิดขึ้นมากน้อยเพียงใดในแต่ละ Session และเกิดขึ้นในวันที่เท่าไรและเวลาไหน รวมถึงแชทบอทสามารถตอบคำถามได้หรือไม่ โดยหากแชทบอทพบคำถามที่ไม่สามารถตอบได้จะปรากฏเครื่องหมายอัศเจรีย์ในวงกลมสีเหลืองดังรูปที่ 4-10 ทำให้ทราบได้ว่ายังต้องมีการปรับปรุงตรงส่วนนี้เพิ่มเติม



รูปที่ 4-10: ข้อมูลที่ได้จากการทดลองใช้งานโดยลูกค้า

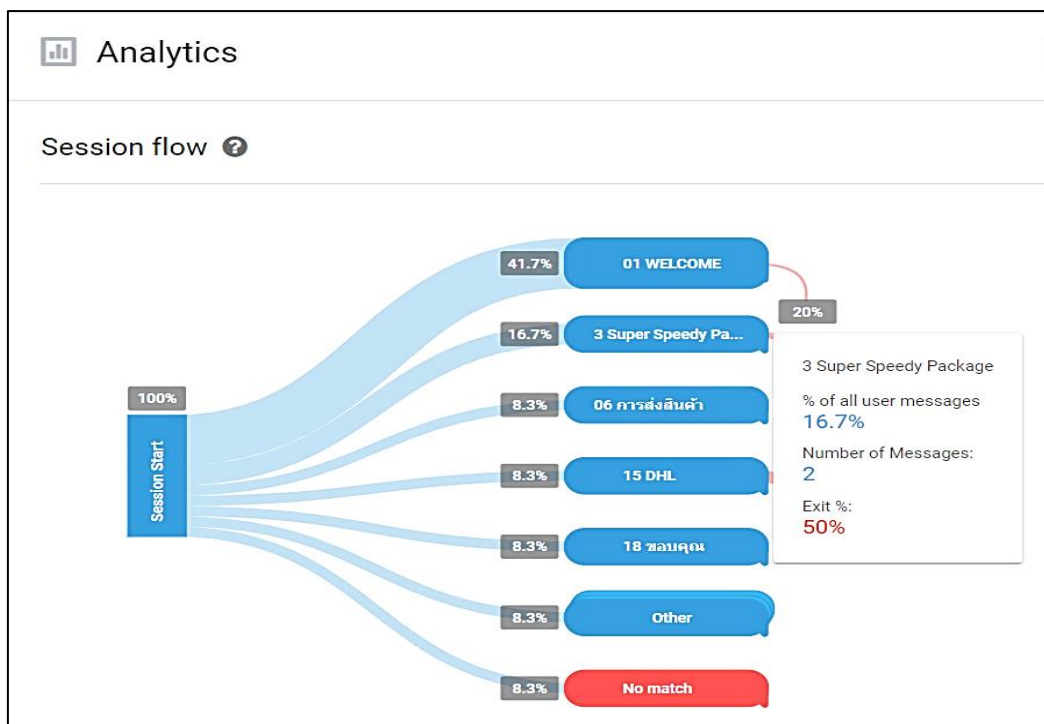
นอกจากนี้ ในส่วนของ Analytics ยังสามารถตรวจสอบได้อีกด้วยว่าภายในช่วงเวลาที่กำหนด ลูกค้ามีการใช้งานและเลิกใช้งานโดยออกจากแต่ละ Intent มากน้อยเพียงใด ดังรูปที่ 4-11

Intents ?			
Intent	Sessions	Count	Exit %
Welcome	17	42	21.43%
3 สอบถามอื่น ๆ	4	15	6.67%
การติดต่อ	7	14	28.57%
ตกลง	5	10	10%

รูปที่ 4-11: จำนวนการใช้งานแต่ละ Intent

และในส่วนของ Session Flow จะมีรูปที่สามารถเห็นถึงภาพรวมได้ว่าตลอดการใช้งานตั้งแต่ต้นจนจบของลูกค้า นั้นมีลักษณะอย่างไร และออกจากการสนทนาที่ Intent ใดเป็นจำนวนเท่าไร ซึ่งการพิจารณาจากประวัติการใช้งานโดยลูกค้าจะสามารถวิเคราะห์ถึงลักษณะความต้องการในด้านข้อมูลของลูกค้า รวมถึงสามารถช่วยในการพัฒนาปรับปรุงแชทบอตได้อีกด้วย ดังจะเห็นได้จากในรูปที่

4-12



รูปที่ 4-12: Session Flow แสดงภาพรวมการใช้งานตลอด Session ของลูกค้า

ขั้นตอนการดำเนินงาน (Working Processes)

- 1) รวบรวมข้อมูลจากการใช้งานของลูกค้าเพื่อใช้เป็นข้อมูลนำเข้า
- 2) จัดทำสรุปข้อมูลเพื่อพิจารณาทั้งในด้านคำถามของลูกค้ารวมถึงประสิทธิภาพการทำงานของแช็ตบอต
- 3) ทำความเข้าใจและวิเคราะห์ข้อความต่าง ๆ ที่ลูกค้าโต้ตอบกับแช็ตบอต รวมทั้งศึกษา ลักษณะแนวโน้มความต้องการของลูกค้าในแต่ละช่วงเวลาเพื่อเป็นแนวทางในการวางแผนดำเนินธุรกิจต่อไป
- 4) ปรับปรุงประสิทธิภาพการทำงานของแช็ตบอตเพื่อให้สามารถรวบรวมข้อมูลจากลูกค้าและบริการลูกค้าได้ดียิ่งขึ้นในอนาคต

4.3.3 ระบบทำนายการสมัครใช้บริการเว็บไซต์ (Website Subscription Prediction System) ภาพรวมของระบบ (System Overview)

ระบบทำนายการสมัครใช้บริการเว็บไซต์ (Website Subscription Prediction System) เป็นระบบที่พัฒนาขึ้นเพื่อใช้ในการทำนายว่าจากข้อมูลกลุ่มผู้ที่เข้าใช้งานเว็บไซต์ของบริษัทนั้น ผู้ใช้ที่

มีลักษณะและการใช้งานแบบใดที่มีแนวโน้มจะกดสมัครใช้บริการกับบริษัทในท้ายที่สุด เพื่อให้สามารถมองเห็นได้ว่าปัจจัยใดที่น่าจะส่งผลต่อการสมัครใช้บริการของลูกค้าได้บ้าง และดำเนินการเพื่อให้อัตราการสมัครใช้งานจากผู้เข้ามาเยี่ยมชมเว็บไซต์นั้นเพิ่มสูงขึ้น

ผู้ใช้ (Users)

- 1) ผู้บริหาร
- 2) ฝ่ายการตลาด
- 3) ฝ่ายขาย
- 4) ฝ่ายเทคโนโลยีสารสนเทศ

ข้อมูลนำเข้า (Input Data)

เป็นข้อมูลของผู้เข้าชมเว็บไซต์ทั้งหมดจำนวน 22,006 รายการ โดยมีตัวแปรที่นำมาใช้ทั้งหมด 4 ตัวแปรด้วยกัน ดังตารางที่ 4-3

ตารางที่ 4-3: ข้อมูลนำเข้าสำหรับการทำนายการสมัครใช้บริการ

ลำดับ	ชื่อตัวแปร	คำอธิบาย	หน่วย
1	Age	อายุของผู้ใช้	ปี
2	Days_Since_LastVisit	จำนวนวันตั้งแต่เข้าเว็บไซต์ครั้งก่อนจนถึงครั้งล่าสุด	วัน
3	Duration	ระยะเวลาที่เข้าใช้งานเว็บไซต์	นาที
4	Pages_viewed	จำนวนหน้าเว็บไซต์ที่เปิดใช้งาน	หน้า

ขั้นตอนการดำเนินงาน (Working Processes)

ในการดำเนินงานจะนำหลักการของ Cross-Industry Standard Process for Data Mining (CRISP-DM) ซึ่งเป็นแนวทางในการดำเนินงานในการวิเคราะห์ข้อมูลขั้นสูงมาใช้ โดยประกอบไปด้วย 6 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 การทำความเข้าใจทางธุรกิจ (Business Understanding)

ทำความเข้าใจเกี่ยวกับธุรกิจโลจิสติกส์และศึกษาขั้นตอนการดำเนินงานของหน่วยงาน รวบรวมปัญหาที่เกิดขึ้นในปัจจุบัน และความต้องการจากผู้บริหารและฝ่ายการตลาด เพื่อเป็นแนวทางในการพัฒนาระบบการทำนาย จากนั้นนำความต้องการของผู้ใช้งานและกลุ่มเป้าหมายมาวิเคราะห์ และสรุปเป็นความต้องการของระบบ พร้อมทั้งกำหนดขอบเขตของโครงการ เพื่อให้ครอบคลุมการแก้ไขปัญหาที่เกิดขึ้น

โดยจากการทำความเข้าใจในประเด็นเรื่อง “การทำนายการสมัครใช้บริการเว็บไซต์” แล้วสรุปได้ว่าผู้ใช้งานต้องการได้ข้อมูลแนวโน้มการสมัครใช้บริการเว็บไซต์เพื่อให้ทราบถึงลักษณะการใช้งานรวมถึงลักษณะเบื้องต้นของผู้ที่อาจกลายเป็นลูกค้าของบริษัทได้ในอนาคต เนื่องจากสำหรับธุรกิจโลจิสติกส์ในปัจจุบันต้องเน้นการแข่งขันเพื่อตอบสนองความต้องการของลูกค้าด้วยความรวดเร็วและสอดคล้องกับความต้องการ รวมถึงลักษณะของกลุ่มเป้าหมายให้ได้มากที่สุด ข้อมูลที่ได้จากการทำนายจึงจะช่วยสร้างความเข้าใจเชิงลึก (Insights) เกี่ยวกับกลุ่มผู้ใช้เว็บไซต์ของบริษัทซึ่งมีลักษณะที่แตกต่างกัน และเมื่อนำมาวิเคราะห์ต่อยอดก็จะส่งผลให้บริษัทสามารถเพิ่มโอกาสขยายฐานลูกค้าใหม่ผ่านทางเว็บไซต์ รวมถึงวางแผนทางการตลาดหรือกลยุทธ์การดำเนินธุรกิจได้ดียิ่งขึ้น

ขั้นตอนที่ 2 การทำความเข้าใจข้อมูล (Data Understanding)

รวบรวมข้อมูลทั้งหมดที่จำเป็นในการพัฒนาระบบการทำนาย และนำเข้าข้อมูลสู่ฐานข้อมูล (Data Base) จากนั้นจึงทำ Data Exploration เพื่อทำความเข้าใจกับข้อมูล และออกแบบรูปแบบข้อมูลที่เหมาะสมในการจัดทำ Predictive Model

โดยสำหรับระบบนี้พบว่าข้อมูลที่สามารถนำมาใช้ในการทำนายการสมัครใช้บริการเว็บไซต์ได้ดังนี้

- อายุของผู้ใช้ (Age)
- จำนวนวันตั้งแต่เข้าเว็บไซต์ครั้งก่อนจนถึงครั้งล่าสุด (Days_Since_LastVisit)
- ระยะเวลาที่เข้าใช้งานเว็บไซต์ (Duration)
- จำนวนหน้าเว็บไซต์ที่เปิดใช้งาน (Pages_viewed)

ขั้นตอนที่ 3 การเตรียมข้อมูล (Data Preparation)

- 1) **Data Cleansing** โดยตรวจสอบและลบข้อมูลที่อาจแปลกแยกจากข้อมูลอื่น (Outlier) ลบข้อมูลที่ผิดปกติและตรวจสอบข้อมูลที่หายไป เช่น ในกรณีตัวแปรระยะเวลาที่เข้าใช้งานเว็บไซต์ (Duration) บาง Record มีระยะเวลานานเกิน 500 นาที ซึ่งอาจเป็นไปได้ว่าเกิดความผิดพลาดของข้อมูล เนื่องจากมีค่าสูงผิดปกติและมีข้อมูลลักษณะนี้อยู่เป็นจำนวนน้อย จึงต้องทำการลบออกเพื่อไม่ให้ส่งผลกระทบต่อประสิทธิภาพในการทำนายของโมเดล นอกจากนี้ยังพบข้อมูลบาง Record ที่ตัวแปรบางส่วนขาดหายไป แต่มีจำนวนรวมแล้วไม่ถึง 10 Records ซึ่งอาจเกิดจากการพิมพ์ข้อมูลที่ตกหล่น จึงต้องทำการลบข้อมูลใน Record นั้น ๆ เพื่อไม่ให้ส่งผลต่อการสร้างโมเดลในขั้นตอนต่อไป
- 2) **Set Data Role** ตั้งค่าให้ข้อมูลในคอลัมน์ที่ต้องการใช้โมเดลทำนายเปลี่ยนจาก Regular เป็น Label ซึ่งในที่นี้คือข้อมูลในคอลัมน์ Subscription (การสมัครใช้บริการของลูกค้า)
- 3) **Data Sampling** เนื่องจากข้อมูลมีความ Imbalance จึงต้องทำการสุ่มข้อมูลที่จะนำมาใช้ในการ Train โมเดลโดยใช้เทคนิค Bootstrapping เพื่อทำการสุ่มชุดข้อมูลย่อยขึ้นมาในแต่ละครั้งแบบไม่ซ้ำกัน จากจำนวนของชุดข้อมูลตัวอย่าง (Sample Size) ซึ่งเป็นข้อมูลนำเข้าทั้งหมดจำนวน 22,006 Records
- 4) **Data Splitting** แบ่งข้อมูล Train / Validate / Test Set ตามรูปแบบของ Predictive Model โดยแบ่งตามอัตราส่วนที่ส่งผลให้โมเดลสามารถทำนายได้อย่างมีประสิทธิภาพสูงสุด

ขั้นตอนที่ 4 การสร้างโมเดล (Modeling)

สำหรับโมเดลในการทำนายข้อมูลนั้นจะสร้างขึ้นทั้งหมด 2 โมเดลด้วยกัน เพื่อให้สามารถนำผลลัพธ์ที่ได้ไปใช้ประโยชน์ในฝ่ายงานที่แตกต่างกันได้ โดยปัจจัยหลักในการพิจารณาคัดเลือกรูปแบบโมเดลที่จะสร้างนั้นมาจากคุณลักษณะเฉพาะของแต่ละโมเดลผนวกกับลักษณะของชุดข้อมูลที่จะนำมาใช้ในการทำนาย สำหรับระบบทำนายการสมัครใช้บริการเว็บไซต์แล้ว โมเดลซึ่งน่าจะมีความเหมาะสมที่สุดคือเทคนิคการทำนายด้วย Decision Tree Model และ Neural Network Model ซึ่งมีรายละเอียดดังต่อไปนี้

พารามิเตอร์ของโมเดล Decision Tree ในโปรแกรม RapidMiner

- 1) **Criterion:** เป็นส่วนที่ใช้กำหนดเกณฑ์ในการเลือกแอททริบิวต์ที่ต้นไม้อาจทำการแตกกิ่ง (Split) โดยจะมีการปรับค่าของการแตกกิ่งให้เหมาะสมตามเกณฑ์ที่เลือก ซึ่งจะสามารถเลือกได้หนึ่งเกณฑ์จาก 5 ตัวเลือกต่อไปนี้:

(1.) **Information_gain:** เป็นตัวชี้วัดการแบ่งข้อมูลออกเป็นชุดข้อมูลย่อย โดยหลักการก็คือยิ่งแอททริบิวต์ใดมี Information Gain สูงจะยิ่งเหมาะกับการนำมาใช้งาน เนื่องจากมาตรวัดนี้จะใช้วิธีคำนวณค่า Entropy ของแอททริบิวต์ทั้งหมดซึ่งเป็นการวัดความแตกต่างหรือการกระจายตัวของข้อมูล (Entropy มีค่าระหว่าง 0-1 โดยจะมีค่าสูงขึ้นเมื่อข้อมูลมีความแตกต่างกันมาก และในทางตรงข้ามถ้าข้อมูลมีความคล้ายกันมากค่า Entropy ก็ต่ำ) ดังนั้น Information Gain จะทำงานโดยวัดค่า Entropy ก่อนการแบ่งข้อมูลออกตามคลาสและหลังการแบ่งว่ามีประสิทธิภาพดีขึ้นหรือไม่ ถ้ามีประสิทธิภาพดีขึ้นค่า Information Gain จะมีค่าสูง (โดยที่ค่า Entropy น้อยที่สุดเพื่อให้ข้อมูลคล้ายกัน) และแอททริบิวต์นั้นจะถูกเลือกมาใช้ในการแตกกิ่ง แต่วิธีนี้จะเกิด Bias ในกรณีของการเลือก Attributes ที่มีจำนวนมาก จึงเป็นเกณฑ์ที่ใช้ได้ดีสำหรับการทำงานกับเงื่อนไขที่ไม่ได้ซับซ้อนมากนัก (Boonyang, 2561)

(2.) **Gain_ratio:** เป็นตัวชี้วัดการแบ่งชุดข้อมูลออกเป็นชุดข้อมูลย่อยที่พัฒนามาจาก Information Gain เนื่องจากเมื่อใช้ค่า Information Gain ในการแบ่งชุดข้อมูลจะส่งผลให้เกิดความเอนเอียง (Bias) ขึ้นเมื่อ แอททริบิวต์ที่ทำการพิจารณามีค่าที่เกิดขึ้นเป็นจำนวนมาก ตัวอย่างเช่น แอททริบิวต์รายการสินค้าที่มีค่าเป็นรหัสสินค้าต่าง ๆ (Product_ID) ซึ่งหากทำการแบ่งชุดข้อมูลตามแอททริบิวต์รายการสินค้าจะทำให้มีชุดข้อมูลย่อยเป็นจำนวนมาก (ชุดข้อมูลย่อยจะมีจำนวนเท่ากับรายการสินค้าทั้งหมดที่มีในชุดข้อมูล) โดยแต่ละชุดข้อมูลย่อยจะมีข้อมูลเพียงหนึ่งเรคคอร์ดเท่านั้นและจะทำให้ชุดข้อมูลย่อยนั้น ๆ มีข้อมูลที่มีหมวดหมู่ของข้อมูลเหมือนกันทั้งหมด (เพราะมีเพียงหนึ่งเรคคอร์ดจึงเหมือนกันทั้งหมด) ผลที่ตามมาคือจะทำให้ค่า Information Gain ของแอททริบิวต์รายการสินค้านั้นมีค่าสูง และ

แอททริบิวต์ที่มีลักษณะคล้ายกับแอททริบิวต์รายการสินค้ามักจะถูกเลือกเพื่อใช้ในการแบ่งชุดข้อมูลเสมอ จากปัญหานี้จึงได้มีนักวิจัยที่พยายามจะลดทอนความเอนเอียงลง โดยพัฒนาตัวชี้วัดการแบ่งข้อมูลใหม่ที่มีชื่อว่า Gain ratio ขึ้นมา ซึ่งเป็นวิธีคัดเลือกเพื่อให้ได้ตัวแปรที่เป็นตัวแบ่งข้อมูลออกเป็นกลุ่มย่อยที่มีสมาชิกภายในกลุ่มเป็นชนิดเดียวกันมากที่สุด (Homogeneous) โดยมาตรวัด Gain Ratio นั้นเป็นอัตราส่วนของค่าเกน (Gain หรือ Information Gain) กับค่าสารสนเทศการแบ่งกลุ่ม (Split Info) เพื่อเป็นการลดอิทธิพลของตัวแปรที่มีค่าหลายค่า ซึ่งผลที่ได้จากการใช้เทคนิคนี้จะได้ลำดับของตัวแปรซึ่งตัวแปรที่อยู่ลำดับแรก ๆ จะถือว่ามีอิทธิพลในการพยากรณ์ตัวแปรเป้าหมายมากกว่าตัวแปรในลำดับถัดไป ส่งผลให้สามารถพิจารณาเลือกจำนวนตัวแปรที่เหมาะสมได้อย่างมีประสิทธิภาพ (นิภาพร ชนะมาร, 2557)

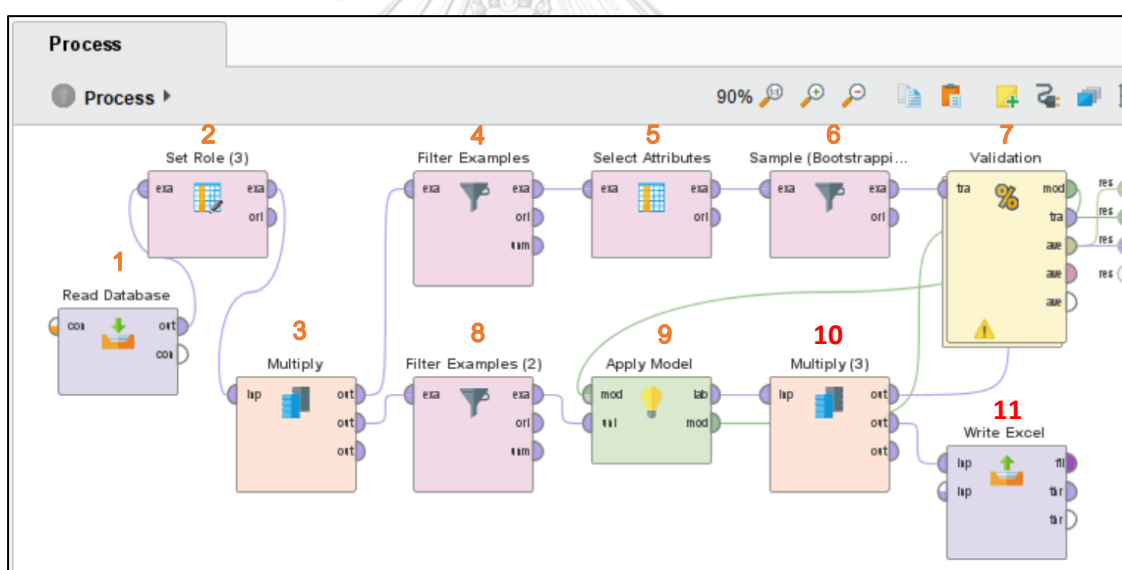
- (3.) **Gini_index:** เป็นตัวชี้วัดที่จะพิจารณาความไม่บริสุทธิ์ของชุดข้อมูล (การวัดระดับความบริสุทธิ์ของ Node คือเมื่อค่า Gini Index เท่ากับศูนย์แปลว่า Node บริสุทธิ์) โดยชี้วัดถึงความไม่สม่ำเสมอในการกระจายตัวของ Label (แอททริบิวต์ที่เป็นคำตอบ) ดังนั้นการแตกกิ่งตรงแอททริบิวต์ที่เลือกจึงส่งผลดีเนื่องจากทำให้ค่าเฉลี่ย Gini Index ของชุดข้อมูลย่อยที่ได้นั้นลดลง (โกเมศ อัมพวัน, 2561)
- (4.) **Accuracy:** เป็นการแตกกิ่งโดยเลือกแอททริบิวต์ที่ทำให้ค่าความแม่นยำ (Accuracy) ของต้นไม้ทั้งต้นมีค่าสูงสุด
- (5.) **Least_square:** เป็นการแตกกิ่งโดยเลือกแอททริบิวต์ที่จะช่วยลดระยะห่างยกกำลังสอง (Squared Distance) ระหว่างค่าเฉลี่ยของค่าข้อมูลในโหนดโดยพิจารณาจากค่าที่แท้จริง
- 2) **Maximal_depth:** เนื่องจากความลึกของต้นไม้จะแตกต่างกันออกไปโดยขึ้นอยู่กับขนาดและลักษณะของชุดข้อมูล พารามิเตอร์นี้จึงถูกใช้เพื่อจำกัดความลึกของต้นไม้ โดยหากตั้งค่าเป็น “-1” จะไม่มีการกำหนดความลึกของต้นไม้ และในกรณีนี้ต้นไม้จะถูกสร้างขึ้นจนกว่าจะถูกหยุดด้วยเกณฑ์อื่น ๆ แต่หากตั้งค่าเป็น “1” จะเป็นการสร้างต้นไม้ที่มีเพียงโหนดเดียวเท่านั้น

- 3) **Apply_pruning:** โมเดล Decision Tree สามารถถูกตัดแต่งกิ่ง (Pruned) ได้หลังจากถูกสร้างขึ้นแล้ว ซึ่งหากมีการคัดเลือกตัวเลือกนี้ไว้กิ่งไม้บางกิ่งจะถูกแทนที่ด้วยใบไม้ตามที่มีการตั้งค่าไว้ในพารามิเตอร์ Confidence
- 4) **Confidence:** พารามิเตอร์นี้จะเป็นส่วนที่ใช้กำหนดระดับความเชื่อมั่น (Confidence level) ที่ใช้สำหรับการคำนวณข้อผิดพลาดของการตัดแต่งกิ่ง
- 5) **Apply_prepruning:** พารามิเตอร์นี้เป็นส่วนที่ใช้กำหนดว่าควรมีเกณฑ์การหยุดสร้างต้นไม้มากกว่าเรื่องของความลึกสูงสุดสูงสุด (Maximal depth) หรือไม่ในระหว่างการสร้างโมเดลต้นไม้ ซึ่งหากมีการคัดเลือกไว้ พารามิเตอร์ Minimal gain, Minimal leaf size, Minimal size for split และ Number of prepruning alternatives จะถูกนำมาใช้เป็นเกณฑ์ในการหยุดสร้างต้นไม้
- 6) **Minimal_gain:** เนื่องจากค่า Gain (ซึ่งเป็นค่าที่ใช้บอกว่าแอททริบิวต์ที่ทำหน้าที่เป็น Root node สามารถจำแนกข้อมูลได้ดีมากน้อยเพียงใด) จะถูกคำนวณก่อนที่จะมีการแตกกิ่ง การแตกกิ่งจึงจะเกิดขึ้นเฉพาะเมื่อโหนดนั้น ๆ มีค่า Gain สูงกว่าค่า Minimal Gain ที่ถูกกำหนดไว้ ดังนั้นการตั้งค่า Minimal Gain ยิ่งสูงจึงยิ่งส่งผลให้เกิดการแตกกิ่งน้อยและได้ผลลัพธ์เป็นต้นไม้ที่มีขนาดเล็ก ขณะที่การตั้งค่า Minimal Gain ที่สูงจนเกินไปจะเป็นตัวขัดขวางการแตกกิ่งและเป็นการสร้างต้นไม้ที่มีเพียงโหนดเดียว
- 7) **Minimal_leaf_size:** ขนาดของใบไม้ (Leaf Size) หมายถึงจำนวนข้อมูลในแต่ละกลุ่มข้อมูลที่มีการแบ่งย่อยออกมาแล้ว ดังนั้นต้นไม้จึงจะถูกสร้างในลักษณะที่ใบไม้ทุกใบมีจำนวนข้อมูลขั้นต่ำใน Leaf Node ตามที่มีการกำหนดไว้ การหยุด Split Node นั้น ๆ หากมีจำนวนข้อมูลต่ำกว่าที่กำหนดจะเป็นการลด Overfitting โดยมักกำหนดค่า Min_leaf_size ให้สอดคล้องตามขนาดของ Data set (เช่น ตั้งแต่ 2-100 ตามขนาดข้อมูล)
- 8) **Minimal_size_for_split:** หมายถึงขนาดของโหนดหรือจำนวนของข้อมูลในหน่วยนั้น ๆ ซึ่งจะมีเฉพาะโหนดที่มีขนาดมากกว่าหรือเท่ากับขนาดต่ำสุดที่มีการตั้งค่าไว้เท่านั้นที่จะมีการแตกกิ่งออกไป

- 9) **Number_of_prepruning_alternatives:** เมื่อการ Prepruning นั้นเป็นการตัดขบวนการแตกกิ่งที่บางโหนด พารามิเตอร์นี้จะเป็นตัวปรับจำนวนโหนดทางเลือก (Alternative Nodes) ที่จะถูกทดสอบสำหรับการแตกกิ่ง และจะทำงานเมื่อมีการ Prepruning ควบคุมเกี่ยวกับกระบวนการสร้างต้นไม้ ซึ่งอาจทำการปิดกั้นการแตกกิ่งที่บางโหนดหากการแตกกิ่งที่โหนดนั้น ๆ ไม่ได้ช่วยเพิ่มประสิทธิภาพในการจำแนกแบ่งกลุ่มของต้นไม้ทั้งต้น จึงจะส่งผลให้เกิดการลองแตกกิ่งที่โหนดทางเลือก (Alternative Nodes) ในกรณีเช่นนี้

ขั้นตอนการสร้างโมเดล Decision Tree ในโปรแกรม RapidMiner

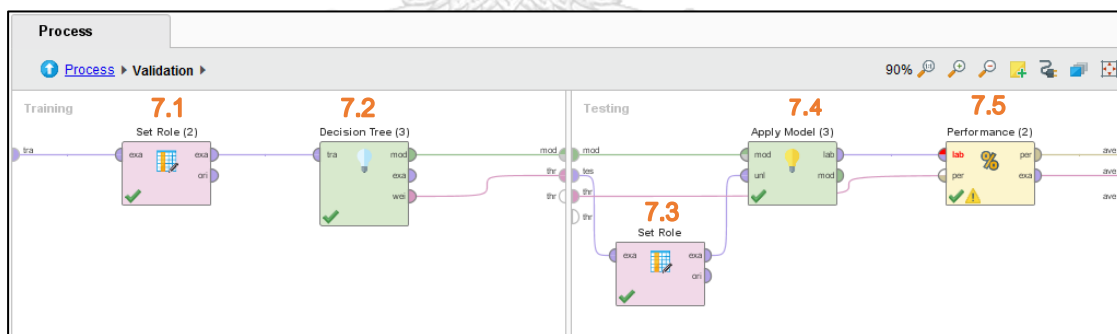
Process ทั้งหมดสำหรับการทำนายการสมัครใช้บริการเว็บไซต์ด้วยโมเดล Decision Tree สามารถแสดงได้ดังรูปที่ 4-13 และ 4-14



รูปที่ 4-13: Process การทำนายการสมัครใช้บริการเว็บไซต์ด้วยโมเดล Decision Tree

- 1) Read Database: เชื่อมต่อกับฐานข้อมูลเพื่อเข้าถึงชุดข้อมูลสำหรับการสร้างโมเดล
- 2) Set Role: กำหนดให้ตัวแปร Subscription (สถานะการสมัครใช้บริการของลูกค้า) เป็น Label ที่ต้องการทำนาย โดยกรณีในตัวแปรที่มีค่าเท่ากับ 0 หมายถึง User ไม่ได้กดสมัครใช้บริการ แต่หากตัวแปรที่มีค่าเท่ากับ 1 จะหมายถึง User กดสมัครใช้บริการ
- 3) Multiply: เนื่องจากต้องการนำ Output ที่มีการ Set Role แล้วไปใช้แยกส่วนกัน

- 4) Filter Example: เลือกเฉพาะข้อมูลส่วนที่ Label ไม่ Missing เพื่อนำไปใช้เป็น Training Set
- 5) Select Attributes: เลือกตัวแปรที่จะนำมาใช้ทำนาย ซึ่งในที่นี้เลือกไว้ทั้งหมด 4 ตัวแปร
- 6) Sample (Bootstrapping): ทำการสุ่มข้อมูลที่จะนำมาใช้ เนื่องจากชุดข้อมูลมีความ Imbalance
- 7) Validation: ตรวจสอบค่าความแม่นยำของโมเดล Decision Tree ที่สร้างขึ้นภายในโอเพอร์เรเตอร์นี้
- 8) Filter Example: เลือกเฉพาะข้อมูลส่วนที่ Label is Missing เพื่อให้โมเดลทำนายผลต่อไป
- 9) Apply Model: นำโมเดลที่สร้างขึ้นใหม่มาใช้ทำนายข้อมูลส่วนที่เลือกไว้ในขั้นตอนที่ 8
- 10) Multiply: เนื่องจากต้องการนำ Output ที่มีการ Apply Model แล้วไปใช้แยกส่วนกัน
- 11) Write Excel: เป็นการ Export ข้อมูลที่ได้หลังการ Apply Model แล้วออกมาเป็นไฟล์ Excel



รูปที่ 4-14: Sub-Process ภายในโอเพอร์เรเตอร์ Validation

จากรูปที่ 4-14 อธิบายได้ดังนี้

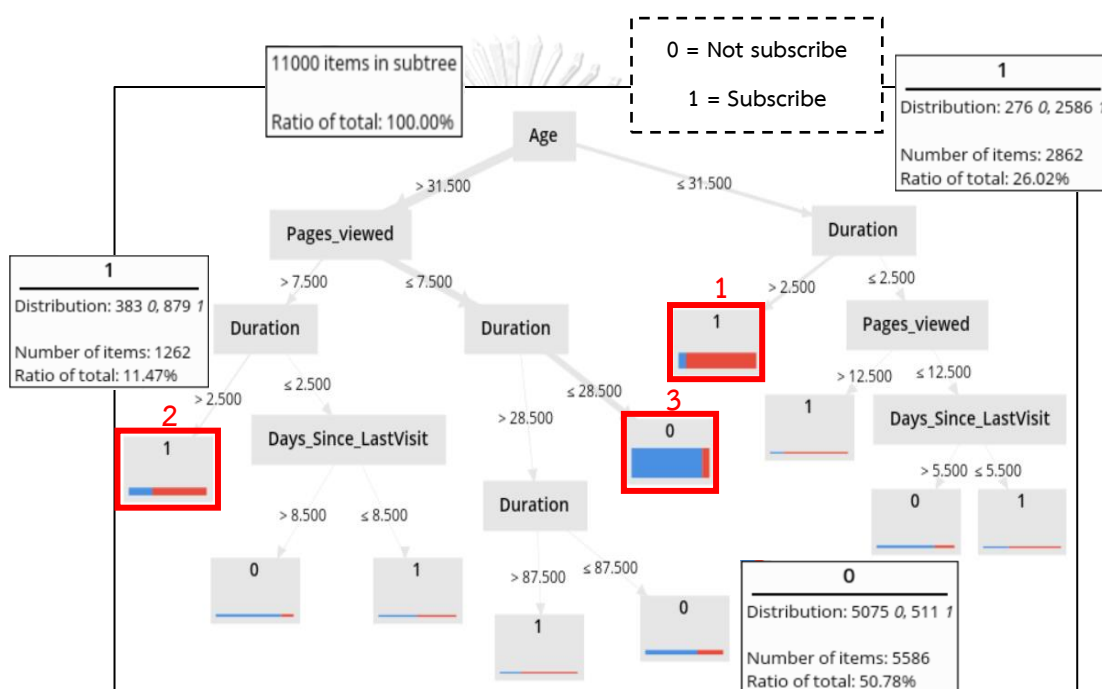
- 7.1) Set Role: กำหนดแอททริบิวต์ที่จะทำหน้าที่เป็น Label ซึ่งในที่นี้คือแอททริบิวต์ Subscription
- 7.2) Decision Tree: เป็นการสร้างโมเดล Decision Tree ขึ้นจาก Training Set เพื่อใช้สำหรับทำนายข้อมูลต่อไป

7.3) Set Role: กำหนดแอททริบิวต์ที่จะทำหน้าที่เป็น Label ซึ่งในที่นี้คือแอททริบิวต์ Subscription

7.4) Apply Model: นำโมเดล Decision Tree ที่ได้สร้างขึ้นมาใช้กับข้อมูลที่เป็น Test Set

7.5) Performance: ตรวจสอบค่าความแม่นยำของโมเดล Decision Tree ที่ได้สร้างขึ้น

ผลลัพธ์ที่ได้คือโมเดล Decision Tree ซึ่งสามารถนำไปวิเคราะห์หรือหาคำอธิบายต่อได้ ดังรูปที่ 4-15



รูปที่ 4-15: ผลลัพธ์โมเดล Decision Tree ที่ได้สร้างขึ้นจากข้อมูล Training Set

เบื้องต้นจากผลลัพธ์ของโมเดล Decision Tree พบว่า “อายุ” เป็นตัวแปรที่สามารถจำแนกข้อมูลได้ดีที่สุด และรองลงมาจะมีตัวแปรในเรื่องของลักษณะการใช้งานคือ จำนวนหน้าเว็บไซต์ที่เปิด (Page Viewed) รวมถึงระยะเวลาในการใช้งานเว็บไซต์ (Duration)

นอกจากนี้สำหรับ Leaf Node ที่มีความน่าสนใจนั้นมีอยู่สามโหนดด้วยกัน โดยพบว่ากลุ่มหมายเลข 1 เป็นโหนดที่ผู้เข้าชมเว็บไซต์มีแนวโน้มสูงที่จะกดสมัครใช้บริการกับทางบริษัท โดยคิดเป็นสัดส่วนถึง 26.02 เปอร์เซ็นต์ของข้อมูลทั้งหมดในโมเดล โดยเป็นกลุ่มที่อายุน้อยกว่า 31 ปีและใช้เวลาอยู่ในเว็บไซต์นานกว่า 2.5 นาที ซึ่งมีสัดส่วนของผู้ที่กดสมัครใช้บริการมากถึง 90 เปอร์เซ็นต์

สำหรับข้อมูลที่มาตกในโหนดนี้ โดยจะเห็นได้ว่าบริเวณแถบสีแดงซึ่งหมายถึงการกดสมัครใช้บริการ มีพื้นที่มากกว่าสีฟ้า ดังนั้นจึงมีโอกาสสูงที่ผู้เข้าชมเว็บไซต์กลุ่มนี้จะกดสมัครใช้บริการ

ส่วนอีกด้านคือถ้าเป็นกลุ่มที่อายุมากกว่า 31 ปี และมีจำนวนหน้าเว็บไซต์ หรือ Page Viewed ที่เปิดมากกว่า 7 หน้า รวมถึงใช้เวลาอยู่ในเว็บไซต์นานกว่า 2 นาที ซึ่งคิดเป็น 11.47 เปอร์เซ็นต์ของข้อมูลทั้งหมดในโมเดล ก็เป็นกลุ่มที่มีโอกาสสูงที่จะกดสมัครใช้บริการเช่นกัน เนื่องจากถ้าเทียบกันแล้ว กลุ่มนี้จะมีสัดส่วนผู้ที่กดสมัครใช้บริการมากถึงประมาณ 70 เปอร์เซ็นต์

แต่ในอีกด้านหนึ่งสำหรับกลุ่มที่มีแนวโน้มสูงว่าจะไม่กดสมัครใช้บริการ ซึ่งมีสัดส่วนมากถึง 50.78 เปอร์เซ็นต์ ก็จะเป็นกลุ่มที่อายุมากกว่า 31 ปี และเปิดเว็บไซต์ไม่เกิน 8 หน้า รวมถึงใช้เวลาในเว็บไซต์ไม่เกินประมาณ 29 นาทีเท่านั้น โดยจากข้อมูลทั้งหมดที่มาตกในโหนดนี้ก็สะท้อนว่ามีสัดส่วนของข้อมูลที่เป็นหมายเลข 0 ก็คือไม่กดสมัครใช้บริการ สูงถึง 90 เปอร์เซ็นต์ จึงสะท้อนถึงความน่าจะเป็นที่สูงกว่าผู้เข้าชมเว็บไซต์กลุ่มนี้จะไม่กดสมัครใช้บริการ

ดังนั้นจากผลลัพธ์ที่ได้จึงบ่งชี้ว่าเพื่อให้สามารถจูงใจผู้ใช้งานได้ บริษัทอาจยังต้องปรับปรุงเว็บไซต์เพิ่มเติม ตัวอย่างเช่น

- 1) ขยายฐานลูกค้าในกลุ่มที่ค่อนข้างเป็นวัยผู้ใหญ่ (คืออายุเกินกว่าประมาณ 32 ปี) เนื่องจากอาจมีลักษณะเป็นกลุ่มที่ต้องใช้เวลาในการศึกษาข้อมูลรายละเอียดต่าง ๆ ก่อนจะตัดสินใจสมัครใช้บริการ ดังนั้นผู้เข้าชมเว็บไซต์ที่กดสมัครใช้บริการจึงมักเป็นกลุ่มที่ใช้เวลาอยู่ในเว็บไซต์นานพอสมควร บริษัทอาจปรับปรุงเว็บไซต์โดยเน้นด้านการให้ข้อมูลที่กระชับเข้าใจง่ายไม่ต้องคลิกเพื่อค้นหารายละเอียดหลายหน้า หลายครั้ง หรือรวบรวมคำถามที่พบบ่อยและคำถามหมวดหมู่ที่มักพบร่วมกัน รวมถึงข้อมูลหลัก ๆ ที่จำเป็นสำหรับลูกค้าใหม่ให้อยู่ในจุดที่เด่นชัด มองเห็นได้ง่าย เพื่อให้สะดวกต่อผู้เข้าชมในการศึกษาข้อมูล
- 2) ผู้เข้าชมกลุ่มที่มีจำนวน Page Viewed น้อยอาจมีข้อติดขัดเรื่องการใช้งาน หรือเมื่อเข้ามาแล้วพบว่าหน้าเว็บไซต์ไม่น่าสนใจหรือไม่น่าเชื่อถือมากพอ จึงอาจต้องมีการทำวิจัยหรือสอบถามความคิดเห็นเพื่อขยายผลต่อไปว่าสาเหตุเกิดจากอะไรได้บ้าง ในกรณีนี้บริษัทสามารถนำเอาแช็ตบอตเข้ามาช่วยในสอบถามความคิดเห็นของผู้ใช้งานได้ เป็นอีกทางเลือกหนึ่งที่มีความน่าสนใจนอกเหนือจากการส่งอีเมลล์หรือโทรศัพท์หาลูกค้าเพื่อ

สอบถามโดยตรง ซึ่งอาจมีความเป็นทางการมากกว่า ทำให้ไม่ได้รับความคิดเห็นที่แท้จริงจากผู้เข้าชมเว็บไซต์

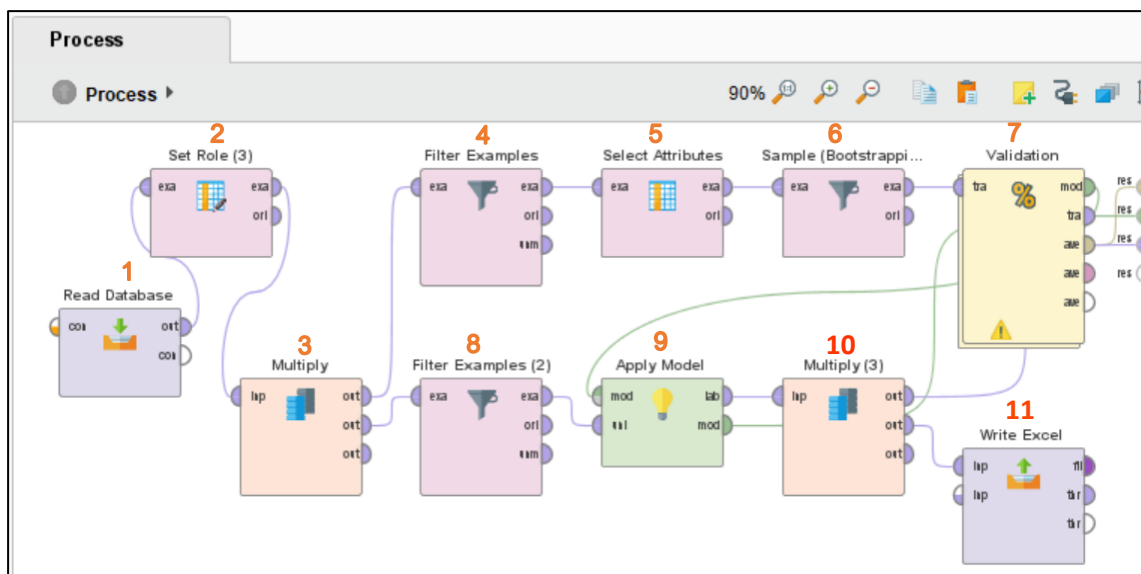
พารามิเตอร์ของโมเดล Neural Network ในโปรแกรม RapidMiner

- 1) **Hidden_layers:** พารามิเตอร์นี้ใช้สำหรับกำหนดชื่อและขนาดของ Hidden Layers ทั้งหมด ผู้พัฒนาสามารถกำหนดโครงสร้างของ Neural Network ได้ด้วยพารามิเตอร์นี้ และจะสามารถแสดงผลออกมาเป็นภาพของโมเดลได้ หากมีการระบุเพียงเลเยอร์เดียว โดยไม่มีโหนด จะส่งผลให้โหนดอินพุตเชื่อมต่อโดยตรงกับโหนดเอาต์พุตและจะไม่มี Hidden layers ด้านใน
- 2) **Training_cycles:** เป็นพารามิเตอร์ที่ใช้ระบุถึงจำนวนรอบของการ Train โมเดล Neural Network โดยในกระบวนการส่งค่าย้อนกลับ (Back Propagation) ค่าเอาต์พุต จะถูกเปรียบเทียบกับค่าตอบที่ถูกต้องเพื่อคำนวณค่าของข้อผิดพลาด จากนั้นข้อผิดพลาดจะถูกป้อนกลับผ่าน Network และอัลกอริทึมจะอาศัยข้อมูลนี้เพื่อทำการปรับน้ำหนักของแต่ละจุดเชื่อมต่อเพื่อลดค่าของข้อผิดพลาดทีละน้อย กระบวนการนี้จะถูกวนซ้ำเป็นจำนวน n ครั้ง และผู้พัฒนาจะสามารถกำหนดค่า n ได้โดยใช้พารามิเตอร์นี้
- 3) **Learning_rate:** พารามิเตอร์นี้ใช้สำหรับกำหนดว่าต้องการปรับน้ำหนักในแต่ละขั้นตอนมากแค่ไหน และไม่ควรตั้งค่าเป็น 0
- 4) **Momentum:** อัตราในการรักษาระดับ Momentum ของการเรียนรู้ เป็นค่าคงที่ซึ่งช่วยในเรื่องของการปรับน้ำหนัก เพื่อเหนี่ยวนำค่า Output ที่ได้ในแต่ละรอบ หรือเป็นตัวช่วยในการควบคุม Learning Rate อีกทีหนึ่ง กล่าวคือเมื่อ Learning Rate สูงจะทำให้การ Run โมเดลจบเร็วขึ้น แต่หากสูงมากเกินไปจะเกิดการแกว่งของผลแตกต่างกันระหว่างค่าทำนายจริงที่ได้ คือค่าแตกต่างกันจะไม่ลดน้อยลงเรื่อย ๆ แบบที่ควรจะเป็น ดังนั้นเพื่อเป็นการแก้ปัญหาการแกว่งขึ้นลงนี้จึงเกิดการกำหนดค่า Momentum Rate (Mr. Automated) ซึ่งมีค่าอยู่ระหว่าง 0-1 ขึ้นมา
- 5) **Decay:** เป็นพารามิเตอร์ที่ใช้ในการกำหนดว่าควรลดอัตราการเรียนรู้ (Learning Rate) ในระหว่างการ Train โมเดลหรือไม่

- 6) **Shuffle:** เป็นพารามิเตอร์ที่กำหนดว่าควรจะมีการสับเปลี่ยนตำแหน่งของข้อมูลนำเข้า (Input Data) ก่อนการ Train โมเดลหรือไม่ และแม้ว่าการ Shuffle จะทำให้ต้องใช้หน่วยความจำเพิ่ม แต่ก็เป็นที่ควรทำหากข้อมูลที่นำมาใช้มีการเรียงลำดับมาก่อน
- 7) **Normalize:** เนื่องจากโอเพอร์เรเตอร์ Neural Net ทำงานโดยใช้ฟังก์ชัน Sigmoid (ฟังก์ชันที่มีค่า Output อยู่ระหว่าง 0 – 1 จึงเหมาะที่จะถูกใช้ในงานที่ต้องการ Output เป็นความน่าจะเป็น (Probability) หรือใช้เป็น Output ว่า 1=Yes, 0=No) ดังนั้นจึงควรมีการ Scale ค่าของแอททริบิวต์ให้อยู่ในช่วง -1 ถึง +1 ซึ่งสามารถทำได้ผ่านพารามิเตอร์ Normalize โดยการทำให้ Normalize จะเกิดขึ้นก่อนการ Train โมเดล และแม้ว่าจะเป็นการเพิ่ม Runtime แต่ก็มีผลจำเป็นจำเป็นในหลาย ๆ กรณี (Keng Surapong, 2562)
- 8) **Error_epsilon:** การทำ Optimization (เปลี่ยนแปลงค่าน้ำหนักและค่า Bias ที่เชื่อมกับ Neuron) จะหยุดลงถ้าระดับข้อผิดพลาดจากการ Train อยู่ต่ำกว่าค่า Epsilon นี้
- 9) **Use_local_random_seed:** ใช้ในการกำหนดว่าจะมีการตั้งค่าตัวเลขเริ่มต้นสำหรับการสุ่มข้อมูลหรือไม่
- 10) **Local_random_seed:** พารามิเตอร์นี้ใช้ในการกำหนดตัวเลขในการสุ่มข้อมูล ซึ่งจะใช้ได้เฉพาะในกรณีที่มีการเปิดใช้งานพารามิเตอร์ Use local random seed

ขั้นตอนการสร้างโมเดล Neural Network ในโปรแกรม RapidMiner

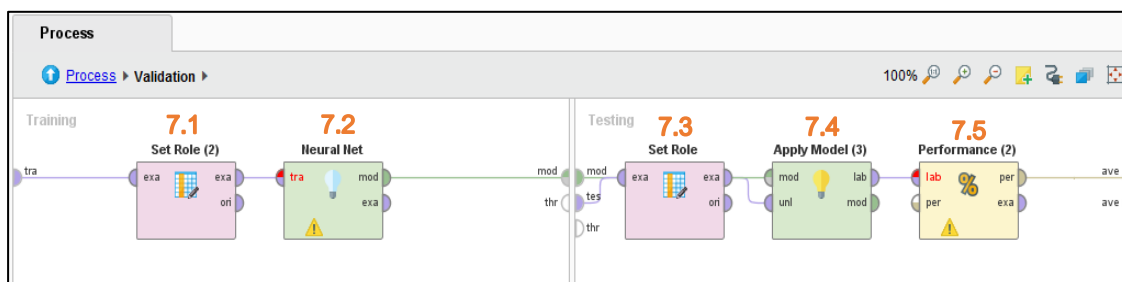
Process ทั้งหมดสำหรับการทำนายการสมัครใช้บริการเว็บไซต์ด้วยโมเดล Neural Network สามารถแสดงได้ดังรูปที่ 4-16 และ 4-17



รูปที่ 4-16: Process การทำนายการสมัครใช้บริการเว็บไซต์ด้วยโมเดล Neural Network

- 1) Read Database: เชื่อมต่อกับฐานข้อมูลเพื่อเข้าถึงชุดข้อมูลสำหรับการสร้างโมเดล
- 2) Set Role: กำหนดให้ตัวแปร Subscription (สถานะการสมัครใช้บริการของลูกค้า) เป็น Label ที่ต้องการทำนาย โดยกรณีที่ตัวแปรนี้มีค่าเท่ากับ 0 หมายถึง User ไม่ได้กดสมัครใช้บริการ แต่หากตัวแปรนี้มีค่าเท่ากับ 1 จะหมายถึง User กดสมัครใช้บริการ
- 3) Multiply: เนื่องจากต้องการนำ Output ที่มีการ Set Role แล้วไปใช้แยกส่วนกัน
- 4) Filter Example: เลือกเฉพาะข้อมูลส่วนที่ Label ไม่ Missing เพื่อนำไปใช้เป็น Training Set
- 5) Select Attributes: เลือกตัวแปรที่จะนำมาใช้ทำนาย ซึ่งในที่นี้เลือกไว้ทั้งหมด 4 ตัวแปรด้วยกัน
- 6) Sample (Bootstrapping): ทำการสุ่มข้อมูลที่นำมาใช้ เนื่องจากข้อมูลมีความ Imbalance
- 7) Validation: ตรวจสอบค่าความแม่นยำของโมเดล Neural Network ที่ได้สร้างขึ้นภายในโอเพอร์เรเตอร์นี้
- 8) Filter Example: เลือกเฉพาะข้อมูลส่วนที่ Label is Missing เพื่อให้โมเดลทำนายผลต่อไป
- 9) Apply Model: นำโมเดลที่สร้างขึ้นใหม่มาใช้ทำนายข้อมูลส่วนที่เลือกไว้ในขั้นตอนที่ 8

- 10) Multiply: เนื่องจากต้องการนำ Output ที่มีการ Apply Model แล้วไปใช้แยกส่วนกัน
- 11) Write Excel: เป็นการ Export ข้อมูลที่ได้หลังการ Apply Model แล้วออกมาเป็นไฟล์ Excel



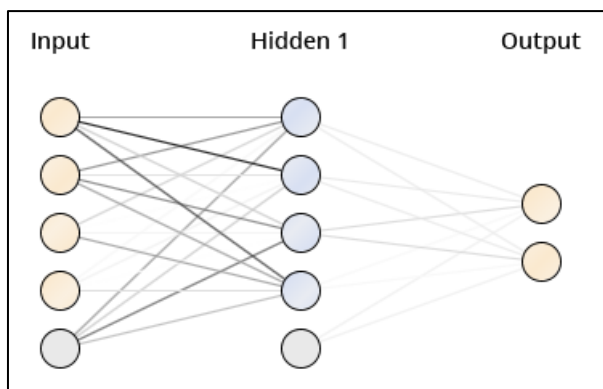
รูปที่ 4-17: Sub-Process ภายในโอเพอร์เรเตอร์ Validation

จากรูปที่ 4-17 อธิบายได้ดังนี้

- 7.1 Set Role: กำหนดแอททริบิวต์ที่จะทำหน้าที่เป็น Label ซึ่งในที่นี้คือแอททริบิวต์ Subscription
- 7.2 Neural Net: เป็นการสร้างโมเดล Neural Network ขึ้นจาก Training Set เพื่อใช้สำหรับทำนายข้อมูลที่ต้องการต่อไป
- 7.3 Set Role: กำหนดแอททริบิวต์ที่จะทำหน้าที่เป็น Label ซึ่งในที่นี้คือแอททริบิวต์ Subscription
- 7.4 Apply Model: นำโมเดล Neural Network ที่ได้สร้างขึ้นมาใช้กับข้อมูลที่เป็น Test Set
- 7.5 Performance: เป็นการตรวจสอบค่าความแม่นยำของโมเดล Neural Network ที่ได้สร้างขึ้น

ผลลัพธ์ที่ได้คือโมเดล Neural Network ซึ่งสามารถนำไปใช้ทำนายข้อมูลที่ต้องการได้ต่อไป

ดังรูปที่ 4-18



รูปที่ 4-18: โมเดล Neural Network ที่ได้สร้างขึ้นจากข้อมูล Training Set

ขั้นตอนที่ 5 การประเมินผล (Evaluation)

การวัดประสิทธิภาพของ Predictive Model สามารถทำได้โดยดูจากค่าความถูกต้อง (Accuracy) ของโมเดลที่ถูกสร้างขึ้น ทั้งนี้การ Evaluate โมเดลที่สร้างขึ้นนั้นสามารถพิจารณาได้จาก Confusion Matrix ตามที่ได้อธิบายรายละเอียดไว้ในหัวข้อที่ 2.3 และหลังจากที่ได้ทำการประเมินผลแล้วพบว่าค่าความแม่นยำของโมเดล Decision Tree มีรายละเอียดดังรูปที่ 4-19 และเป็นโมเดลที่มีลักษณะเด่นคือสามารถนำมาอธิบายย้อนกลับได้ มีความยืดหยุ่นในการรับตัวแปรได้ทั้งประเภท Nominal และ Numerical รวมถึงใช้เวลาในการรันโมเดลน้อยกว่า Neural Network Model

accuracy: 83.07%			
	true 0	true 1	class precision
pred. 0	378	57	86.90%
pred. 1	70	245	77.78%
class recall	84.38%	81.13%	

รูปที่ 4-19: ค่าความแม่นยำของโมเดล Decision Tree ในการทำนายการสมัครใช้บริการเว็บไซต์

ส่วนโมเดล Neural Network นั้นเป็น Black Box Model ที่ไม่สามารถอธิบายย้อนกลับได้ และใช้เวลาในการรันโมเดลนานกว่า Decision Tree รวมถึงมีข้อจำกัดในรับเฉพาะตัวแปรประเภท Numerical เท่านั้น แต่ก็ยังเป็นโมเดลที่มีจุดเด่นในเรื่องของความแม่นยำ โดยค่าความแม่นยำของโมเดล Neural Network มีรายละเอียดดังรูปที่ 4-20

accuracy: 85.33%			
	true 0	true 1	class precision
pred. 0	1624	276	85.47%
pred. 1	164	936	85.09%
class recall	90.83%	77.23%	

รูปที่ 4-20: ค่าความแม่นยำของโมเดล Neural Network ในการทำนายการสมัครใช้บริการเว็บไซต์

ขั้นตอนที่ 6 การใช้งาน (Deployment)

สำหรับในด้านของการใช้งานนั้น บริษัทสามารถนำเอาโมเดลที่สร้างขึ้นไปใช้ในการทำนายข้อมูลเพื่อเป็นประโยชน์ต่อการดำเนินธุรกิจในอนาคต เช่น ในเรื่องของการทำนายว่าผู้เข้าชมเว็บไซต์ลักษณะใดมีแนวโน้มจะสมัครใช้บริการจัดส่งพัสดุของบริษัท โดยบริษัทสามารถนำผลลัพธ์จากทั้งสองโมเดลไปใช้งานในวัตถุประสงค์ที่แตกต่างกันได้ตามความเหมาะสมและลักษณะเด่นของโมเดล กล่าวคือสำหรับ Decision Tree Model สามารถนำมาทำความเข้าใจและอธิบายย้อนกลับได้จึงสามารถนำมาวิเคราะห์เพิ่มเติมถึงเหตุผลที่มาที่ไปหรือลักษณะเฉพาะของผู้ที่มีแนวโน้มจะเป็นลูกค้าในอนาคตได้ ส่วน Neural Network Model นั้นให้ผลลัพธ์ที่เน้นเรื่องความแม่นยำแต่ไม่สามารถอธิบายย้อนกลับได้ ดังนั้นจึงอาจนำไปใช้ในกรณีที่ไม่ได้ต้องการคำอธิบายหรือรายละเอียดต่าง ๆ เกี่ยวกับผู้เข้าชมเว็บไซต์มากนัก แต่อาจเน้นในเรื่องของการวิเคราะห์เชิงสถิติหรือตัวเลขในภาพรวมมากกว่า

4.3.4 ระบบจัดกลุ่มผู้ใช้งานเว็บไซต์ (Website Users Clustering System)

ภาพรวมของระบบ (System Overview)

ระบบจัดกลุ่มผู้ใช้งานเว็บไซต์ (Website Users Clustering System) เป็นระบบที่พัฒนาขึ้นเพื่อให้บริษัทสามารถมองเห็นถึงภาพรวมของกลุ่มผู้เข้าชมเว็บไซต์ทั้งหมดได้ว่ามีลักษณะอย่างไร ซึ่งจะช่วยให้นำไปวิเคราะห์ต่อยอดถึงแผนทางการตลาดให้เหมาะกับลูกค้าแต่ละกลุ่มได้มากที่สุด เพื่อเพิ่มโอกาสในการเข้าถึงลูกค้ากลุ่มใหม่ ๆ ได้มากขึ้นอย่างไม่เคยทำได้ในอดีต

ผู้ใช้ (Users)

- 1) ผู้บริหาร
- 2) ฝ่ายการตลาด
- 3) ฝ่ายเทคโนโลยีสารสนเทศ

ข้อมูลนำเข้า (Input Data)

เป็นข้อมูลของผู้เข้าชมเว็บไซต์ของบริษัททั้งหมดจำนวน 22,006 รายการ โดยมีตัวแปรที่นำมาใช้ทั้งหมด 12 ตัวแปร ดังตารางที่ 4-4

ตารางที่ 4-4: ข้อมูลนำเข้าสำหรับการจัดกลุ่มผู้ใช้งานเว็บไซต์

ลำดับ	ชื่อตัวแปร	คำอธิบาย	หน่วย
1	Age	อายุของผู้ใช้	ปี
2	Gender	เพศของผู้ใช้	-
	1	เพศชาย	
	2	เพศหญิง	
3	Days_Since_LastVisit	จำนวนวันนับจากที่เข้าใช้งานเว็บไซต์ครั้ง	วัน
4	Duration	ระยะเวลาที่เข้าใช้งานเว็บไซต์	นาที
5	Pages_viewed	จำนวนหน้าเว็บไซต์ที่เปิดใช้งาน	หน้า
6	Parcel_Cat	จำนวนประเภทของพัสดุที่ส่ง	ประเภท
7	Income	ช่วงรายได้ต่อเดือน	บาท
	1	ต่ำกว่า 20,000	
	2	20,000 - 29,999	
	3	30,000 - 39,999	
	4	40,000 - 50,000	
	5	สูงกว่า 50,000	
8	References	ช่องทางที่ทำให้รู้จักบริษัท	-
	1	เพื่อน / คนรู้จัก	
	2	โฆษณาตามเว็บไซต์ / Search engine	
	3	สื่อ Social Media	
9	Region	ภูมิภาคที่อยู่ของผู้ใช้	-
	1	Bangkok (กรุงเทพฯ)	
	2	Central (ภาคกลาง)	
	3	Eastern (ภาคตะวันออก)	
	4	Northeastern (ภาคตะวันออกเฉียงเหนือ)	
	5	Northern (ภาคเหนือ)	
	6	Southern (ภาคใต้)	
	7	Western (ภาคตะวันตก)	
10	Education	ระดับการศึกษา	-

ลำดับ	ชื่อตัวแปร	คำอธิบาย	หน่วย
	1	ระดับมัธยมปลายหรือต่ำกว่า	
	2	นิสิต/นักศึกษาที่ยังไม่จบการศึกษา	
	3	จบการศึกษาระดับอนุปริญญา	
	4	จบการศึกษาระดับปริญญาตรี	
	5	จบการศึกษาระดับปริญญาโทหรือสูงกว่า	
11	Total_Amount	ราคารวมของค่าจัดส่งพัสดุ	บาท
12	Subscription	สถานะสมัครใช้หรือไม่ใช้บริการ	-
	1	สมัครใช้บริการ	
	0	ไม่สมัครใช้บริการ	

ขั้นตอนการดำเนินงาน (Working Processes)

ใช้หลักของกระบวนการหาความรู้แบบ Cross-Industry Standard Process for Data Mining (CRISP-DM) ซึ่งเป็นแนวทางในการดำเนินงานในการวิเคราะห์ข้อมูลขั้นสูง ซึ่งประกอบไปด้วย 6 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 การทำความเข้าใจทางธุรกิจ (Business Understanding)

ทำความเข้าใจเกี่ยวกับธุรกิจโลจิสติกส์และศึกษาขั้นตอนการดำเนินงานของหน่วยงานรวบรวมปัญหาที่เกิดขึ้นในปัจจุบัน และความต้องการจากผู้บริหารและฝ่ายการตลาด เพื่อเป็นแนวทางในการพัฒนาระบบการจัดกลุ่ม นำความต้องการของผู้ใช้งานและกลุ่มเป้าหมายมาวิเคราะห์และสรุปเป็นความต้องการของระบบ พร้อมทั้งกำหนดขอบเขตของโครงการ เพื่อให้ครอบคลุมการแก้ไขปัญหาที่เกิดขึ้น โดยจากการทำความเข้าใจในประเด็นเรื่อง “การจัดกลุ่มผู้ใช้งานเว็บไซต์” แล้วสรุปได้ว่าผู้ใช้งานต้องการทราบข้อมูลเกี่ยวกับการจัดกลุ่มผู้ที่เข้าชมเว็บไซต์ของบริษัทว่าสามารถแบ่งได้เป็นกี่กลุ่ม และแต่ละกลุ่มมีคุณลักษณะสำคัญอย่างไร เพื่อให้สามารถนำไปใช้ในการเพิ่มโอกาสตอบสนองความต้องการและเพิ่มความพึงพอใจของผู้เข้าชมแต่ละกลุ่มได้อย่างเหมาะสม ซึ่งจะหมายถึงการที่บริษัทสามารถวางแผนทางการตลาดหรือกลยุทธ์การดำเนินธุรกิจได้ดียิ่งขึ้น เพราะสำหรับธุรกิจโลจิสติกส์ในปัจจุบันต้องสามารถเข้าถึงลูกค้าแต่ละกลุ่มโดยเข้าใจถึงลักษณะและความต้องการที่แตกต่างกัน ยิ่งมีความรู้ความเข้าใจเกี่ยวกับลูกค้าแต่ละกลุ่มมากเท่าใด ยิ่งเป็นการสร้างความได้เปรียบและเพิ่มโอกาสสร้างการเติบโตของธุรกิจได้มากยิ่งขึ้น

ขั้นตอนที่ 2 การทำความเข้าใจข้อมูล (Data Understanding)

รวบรวมข้อมูลทั้งหมดที่จำเป็นในการพัฒนาระบบการทำนาย และนำเข้าข้อมูลสู่ฐานข้อมูล (Data Base) จากนั้นจึงทำ Data Exploration เพื่อทำความเข้าใจกับข้อมูล และออกแบบรูปแบบข้อมูลที่เหมาะสมในการจัดทำ Clustering Model

โดยสำหรับระบบนี้พบว่าข้อมูลที่สามารถนำมาใช้ในการจัดกลุ่มผู้ใช้งานเว็บไซต์ของบริษัทได้ดังนี้

- อายุของผู้ใช้ (Age)
- เพศของผู้ใช้ (Gender)
- จำนวนวันนับจากที่เข้าใช้งานเว็บไซต์ครั้งล่าสุด (Days_Since_LastVisit)
- ระยะเวลาที่เข้าใช้งานเว็บไซต์ (Duration)
- จำนวนหน้าเว็บไซต์ที่เปิดใช้งาน (Pages_viewed)
- จำนวนประเภทของพัสดุที่ส่ง (Parcel_Cat)
- ช่วงรายได้ (Income)
- ช่องทางที่ทำให้รู้จักบริษัท (References)
- ภูมิภาคที่อยู่ของผู้ใช้ (Region)
- ระดับการศึกษา (Education)
- สถานะสมัครใช้หรือไม่ใช้บริการ (Subscription)
- ราคารวมของค่าจัดส่งพัสดุ (Total_Amount)

ขั้นตอนที่ 3 การเตรียมข้อมูล (Data Preparation)

- 1) **Data Cleansing** โดยตรวจสอบและลบข้อมูลที่อาจแปลกแยกจากข้อมูลอื่น (Outlier) ลบข้อมูลที่ผิดปกติและตรวจสอบข้อมูลที่หายไป เช่น สำหรับตัวแปรภูมิภาคที่อยู่ของผู้ใช้ (Region) ซึ่งควรมีค่าอยู่ระหว่าง 1-7 แต่กลับพบว่ามียังบาง Record ซึ่งมีค่าสูงเกินกว่า 7 ซึ่งอาจเกิดจากความผิดพลาดระหว่างการคีย์ข้อมูล (Human Error) หรือการถ่ายโอนข้อมูลจากต้นทาง นอกจากนี้ยังพบว่าข้อมูลบางส่วนขาดหายไป ซึ่งส่งผลให้ข้อมูลใน Record นั้น ๆ ไม่เหมาะต่อการนำมาจัดกลุ่ม จึงได้ทำการลบข้อมูลลักษณะดังกล่าวนี้

ออกเนื่องจากมีจำนวนน้อยกว่า 20 Records จึงไม่ส่งผลกระทบต่อการจัดกลุ่มข้อมูลในภาพรวม

- 2) **Data Transformation** ทำการปรับ Format ของข้อมูล ทั้ง Discrete หรือ Continuous ให้เหมาะสมกับโมเดลที่ต้องการพัฒนา เช่น ข้อมูลในคอลัมน์ของ Gender, Income, Education, References และ Region เดิมได้ถูกกำหนดตามค่าเริ่มต้น (Default) ให้เป็น General หรือ Number ในไฟล์ข้อมูล Excel ตั้งต้น เนื่องจากเก็บค่าเป็นตัวเลข เมื่อต้องการนำมาจัดกลุ่มจึงต้องมีการปรับเปลี่ยน Format ให้เป็น Text เพื่อให้เมื่อนำข้อมูลเข้าสู่การวิเคราะห์จัดกลุ่มจะได้มีความถูกต้องเหมาะสมกับลักษณะของตัวแปรที่เป็น Discrete ไม่ใช่ Continuous
- 3) **Data Transformation** ทำการ Normalize หรือ Scaling เพื่อปรับข้อมูลในอยู่ขอบเขตที่ต้องการและเหมาะสมต่อการจัดกลุ่ม โดยใช้ Z-Transformation Method ซึ่งให้ผลลัพธ์ที่ดีที่สุด

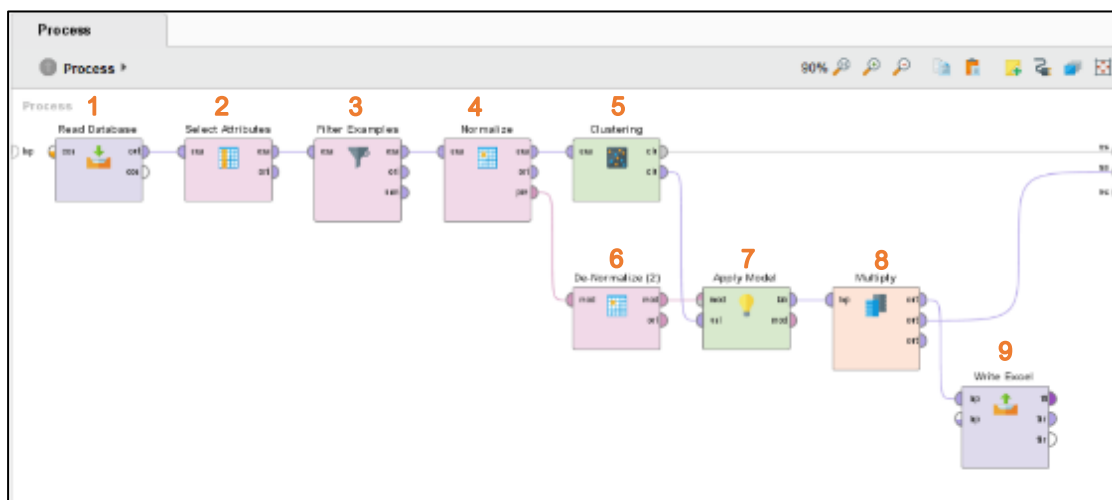
ขั้นตอนที่ 4 การสร้างโมเดล (Modeling)

สำหรับหลักการพื้นฐานของทำ Clustering มีพารามิเตอร์ที่ต้องกำหนด คือจำนวนกลุ่มที่ต้องการแบ่งหรือจำนวนคลัสเตอร์ซึ่งแทนด้วยตัวแปร K และมีขั้นตอนการทำงานเบื้องต้นดังต่อไปนี้ (Saetang, 2561)

- 1) เลือกจำนวนของคลัสเตอร์ (K)
- 2) สุ่มเลือกจุดศูนย์กลาง (Centroid) ขึ้นมาตามจำนวนคลัสเตอร์
- 3) ระบบกำหนดให้ข้อมูลอยู่ในคลัสเตอร์ที่ใกล้ที่สุด
- 4) ระบบคำนวณหาจุดศูนย์กลางแต่ละคลัสเตอร์ใหม่
- 5) ทำซ้ำข้อ 3 และ 4 จนกระทั่ง Centroid ไม่มีการเปลี่ยนแปลง

และสำหรับโมเดลที่เลือกใช้ในโครงการพิเศษนี้เพื่อจัดกลุ่มลูกค้าประเภทต่าง ๆ ที่มีลักษณะใกล้เคียงกันให้อยู่ในกลุ่มเดียวกันนั้นเป็นวิธีการวิเคราะห์ข้อมูลด้วยการจัดกลุ่ม (K-Medoids Clustering) โดยเป็นอัลกอริทึมที่ได้รับการพัฒนาสืบเนื่องจาก K-Means Clustering ซึ่งมีข้อจำกัดที่ค่าคุณลักษณะต่าง ๆ ต้องเป็นเชิงตัวเลข (Numerical) เมื่อนำมาใช้กับข้อมูลที่ไม่ใช่เชิงตัวเลขจึงไม่สามารถคำนวณค่าเฉลี่ยได้ จึงเกิดอัลกอริทึม K-Medoids Clustering สำหรับจัดกลุ่มข้อมูลวัตถุที่มี

ค่าคุณลักษณะเป็นประเภทเชิงกลุ่ม (Categorical) ขึ้นมา โดย Process หรือขั้นตอนของระบบการ จัดกลุ่ม (Clustering Model) ในโปรแกรม RapidMiner นั้นมีรายละเอียดดังรูปที่ 4-21



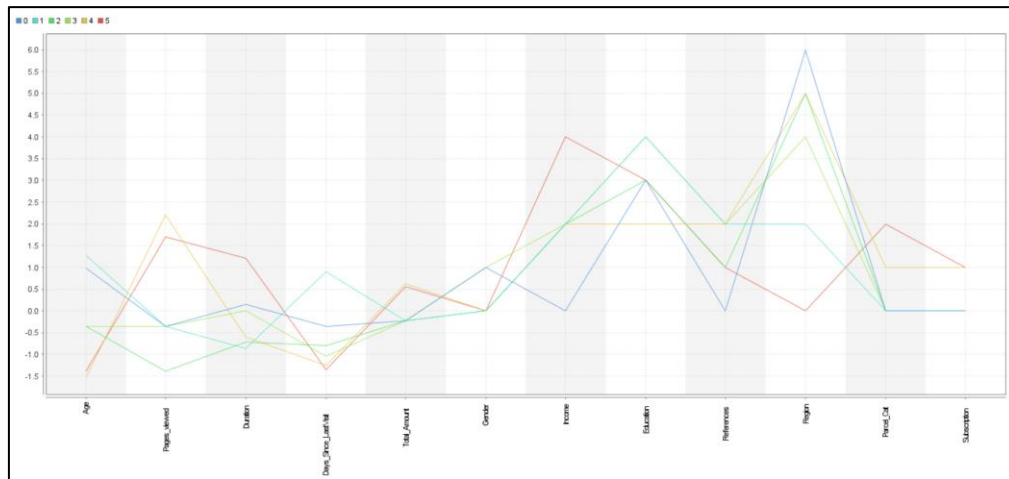
รูปที่ 4- 21: Process การจัดกลุ่มข้อมูลผู้ใช้งานเว็บไซต์ของบริษัทด้วยการจัดกลุ่มแบบ K-Medoids

Clustering

- 1) Read Database: เชื่อมต่อกับฐานข้อมูลเพื่อเข้าถึงชุดข้อมูลสำหรับการสร้างโมเดล
- 2) Select Attributes: เลือกตัวแปรที่จะนำมาจัดกลุ่ม ซึ่งในที่นี้เลือกไว้ทั้งหมด 12 ตัวแปร ดังที่กล่าวถึงในตารางที่ 4-4
- 3) Filter Example: เลือกเฉพาะข้อมูลส่วนที่ Label ไม่ Missing เพื่อให้สามารถจัดกลุ่มได้
- 4) Normalize: เป็นส่วนสำคัญเพื่อให้ข้อมูลอยู่ในฐานเดียวกันและทำการจัดกลุ่มได้อย่างมีประสิทธิภาพ
- 5) Clustering: ทำการจัดกลุ่มข้อมูลแบบ K-Medoids เนื่องจากลักษณะของข้อมูลมีทั้ง Numerical และ Non-numerical
- 6) De-Normalize: เป็นการแปลงค่าของข้อมูลให้กลับเป็นค่าปกติเพื่อให้ผู้ใช้ระบบสามารถอ่านค่าได้เข้าใจ
- 7) Apply Model: นำ Clustering Model ที่สร้างมาใช้กับข้อมูลที่ De-Normalize แล้ว
- 8) Multiply: เนื่องจากต้องการ Copy ข้อมูลที่มีการจัดกลุ่มแล้วไปใช้ใน 2 ส่วนคือการ แสดงผลและ Export ออกมาเป็น Excel File

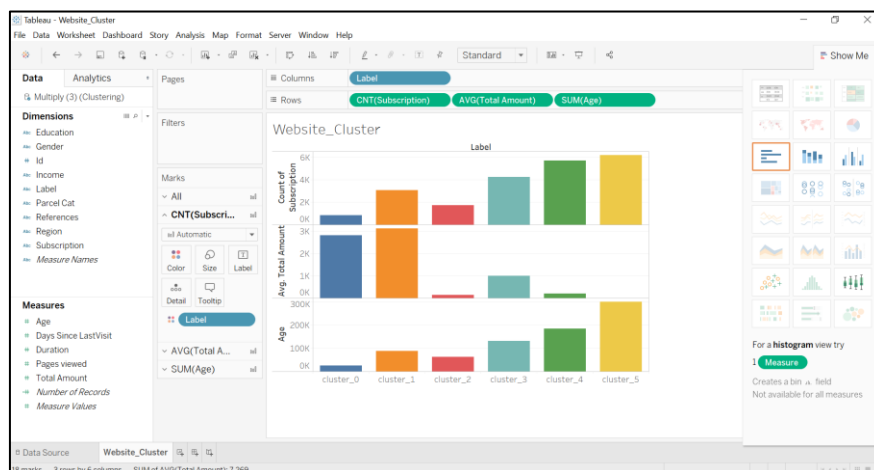
9) Write Excel: เป็นการ Export Output ที่ทำการจัดกลุ่มและ De-Normalize แล้วออกมาเพื่อทำการวิเคราะห์รายละเอียดหรือทำ Visualization ต่อไป

หลังเสร็จสิ้น Process ของการจัดกลุ่มแล้ว ผู้ใช้งานระบบสามารถตรวจสอบผลลัพธ์เบื้องต้นในลักษณะของกราฟได้จากส่วนของ Results ซึ่งจะแสดงผลดังรูปที่ 4-22



รูปที่ 4-22: Visualization เบื้องต้นของผลลัพธ์การจัดกลุ่มข้อมูลในโปรแกรม RapidMiner

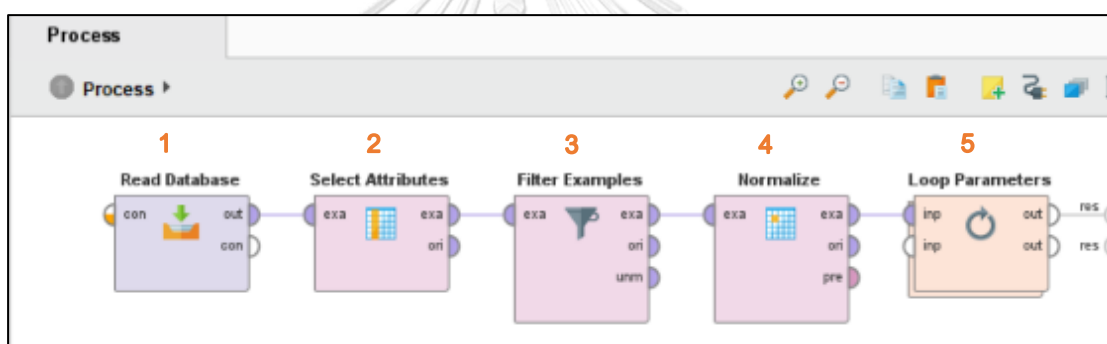
นอกจากนี้ผู้ใช้งานสามารถบันทึกหรือ Export ไฟล์ผลลัพธ์ที่ได้จากโปรแกรม RapidMiner เพื่อนำไปวิเคราะห์รายละเอียด หรือทำ Visualization เพิ่มเติมในเชิงลึกได้ เช่น การใช้โปรแกรม Tableau เพื่อดูอายุเฉลี่ยของแต่ละ Cluster หรือดูว่า Cluster ใดมีค่าเฉลี่ยในการใช้จ่ายเงินเพื่อจัดส่งพัสดุมากที่สุด รวมถึงดูจำนวนผู้ที่กดสมัครใช้บริการผ่านเว็บไซต์ของแต่ละ Cluster ว่ามีจำนวนมากน้อยเท่าใด ฯลฯ ได้ดังรูปที่ 4-23



รูปที่ 4-23: หน้าจอการทำ Visualization ผลลัพธ์เพิ่มเติมด้วยโปรแกรม Tableau

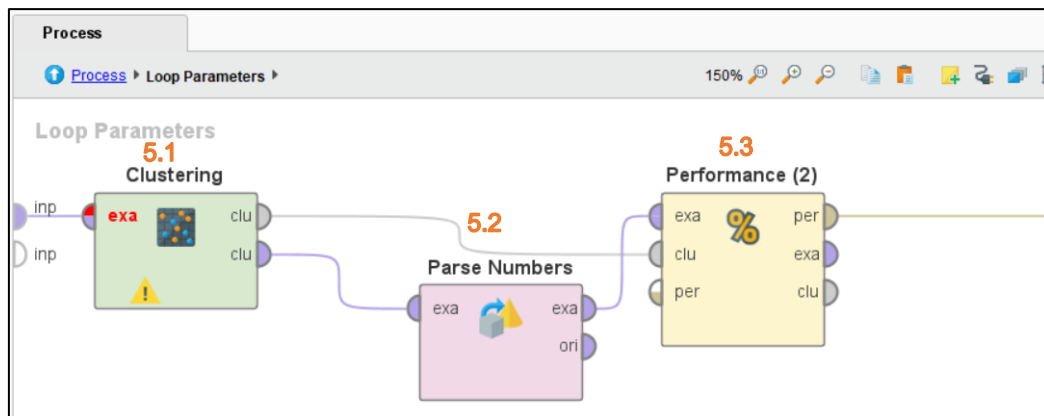
ขั้นตอนที่ 5 การประเมินผล (Evaluation)

การวัดประสิทธิภาพของโมเดลสามารถทำได้โดยใช้วิธีพิจารณาค่า K หรือจำนวนกลุ่มที่เหมาะสม (K-Medoids Criterion Approach) เนื่องจากชุดข้อมูลที่นำมาใช้จัดกลุ่มมีทั้งตัวแปรประเภท Numerical และ Non-numerical จึงต้องอาศัยวิธีพิจารณาเปรียบเทียบจากค่า Davies-Bouldin Index ซึ่งเป็นเกณฑ์การวัดคุณภาพการจัดกลุ่มที่เป็นอิสระต่อจำนวนของกลุ่มที่จะวิเคราะห์และไม่ขึ้นอยู่กับวิธีในการแบ่งกลุ่ม ดังนั้นการคำนวณค่าของ Davies-Bouldin Index แบบง่าย ๆ คือเป็นอัตราส่วนระหว่างผลรวมของการกระจายตัวของข้อมูลในกลุ่ม และระยะห่างระหว่างกลุ่ม โดยในการแบ่งกลุ่มที่ดีนั้น การกระจายตัวในกลุ่มจะต้องน้อย และระยะห่างระหว่างแต่ละกลุ่มจะต้องมาก การทำให้ค่าของ Davies Bouldin Index เล็กที่สุดจึงจะทำให้ได้การแบ่งแยกของกลุ่มที่ดีที่สุด (วีรศักดิ์ ช่องงูเหลือม, 2555) โดยขั้นตอนในการพิจารณาค่า K ที่เหมาะสมนั้นมีดังรูปที่ 4-24 และ 4-25



รูปที่ 4-24: Process การหาค่า K ที่เหมาะสมสำหรับการจัดกลุ่มข้อมูลผู้ใช้เว็บไซต์

- 1) Read Database: เชื่อมต่อกับฐานข้อมูลเพื่อเข้าถึงชุดข้อมูลสำหรับการสร้างโมเดล
- 2) Select Attributes: เลือกตัวแปรที่จะนำมาจัดกลุ่ม ซึ่งในที่นี้เลือกไว้ทั้งหมด 12 ตัวแปร
- 3) Filter Example: เลือกเฉพาะข้อมูลส่วนที่ Label ไม่ Missing เพื่อนำมาใช้จัดกลุ่ม
- 4) Normalize: เป็นส่วนสำคัญเพื่อให้ข้อมูลอยู่ในฐานเดียวกันและทำการจัดกลุ่มได้อย่างมีประสิทธิภาพ
- 5) Loop Parameter: เป็นการวนลูประจัดกลุ่มข้อมูลและวัดค่า Davies-Bouldin Index ของ Cluster ด้วยโอเพอร์เรเตอร์ Performance เมื่อ Set ค่า K ให้มีค่าแตกต่างกัน คือ 5 - 22 และเก็บข้อมูลเอาไว้เพื่อนำมาใช้พิจารณาว่าควรแบ่งข้อมูลออกเป็นจำนวนกี่กลุ่ม

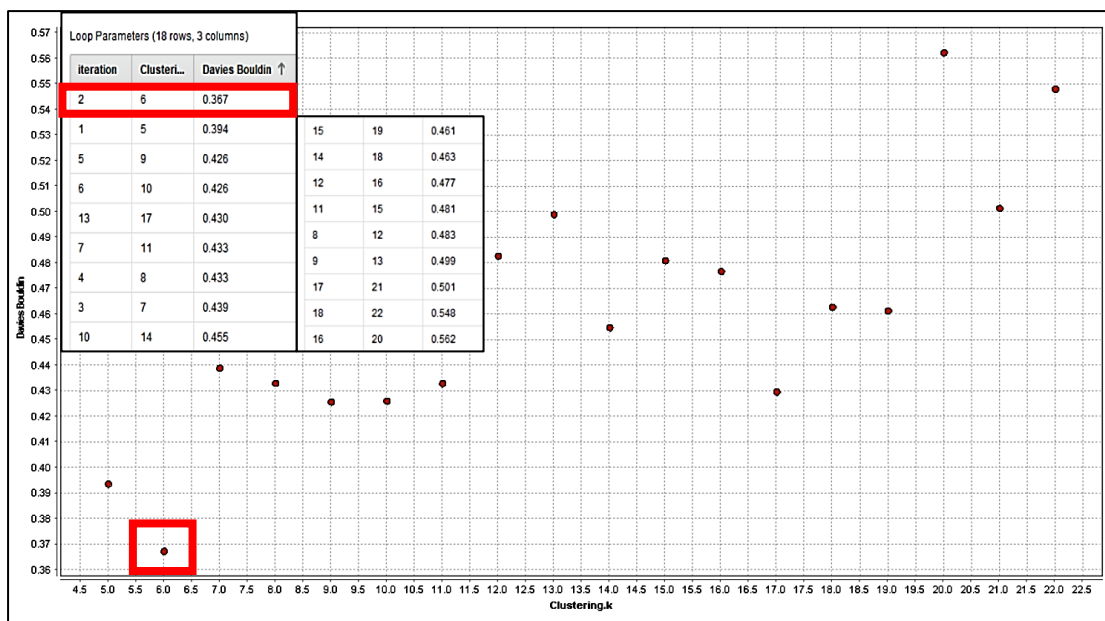


รูปที่ 4-25: Sub-Process ภายในโอเพอร์เรเตอร์ Loop Parameter

จากรูปที่ 4-25 อธิบายได้ดังนี้

- 5.1 Clustering: จัดกลุ่มข้อมูลผู้ใช้งานเว็บไซต์ของบริษัทโดยใช้วิธีจัดกลุ่มแบบ K-Medoids Clustering
- 5.2 Parse Numbers: เปลี่ยนประเภทของตัวแปรจาก Nominal เป็น Numerical เพื่อให้สามารถวัดประสิทธิภาพของการจัดกลุ่มได้ด้วยค่า Davies-Bouldin Index
- 5.3 Cluster Distance Performance: วัดระยะห่างระหว่างข้อมูล (Distance) ภายใน Cluster เดียวกันโดยใช้มาตรวัดของ Davies-Bouldin

ผลลัพธ์ที่ได้คือค่า Davies-Bouldin Index ภายใน Cluster เมื่อมีการ Set ค่า K ให้เป็นตัวเลขที่แตกต่างกัน ซึ่งในกรณีนี้ค่า K ที่เหมาะสมคือ 6 เนื่องจากเป็นค่าที่จะทำให้ข้อมูลภายใน Cluster เดียวกันมีระยะห่างจากกันน้อยที่สุด หรือหมายถึงการที่ข้อมูลมีความเกาะกลุ่มกันมากที่สุด ดังรูปที่ 4-27



รูปที่ 4-26: ผลลัพธ์จากการวน Loop เพื่อหาค่า K ที่เหมาะสมต่อการจัดกลุ่มข้อมูล

ขั้นตอนที่ 6 การใช้งาน (Deployment)

หลังจากที่ได้จัดกลุ่มข้อมูลและ Export Output ออกมาแล้ว บริษัทสามารถนำเอาข้อมูลที่เป็นผลลัพธ์มาใช้งานจริงกับธุรกิจ โดยผลลัพธ์ที่ได้จะสะท้อนถึงลักษณะของผู้ใช้งานเว็บไซต์ในแต่ละ Cluster ว่ามีอายุ เพศ ระดับการศึกษา หรือลักษณะอื่น ๆ เป็นอย่างไร ดังนั้นสำหรับการนำมาใช้ให้เป็นประโยชน์ต่อธุรกิจจึงอาจนำเอาผลลัพธ์ดังกล่าวมาเป็นวัตถุดิบในการพัฒนาต่อยอดธุรกิจ รวมถึงกำหนดทิศทางของการทำการตลาด ผนวกกับการติดตามรวบรวมผลที่ได้จากการดำเนินการเพื่อใช้ในการปรับปรุงกลยุทธ์การดำเนินงานต่อไป เช่น อาจมีการจัดโปรโมชั่น มอบโค้ดส่วนลดสำหรับลูกค้าที่ใช้บริการส่งพัสดุบ่อย ใช้เป็นปัจจัยในการคัดเลือกพรีเซนเตอร์คนใหม่ของบริษัท ใช้เพื่อปรับปรุงรูปแบบของเว็บไซต์ หรือปรับ Wording ให้ตรงตามลักษณะของลูกค้า เช่น วัยรุ่น วัยผู้ใหญ่ ระดับการศึกษา หรือวางแผนทางการตลาดอื่น ๆ ให้เหมาะสมกับลักษณะและความสนใจของผู้ใช้งานแต่ละกลุ่มให้มากที่สุด

และจากผลลัพธ์ของการจัดกลุ่มด้วยโมเดล K-Medoid Clustering เมื่อมีการวิเคราะห์เพิ่มเติมในรายละเอียดแล้วก็พบว่ากลุ่มที่มีความน่าสนใจคือ Cluster ที่ 4 และ 5 ซึ่งมีจำนวนคิดเป็นสัดส่วนประมาณ 20 เปอร์เซ็นต์ของข้อมูลผู้เข้าชมเว็บไซต์ทั้งหมด โดยเป็นกลุ่มที่สร้างรายได้ให้บริษัทสูงสุด และเป็นกลุ่มที่มีลักษณะเด่นคือ

1. มีอายุน้อยกว่า Cluster อื่น ๆ โดยอายุเฉลี่ยนั้นอยู่ที่ประมาณ 27 - 28 ปี
2. มีสัดส่วนประมาณ 60 เปอร์เซ็นต์ที่เป็นผู้หญิง
3. ระดับของการศึกษาโดยส่วนมากอยู่ในช่วงปริญญาตรี
4. มีจำนวนหน้าเว็บไซต์ (Page Viewed) ที่เปิดค่อนข้างสูง และใช้เวลาในเว็บไซต์ (Duration) ค่อนข้างนาน
5. เข้าใช้งานเว็บไซต์ค่อนข้างบ่อย (Days since last visit ต่ำ) ซึ่งสะท้อนว่าอาจมีการเคลื่อนไหวรวมถึงตรวจสอบข้อมูลต่าง ๆ อยู่เป็นประจำ
6. มี Income อยู่ในช่วงที่ไม่สูงมากคือไม่เกินประมาณ 40,000 บาทต่อเดือน จึงคาดว่าอาจเป็นกลุ่ม SME ที่มีการส่งสินค้าไปขายต่างประเทศ เนื่องจากประเภทของพัสดุที่ส่งไม่ได้หลากหลายมากนัก คืออยู่ที่ประมาณไม่เกิน 3 ประเภทเท่านั้น

และจากผลลัพธ์ของการจัดกลุ่ม บริษัทสามารถนำมาส่งเสริมในส่วนของ E-Marketing ได้คือ มีการใช้เครื่องมือทางอิเล็กทรอนิกส์ เครื่องมือดิจิทัลเข้ามาช่วยในการทำการตลาด เพื่อช่วยให้ลูกค้าสามารถเข้าถึงบริการของบริษัทได้โดยง่าย และมองว่ามีความน่าสนใจ สะดวกต่อการใช้งานเพื่อเพิ่มสัดส่วนลูกค้าในส่วนที่สร้างรายได้สูงให้กับบริษัทจาก 20 เปอร์เซ็นต์ให้ขยายตัวมากขึ้น เนื่องจากลักษณะลูกค้าเป็นกลุ่มคนอายุน้อยที่มีพื้นฐานในเรื่องของ Social Media และ Internet อยู่แล้ว นอกจากนี้ยังช่วยให้บริษัทสามารถเก็บข้อมูลลูกค้า สื่อสารกับกลุ่มเป้าหมายได้อย่างเฉพาะเจาะจงตามลักษณะของแต่ละ Cluster ทั้งยังเป็นแนวทางที่ใช้ต้นทุนน้อยแต่ทรงประสิทธิภาพ และสามารถวัดผลได้ทันทีอีกด้วย (Pawoot, 2551) ซึ่งกุญแจสำคัญหลัก ๆ ที่จะช่วยให้บริษัทสามารถนำเสนอสินค้าและบริการได้ตรงกับสิ่งที่ลูกค้าต้องการก็คือการทำความเข้าใจลูกค้าให้ดีเสียก่อนจากเครื่องมือในการวิเคราะห์ลูกค้า อันได้แก่

- Customer Persona: เพื่อรู้ว่าลูกค้าต้องการอะไร และอยากแก้ปัญหาใดบ้าง
- Digital Touchpoint: เพื่อรู้ว่าควรทำการตลาดช่องทางไหน ในเวลาใด
- Customer Journey: เพื่อรู้ว่าควรให้ความสำคัญช่องทางใดมากที่สุด

โดยจากลักษณะและรูปแบบการใช้งานของทั้ง Cluster ที่ 4 และ 5 ข้างต้น ถือเป็นการช่วยให้บริษัททราบได้ถึง Persona ของลูกค้า ซึ่งเป็นสิ่งสำคัญที่จะช่วยให้บริษัทสามารถเพิ่ม Engagement ของลูกค้า โดยสามารถจัดทำคอนเทนต์ที่ตรงใจลูกค้าได้ และนำไปสู่การเพิ่มอัตราการใช้บริการได้ในท้ายที่สุด

และนอกจากเรื่องของ Content แล้ว บริษัทยังอาจจัดทำ Digital Touchpoint เพิ่มเติม ซึ่งเป็นเครื่องมือที่ใช้วิเคราะห์ช่องทางที่กลุ่มเป้าหมายใช้งานในแต่ละวัน แต่ละช่วงเวลา โดยไม่ได้เจาะจง

ว่าเป็นช่องทางที่ลูกค้าใช้เพื่อจะตัดสินใจซื้อ โดย Digital Touchpoint จะช่วยให้รู้ว่าช่องทางไหนที่กลุ่มเป้าหมายใช้งานบ่อยที่สุด และในเวลาใด ซึ่งสามารถนำไปใช้ต่อยอดในการเลือกช่องทางและเวลาในการนำเสนอคอนเทนต์ให้กลุ่มเป้าหมายพบเจอบริษัทได้ง่ายมากขึ้น เพื่อเพิ่มโอกาสที่ลูกค้าจะซื้อสินค้ามากขึ้นด้วย เช่น อาจพบว่าช่องทางที่กลุ่มเป้าหมายใช้งานบ่อยที่สุดคือ Line ช่องทางที่ใช้รองลงมาในวันธรรมดาคือ Instagram และ Facebook อีกทั้งมีการใช้งาน Youtube มากที่สุดในวันหยุด ซึ่งช่วงเวลาในการใช้งานแต่ละช่องทางก็ต่างกัน เช่น ใช้งาน Instagram วันธรรมดาในช่วงเวลา 8.00-10.00 น. และ 12.00-14.00 น. และใช้งาน Facebook วันธรรมดาในช่วงเวลา 8.00-10.00 น. และ 20.00-22.00 น. เป็นต้น ดังนั้น จากตัวอย่างที่วิเคราะห์ หากบริษัทต้องการให้กลุ่มเป้าหมายมีโอกาสเห็นโพสต์ของบริษัทมากที่สุด จึงควรทำโฆษณาหรือคอนเทนต์ผ่าน Instagram ในช่วงเวลา 8.00 – 10.00 น. หรือ 12.00-14.00 น. และ Facebook ในวันธรรมดา ช่วง 8.00-10.00 น. และ 20.00-22.00 น เป็นต้น และเนื่องจากกลุ่มเป้าหมายใช้งาน Line บ่อยที่สุดในทุก ๆ คอนเทนต์จึงควรใส่ช่องทางติดต่อผ่าน Line เข้าไปด้วยเสมอ เพราะมีโอกาสสูงที่ลูกค้าจะติดต่อมาทางช่องทางนี้ ซึ่งเป็นช่องทางที่ลูกค้ามีความคุ้นชินมากที่สุดนั่นเอง

แต่เนื่องจาก Digital Touchpoint ไม่ได้เจาะจงถึงช่องทางที่มีผลต่อการตัดสินใจซื้อ จึงต้องมีเครื่องมือถัดมาที่เรียกว่า Customer Journey เพื่อวิเคราะห์ลำดับการเดินทางของกลุ่มเป้าหมาย ตั้งแต่เริ่มต้น จนตัดสินใจซื้อ ยกตัวอย่างเช่น ลำดับขั้นตอนการเดินทางของกลุ่มเป้าหมายเมื่อจะซื้อสินค้าหรือบริการ

1. เริ่มต้นจากการหาข้อมูลรีวิวผ่าน Search Engine
2. จากนั้นอ่านรีวิวเพิ่มเติมต่อจากเพจที่เกี่ยวข้อง
3. ศึกษาโปรโมชั่น วันและเวลา
4. เปรียบเทียบราคาหลาย ๆ แปรณต์
5. Search อีกครั้งเกี่ยวกับปัจจัยอื่นที่เกี่ยวข้อง

และเมื่อผ่านการเดินทางทั้ง 5 ขั้นตอน แล้วจึงเกิดการตัดสินใจซื้อ ด้วยเหตุนี้การวิเคราะห์ Customer Journey จึงจะช่วยให้บริษัทสามารถมองเห็นได้ว่า ช่องทางใดที่มีผลต่อการตัดสินใจ เพราะใช้เป็นแหล่งอ้างอิงเมื่อจะซื้อสินค้าและบริการบางอย่าง เป็นการวิเคราะห์ที่ช่วยคัดกรองว่า บริษัทควรทำโฆษณา และให้ความสำคัญกับช่องทางใดมากที่สุดเพื่อให้กลุ่มเป้าหมายตัดสินใจซื้อ (Kankate, 2562)

4.3.5 ระบบทำนายการเลิกใช้บริการของลูกค้า (Customer Churn Prediction System)

ภาพรวมของระบบ (System Overview)

ระบบทำนายการเลิกใช้บริการของลูกค้า (Customer Churn Prediction System) เป็นระบบที่พัฒนาขึ้นเพื่อใช้ในการทำนายว่าจากลูกค้าปัจจุบันของบริษัท ลูกค้ารายใดมีแนวโน้มจะเลิกใช้บริการของบริษัท (Churn) และลูกค้ารายใดที่จะยังคงใช้บริการกับทางบริษัทต่อไป (Loyal) เพื่อให้บริษัทสามารถแก้ปัญหาและลดอัตราการเลิกใช้บริการ (Churn Rate) ให้ลดลงเหลือน้อยที่สุด

ผู้ใช้ (Users)

- 1) ผู้บริหาร
- 2) ฝ่ายการตลาด
- 3) ฝ่ายบริการลูกค้า

ข้อมูลนำเข้า (Input Data)

เป็นข้อมูลของลูกค้าจำนวน 7,152 รายการ โดยมีตัวแปรที่นำมาใช้ทั้งหมด 7 ตัวแปร ดังตารางที่ 4-5

ตารางที่ 4-5: ข้อมูลนำเข้าสำหรับการทำนายการเลิกใช้บริการของลูกค้า

ลำดับ	ชื่อตัวแปร	คำอธิบาย	หน่วย
1	Age	อายุของลูกค้า	ปี
2	Gender	เพศของลูกค้า	-
	1	เพศชาย	
	2	เพศหญิง	
3	Timespan	จำนวนเดือนตั้งแต่การเริ่มสมัครใช้บริการ	เดือน
4	Destination	ประเทศปลายทางของการจัดส่งพัสดุ	-
5	Region	ภูมิภาคที่อยู่ของลูกค้า	-
	1	Bangkok (กรุงเทพฯ)	
	2	Central (ภาคกลาง)	
	3	Eastern (ภาคตะวันออก)	
	4	Northeastern (ภาคตะวันออกเฉียงเหนือ)	
	5	Northern (ภาคเหนือ)	
	6	Southern (ภาคใต้)	
	7	Western (ภาคตะวันตก)	
6	Payment Method	รูปแบบการชำระเงิน	-
	1	Cash	
	2	Credit Card	

ลำดับ	ชื่อตัวแปร	คำอธิบาย	หน่วย
7	Days Since LastVisit	จำนวนวันนับจากที่เข้าใช้งานเว็บไซต์ครั้ง	วัน

7.5.4 ขั้นตอนการดำเนินงาน (Working Processes)

ใช้หลักของกระบวนการหาความรู้แบบ Cross-Industry Standard Process for Data Mining (CRISP-DM) ซึ่งเป็นแนวทางในการดำเนินงานในการวิเคราะห์ข้อมูลขั้นสูง ซึ่งประกอบไปด้วย 6 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 การทำความเข้าใจทางธุรกิจ (Business Understanding)

ทำความเข้าใจเกี่ยวกับธุรกิจโลจิสติกส์และศึกษาขั้นตอนการดำเนินงานของหน่วยงาน รวบรวมปัญหาที่เกิดขึ้นในปัจจุบัน และความต้องการจากผู้บริหารและฝ่ายการตลาด เพื่อเป็นแนวทางในการพัฒนาระบบ นำความต้องการของผู้ใช้งานและกลุ่มเป้าหมายมาวิเคราะห์ และสรุปเป็นความต้องการของระบบ พร้อมทั้งกำหนดขอบเขตของโครงการ เพื่อให้ครอบคลุมการแก้ไขปัญหาที่เกิดขึ้น โดยจากการทำความเข้าใจในประเด็นเรื่อง “การทำนายการเลิกใช้บริการของลูกค้า” แล้วสรุปได้ว่าผู้ใช้ต้องการได้ข้อมูลแนวโน้มการเลิกใช้บริการของลูกค้าเพื่อให้ทราบถึงปัจจัยที่ส่งผลกระทบต่อ และสามารถดำเนินการป้องกันหรือยับยั้งการเลิกใช้บริการของลูกค้า เนื่องจากการสูญเสียลูกค้าไปย่อมส่งผลกระทบต่อการค้าเงินธุรกิจ โดยเฉพาะอย่างยิ่งในปัจจุบันซึ่งมีอัตราการแข่งขันสูง และการเข้ามาของเทคโนโลยีส่งผลให้ลูกค้าสามารถเปลี่ยนไปใช้บริการของคู่แข่งได้สะดวกขึ้นกว่าในอดีต การรักษารฐานลูกค้าเดิมเอาไว้ให้ได้จึงเป็นสิ่งสำคัญอย่างมาก

ขั้นตอนที่ 2 การทำความเข้าใจข้อมูล (Data Understanding)

รวบรวมข้อมูลทั้งหมดที่จำเป็นในการพัฒนาระบบการทำนาย และนำเข้าข้อมูลสู่ฐานข้อมูล (Data Base) จากนั้นจึงทำ Data Exploration เพื่อทำความเข้าใจกับข้อมูล และออกแบบรูปแบบข้อมูลที่เหมาะสมในการจัดทำ Predictive Model

โดยสำหรับระบบนี้พบว่ามึข้อมูลที่สามารถนำมาใช้ในการทำนายการเลิกใช้บริการของลูกค้าได้ดังนี้

- อายุของลูกค้า (Age)
- เพศของลูกค้า (Gender)

- ภูมิภาคที่อยู่ของลูกค้า (Region)
- รูปแบบการชำระเงิน (Payment Method)
- จำนวนวันนับจากที่เข้าใช้งานเว็บไซต์ครั้งล่าสุด (Days_Since_LastVisit)
- จำนวนเดือนตั้งแต่การเริ่มสมัครใช้บริการ (Timespan)
- ประเทศปลายทางของการจัดส่งพัสดุ (Destination)

ขั้นตอนที่ 3 การเตรียมข้อมูล (Data Preparation)

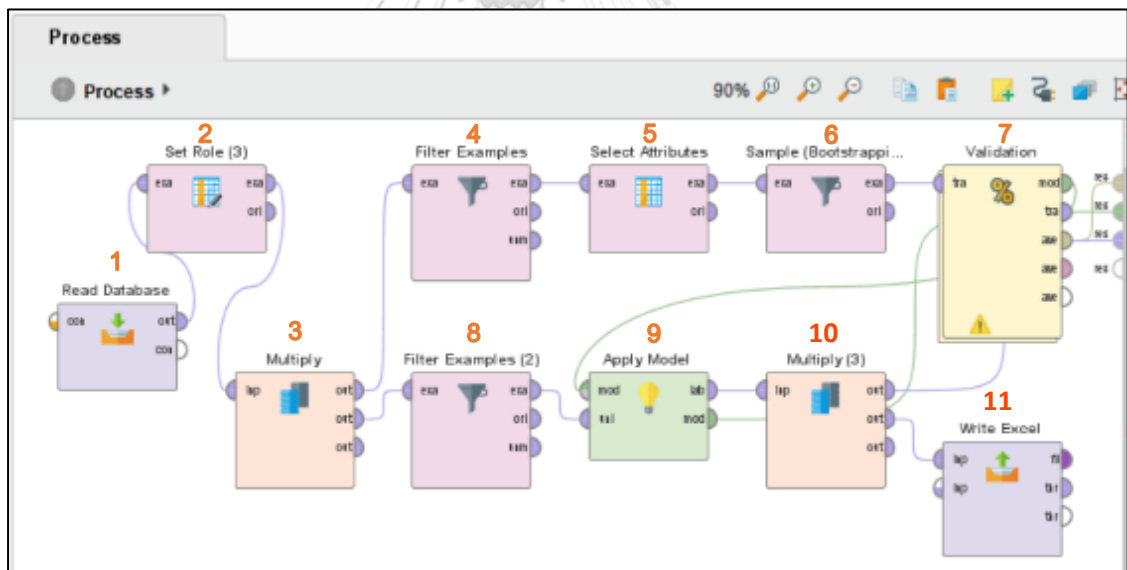
- 1) **Data Cleansing** โดยตรวจสอบและลบข้อมูลที่อาจแปลกแยกจากข้อมูลอื่น (Outlier) ลบข้อมูลที่ผิดรูปแบบและตรวจสอบข้อมูลที่หายไป เช่น ในกรณีตัวแปรเพศของลูกค้า (Gender) ซึ่งควรมีเฉพาะค่า 1 และ 2 แต่กลับพบว่าในบาง Record มีค่าอื่น ๆ ปรากฏอยู่ ซึ่งอาจเป็นไปได้ว่าเกิดความผิดพลาดในการบันทึกข้อมูลหรือความคลาดเคลื่อนระหว่างการถ่ายโอน อย่างไรก็ตาม เนื่องจากข้อมูลลักษณะนี้มีจำนวนน้อย จึงทำการลบทิ้งเพื่อไม่ให้ส่งผลต่อการสร้างโมเดลหรือประสิทธิภาพในการทำนายของโมเดล นอกจากนี้ยังพบว่ามีข้อมูลบาง Record ที่ตัวแปรบางส่วนขาดหายไป แต่มีจำนวนรวมแล้วไม่ถึง 15 Records ซึ่งอาจเกิดจากการบันทึกข้อมูลที่ตกหล่น จึงต้องทำการลบข้อมูลใน Record นั้น ๆ ออกเพื่อไม่ให้ส่งผลต่อการสร้างโมเดลในขั้นตอนต่อไป
- 2) **Set Data Role** ตั้งค่าให้ข้อมูลในคอลัมน์ที่ต้องการใช้โมเดลทำนายเปลี่ยนจาก Regular เป็น Label ซึ่งสำหรับกรณีนี้คือข้อมูลในคอลัมน์ Churn (การเลิกใช้บริการของลูกค้า)
- 3) **Data Sampling** เนื่องจากข้อมูลมีความ Imbalance จึงต้องทำการสุ่มข้อมูลที่จะนำมาใช้ในการ Train โมเดลโดยอาศัยเทคนิค Bootstrapping เพื่อทำการสุ่มข้อมูลชุดย่อย ๆ ขึ้นมาในแต่ละครั้งโดยไม่ซ้ำกัน ตามอัตราส่วนของชุดข้อมูลตัวอย่าง (Sample Ratio) ที่ได้กำหนดไว้ สำหรับชุดข้อมูลตัวอย่าง (Sample Size) ซึ่งเป็นข้อมูลนำเข้าทั้งหมดจำนวน 7,152 Records
- 4) **Data Splitting** แบ่งข้อมูล Train / Validate / Test ตามรูปแบบของ Predictive Model โดยแบ่งตามอัตราส่วน (Split Ratio) ที่ส่งผลให้โมเดลสามารถทำนายได้อย่างมีประสิทธิภาพสูงสุด

ขั้นตอนที่ 4 การสร้างโมเดล (Modeling)

เนื่องจากวัตถุประสงค์ของการทำนายคือเพื่อให้ทราบว่าลูกค้าที่มีแนวโน้มจะเลิกใช้บริการ (Churn) นั้นคือใครและมีลักษณะอย่างไร หลังจากพิจารณาลักษณะของชุดข้อมูลแล้วจึงจะมีการนำ Decision Tree Model มาใช้เพื่อให้ทราบถึงคุณลักษณะของลูกค้า และใช้ Random Forest Model เพื่อเน้นด้านความแม่นยำในการระบุว่าลูกค้าคนใดบ้างที่มีแนวโน้มเลิกใช้บริการ โดยรายละเอียด จุดเด่น จุดด้อย และพารามิเตอร์ของ Decision Tree Model มีดังที่ได้อธิบายไว้ในหัวข้อที่ 2.3 และ 4.3.3 ตามลำดับ

ขั้นตอนการสร้างโมเดล Decision Tree ในโปรแกรม RapidMiner เพื่อทำนายการเลิกใช้บริการของลูกค้า

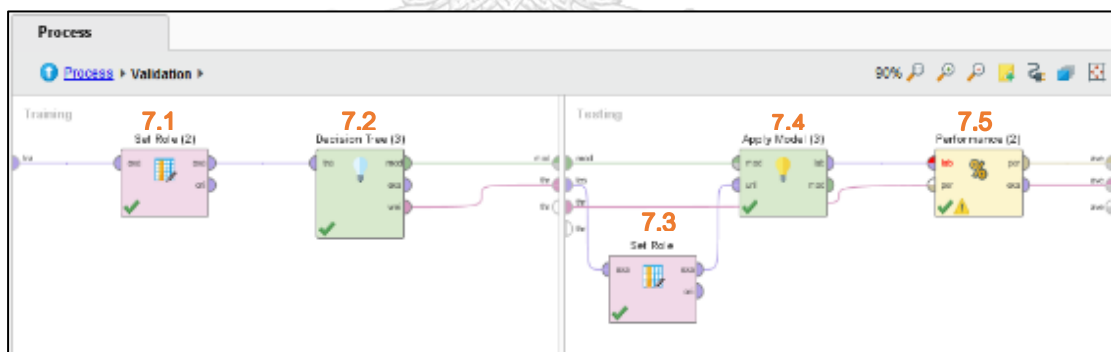
Process ทั้งหมดสำหรับการทำนายการเลิกใช้บริการของลูกค้าด้วยโมเดล Decision Tree สามารถแสดงได้ดังรูปที่ 4-27 และ 4-28



รูปที่ 4-27: Process การทำนายการเลิกใช้บริการของลูกค้าด้วยโมเดล Decision Tree

- 1) Read Database: เชื่อมต่อกับฐานข้อมูลเพื่อเข้าถึงชุดข้อมูลสำหรับการสร้างโมเดล
- 2) Set Role: กำหนดให้ตัวแปร Churn (สถานะการเลิกใช้บริการของลูกค้า) ให้ทำหน้าที่เป็น Label
- 3) Multiply: เนื่องจากต้องการนำ Output ที่มีการ Set Role แล้วไปใช้แยกส่วนกัน

- 4) Filter Example: เลือกเฉพาะข้อมูลส่วนที่ Label ไม่ Missing เพื่อนำไปใช้เป็น Training Set
- 5) Select Attributes: เลือกตัวแปรที่จะนำมาใช้ทำนาย ซึ่งในที่นี้เลือกไว้ทั้งหมด 7 ตัวแปรด้วยกัน
- 6) Sample (Bootstrapping): ทำการสุ่มข้อมูลที่จะนำมาใช้ เนื่องจากข้อมูลมีความ Imbalance
- 7) Validation: ตรวจสอบค่าความแม่นยำของโมเดล Decision Tree ที่สร้างขึ้นภายในโอเปอร์เรเตอร์นี้
- 8) Filter Example: เลือกเฉพาะข้อมูลส่วนที่ Label is Missing เพื่อให้โมเดลทำนายผลต่อไป
- 9) Apply Model: นำโมเดลที่สร้างขึ้นใหม่มาใช้ทำนายข้อมูลส่วนที่เลือกไว้ในขั้นตอนที่ 8
- 10) Multiply: เนื่องจากต้องการนำ Output ที่มีการ Apply Model แล้วไปใช้แยกส่วนกัน
- 11) Write Excel: เป็นการ Export ข้อมูลที่ได้หลังการ Apply Model แล้วออกมาเป็นไฟล์ Excel



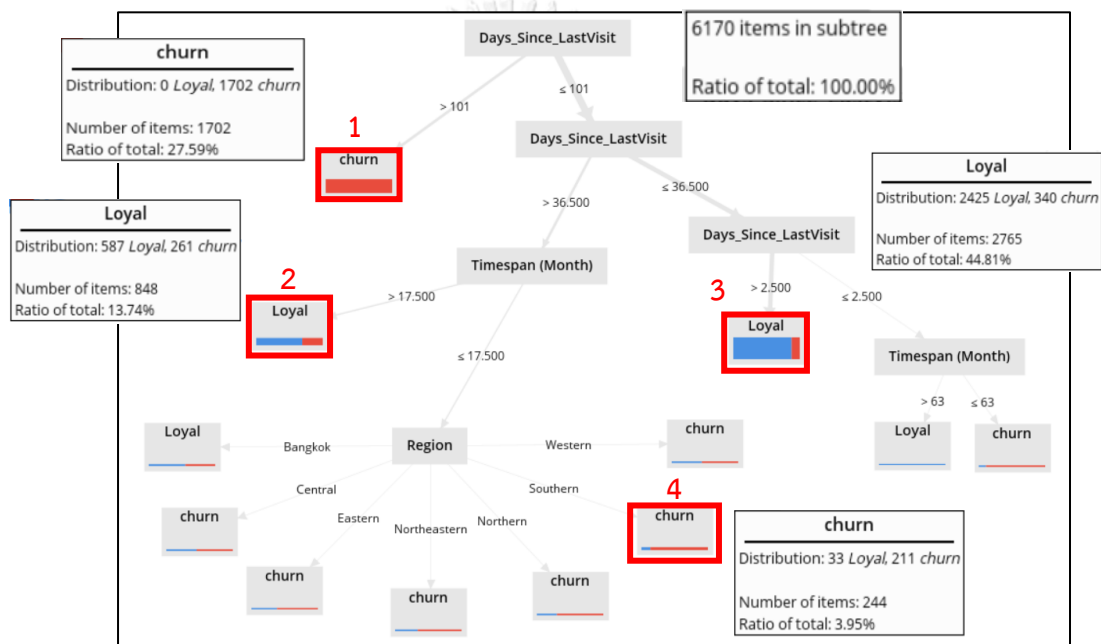
รูปที่ 4-28: Sub-Process ภายในโอเปอร์เรเตอร์ Validation

จากรูปที่ 4-28 อธิบายได้ดังนี้

- 7.1 Set Role: กำหนดแอททริบิวต์ที่จะทำหน้าที่เป็น Label ซึ่งในที่นี้คือแอททริบิวต์ Churn
- 7.2 Decision Tree: เป็นการสร้างโมเดล Decision Tree ขึ้นจาก Training Set เพื่อใช้สำหรับทำนายข้อมูลต่อไป

- 7.3 Set Role: กำหนดแอททริบิวต์ที่จะทำหน้าที่เป็น Label ซึ่งในที่นี้คือแอททริบิวต์ Churn
- 7.4 Apply Model: นำโมเดล Decision Tree ที่ได้สร้างขึ้นมาใช้กับข้อมูลที่เป็น Test Set
- 7.5 Performance: เป็นการตรวจสอบค่าความแม่นยำของโมเดล Decision Tree ที่ได้สร้างขึ้น

ผลลัพธ์ที่ได้คือโมเดล Decision Tree ซึ่งสามารถนำไปวิเคราะห์หรือหาคำอธิบายต่อได้ ดังรูปที่ 4-29



รูปที่ 4-29: ผลลัพธ์โมเดล Decision Tree ที่ได้สร้างขึ้นจากข้อมูล Training Set

จากผลลัพธ์เบื้องต้นที่ได้จากโมเดล Decision Tree นั้นมีทั้งผลที่เป็นไปตามความคาดหมาย และผลลัพธ์ที่มีความน่าสนใจเพราะแตกต่างจากสมมติฐานที่ตั้งไว้ โดยเห็นได้ว่าตัวแปรหลัก ๆ นั้นมีเรื่องของเงื่อนไขเวลาเป็นสำคัญ ดังนั้นสิ่งที่สามารถคาดเดาได้คือ Account ของลูกค้าที่ขาดการเคลื่อนไหวหรือไม่ได้ใช้บริการเป็นเวลานาน (Days_Since_LastVisit สูง) มีแนวโน้มที่จะเลิกใช้บริการ (Churn) สูง ดังนั้นต่อไปบริษัทอาจวางแนวทางในการดำเนินงาน เช่น หากพบ Account ใดของ ลูกค้าที่ไม่มีการเคลื่อนไหวเป็นเวลานานก็อาจดำเนินการเพื่อกระตุ้นให้ลูกค้ากลับมาใช้บริการ เช่น ส่งโปรโมชั่นจูงใจต่าง ๆ เพื่อเป็นการดำเนินการเบื้องต้นและป้องกันก่อนที่ลูกค้าจะเลิกใช้บริการ

และสำหรับ 4 Leaf Nodes ที่มีความน่าสนใจ คือบริเวณโหนดหมายเลข 1 เป็นลูกค้ากลุ่มที่มีโอกาสสูงว่าจะเลิกใช้บริการซึ่งมีสัดส่วนมากถึง 27.59 เปอร์เซ็นต์ของข้อมูลนำเข้าทั้งหมดในโมเดล โดยเป็นกลุ่มที่มีลักษณะคือเข้ามาใช้งานครั้งล่าสุดเกินกว่า 101 วันแล้ว และจากข้อมูลทั้งหมดที่มาจากในโหนดนี้ก็พบว่าเป็นกลุ่มที่ Churn ทั้งหมด

ส่วนอีกด้านสำหรับกลุ่มหมายเลข 2 เป็นลูกค้ากลุ่มที่มีโอกาสสูงว่าจะยังใช้บริการต่อไปโดยคิดเป็น 13.74 เปอร์เซ็นต์ของข้อมูลทั้งหมดในโมเดล และเมื่อเทียบกันแล้วพบว่ามีลูกค้าจำนวนมากถึง 69 เปอร์เซ็นต์ที่ Loyal จึงมีความน่าจะเป็นสูงที่ลูกค้าในกลุ่มนี้จะยังคงใช้บริการต่อไป โดยมีลักษณะคือเป็นกลุ่มที่เข้ามาใช้งานครั้งล่าสุดอยู่ในช่วงประมาณ 37 - 101 วัน และสมัครเป็นสมาชิกมาได้ระยะหนึ่งแล้วคือนานกว่า 17 เดือน (หรือประมาณหนึ่งปีครึ่ง)

เช่นเดียวกับลูกค้าในกลุ่มที่ 3 ซึ่งมีแนวโน้มว่าจะยังใช้บริการต่อไปเช่นกันจำนวน 44.81 เปอร์เซ็นต์โดยเป็นกลุ่มที่เข้ามาใช้งานครั้งล่าสุดอยู่ในช่วง 3-36 วัน และสำหรับโหนดนี้ก็มีสัดส่วนของกลุ่มที่ Churn กับ Loyal ต่างกันอย่างชัดเจนคือจะมีกลุ่มที่ Loyal คิดเป็น 87 เปอร์เซ็นต์ของข้อมูลทั้งหมดที่มาจากในโหนดนี้

และอีกจุดหนึ่งที่มีความน่าสนใจคือในด้านภูมิภาคที่อยู่ของลูกค้า ซึ่งผลลัพธ์ที่ออกมา นั้นแตกต่างจากที่คาดการณ์ไว้ว่าลูกค้าในพื้นที่กรุงเทพมหานครหรือภาคกลางน่าจะมีปัญหาเลิกใช้บริการจำนวนมากเนื่องจากมีบริษัทที่ทำธุรกิจลักษณะใกล้เคียงกันมาก ขณะที่จำนวนสัดส่วนลูกค้าน้อย ทำให้การแข่งขันสูง แต่ผลที่ออกมา กลับกลายเป็นว่าภาคใต้ (Southern) ใน Leaf Node หมายเลข 4 มีอัตราเลิกใช้บริการมากกว่าภาคอื่น ๆ โดยกลุ่มที่ Churn คิดเป็นสัดส่วนมากถึงประมาณ 86 เปอร์เซ็นต์ ดังนั้นสำหรับใน Leaf Node นี้ ทางบริษัทอาจจะมีการเก็บข้อมูลหรือมีการศึกษาวิเคราะห์เพิ่มเติมถึงสาเหตุต่อไปว่าสาเหตุอาจเกิดจากปัจจัยใดบ้างที่ส่งผลกระทบต่อการใช้บริการของลูกค้า เพื่อเร่งหาแนวทางแก้ไข เช่น อาจเป็นปัจจัยภายนอกอย่างการที่มีบริษัทคู่แข่งเข้าไปเปิดบริการเพิ่มในพื้นที่นั้น ๆ แล้วดึงลูกค้าไป เกิดจากปัจจัยในด้านที่เกี่ยวกับลูกค้าเอง หรือเกิดจากปัจจัยภายในของตัวบริษัทเองอย่างการให้บริการที่อาจเกิดปัญหา พนักงานขาดความกระตือรือร้น ใส่ใจลูกค้าน้อยลง โปรโมชันไม่น่าสนใจเท่าของบริษัทคู่แข่ง รวมถึงการอิมตัวของตลาดการส่งออกในพื้นที่นั้น ๆ ฯลฯ

พารามิเตอร์ของโมเดล Random Forest ในโปรแกรม RapidMiner

- 1) **Number_of_trees:** พารามิเตอร์นี้ใช้ในการระบุจำนวนของต้นไม้ที่จะสร้าง และต้นไม้แต่ละต้นจะมีการเลือกชุดข้อมูลย่อยด้วยวิธีการ Bootstrapping ซึ่งจำนวนต้นไม้ใน Random Forest ที่มากขึ้นจะทำให้ Performance ของโมเดลดีขึ้นจนถึงจุดหนึ่งที่ Performance เริ่มจะนิ่ง จนจำนวนต้นไม้ไม่ส่งผลต่อ Performance ของโมเดลอีกต่อไป และการมีต้นไม้จำนวนมากอาจส่งผลให้เวลาในการรันโมเดลยาวนานขึ้นด้วย ดังนั้นจำนวนต้นไม้ที่แนะนำในช่วงของการทดสอบโมเดลเบื้องต้นจึงควรกำหนดค่าน้อย ๆ ก่อน (เช่น 50–100 ต้น) และเมื่อเป็นการ Train อย่างจริงจังแล้วจึงอาจกำหนดค่าเป็น 1000 ต้นขึ้นไป
- 2) **Criterion:** เป็นส่วนที่ใช้เลือกเกณฑ์สำหรับระบุแอททริบิวต์ในการแตกกิ่ง (Split) ของต้นไม้ โดยจะมีการปรับค่าของการแตกกิ่งให้เหมาะสมตามเกณฑ์ที่เลือก ซึ่งจะสามารถเลือกได้หนึ่งเกณฑ์จาก 5 ตัวเลือก ต่อไปนี้:

(1.) **Information_gain:** เป็นตัวชี้วัดการแบ่งข้อมูลออกเป็นชุดข้อมูลย่อย โดยหลักการก็คือยิ่งแอททริบิวต์ใดมี Information Gain สูงจะยิ่งเหมาะกับการนำมาใช้งาน เนื่องจากมาตรวัดนี้จะใช้วิธีคำนวณค่า Entropy ของแอททริบิวต์ทั้งหมดซึ่งเป็นการวัดความแตกต่างหรือการกระจายกระจายของข้อมูล (Entropy มีค่าระหว่าง 0-1 โดยจะมีค่าสูงขึ้นเมื่อข้อมูลมีความแตกต่างกันมาก และในทางตรงข้ามถ้าข้อมูลมีความคล้ายกันมาก ค่า Entropy ก็ต่ำ) ดังนั้น Information Gain จะทำงานโดยวัดค่า Entropy ก่อนการแบ่งข้อมูลออกตามคลาสและหลังการแบ่งว่ามีประสิทธิภาพดีขึ้นหรือไม่ ถ้ามีประสิทธิภาพดีขึ้น ค่า Information Gain จะมีค่าสูง (โดยที่ค่า Entropy น้อยที่สุดเพื่อให้ข้อมูลคล้ายกัน) และแอททริบิวต์นั้นจะถูกเลือกมาใช้ในการแตกกิ่ง แต่วิธีนี้จะเกิด Bias ในกรณีของการเลือก Attributes ที่มีจำนวนมาก จึงเป็นเกณฑ์ที่ใช้ได้ดีสำหรับการทำงานกับเงื่อนไขที่ไม่ได้ซับซ้อนมากนัก

(2.) **Gain_ratio:** เป็นตัวชี้วัดการแบ่งชุดข้อมูลออกเป็นชุดข้อมูลย่อยที่พัฒนามาจาก Information Gain เนื่องจากเมื่อใช้ค่า Information Gain ในการแบ่งชุดข้อมูลจะส่งผลให้เกิดความเอนเอียง (Bias) ขึ้นเมื่อแอททริบิวต์ที่ทำการพิจารณามีค่าที่เกิดขึ้นเป็นจำนวนมาก ตัวอย่างเช่น แอททริบิวต์รายการสินค้าที่มีค่าเป็นรหัสสินค้าต่าง ๆ

(Product_ID) ซึ่งหากทำการแบ่งชุดข้อมูลตามแอททริบิวต์รายการสินค้าจะทำให้มีชุดข้อมูลย่อยเป็นจำนวนมาก (ชุดข้อมูลย่อยจะมีจำนวนเท่ากับรายการสินค้าทั้งหมดที่มีในชุดข้อมูล) โดยแต่ละชุดข้อมูลย่อยจะมีข้อมูลเพียงหนึ่งเรคคอร์ดเท่านั้นและจะทำให้ชุดข้อมูลย่อยนั้น ๆ มีข้อมูลที่มีหมวดหมู่ของข้อมูลเหมือนกันทั้งหมด (เพราะมีเพียงหนึ่งเรคคอร์ดจึงเหมือนกันทั้งหมด) ผลที่ตามมาคือจะทำให้ค่า Information Gain ของแอททริบิวต์รายการสินค้านั้นมีค่าสูง และแอททริบิวต์ที่มีลักษณะคล้ายกับแอททริบิวต์รายการสินค้ามักจะถูกเลือกเพื่อใช้ในการแบ่งชุดข้อมูลเสมอ (โกเมศ อัมพวัน, 2561)

จากปัญหาข้างต้นจึงได้มีนักวิจัยพยายามที่จะลดทอนความเอนเอียงลง โดยพัฒนาตัวชี้วัดการแบ่งข้อมูลใหม่ที่มีชื่อว่า Gain Ratio ขึ้นมา ซึ่งเป็นวิธีคัดเลือกเพื่อให้ได้ตัวแปรที่เป็นตัวแบ่งข้อมูลออกเป็นกลุ่มย่อยที่มีสมาชิกภายในกลุ่มเป็นชนิดเดียวกันมากที่สุด (Homogeneous) โดยมาตรวัด Gain Ratio นั้นเป็นอัตราส่วนของค่าเกน (Gain หรือ Information Gain) กับค่าสารสนเทศการแบ่งกลุ่ม (Split Info) เพื่อเป็นการลดอิทธิพลของตัวแปรที่มีค่าหลายค่า ซึ่งผลที่ได้จากการใช้เทคนิคนี้จะได้ลำดับของตัวแปรซึ่งตัวแปรที่อยู่ลำดับแรก ๆ จะถือว่ามามีอิทธิพลในการพยากรณ์ตัวแปรเป้าหมายมากกว่าตัวแปรในลำดับถัดไป ส่งผลให้สามารถพิจารณาเลือกจำนวนตัวแปรที่เหมาะสมได้อย่างมีประสิทธิภาพ (นิภาพร ชนะมาร, 2557)

- (3.) **Gini_index:** เป็นตัวชี้วัดที่จะพิจารณาความไม่บริสุทธิ์ของชุดข้อมูล (การวัดระดับความบริสุทธิ์ของ Node คือเมื่อค่า Gini Index เท่ากับศูนย์แปลว่า Node บริสุทธิ์) โดยชี้วัดถึงความไม่สม่ำเสมอในการกระจายตัวของ Label (แอททริบิวต์ที่เป็นคำตอบ) ดังนั้นการแตกกิ่งตรงแอททริบิวต์ที่เลือกจึงส่งผลดีเนื่องจากทำให้ค่าเฉลี่ย Gini Index ของชุดข้อมูลย่อยที่ได้นั้นลดลง (Boonyang, 2561)
- (4.) **Accuracy:** เป็นการแตกกิ่งโดยเลือกแอททริบิวต์ที่ทำให้ค่าความแม่นยำ (Accuracy) ของต้นไม้ทั้งต้นมีค่าสูงสุด
- (5.) **Least_square:** เป็นการแตกกิ่งโดยเลือกแอททริบิวต์ที่จะช่วยลดระยะห่างยกกำลังสอง (Squared Distance) ระหว่างค่าเฉลี่ยของค่าข้อมูลในโหนดโดยพิจารณาจากค่าที่แท้จริง

- 3) **Maximal_depth:** เนื่องจากความลึกของต้นไม้จะแตกต่างกันออกไปโดยขึ้นอยู่กับขนาดและลักษณะของชุดข้อมูล พารามิเตอร์นี้จึงถูกใช้เพื่อจำกัดความลึกของต้นไม้ โดยหากตั้งค่าเป็น “-1” จะไม่มีการกำหนดความลึกของต้นไม้ และในกรณีนี้ต้นไม้จะถูกสร้างขึ้นจนกว่าจะถูกหยุดด้วยเกณฑ์อื่น ๆ แต่หากตั้งค่าเป็น “1” จะเป็นการสร้างต้นไม้ที่มีเพียงโหนดเดียวเท่านั้น
- 4) **Apply_prepruning:** พารามิเตอร์นี้เป็นส่วนที่ใช้กำหนดว่าควรมีเกณฑ์การหยุดสร้างต้นไม้มากกว่าเรื่องของความลึกสูงสุดสูงสุด (Maximal Depth) หรือไม่ในระหว่างการสร้างโมเดลต้นไม้ ซึ่งหากกดเลือกไว้ พารามิเตอร์ Minimal gain, Minimal leaf size, Minimal size for split และ Number of prepruning alternatives จะถูกนำมาใช้เป็นเกณฑ์ในการหยุดสร้างต้นไม้
- 5) **Minimal_gain:** เนื่องจากค่า Gain (ซึ่งเป็นค่าที่ใช้บอกว่าแอททริบิวต์ที่ทำหน้าที่เป็น Root Node สามารถจำแนกข้อมูลได้ดีมากน้อยเพียงใด) จะถูกคำนวณก่อนที่จะมีการแตกกิ่ง การแตกกิ่งจึงจะเกิดขึ้นเฉพาะเมื่อโหนดนั้น ๆ มีค่า Gain สูงกว่าค่า Minimal Gain ที่ถูกกำหนดไว้ ดังนั้นการตั้งค่า Minimal Gain ยิ่งสูงจึงยิ่งส่งผลให้เกิดการแตกกิ่งน้อยและได้ผลลัพธ์เป็นต้นไม้ที่มีขนาดเล็ก และการตั้งค่า Minimal Gain ที่สูงจนเกินไปจะเป็นตัวขัดขวางการแตกกิ่งและเป็นการสร้างต้นไม้ที่มีเพียงโหนดเดียว
- 6) **Minimal_leaf_size:** ขนาดของใบไม้ (Leaf Size) หมายถึงจำนวนข้อมูลในแต่ละกลุ่มข้อมูลที่มีการแบ่งย่อยออกมาแล้ว ดังนั้นต้นไม้จึงจะถูกสร้างในลักษณะที่ใบไม้ทุกใบมีจำนวนข้อมูลขั้นต่ำใน Leaf Node ตามที่มีการกำหนดไว้ การหยุด Split Node นั้น ๆ หากมีจำนวนข้อมูลต่ำกว่าที่กำหนดจะเป็นการลด Overfitting โดยมักกำหนดค่า Min_leaf_size ให้สอดคล้องตามขนาดของ Data Set (เช่น ตั้งแต่ 2-100 ตามขนาดข้อมูล)
- 7) **Minimal_size_for_split:** หมายถึงขนาดของโหนดหรือจำนวนของข้อมูลในหน่วยนั้น ๆ ซึ่งจะมีเฉพาะโหนดที่มีขนาดมากกว่าหรือเท่ากับขนาดต่ำสุดที่มีการตั้งค่าไว้เท่านั้นที่จะมีการแตกกิ่งออกไป
- 8) **Number_of_prepruning_alternatives:** เมื่อการ Prepruning นั้นเป็นการขัดขวางการแตกกิ่งที่บางโหนด พารามิเตอร์นี้จะเป็นตัวปรับจำนวนโหนดทางเลือก (Alternative Nodes) ที่จะถูกทดสอบสำหรับการแตกกิ่ง และจะทำงานเมื่อมีการ Prepruning ควบคู่ไปกับกระบวนการสร้างต้นไม้ ซึ่งอาจทำการปิดกั้นการแตกกิ่งที่บางโหนดหากการแตกกิ่งที่โหนดนั้น ๆ ไม่ได้ช่วย

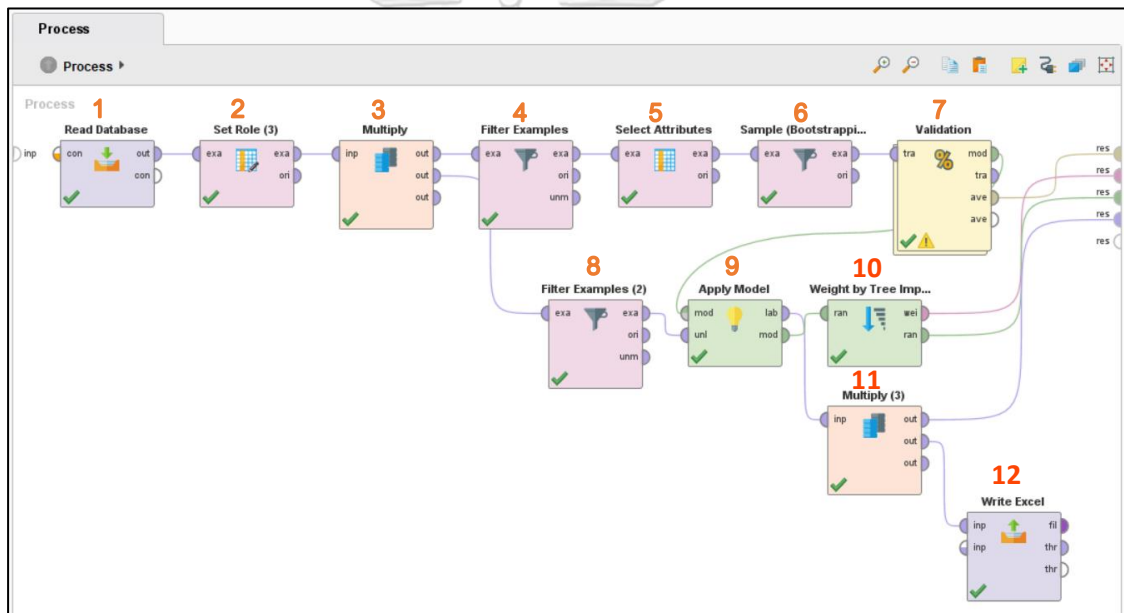
เพิ่มประสิทธิภาพในการจำแนกแบ่งกลุ่มของต้นไม้ทั้งต้น จึงจะส่งผลให้เกิดการลองแตกกิ่งที่ โหนดทางเลือก (Alternative Nodes) ในกรณีเช่นนี้

- 9) **Apply_pruning:** ต้นไม้ของโมเดล Random Forest สามารถถูกตัดแต่งกิ่ง (Pruned) ได้ หลังจากถูกสร้างขึ้นแล้ว ซึ่งหากมีการคัดเลือกตัวเลือกนี้ไว้กิ่งไม้บางกิ่งจะถูกแทนที่ด้วยใบไม้ ตามที่มีการตั้งค่าไว้ในพารามิเตอร์ Confidence แต่พารามิเตอร์นี้จะไม่สามารถใช้ได้กับเกณฑ์ การแตกกิ่งแบบ “Least_square”
- 10) **Confidence:** พารามิเตอร์นี้จะเป็นส่วนที่ใช้กำหนดระดับความเชื่อมั่น (Confidence level) ที่ใช้สำหรับการคำนวณข้อผิดพลาดของการตัดแต่งกิ่ง
- 11) **Random_splits:** หากกดเปิดใช้พารามิเตอร์นี้จะทำให้เกิดการเลือกแบ่งแอททริบิวต์ตัวเลข แบบสุ่มแทนการปรับให้เหมาะสม และการเปิดใช้งานพารามิเตอร์นี้ในขณะที่ปิดฟังก์ชันตัดแต่ง กิ่ง (Pruning) จะเป็นการกำหนดค่า Random Forest ให้กลายเป็นต้นไม้ที่สุ่มอย่างมาก (หรือ ที่เรียกว่า Extra-Tree) นอกจากนี้ยังเพิ่มความรวดเร็วในกระบวนการสร้างโมเดลอีกด้วย
- 12) **Guess_subset_ratio:** หากเปิดใช้พารามิเตอร์นี้จะมีการกำหนดอัตราส่วนแอททริบิวต์ด้วย สูตร $\text{int}(\log(m) + 1)$ หรือสามารถระบุอัตราส่วนของแอททริบิวต์ที่จะถูกเลือกแบบสุ่มเพื่อ ทำการทดสอบได้ด้วยพารามิเตอร์ Subset Ratio
- 13) **Subset_ratio:** พารามิเตอร์นี้ใช้เพื่อระบุอัตราส่วนของแอททริบิวต์ที่จะถูกเลือกแบบสุ่มเพื่อ ทำการทดสอบ
- 14) **Voting_strategy:** เป็นส่วนที่ใช้ระบุกลยุทธ์การทำนายในกรณีที่ต้นไม้มีการทำนายที่ไม่ สอดคล้องกัน โดยพารามิเตอร์นี้ไม่สามารถใช้ได้กับเกณฑ์การแตกกิ่งแบบ “Least_square”
 - Confidence_vote: เป็นการเลือกคลาสที่มีความเชื่อมั่นสะสม (Accumulated Confidence) สูงสุด
 - Majority_vote: เลือกคลาสที่โมเดลต้นไม้ส่วนใหญ่ทำนายออกมา
- 15) **Use_local_random_seed:** ใช้ในการกำหนดว่าจะมีการตั้งค่าตัวเลขสำหรับการสุ่มข้อมูล หรือไม่
- 16) **Local_random_seed:** หากมีการเปิดใช้งานพารามิเตอร์ Use local random seed พารามิเตอร์นี้จะถูกใช้เพื่อกำหนดตัวเลขในการสุ่มข้อมูล

17) **Enable_parallel_execution**: พารามิเตอร์นี้ใช้สำหรับเปิดกระบวนการทำงานแบบคู่ขนานของการสร้างโมเดลโดยกระจายการสร้างต้นไม้แบบสุ่ม (Random Tree) ไปให้กับกระบวนการทำงานของ CPU (CPU Threads) ทั้งหมดที่สามารถทำได้ ดังนั้นจึงควรปิดกระบวนการทำงานแบบคู่ขนานหากพบว่ามีปัญหาในด้านของหน่วยความจำ

การสร้างโมเดล Random Forest ในโปรแกรม RapidMiner เพื่อทำนายการเลิกใช้บริการของลูกค้า

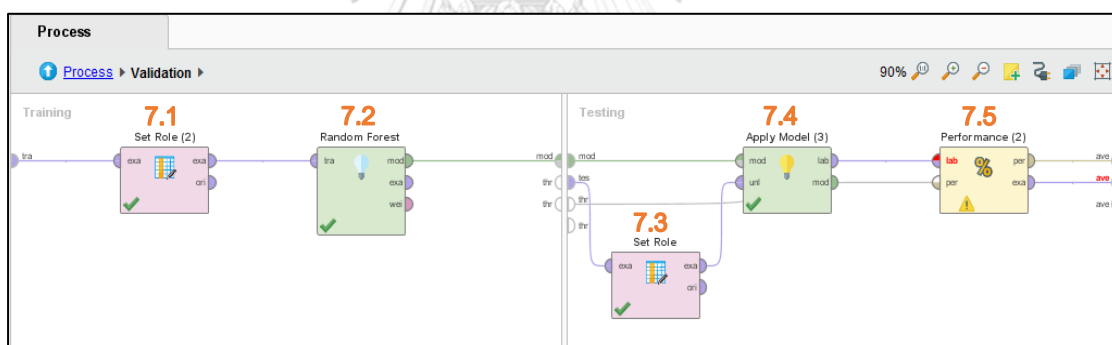
Process ทั้งหมดสำหรับการทำนายการเลิกใช้บริการของลูกค้าด้วยโมเดล Random Forest มีรายละเอียดดังรูปที่ 4-30 และ 4-31



รูปที่ 4-30: Process การทำนายการเลิกใช้บริการของลูกค้าด้วยโมเดล Random Forest

- 1) Read Database: เชื่อมต่อกับฐานข้อมูลเพื่อเข้าถึงชุดข้อมูลสำหรับการสร้างโมเดล
- 2) Set Role: กำหนดให้ตัวแปร Churn (การเลิกใช้บริการของลูกค้า) ทำหน้าที่เป็น Label
- 3) Multiply: เนื่องจากต้องการนำ Output ที่มีการ Set Role แล้วไปใช้แยกส่วนกัน
- 4) Filter Example: เลือกเฉพาะข้อมูลส่วนที่ Label ไม่ Missing เพื่อใช้เป็น Training Set
- 5) Select Attributes: เลือกตัวแปรที่จะนำมาใช้ทำนาย ซึ่งในที่นี้เลือกไว้ทั้งหมด 7 ตัวแปร

- 6) Sample (Bootstrapping): ทำการสุ่มข้อมูลที่นำมาใช้ เนื่องจากข้อมูลมีความ Imbalance
- 7) Validation: ตรวจสอบความแม่นยำของโมเดล Random Forest ที่สร้างขึ้นภายในโอเพอร์เรเตอร์นี้
- 8) Filter Example: เลือกเฉพาะข้อมูลส่วนที่ Label is Missing เพื่อให้โมเดลทำนายผลต่อไป
- 9) Apply Model: นำโมเดลที่สร้างขึ้นใหม่มาใช้ทำนายข้อมูลส่วนที่ได้เลือกไว้ในขั้นตอนที่ 8
- 10) Weight by Tree Importance: ตรวจสอบค่าน้ำหนักของตัวแปรซึ่งสะท้อนถึงความสำคัญของตัวแปรแต่ละตัวในการสร้างโมเดล
- 11) Multiply: เนื่องจากต้องการนำ Output ที่มีการ Apply Model แล้วไปใช้แยกส่วนกัน
- 12) Write Excel: เป็นการ Export ข้อมูลที่ได้หลังการ Apply Model แล้วออกมาเป็นไฟล์ Excel



รูปที่ 4-31: Sub-Process ภายในโอเพอร์เรเตอร์ Validation

จากรูปที่ 4-31 อธิบายได้ดังนี้

- 7.1 Set Role: กำหนดแอททริบิวต์ที่จะทำหน้าที่เป็น Label ซึ่งในที่นี้คือแอททริบิวต์ Churn
- 7.2 Random Forest: เป็นการสร้างโมเดล Random Forest ขึ้นจาก Training Set เพื่อใช้สำหรับทำนายข้อมูลที่ต้องการต่อไป
- 7.3 Set Role: กำหนดแอททริบิวต์ที่จะทำหน้าที่เป็น Label ซึ่งในที่นี้คือแอททริบิวต์ Churn

7.4 Apply Model: นำโมเดล Random Forest ที่ได้สร้างขึ้นมาใช้กับข้อมูลที่เป็น Test set

7.5 Performance: เป็นการตรวจสอบค่าความแม่นยำของโมเดล Random Forest ที่ได้สร้างขึ้น

ขั้นตอนที่ 5 การประเมินผล (Evaluation)

การวัดประสิทธิภาพของ Predictive Model สามารถทำได้โดยดูจากค่าความถูกต้อง (Accuracy) ของทั้งสองโมเดลที่ถูกสร้างขึ้น เพื่อให้สามารถเลือกใช้ได้อย่างเหมาะสมตามวัตถุประสงค์ และลักษณะเด่นของแต่ละโมเดล โดยอาจมีการปรับปรุงโมเดลเพิ่มเติมอย่างต่อเนื่อง เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด และการ Evaluate โมเดลที่สร้างขึ้นนั้นสามารถพิจารณาได้จาก Confusion Matrix ซึ่งถือเป็นเครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนาย หรือ Prediction จากโมเดล โดยมีแนวคิดในการวัดว่าสิ่งที่โมเดลทำนายกับสิ่งที่เกิดขึ้นจริงนั้นมีสัดส่วนเป็นอย่างไร ซึ่งรายละเอียดเกี่ยวกับ Confusion Matrix มีดังที่อธิบายไว้ในหัวข้อที่ 2.3

และหลังจากที่ได้ทำการประเมินผลแล้วพบว่าค่าความแม่นยำของโมเดล Decision Tree มีรายละเอียดดังรูปที่ 4-32

accuracy: 84.84%			
	true Loyal	true churn	class precision
pred. Loyal	6963	1232	84.97%
pred. churn	951	5249	84.66%
class recall	87.98%	80.99%	

รูปที่ 4-32: ค่าความแม่นยำของโมเดล Decision Tree ในการทำนายการเลิกใช้บริการของลูกค้า

ค่าความแม่นยำของโมเดล Random Forest มีรายละเอียดดังรูปที่ 4-33

accuracy: 88.60%			
	true Loyal	true churn	class precision
pred. Loyal	263	41	86.51%
pred. churn	16	180	91.84%
class recall	94.27%	81.45%	

รูปที่ 4-33: ค่าความแม่นยำของโมเดล Random Forest ในการทำนายการเลิกใช้บริการของลูกค้า

ขั้นตอนที่ 6 การใช้งาน (Deployment)

สำหรับการใช้งาน สามารถนำเอาโมเดลที่สร้างขึ้นไปใช้ในการทำนายข้อมูลเพื่อเป็นประโยชน์ต่อการดำเนินธุรกิจในอนาคต เช่น ในด้านการทำนายว่าลูกค้าลักษณะใดมีแนวโน้มจะยกเลิกการใช้บริการจัดส่งพัสดุของบริษัท โดยผลลัพธ์จาก Decision Tree Model สามารถนำมาทำความเข้าใจและอธิบายย้อนกลับได้จึงสามารถนำมาวิเคราะห์เพิ่มเติมถึงเหตุผลที่มาที่ไปหรือลักษณะเฉพาะของลูกค้าที่มีแนวโน้มยกเลิกบริการในอนาคตได้ ส่วน Random Forest Model นั้นให้ผลลัพธ์ที่เน้นเรื่องความแม่นยำแต่ไม่สามารถแปลผลย้อนกลับได้ ดังนั้นจึงอาจนำไปใช้ในกรณีที่ไม่ได้ต้องการคำอธิบายหรือรายละเอียดต่าง ๆ เกี่ยวกับลูกค้ามากนัก แต่อาจเน้นในเรื่องของตัวเลขในภาพรวม เพื่อให้ทราบถึงสถิติ หรือมุ่งเรื่องความแม่นยำในการทำนายเพื่อให้สามารถยับยั้งการเลิกใช้บริการของลูกค้าให้ได้มากที่สุด รวมถึงสามารถตรวจสอบถึงค่าน้ำหนักของตัวแปรที่นำเข้าสู่โมเดลดังรูปที่ 4-34 ซึ่งจะช่วยให้สามารถทราบถึงความสำคัญของตัวแปรแต่ละตัวได้อย่างชัดเจน

attribute	weight ↓
Days_Since_LastVisit	34.576
Age	28.625
Timespan (Month)	17.390
Gender	9.271
Payment Method	8.322
Region	0.462

รูปที่ 4-34: ค่าน้ำหนักซึ่งแสดงถึงความสำคัญของตัวแปรที่นำเข้าสู่โมเดล

บทที่ 5

บทสรุป ปัญหา และข้อเสนอแนะ

ในบทนี้จะกล่าวถึงบทสรุป ปัญหา และข้อเสนอแนะในการพัฒนาโครงการ “ระบบแช็ตบอต และการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” เพื่อเป็นประโยชน์ต่อผู้ที่สนใจในการพัฒนา ระบบสารสนเทศที่ใกล้เคียงกัน หรือผู้ที่ต้องการนำระบบสารสนเทศนี้ไปพัฒนาต่อไป

5.1 บทสรุป

การพัฒนาโครงการ “ระบบแช็ตบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” เริ่มต้นจากการเข้าสัมภาษณ์เจ้าของกิจการ เพื่อศึกษาวิธีการดำเนินงาน และปัญหาที่เกิดขึ้นในปัจจุบัน รวมถึงการเก็บรวบรวมข้อมูล และความต้องการทางด้านสารสนเทศ จากนั้นจึงทำการพัฒนาระบบแช็ตบอต รวมถึงจัดการข้อมูลให้อยู่ในรูปแบบเดียวกัน ทำการสมมุติข้อมูลเพิ่มเติมตามความเหมาะสม และนำเข้าสู่ฐานข้อมูล หลังจากนั้นจึงนำข้อมูลเหล่านี้มาพัฒนาเป็นโมเดลการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจ เพื่อให้สามารถตอบสนองต่อความต้องการของเจ้าของกิจการและผู้ใช้งานที่เกี่ยวข้อง ในการวางแผนกลยุทธ์การดำเนินงานขององค์กรต่อไป

การพัฒนาระบบแช็ตบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ สามารถบรรลุตามวัตถุประสงค์ที่กำหนดไว้ ดังนี้

- 1) เพื่อพัฒนาแช็ตบอตสำหรับให้ข้อมูลและรับเรื่องเกี่ยวกับปัญหาต่าง ๆ ของลูกค้าไว้ในเบื้องต้น ซึ่งจะนำไปสู่การช่วยเพิ่มประสบการณ์ของลูกค้า (Customer Experience) ที่มาใช้บริการ รวมถึงช่วยแบ่งเบาภาระงานของพนักงานฝ่ายขายและฝ่ายบริการลูกค้า ให้สามารถทำงานในส่วนอื่นที่มีความสำคัญหรือซับซ้อนกว่าได้มากยิ่งขึ้น
- สำหรับวัตถุประสงค์ข้อนี้ โครงการ “ระบบแช็ตบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” ได้พัฒนาระบบแช็ตบอตโดยใช้ Platform อย่าง DialogFlow ซึ่งสามารถเป็นหนึ่งในเครื่องมือที่ช่วยให้บริการลูกค้าจำนวนมากได้อย่างรวดเร็วและแม่นยำภายใต้ต้นทุนที่ไม่สูง ไม่ว่าจะเป็นในด้านของการให้ข้อมูล หรือการ

ตอบคำถามข้อสงสัยต่าง ๆ ของลูกค้าในเบื้องต้น ทั้งยังช่วยลดภาระงานของพนักงานได้อย่างมากในระยะยาวอีกด้วย

- 2) เพื่อนำข้อมูลที่ได้จากแชทบอทมาวิเคราะห์เพิ่มเติมและเป็นอีกหนึ่งช่องทางในการรวบรวมอีเมลเพื่อใช้ในการติดต่อกลุ่มลูกค้าใหม่ ๆ เนื่องจากในปัจจุบันการให้พนักงานโทรศัพท์คุยกับลูกค้าเพียงอย่างเดียวนั้นส่งผลให้การเก็บข้อมูลค่อนข้างทำได้ยาก สำหรับวัตถุประสงค์ข้อนี้ ทางโครงการได้พัฒนาระบบแชทบอทโดยใช้ Platform อย่าง DialogFlow ซึ่งสามารถช่วยจัดเก็บข้อมูลเกี่ยวกับคำถามของลูกค้าได้อย่างครบถ้วนและเป็นมาตรฐาน จึงเป็นการเพิ่มองค์ความรู้ในเชิงลึกเกี่ยวกับคำถามของลูกค้า ซึ่งจะช่วยพัฒนาประสิทธิภาพในการทำงานของบริษัทต่อไปได้ในอนาคต
- 3) เพื่อจัดทำระบบทำนายข้อมูลการเลิกใช้บริการของลูกค้าเดิม และการสมัครใช้บริการของกลุ่มลูกค้าใหม่ จากนั้นจึงนำมาใช้ปรับปรุงต่อยอดการให้บริการและขยายฐานลูกค้าใหม่ โดยที่ยังคงรักษากลุ่มลูกค้าเดิมเอาไว้ สำหรับวัตถุประสงค์ข้อนี้ โครงการนี้ได้พัฒนาระบบการวิเคราะห์ข้อมูลขั้นสูงโดยใช้โปรแกรม RapidMiner ซึ่งช่วยให้สามารถทำนายแนวโน้มและสร้างองค์ความรู้เกี่ยวกับการเลิกใช้บริการ รวมถึงการกวดสมัครใช้บริการของลูกค้า ส่งผลให้สามารถหาแนวทางแก้ไขหรือป้องกันการเลิกใช้บริการได้อย่างทันที่ รวมถึงเพิ่มโอกาสให้ผู้เข้าชมเว็บไซต์กวดสมัครใช้บริการของบริษัท ด้วยการจัดทำโปรโมชันให้เหมาะสมกับลักษณะหรือความต้องการของลูกค้าต่อไป
- 4) เพื่อจัดกลุ่มผู้ที่เข้ามาใช้บริการเว็บไซต์ของบริษัทตามปัจจัยทางด้านอายุ เพศ ระดับการศึกษา ลักษณะการใช้งาน ฯลฯ ผู้บริหารและทีมงานของบริษัทจึงจะสามารถวางแผนและตัดสินใจดำเนินการทางการตลาดได้อย่างครอบคลุมและเหมาะสมในระยะยาว สำหรับวัตถุประสงค์ข้อนี้ โครงการนี้ได้พัฒนาระบบการวิเคราะห์ข้อมูลขั้นสูงโดยใช้โปรแกรม RapidMiner ซึ่งช่วยให้ผู้บริหารรวมถึงผู้ใช้งานอื่น ๆ ที่เกี่ยวข้องสามารถมองเห็นภาพรวม ลักษณะเชิงลึก และพฤติกรรมของกลุ่มผู้ที่เข้าชมเว็บไซต์ของบริษัท

เพื่อใช้เป็นข้อมูลในการทำนายแนวโน้มการเพิ่มขึ้นของลูกค้าใหม่ วางแผนกลยุทธ์ รวมถึงเพิ่มสัดส่วนจำนวนลูกค้าใหม่ที่จะสมัครใช้บริการกับทางบริษัทได้

โครงการ “ระบบแชทบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” ประกอบด้วยระบบต่าง ๆ ดังต่อไปนี้

- 1) ระบบแชทบอต (Chatbot System)
- 2) ระบบวิเคราะห์ข้อมูลจากการใช้งานแชทบอต (Data Analysis System from Chatbot Usage)
- 3) ระบบทำนายการสมัครใช้บริการเว็บไซต์ (Website Subscription Prediction System)
- 4) ระบบจัดกลุ่มผู้ใช้งานเว็บไซต์ (Website Users Clustering System)
- 5) ระบบทำนายการเลิกใช้บริการของลูกค้า (Customer Churn Prediction System)

เทคโนโลยีที่ใช้ในโครงการนี้คือ Platform สำหรับพัฒนาแชทบอตของ DialogFlow ร่วมกับ LINE Messaging API และโปรแกรมสำหรับการวิเคราะห์ข้อมูลขั้นสูงอย่าง RapidMiner Studio ซึ่งทั้งหมดล้วนเป็นเครื่องมือที่มีความทันสมัย มีฟังก์ชันการใช้งานที่หลากหลาย ช่วยให้ผู้ใช้สามารถออกแบบและพัฒนาระบบแชทบอตและระบบสารสนเทศเพื่อการวิเคราะห์ข้อมูลภายในบริษัทได้อย่างรวดเร็วตรงตามความต้องการของผู้ใช้ นอกจากนี้ รูปแบบของการใช้งานทั้งแชทบอตและโมเดลต่าง ๆ ยังมีการแสดงผลในรูปแบบที่เรียบง่ายสะดวกต่อการเรียกใช้งาน ทั้งยังรองรับการเชื่อมต่อฐานข้อมูลที่หลากหลาย ส่งผลให้ผู้บริหารสามารถนำไปใช้ในการบริหารงานได้อย่างมีประสิทธิภาพ

สำหรับข้อมูลที่ใช้ในการวิเคราะห์ในโครงการนี้เป็นข้อมูลย้อนหลังในปี พ.ศ. 2561 จนกระทั่งถึงช่วงปี พ.ศ. 2562 ซึ่งจำเป็นต้องมีการสร้างข้อมูลเพิ่มเติมในบางกรณีที่มีข้อมูลไม่เพียงพอต่อการวิเคราะห์ ซึ่งอาจทำให้ผลการวิเคราะห์มีความคลาดเคลื่อนจากข้อมูลจริงได้

5.2 ปัญหา

ปัญหาที่พบในการพัฒนาโครงการ “ระบบแชทบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” มีดังนี้

- 1) ปัญหาในด้านจำนวนของข้อมูล

ปัญหาที่พบ ข้อมูลที่เก็บรวบรวมมาได้นั้นมีจำนวนค่อนข้างจำกัด ส่งผลให้อาจขาดความครอบคลุม และอาจส่งผลกระทบต่อผลลัพธ์หรือประสิทธิภาพของโมเดลที่พัฒนา

แนวทางแก้ไข ผู้พัฒนาโครงการได้จัดการข้อมูล โดยสมมติข้อมูลเพิ่มเติมเพื่อให้มีความเหมาะสมต่อการสร้างโมเดล รวมถึงเพื่อให้สามารถใช้ในการศึกษาโปรแกรมในการวิเคราะห์ข้อมูลขั้นสูงได้อีกทางหนึ่ง

2) ปัญหาด้านการเตรียมข้อมูล

ปัญหาที่พบ ข้อมูลบางส่วนที่ได้มานั้นอยู่ในรูปแบบที่แตกต่างกัน เช่น บางส่วนอยู่ในรูปแบบของไฟล์ Microsoft Excel บางส่วนเป็นการจดบันทึกด้วยมือในเอกสารที่เป็นกระดาษ ซึ่งส่งผลให้ยากต่อการนำมาวิเคราะห์หรือใช้งาน รวมถึงเกิดความซ้ำซ้อนในบางข้อมูล

แนวทางแก้ไข ผู้พัฒนาโครงการตรวจสอบ Clean ข้อมูลทั้งหมดให้อยู่ในรูปแบบที่เหมาะสมต่อการนำมาใช้ในการวิเคราะห์ และนำข้อมูลทั้งหมดเข้าสู่ฐานข้อมูลคือ Microsoft SQL Server 2019 Express

3) ปัญหาด้านคุณภาพของข้อมูล

ปัญหาที่พบ ข้อมูลบางส่วนมี Missing Value เนื่องจากเป็นข้อมูลที่ถูกพนักงานบันทึกไว้ระหว่างการทำงานและเก็บไว้โดยอาจขาดการตรวจสอบ

แนวทางแก้ไข เนื่องจากข้อมูลที่เป็น Missing Value นั้นมีจำนวนไม่มาก และผู้พัฒนาโครงการได้ศึกษาความสัมพันธ์ของข้อมูลและเห็นว่าเป็นข้อมูลส่วนน้อยซึ่งจะไม่ส่งผลกระทบต่อวิเคราะห์ จึงสามารถตัดข้อมูลส่วนนี้ออกได้

ปัญหาที่พบ ชุดข้อมูลทั้งสองชุดที่นำมาใช้นั้นเป็นชุดข้อมูลที่ไม่สมดุล (Imbalanced Data) กล่าวคือจำนวนข้อมูลส่วนที่เป็น Target ซึ่งจะทำหน้าที่เป็น Label ในการทำนายของข้อมูลทั้งสองชุดค่อนข้างแตกต่างกันมาก ซึ่งอาจส่งผลกระทบต่อความแม่นยำของโมเดลในการทำนาย

แนวทางแก้ไข มีการใช้โอเพอร์เรเตอร์สุ่มข้อมูลด้วยเทคนิค Bootstrapping เพื่อสร้างหลาย ๆ โมเดลขึ้นมาเทรนชุดข้อมูลแบบสุ่ม เพื่อให้ไม่เกิดการ Overfit กับข้อมูลมากเกินไป

5.3 ข้อเสนอแนะ

จากการพัฒนาโครงการ “ระบบแช็ตบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ” ผู้พัฒนาได้พบปัญหาต่าง ๆ มากมาย ผู้พัฒนาโครงการจึงขอเสนอแนะแนวทางการจัดทำโครงการ โดยแบ่งออกเป็น 2 ส่วน ดังต่อไปนี้

ข้อเสนอแนะสำหรับการพัฒนาโครงการเพิ่มเติม

- แช็ตบอตที่พัฒนาด้วย DialogFlow และเชื่อมต่อกับ LINE Messaging API สามารถทำการพัฒนาปรับแต่งให้มีความสามารถหรือฟังก์ชันที่ซับซ้อนมากขึ้นได้ตามความเหมาะสมกับการดำเนินธุรกิจหรือความต้องการที่เปลี่ยนแปลงไป โดยต้องอาศัยการเขียนโปรแกรมเพิ่มเติม
- แช็ตบอตที่พัฒนาด้วย DialogFlow สามารถนำไปเชื่อมต่อกับ Platform อื่น ๆ ได้อีกมากมาย เช่น Facebook Messenger, Google Assistance, ฯลฯ และสำหรับบาง Platform สามารถเขียนโปรแกรมให้แช็ตบอตรับ Input ในรูปแบบเสียงของผู้ใช้งานได้
- ข้อมูลเกี่ยวกับการใช้งาน LINE แช็ตบอตสามารถเขียนโปรแกรมเพื่อดึงมาทำการวิเคราะห์ลักษณะของ NLP รวมถึงการเก็บข้อมูลของผู้ใช้งานอย่างละเอียด เนื่องจากในโครงการที่พัฒนาเป็นการเก็บเฉพาะข้อความที่ผู้ใช้พิมพ์มาเท่านั้น แต่ไม่ได้เก็บว่าใครเป็นผู้พิมพ์เข้ามา ดังนั้นหากต้องการเก็บข้อมูลโดยละเอียดหรือดึงข้อมูลจากแหล่งอื่นมาให้แช็ตบอตตอบ จึงต้องอาศัยการเขียนโปรแกรมเพิ่มเติม
- สำหรับการวิเคราะห์ข้อมูลขั้นสูง สามารถพัฒนาโมเดลรูปแบบอื่น ๆ เพิ่มเติมมากขึ้น เพื่อให้มีความเหมาะสม ครอบคลุมข้อมูลหลากหลายประเภท รวมถึงข้อมูลที่อาจเพิ่มปริมาณ หรือเปลี่ยนแปลงไปในอนาคต
- เพิ่มปริมาณข้อมูลในการนำมาพัฒนาโมเดล เพื่อให้โมเดลมีประสิทธิภาพในการจัดกลุ่มและทำนายข้อมูลได้อย่างแม่นยำมากยิ่งขึ้น

ข้อเสนอแนะสำหรับการพัฒนาโครงการใหม่

- การศึกษาปัญหาและขั้นตอนการดำเนินงานของบริษัทต้นแบบรวมถึงบริษัทคู่แข่งในธุรกิจเดียวกันให้เข้าใจในรายละเอียดอย่างแท้จริงนั้น ถือเป็นสิ่งที่จำเป็นและมี

ความสำคัญ เนื่องจากจะเป็นสิ่งที่ช่วยให้สามารถพัฒนาระบบได้อย่างมีประสิทธิภาพ
สูงสุด

- ควรทำการศึกษาและเรียนรู้เครื่องมือที่ใช้ในการพัฒนาระบบให้เข้าใจอย่างละเอียด
เนื่องจากการใช้งานแต่ละซอฟต์แวร์มีความแตกต่างกัน รวมถึงการแสดงผลของโมเดล
หรือ Process รูปแบบต่าง ๆ ก็มีลักษณะเฉพาะ ดังนั้น ผู้พัฒนาจึงควรศึกษาหลักการ
เลือกใช้ซอฟต์แวร์ที่เหมาะสม เพื่อเพิ่มประสิทธิภาพในการวิเคราะห์ของระบบต่าง ๆ
มากยิ่งขึ้น
- ในขั้นตอนของการนำเข้าข้อมูลสู่ฐานข้อมูลหรือนำเข้าสู่โปรแกรม RapidMiner ควร
มีการทดสอบความถูกต้องและครบถ้วนของข้อมูลทุกครั้ง เพื่อป้องกันไม่ให้เกิดความ
ผิดพลาดในการพัฒนาโมเดลหรือการวิเคราะห์ข้อมูลในขั้นสุดท้าย





ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

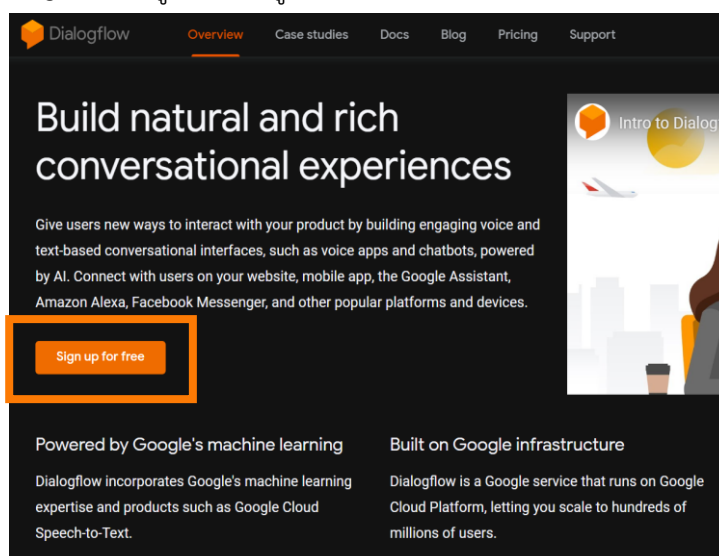
ภาคผนวก ก

การติดตั้งซอฟต์แวร์ที่ใช้งานในระบบ

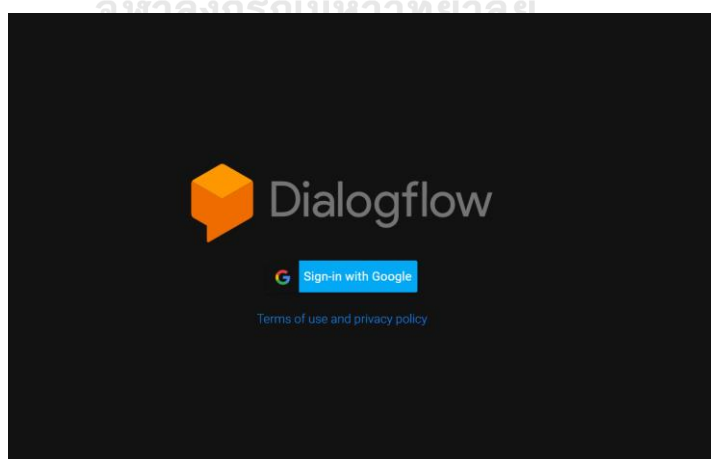
ระบบแชทบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ มีซอฟต์แวร์ที่ใช้งาน คือ DialogFlow และ RapidMiner ซึ่งมีรายละเอียดดังนี้

ก.1 DialogFlow

การเริ่มต้นพัฒนาระบบแชทบอตนั้นสามารถทำได้โดยการเริ่มสมัครใช้งานที่เว็บไซต์ DialogFlow และ Sign-in เข้าสู่ระบบ ดังรูปที่ ก-1 และ ก-2



รูปที่ ก-1: หน้าจอแรกในการ Sign-up เพื่อสร้าง Account ก่อนเริ่มต้นใช้งาน



รูปที่ ก-2: หน้าจอการ Sign-in เข้าสู่ระบบด้วยอีเมลเพื่อเริ่มพัฒนาระบบแชทบอต

ก.2 RapidMiner

การติดตั้งซอฟต์แวร์ที่ใช้งานในระบบวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ ประกอบด้วย 2 ขั้นตอน ดังนี้

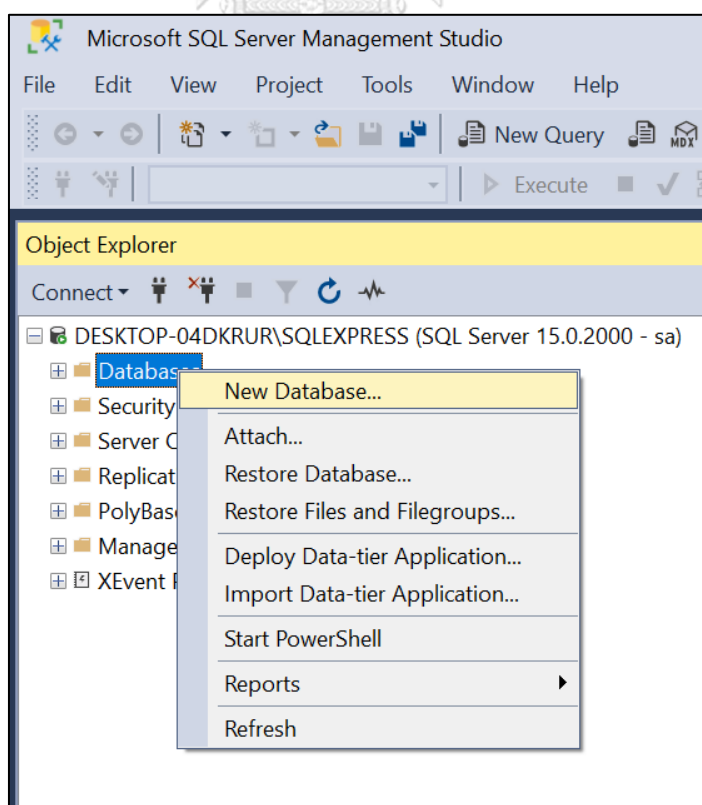
1) การติดตั้งซอฟต์แวร์ที่ใช้งานในระบบ

ผู้พัฒนาระบบจะต้องทำการติดตั้งโปรแกรม Microsoft SQL Server 2019 Express เพื่อใช้เป็นระบบฐานข้อมูลสำหรับการดึงมาใช้ใน Process ของโปรแกรม RapidMiner Studio

2) การนำเข้าข้อมูลสู่ฐานข้อมูล Microsoft SQL Server 2019 Express

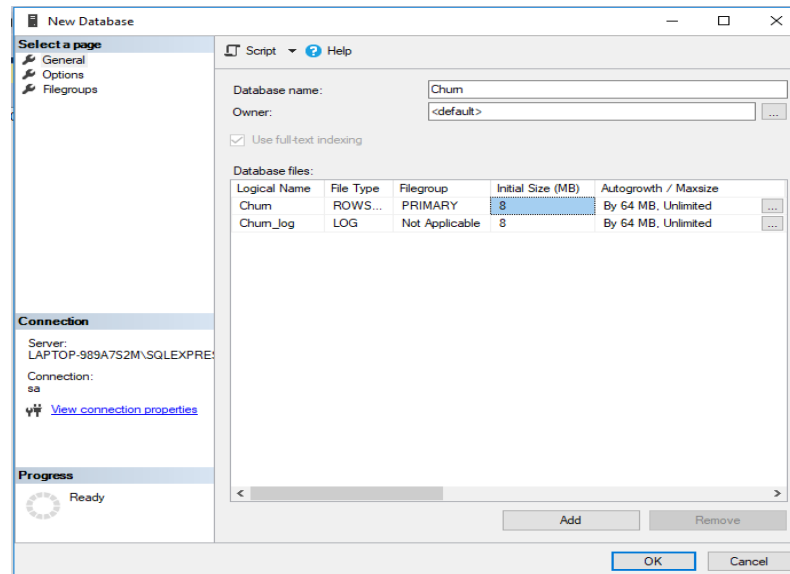
การนำเข้าข้อมูลสู่ฐานข้อมูลด้วยโปรแกรม Microsoft SQL Server 2019 Express มีขั้นตอนดังต่อไปนี้

- (1) การสร้างฐานข้อมูลใหม่โดยคลิกขวาที่ Databases ใน Object Explorer เลือก New Database... ดังรูปที่ ก-3



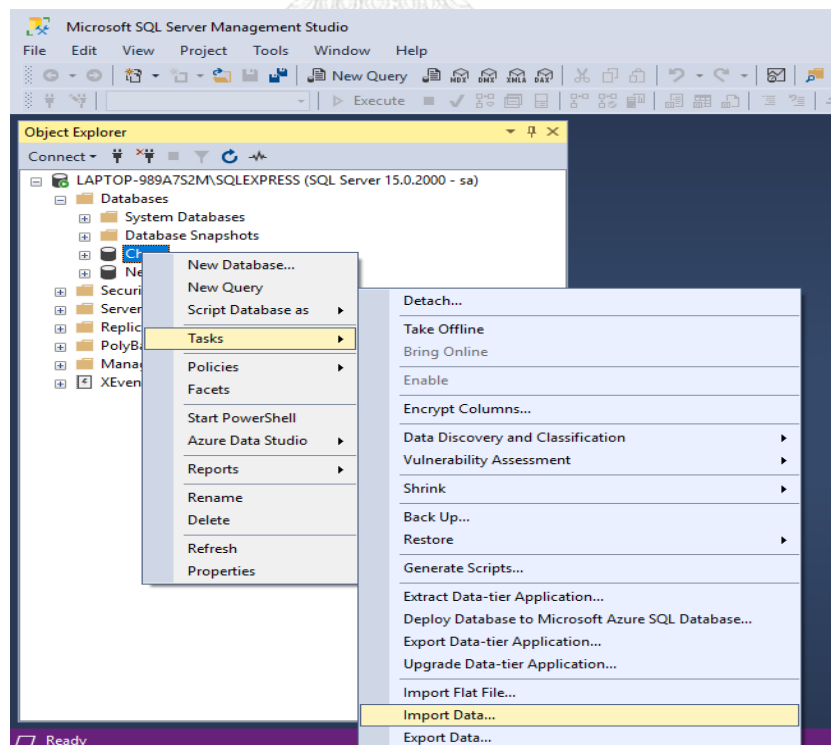
รูปที่ ก-3: หน้าจอเมนูการสร้างฐานข้อมูลใหม่

(2) ตั้งชื่อฐานข้อมูลและกดปุ่ม OK ดังรูปที่ ก-4



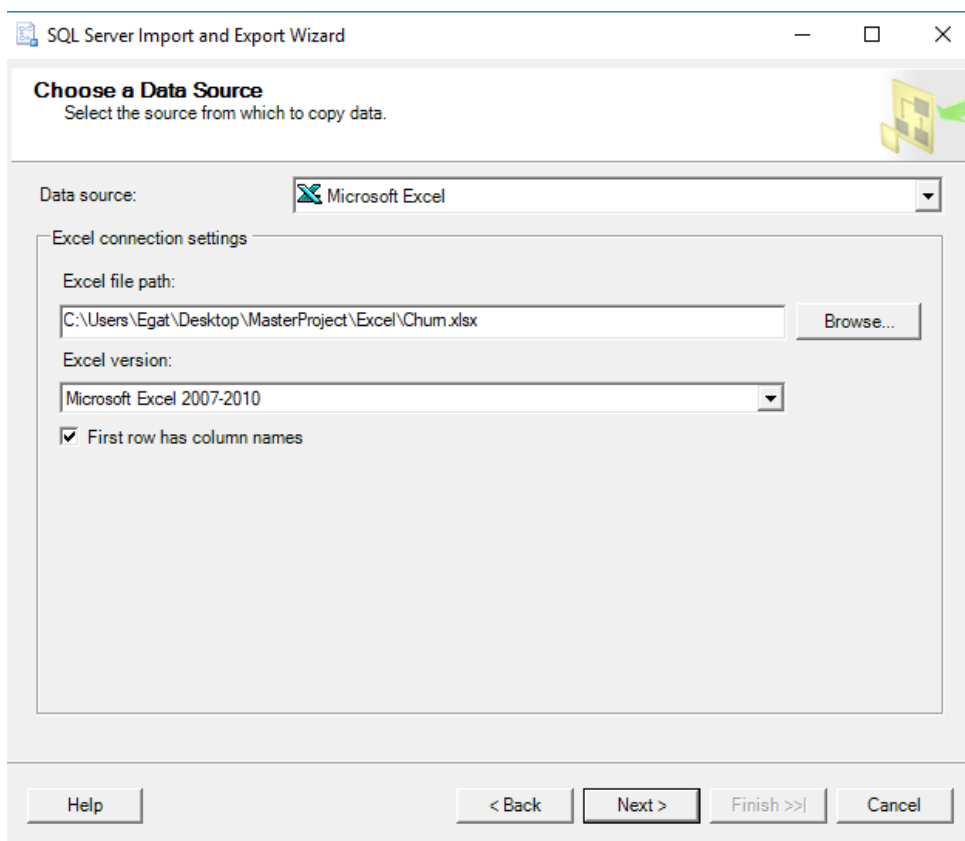
รูปที่ ก-4: หน้าจอการสร้างฐานข้อมูลใหม่

(3) นำเข้าข้อมูลลงฐานข้อมูลโดยคลิกขวาที่ฐานข้อมูลที่สร้างจากขั้นตอนที่ 1 จากนั้นเลือก Tasks และ Import Data... ดังรูปที่ ก-5



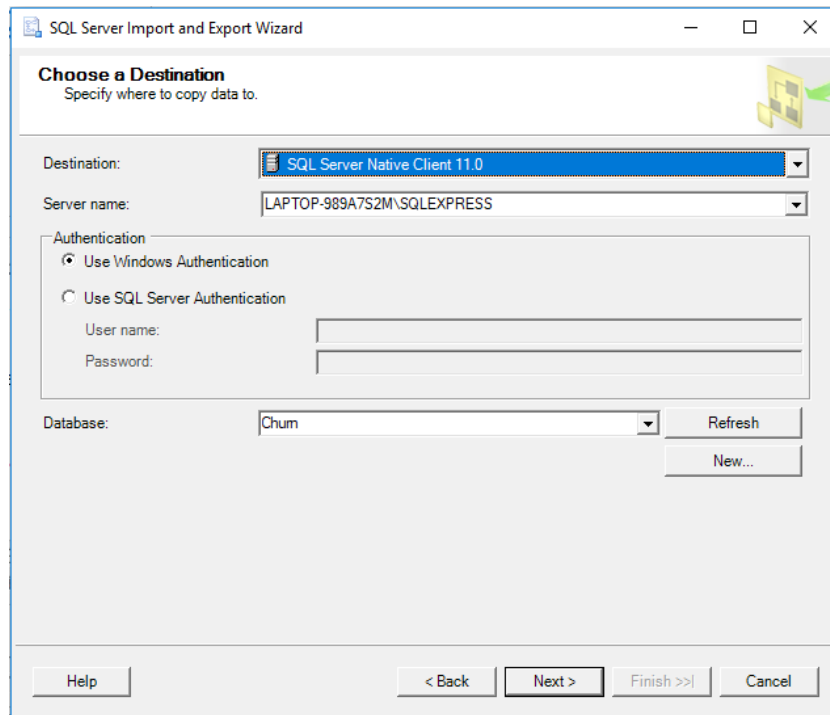
รูปที่ ก-5: หน้าจอเมนูการนำเข้าข้อมูล

- (4) เมื่อปรากฏหน้าต่าง SQL Server Import and Export Wizard ขึ้นมาดังรูปที่ ก-6 เลือก Data Source เป็น Microsoft Excel จากนั้นกดปุ่ม Browse... เลือกไฟล์ Excel ที่ต้องการและเลือกเวอร์ชันของไฟล์ Excel จากนั้นกดปุ่ม Next



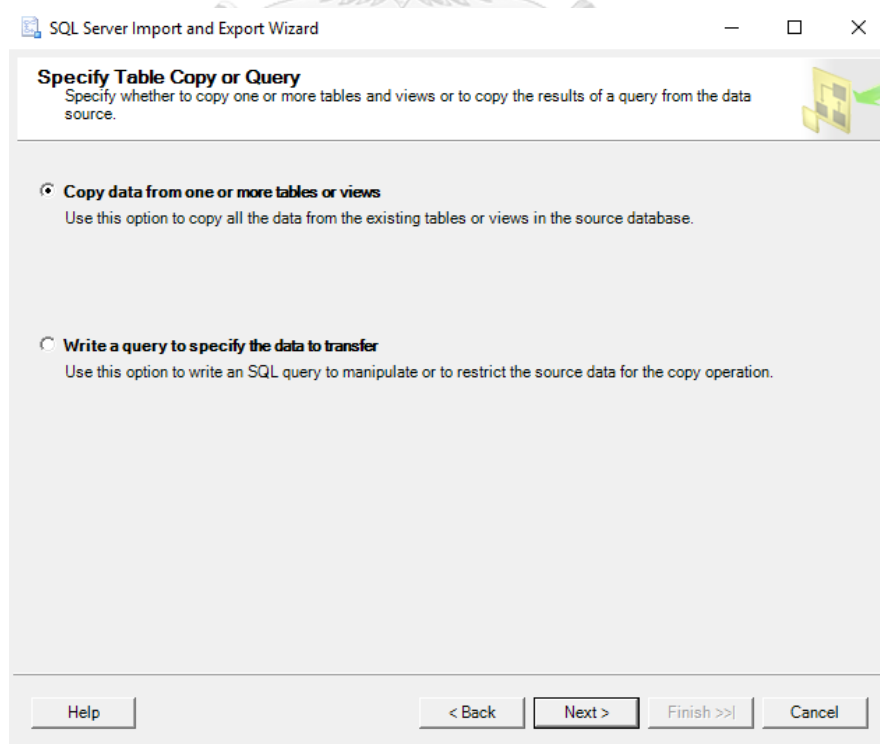
รูปที่ ก-6: หน้าจอการเลือก Data Source

- (5) เลือก Destination เป็น SQL Server Native Client 11.0 จากนั้นเลือก Server Name ที่ต้องการ และเลือกฐานข้อมูลที่ต้องการนำเข้าข้อมูล ดังรูปที่ ก-7



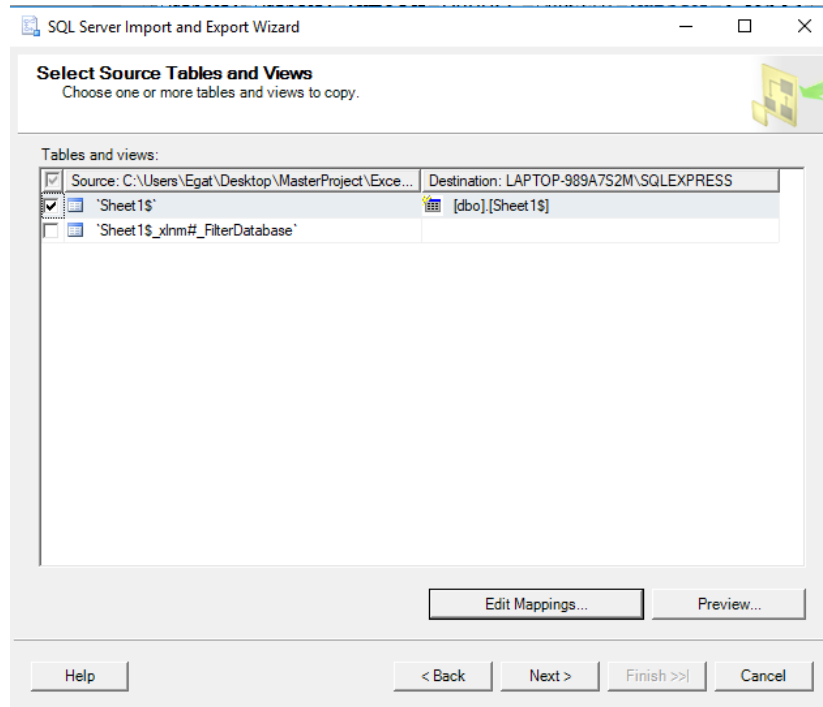
รูปที่ ก-7: หน้าจอการเลือก Destination

(6) เลือก Copy data from one or more table or views และกด Next
ตั้งรูปที่ ก-8



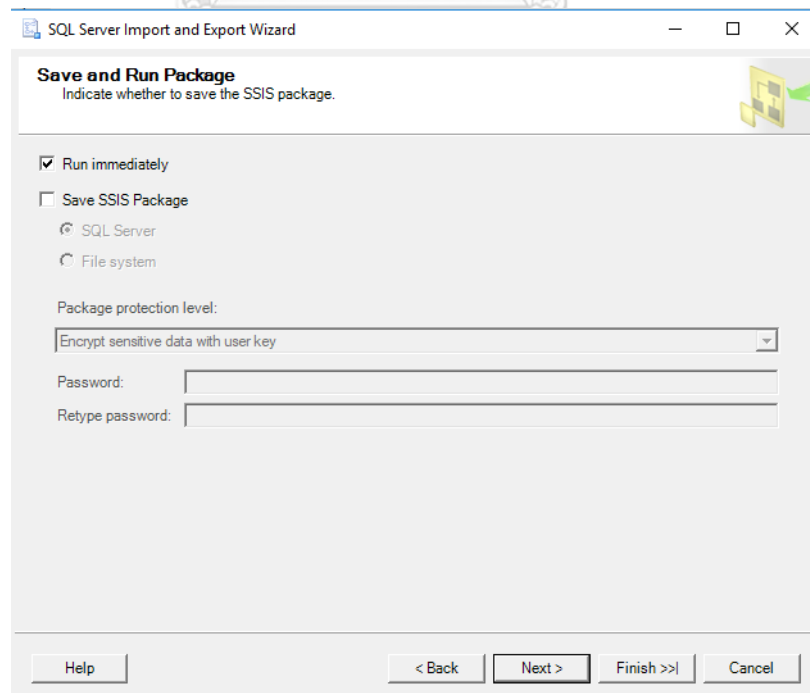
รูปที่ ก-8: หน้าจอการเลือก Copy หรือ Query

(7) เลือก Table ที่เป็นข้อมูลที่ต้องการนำเข้า และกด Next ดังรูปที่ ก-9



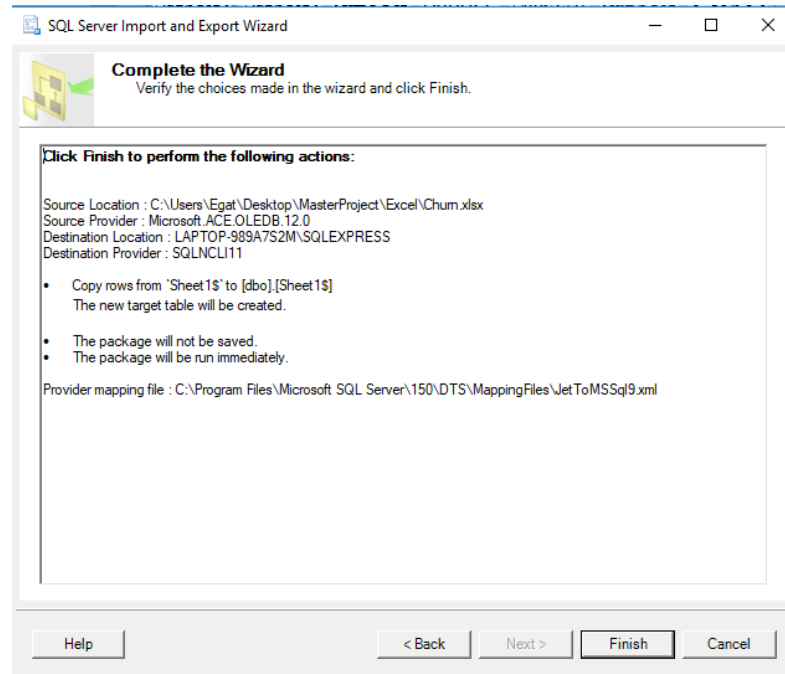
รูปที่ ก-9: หน้าจอการเลือกตารางข้อมูล

(8) เลือก Run immediately และกด Next ดังรูปที่ ก-10



รูปที่ ก-10: หน้าจอการ Save และ Run Package

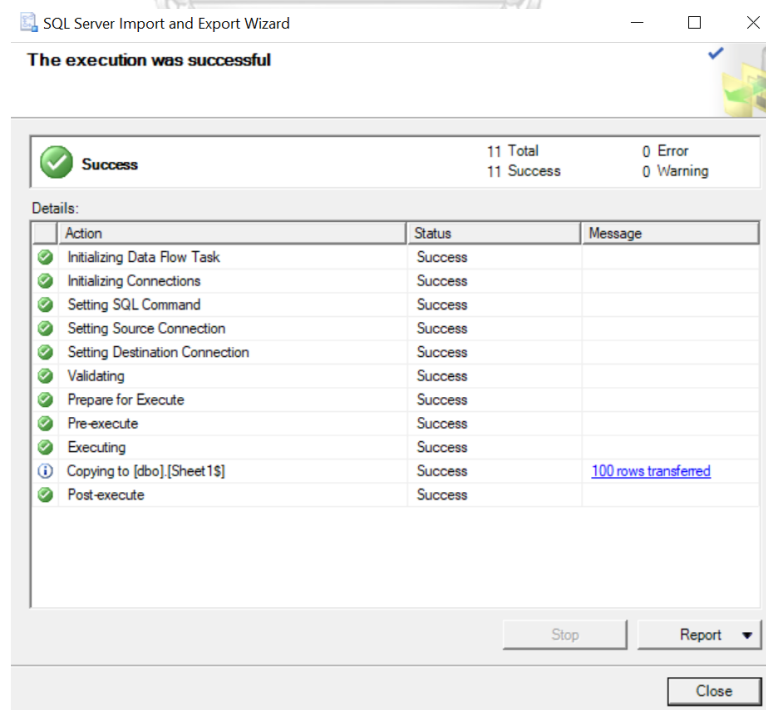
(9) กด Finish ดังรูปที่ ก-11



รูปที่ ก-11: หน้าจอเสร็จสิ้นการนำเข้าข้อมูล

(10) เมื่อนำเข้าข้อมูลสำเร็จกดปุ่ม Close เพื่อปิดหน้าต่าง Wizard ดังรูป ก-

12



รูปที่ ก-12: หน้าจอแสดงผลลัพธ์ในการนำเข้าข้อมูลสำเร็จ

ภาคผนวก ข

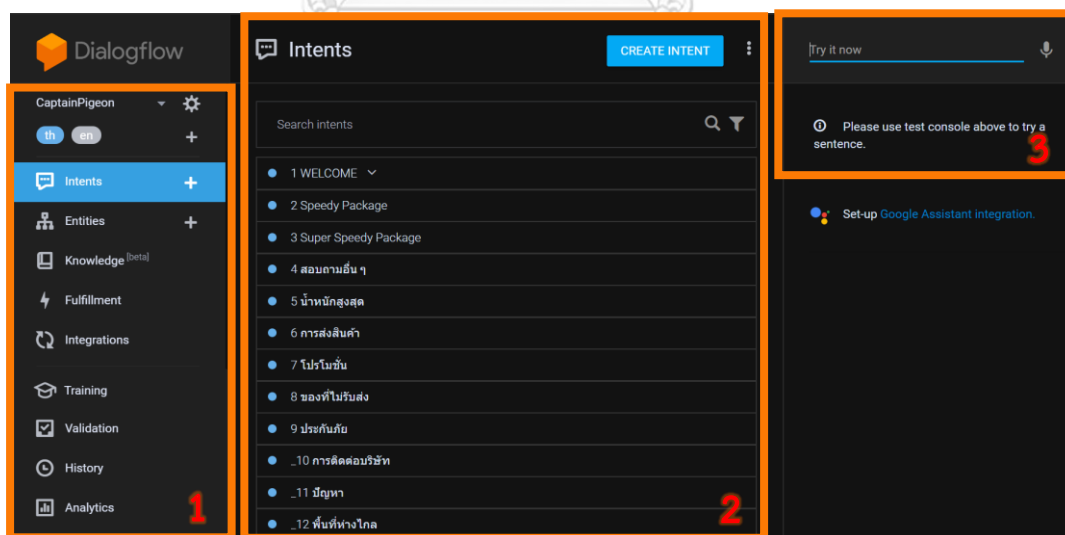
เมนูการทำงานของระบบ

ระบบแช็ตบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ มีเมนูการทำงานทั้งหมด 2 ระบบหลัก คือ ในส่วนของระบบแช็ตบอตด้วย Platform อย่าง DialogFlow และการวิเคราะห์ข้อมูลขั้นสูงด้วยโปรแกรม RapidMiner Studio

ข.1 ระบบแช็ตบอตด้วย Platform ของ DialogFlow

หลังจาก Sign-in เข้าสู่ระบบแล้วจะพบหน้าจอหลักทั้งหมด 3 ส่วนด้วยกัน ได้แก่

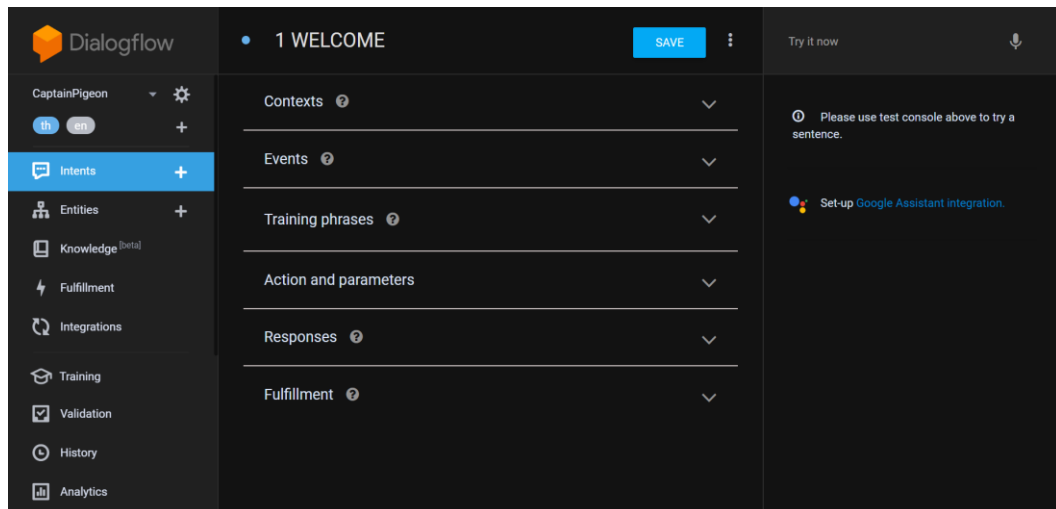
- 1) เมนูหลักซึ่งใช้ในการสร้าง Intent ใหม่ เชื่อมต่อกับ Platform ต่าง ๆ หรือตรวจสอบข้อความที่ผู้ใช้งานพูดคุยได้ตอบกับแช็ตบอต ฯลฯ
- 2) เมื่อย่อยซึ่งจะแสดงผลเกี่ยวกับรายละเอียดของเมนูหลักที่ผู้ใช้งานกดเลือกไว้ เช่น แสดงรายชื่อของ Intent ที่มีการสร้างไว้แล้วและสามารถกดเข้าไปแก้ไขหรือดูรายละเอียดได้ เป็นต้น
- 3) Test Console สำหรับใช้ในการทดสอบพิมพ์ข้อความพูดคุยกับแช็ตบอตเพื่อให้สามารถเห็นผลลัพธ์ของการ Train แช็ตบอตได้ในทันที ดังรูปที่ ข-1



รูปที่ ข-1: หน้าจอหลักทั้ง 3 ส่วนเมื่อเข้าสู่ระบบของ DialogFlow

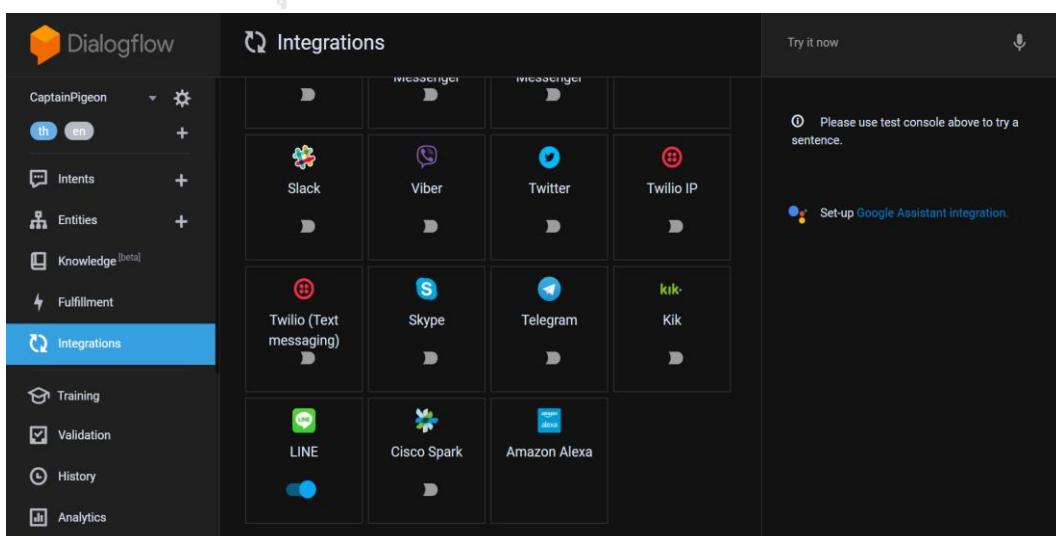
โดยมีขั้นตอนการใช้งาน ดังนี้

- เมื่อกดเลือกที่ Intent หนึ่ง ๆ แล้ว เมนูด้านในของ Intent ที่ถูกเลือกจะประกอบไปด้วย 6 องค์ประกอบซึ่งผู้พัฒนาระบบแชทบอทสามารถเข้าไปทำการปรับแต่งได้ตามรายละเอียดในหัวข้อที่ 4.3.1 ได้แก่ Context, Events, Training Phrases, Action and Parameters, Responses ดังรูปที่ ข-2



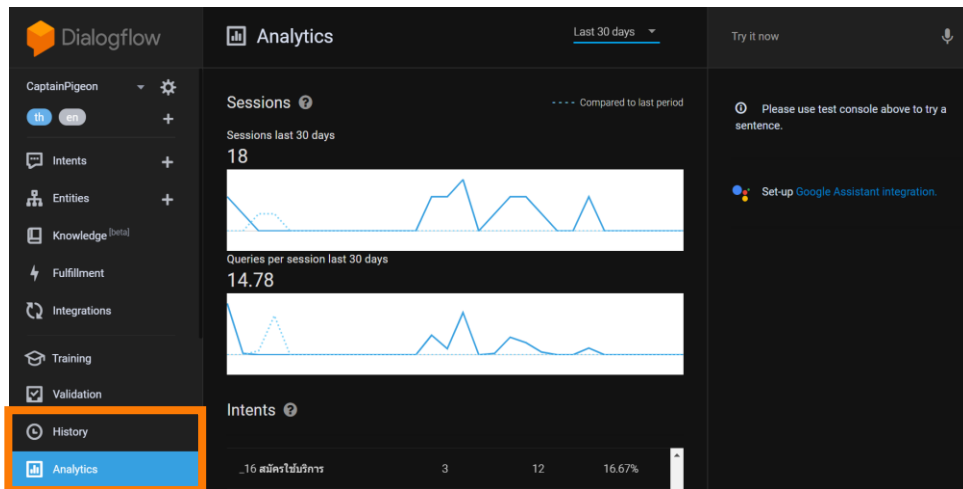
รูปที่ ข-2: หน้าจอการปรับแต่งแชทบอทภายใต้เมนู WELCOME Intent

- หลังจากที่ได้สร้าง Agent หรือแชทบอทโดยกำหนด Intent ต่าง ๆ ให้ครอบคลุมแล้ว ผู้พัฒนาสามารถทำการเชื่อมต่อแชทบอทเข้ากับ Platform ต่าง ๆ ได้ที่เมนู Integrations ดังรูปที่ ข-3



รูปที่ ข-3: หน้าจอการเชื่อมต่อแชทบอทเข้ากับ Platform ต่าง ๆ ด้วยเมนู Integrations

- 3) ผู้ใช้สามารถดูรายละเอียดข้อมูลการใช้งานแชทบอตได้จากเมนู History และ Analytics ซึ่งปรากฏบน Panel หลักทางด้านซ้าย ซึ่งผู้ใช้สามารถตรวจสอบรายละเอียดต่าง ๆ ไม่ว่าจะเป็นข้อความที่แชทบอตยังตอบไม่ได้ จำนวนการใช้งานแต่ละ Intent หรือปริมาณ Session การใช้งานในแต่ละช่วงเวลา ฯลฯ ดังรูปที่ ข-4

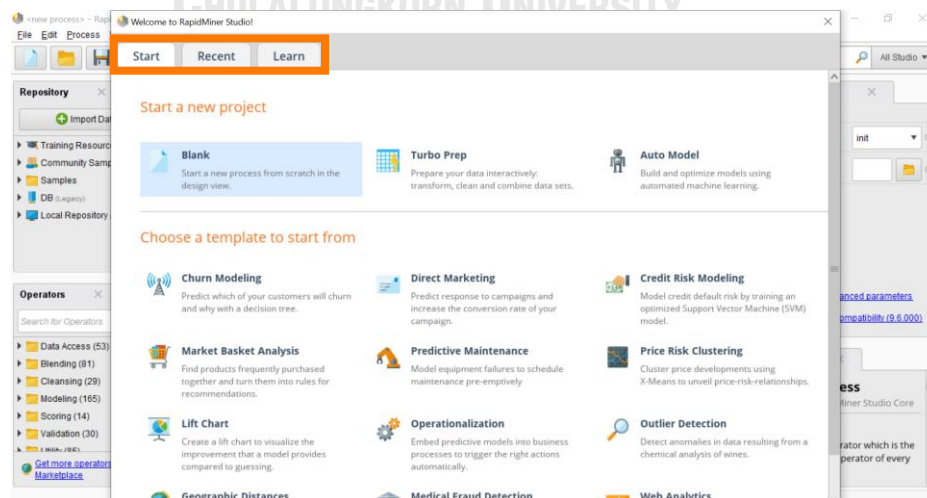


รูปที่ ข-4: หน้าจอแสดงข้อมูลประวัติการใช้งานแชทบอต

ข.2 ระบบวิเคราะห์ข้อมูลขั้นสูงด้วยโปรแกรม RapidMiner Studio

หน้าจอเมนูหลักของระบบ

- 1) เมื่อเริ่มต้นใช้งาน RapidMiner Studio 9.5.1 ผู้ใช้งานจะพบหน้าต่างเริ่มต้นซึ่งประกอบด้วย 3 เมนูหลักดังรูปที่ ข-5



รูปที่ ข-5: หน้าต่างเมนูเริ่มต้นของโปรแกรม RapidMiner Studio 9.5.1

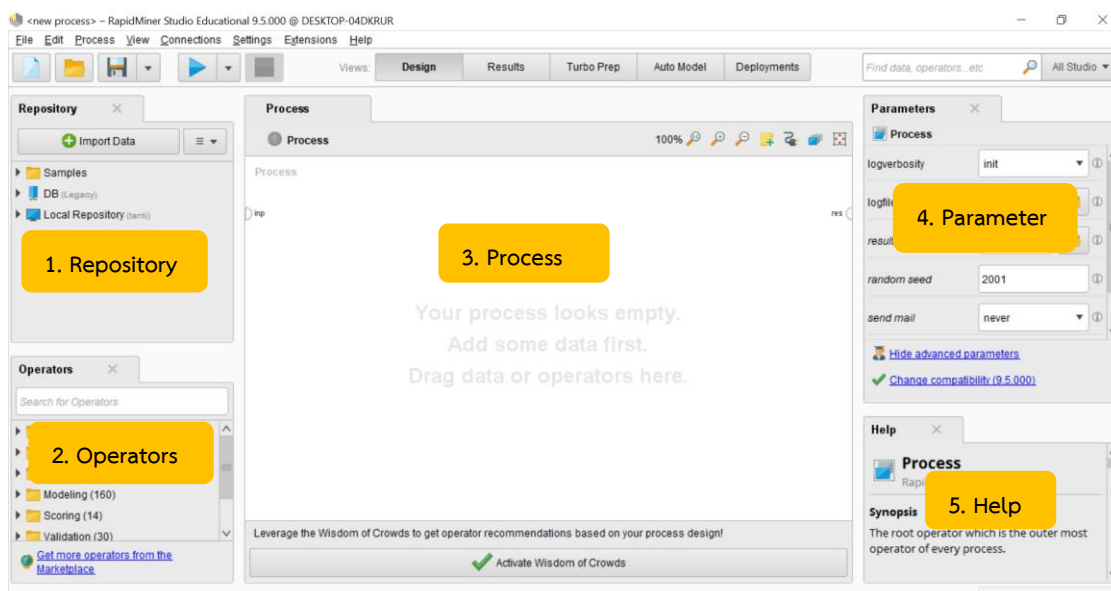
ส่วนที่ 1) **Start** เป็นการเริ่มต้นใช้งาน RapidMiner Studio 9.5.1

- **Start a new project:** เป็นการสร้างโปรเจกต์หรือ Process ใหม่ขึ้นมาโดยสามารถเลือกที่ Blank หรือทำการเตรียมข้อมูลก่อนด้วยการกดเลือก Turbo Prep หรือทดลองสร้างโมเดลที่ดีที่สุดด้วยการเลือกจาก Auto Model
- **Choose a template to start from:** ผู้ใช้สามารถเลือกได้ว่าต้องการเริ่มต้นการใช้งานโดยใช้ Template รูปแบบต่าง ๆ ที่ซอฟต์แวร์มีให้หรือไม่

ส่วนที่ 2) **Recent** เป็นการเปิด Process ที่เคยบันทึกไว้ขึ้นมาใช้งานเพื่อดูหรือแก้ไข โดย Process ที่สร้างไว้แล้วสามารถ Reuse หรือส่งต่อได้

ส่วนที่ 3) **Learn** เป็นส่วนของ Tutorial ต่าง ๆ ที่รวบรวมและแสดงวิธีการใช้งานของ RapidMiner Studio ซึ่งมีให้เลือกทั้งการศึกษาจาก Documents ภายในโปรแกรม และจากเว็บไซต์ของ RapidMiner

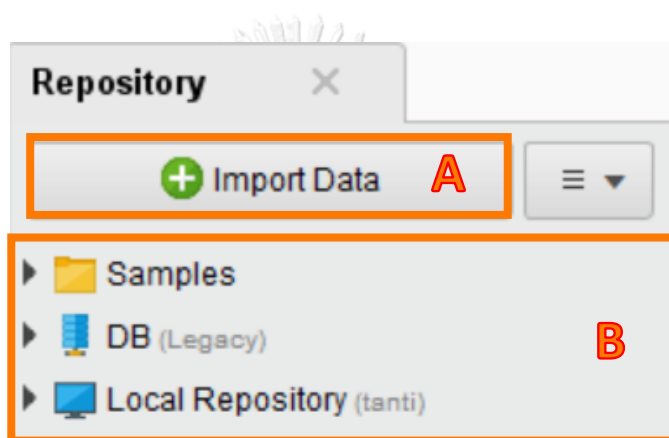
- 2) เมื่อคลิกที่ปุ่ม  Blank ในส่วนของการ Start a new project จะพบหน้าจอหลักในการสร้าง Process ดังรูปที่ ข-6



รูปที่ ข-6: หน้าจอหลักในการสร้าง Process ของโปรแกรม RapidMiner Studio

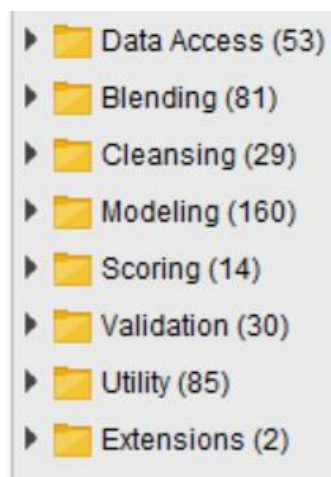
หน้าจอในการสร้าง Process ของโปรแกรม RapidMiner Studio ประกอบด้วย 5 ส่วนหลัก ๆ ได้แก่

- **Repository:** ส่วนนี้เป็นส่วนของการจัดการไฟล์ต่าง ๆ โดย RapidMiner Studio จะจัดการข้อมูลจาก 2 แหล่งคือ DB (ฐานข้อมูล) และ Local (ในเครื่องคอมพิวเตอร์ของผู้ใช้) โดยเก็บไฟล์ Data Set และ Process ต่าง ๆ แยกเก็บคนละโฟลเดอร์กันเพื่อความสะดวกในการเรียกใช้งานครั้งถัดไป ซึ่งในส่วนของ Repository จะแบ่งออกเป็น 2 ส่วน ดังรูปที่ ข-7
 - A ส่วนการเพิ่มข้อมูลและจัดการส่วนต่าง ๆ ของ Repository
 - B แสดง Repository ต่าง ๆ ที่ได้สร้างขึ้นมา หรือมีเตรียมไว้ให้แล้ว

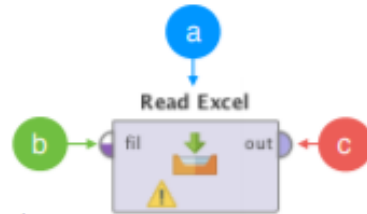


รูปที่ ข-7: ฟังก์ชันในส่วนของ Repository

- **Operators:** เป็นส่วนที่ใช้เก็บโอเปอเรเตอร์ที่ใช้ในการทำงานทั้งหมด ซึ่งจัดเป็นกลุ่ม ๆ โดยกลุ่มที่มีการใช้งานคล้ายคลึงกันจะอยู่ในกลุ่มเดียวกัน และมีทั้งหมด 8 กลุ่มโอเปอเรเตอร์ ได้แก่
 - 1) Data Access
 - 2) Blending
 - 3) Cleansing
 - 4) Modeling
 - 5) Scoring
 - 6) Validation
 - 7) Utility
 - 8) Extensions

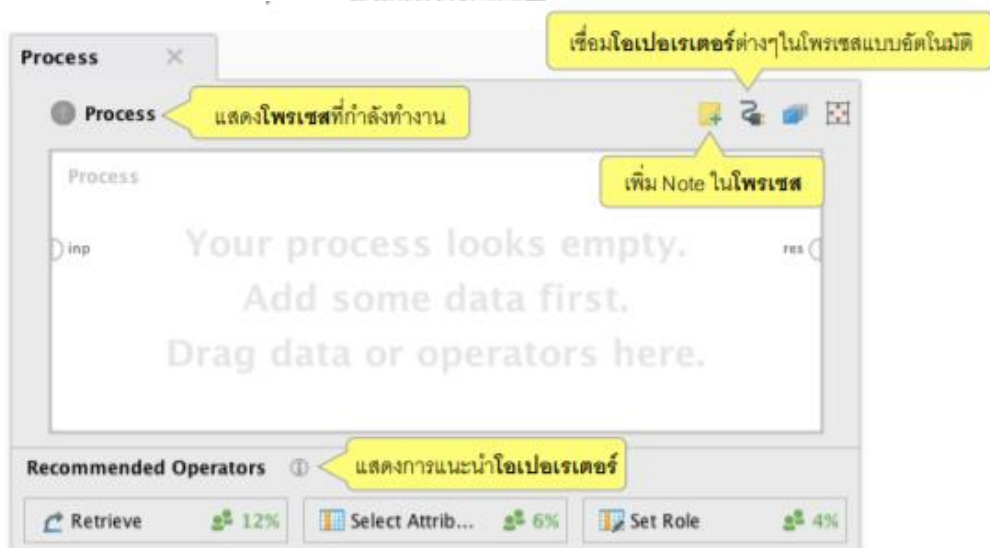


โดยโอเปอร์เรเตอร์แต่ละตัวจะมีองค์ประกอบดังรูปที่ ข-8



รูปที่ ข-8: องค์ประกอบของโอเปอร์เรเตอร์ (Pacharawongsakda, 2560)

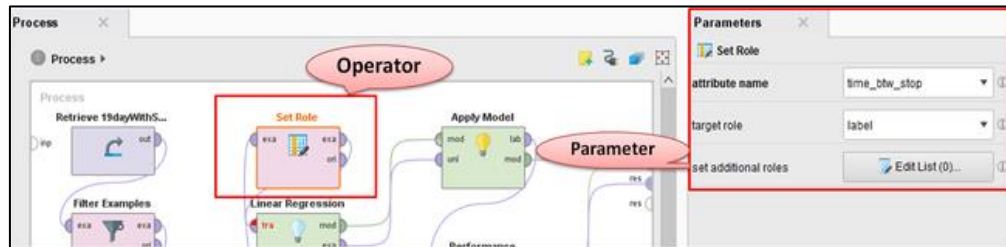
- a ชื่อของโอเปอร์เรเตอร์
- b อินพุต พอร์ต (Input Port) เป็นส่วนของการรับข้อมูลเข้ามาประมวลผล
- c เอาท์พุต พอร์ต (Output Port) เป็นส่วนของการส่งผลลัพธ์ที่ประมวลได้
- **Process:** เป็นพื้นที่หลักของการทำงานในการสร้าง Process สำหรับทำ Machine Learning ของซอฟต์แวร์นี้ โดยจะเป็นส่วนที่นำโอเปอร์เรเตอร์ต่าง ๆ มาประกอบเข้าด้วยกันเพื่อสร้าง Process ขึ้นมาตามวัตถุประสงค์ที่ตั้งไว้ ซึ่งองค์ประกอบหลักในส่วนของ Process มีดังรูปที่ ข-9



รูปที่ ข-9: องค์ประกอบหลักในส่วนของ Process (Pacharawongsakda, 2560)

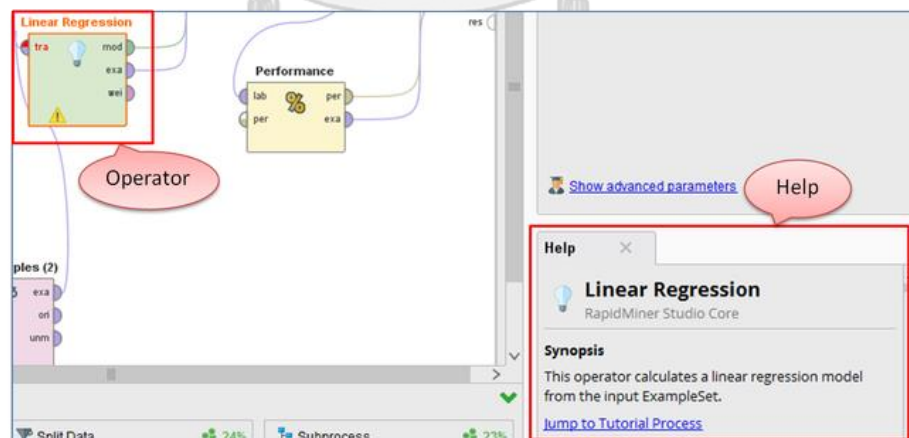
- **Parameters:** เป็นส่วนที่กำหนดรายละเอียดของโอเปอร์เรเตอร์ที่ถูกเลือกใช้งาน เช่น โอเปอร์เรเตอร์ที่ชื่อ Set Role เป็นโอเปอร์เรเตอร์ที่มี Parameter ที่เกี่ยวข้องทั้งหมดสอง Parameter คือ attribute name ที่ใช้กำหนดตัวแปรที่ต้องการเลือก

ซึ่งจากตัวอย่างในรูปที่ ข-10 มีการเลือกตัวแปร time_bt看_stop และใช้ Parameter ที่ชื่อ target role เพื่อระบุว่าให้ตัวแปรที่เลือกไว้ทำหน้าที่เป็น Label



รูปที่ ข-10: ลักษณะของโอเปอร์เรเตอร์และพารามิเตอร์ใน Process (Pacharawongsakda, 2560)

- **Help:** เป็นส่วนช่วยเหลือซึ่งจะแสดงรายละเอียดของ Operator ที่เลือกใช้งานอยู่ ซึ่งประกอบด้วยรายละเอียดเบื้องต้น ความหมายของแต่ละ Parameter และ ตัวอย่างการใช้งานในส่วนท้ายสุด แต่ส่วนช่วยเหลือของ RapidMiner Studio จะบอกเพียงหน้าที่และรายละเอียดคร่าว ๆ ของ Operator ดังนั้นหากต้องการรายละเอียดเพิ่มขึ้นจึงต้องไปที่ Jump to Tutorial Process ซึ่งจะ Link ไปยังเว็บไซต์ที่มีรายละเอียดของเกี่ยวกับโอเปอร์เรเตอร์ที่ใช้อยู่ เช่น โอเปอร์เรเตอร์ชื่อ Linear Regression ใน หน้า Help จะบอกว่าเป็นโอเปอร์เรเตอร์ ใช้คำนวณข้อมูล จากจาก Data set โดยใช้วิธี Linear Regression ดังรูปที่ ข-11



รูปที่ ข-11: ส่วนช่วยเหลือซึ่งแสดงข้อมูลเกี่ยวกับโอเปอร์เรเตอร์ชื่อ Linear Regression

(Pacharawongsakda, 2560)

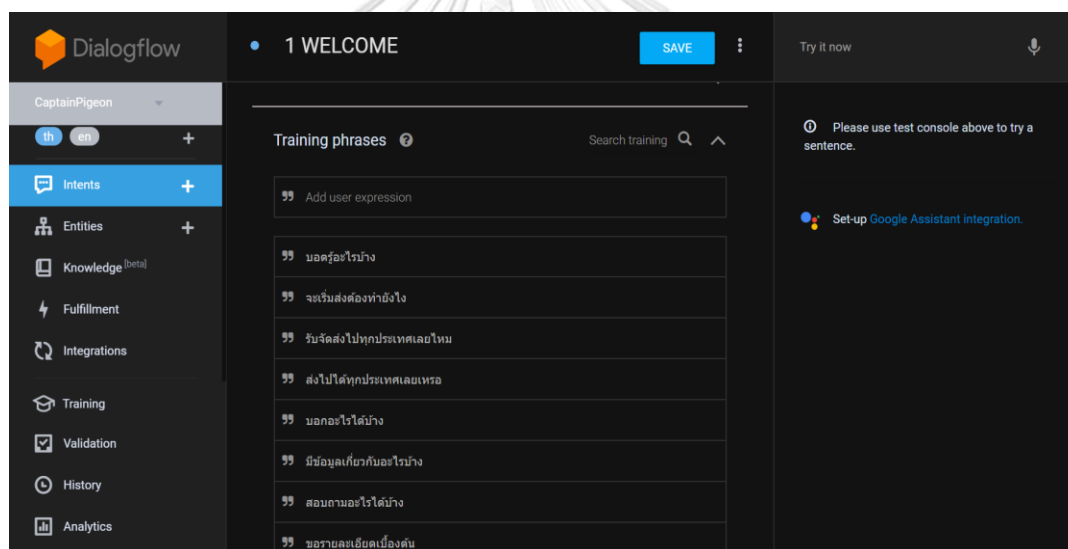
ภาคผนวก ค

ตัวอย่างการใช้งานระบบ

ระบบแช็ตบอตและการวิเคราะห์ข้อมูลขั้นสูงของธุรกิจบริการจัดส่งพัสดุไปต่างประเทศ มีตัวอย่างการใช้งานทั้งหมด 2 ระบบหลักคือในส่วนของระบบแช็ตบอตและการวิเคราะห์ข้อมูลขั้นสูง

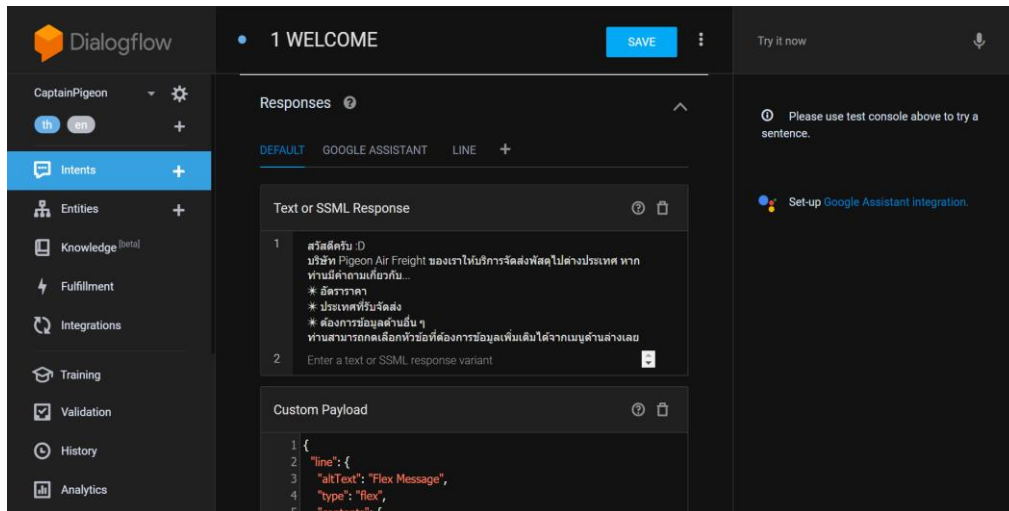
ค.1 ระบบแช็ตบอต

- 1) การปรับแต่งระบบแช็ตบอตสามารถทำได้ผ่านเมนู Intents โดยผู้พัฒนาสามารถกำหนด Training Phrases ให้ครอบคลุมคำถามที่คาดว่าลูกค้าหรือผู้ใช้งานจะพิมพ์เข้ามาพูดคุยกับแช็ตบอต โดยหากยิ่งเพิ่มจำนวนข้อความให้มีความครอบคลุมมาก จะยิ่งเป็นการเพิ่มความสามารถในการตอบคำถามของแช็ตบอตได้มากยิ่งขึ้น ดังรูปที่ ค-1



รูปที่ ค-1: หน้าจอการปรับแต่ง Training Phrases ภายใต้เมนู Intents

- 1) การปรับแต่ง Responses หรือการตอบกลับของแช็ตบอตสามารถทำได้ 2 รูปแบบคือแบบข้อความ (Text หรือ SSML Response) และรูปแบบอื่น ๆ ที่ผู้พัฒนากำหนดเอง (Custom Payload) เช่น การเขียนโค้ดให้แช็ตบอตแสดงผลการตอบกลับในรูปแบบของปุ่มกด, รูปภาพ, Link, เว็บไซต์ ฯลฯ ดังรูปที่ ค-2



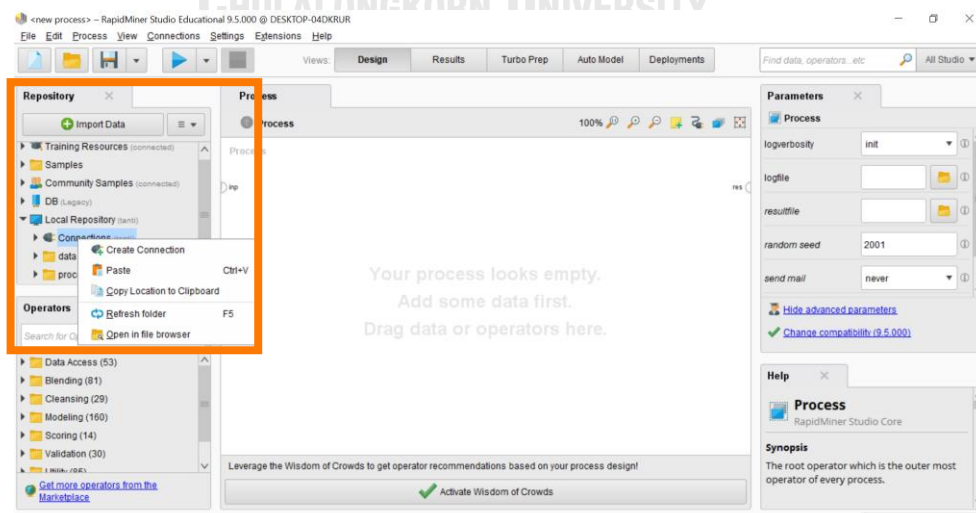
รูปที่ ค-2: การปรับแต่ง Responses ภายใต้เมนู Intents

ค.2 ระบบการวิเคราะห์ข้อมูลขั้นสูงด้วยโปรแกรม RapidMiner Studio

1) การเชื่อมต่อข้อมูลจากฐานข้อมูลเข้าสู่โปรแกรมที่ใช้ในการทำนายข้อมูล

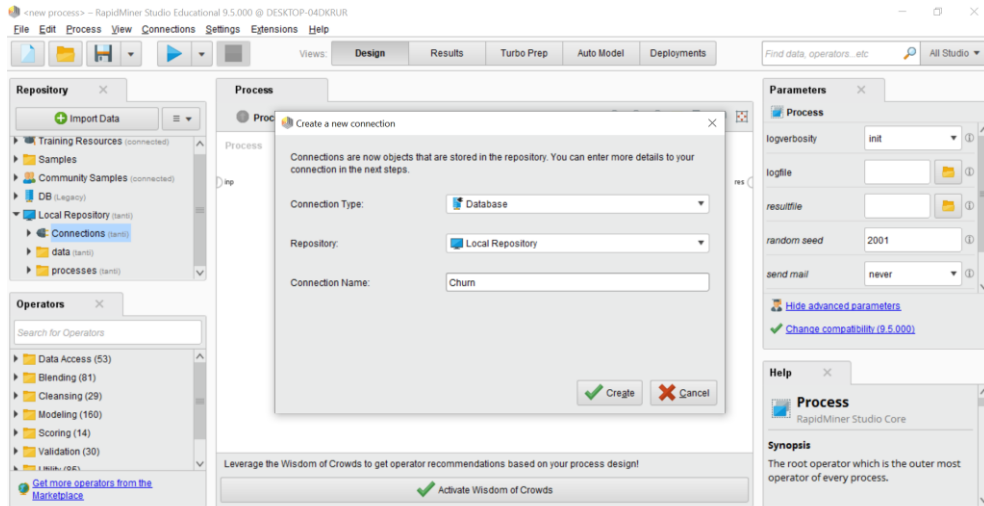
การเชื่อมต่อข้อมูลระหว่างฐานข้อมูลกับ Process การทำนายและจัดกลุ่มข้อมูล โดยนำข้อมูลของลูกค้าเดิมและลูกค้าใหม่ที่สร้างใน Microsoft SQL Server 2019 Express มาเชื่อมต่อเข้ากับโปรแกรม RapidMiner Studio ซึ่งเป็นโปรแกรมที่ใช้สำหรับพัฒนา Process ในการทำนายและจัดกลุ่มข้อมูล มีขั้นตอนดังนี้

- (1) ทำการเชื่อมต่อกับฐานข้อมูลโดยคลิกขวาที่ Connection ในส่วนของ Repository และกดเลือก Create Connection ดังรูปที่ ค-3




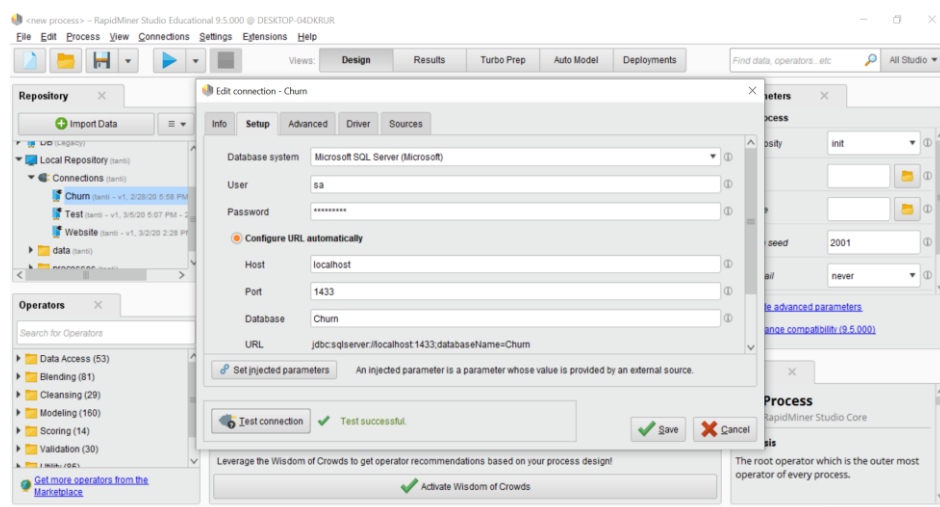
รูปที่ ค-3: การเชื่อมต่อโปรแกรม RapidMiner Studio เข้ากับฐานข้อมูลที่ต้องการ

- (2) ปรากฏหน้าต่างที่ผู้ใช้สามารถตั้งชื่อช่องทางในการเชื่อมต่อได้ ดังรูปที่ ค-4 จากนั้นกด Create




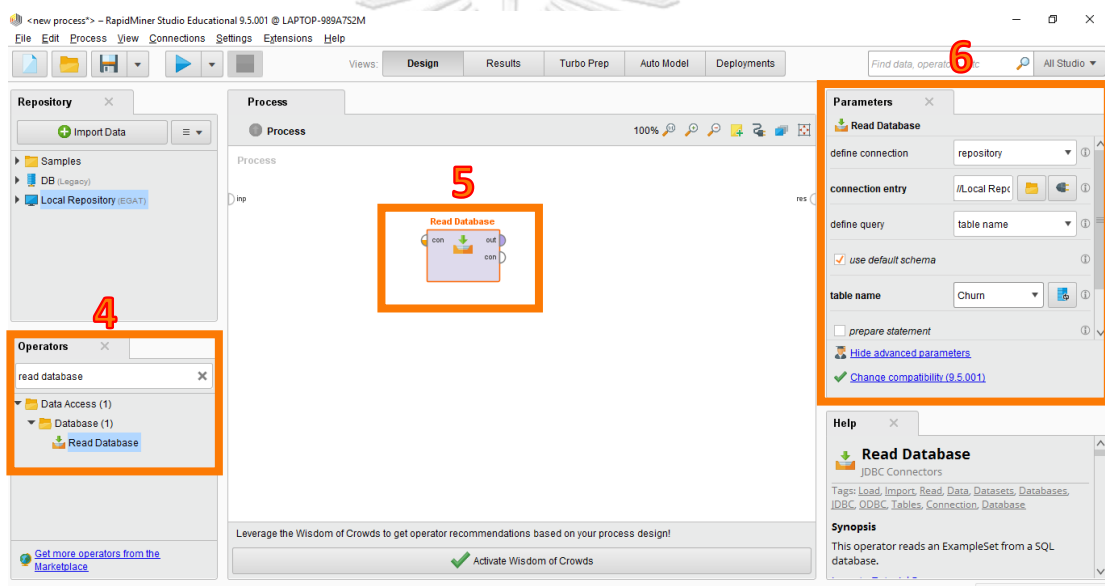
รูปที่ ค-4: หน้าต่างสำหรับตั้งชื่อช่องทางในการเชื่อมต่อ

- (3) หลังจากกด Create แล้วจะปรากฏไอคอน  ช่องทางการเชื่อมต่อในส่วนของ Connections และผู้ใช้จะพบหน้าต่าง Setup ดังรูปที่ ค-5 ซึ่งต้องทำการกรอกข้อมูลให้ตรงกับฐานข้อมูลที่ได้สร้างไว้ และสามารถทดลองเชื่อมต่อกับฐานข้อมูลได้โดยคลิกที่ปุ่ม Test connection เมื่อทดลองเชื่อมต่อสำเร็จโดยพบข้อความ Test successful ให้กดปุ่ม Save เป็นอันเสร็จสิ้นการสร้างช่องทางเชื่อมต่อกับฐานข้อมูล



รูปที่ ค-5: หน้าต่างสำหรับกรอกข้อมูลเพื่อเชื่อมต่อกับฐานข้อมูลที่ได้สร้างไว้

- (4) สำหรับการดึงข้อมูลจากฐานข้อมูลผ่านช่องทางเชื่อมต่อที่กำหนดไว้สามารถทำได้โดยเลือกที่โอเพอร์เรเตอร์ Read Database จากในส่วนของ Operators
- (5) ลากโอเพอร์เรเตอร์มาวางไว้ในพื้นที่ของ Process
- (6) ในส่วนของ Parameters ตั้งค่าตัวเลือก Connection Entry โดยคลิกที่ไอคอน 
- (7) และเลือกช่องทางการเชื่อมต่อที่ได้สร้างไว้ จากนั้นตั้งค่า Define Query โดยกดเลือกให้เป็น Table Name และเลือกชื่อตารางข้อมูลที่ต้องการนำเข้าจากฐานข้อมูลดังรูปที่ ค-6



รูปที่ ค-6: หน้าจอการนำเข้าข้อมูลผ่านช่องทางการเชื่อมต่อที่ได้สร้างขึ้น

2) การแก้ไขรายละเอียดของ Process การทำนายข้อมูล

กรณีที่ผู้พัฒนาระบบต้องการแก้ไขโอเพอร์เรเตอร์หรือพารามิเตอร์ต่าง ๆ รวมถึงปรับเปลี่ยนรายละเอียดใด ๆ ใน Process เพิ่มเติม สามารถทำได้โดยคลิกที่ตัวโอเพอร์เรเตอร์ในส่วนของ Process และทำการปรับเปลี่ยนพารามิเตอร์บริเวณ Panel ด้านข้าง จากนั้นสามารถกดปุ่มรันเพื่อดูผลลัพธ์ได้ดังรูปที่ ค-7

The screenshot displays the RapidMiner Studio interface. The main canvas shows a process flow with three operators: 'Set Role (2)', 'Decision Tree (3)', and 'Apply Model (3)'. The 'Decision Tree (3)' operator is highlighted with an orange box. To its right, the 'Parameters' panel is also highlighted with an orange box, showing the following settings:

- Decision Tree (3) (Decision Tr...
- criterion: gini_index
- maximal depth: 5
- apply pruning
- confidence: 0.3
- apply prepruning
- [Hide advanced parameters](#)
- [Change compatibility \(9.6.000\)](#)

Below the parameters panel, a 'Help' section for 'Decision Tree' is visible, including a synopsis: 'This Operator generates a decisic...'

รูปที่ ค-7: หน้าจอการปรับเปลี่ยน Parameter ของโมเดล Decision Tree



บรรณานุกรม

Affinity Solution. (2561). Chatbot ผู้ช่วยใหม่ ของธุรกิจในยุคดิจิทัล. Retrieved from

<https://www.affinity.co.th/chatbot-%E0%B8%81%E0%B8%B1%E0%B8%9A%E0%B8%98%E0%B8%B8%E0%B8%A3%E0%B8%81%E0%B8%B4%E0%B8%88/?lang=th>

Anak Mirasing. (2560). Chatbot คืออะไร, ดียังไง มารู้จักกันใน 10 นาที. Retrieved from

<https://medium.com/@igroomgrim/chatbot-%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3-%E0%B8%94%E0%B8%B5%E0%B8%A2%E0%B8%B1%E0%B8%87%E0%B9%84%E0%B8%87-%E0%B8%A1%E0%B8%B2%E0%B8%A3%E0%B8%B9%E0%B9%89%E0%B8%81%E0%B8%B1%E0%B8%99%E0%B9%83%E0%B8%99-10-%E0%B8%99%E0%B8%B2%E0%B8%97%E0%B8%B5-3e6165dd34b8>

Boontamthan, M. (2562). สรุป 15 สิ่งที่คุณควรรู้หลัง LINE Redesign. Retrieved from

<https://medium.com/linedevth/%E0%B8%AA%E0%B8%A3%E0%B8%B8%E0%B8%9B-15-%E0%B8%AA%E0%B8%B4%E0%B9%88%E0%B8%87%E0%B8%97%E0%B8%B5%E0%B9%88%E0%B8%84%E0%B8%A7%E0%B8%A3%E0%B8%A3%E0%B8%B9%E0%B9%89%E0%B8%AB%E0%B8%A5%E0%B8%B1%E0%B8%87-redesign-%E0%B9%80%E0%B8%A1%E0%B8%B7%E0%B9%88%E0%B8%AD-line-official-account-%E0%B8%A3%E0%B9%88%E0%B8%A7%E0%B8%A1%E0%B8%A3%E0%B9%88%E0%B8%B2%E0%B8%87%E0%B8%81%E0%B8%B1%E0%B8%9A-line-%E0%B8%89%E0%B8%9A%E0%B8%B1%E0%B8%9A%E0%B8%A2%E0%B9%88%E0%B8%AD-4b07b7f8e0d5>

Boonyang, N. (2561). Information Gain — Data Mining (2). Retrieved from

<https://medium.com/@imkk/information-gain-data-mining-2-12e39b0d0f5a>

Budkod, T. (2561). DATA MINING คืออะไร อยากรู้ต้องอ่าน. Retrieved from

<https://www.glurgeek.com/education/data-mining->

[%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3-](https://www.glurgeek.com/education/data-mining-%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3-)

[%E0%B8%AD%E0%B8%A2%E0%B8%B2%E0%B8%81%E0%B8%A3%E0%B8%B9%E0%B9%89%E0%B8%95%E0%B9%89%E0%B8%AD%E0%B8%87%E0%B8%AD%E0%B9%88/](https://www.glurgeek.com/education/data-mining-%E0%B8%AD%E0%B8%A2%E0%B8%B2%E0%B8%81%E0%B8%A3%E0%B8%B9%E0%B9%89%E0%B8%95%E0%B9%89%E0%B8%AD%E0%B8%87%E0%B8%AD%E0%B9%88/)

Click Next Company Limited. (2555). Application Programming Interface (API). Retrieved from <https://www.smsmkt.com/wp->

[content/themes/smsmkt/assets/doc/SMSMKT-API-HTTP-Thai-Version.pdf](https://www.smsmkt.com/wp-content/themes/smsmkt/assets/doc/SMSMKT-API-HTTP-Thai-Version.pdf)

Coraline Team. (2562). Classification ต่างกับ Regression อย่างไร. Retrieved from

<https://www.coraline.co.th/single-post/Classification-vs-Regression>

Daroontham, W. (2561). เจาะลึก Random Forest !!!— Part 2 of “รู้จัก Decision Tree, Random Forest, และ XGBoost!!!”. Retrieved from

<https://medium.com/@witchapongdaroontham/%E0%B8%B2%E0%B8%B8%81-random-forest-part-2-of-%E0%B8%B1%E0%B8%81-decision-tree-random-forest-%E0%B9%81%E0%B8%A5%E0%B8%B0-xgboost-79b9f41a1c1c>

G-able. (2561). Data Science วิเคราะห์ข้อมูลขั้นสูง เพื่อผลลัพธ์ที่น่าเชื่อถือ พร้อมหาความเชื่อมโยงที่ซ่อนอยู่.

Retrieved from <https://www.g-able.com/products/data-science/>

Gatchalee, P. (2562). Confusion Matrix เครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนายใน Machine learning. Retrieved from [https://medium.com/@pagongatchalee/confusion-](https://medium.com/@pagongatchalee/confusion-matrix-%E0%B8%99machine-learning-fba6e3f9508c)

[matrix-%E0%B8%99machine-learning-fba6e3f9508c](https://medium.com/@pagongatchalee/confusion-matrix-%E0%B8%99machine-learning-fba6e3f9508c)

Jirawatee. (2562). อัปเดต 11 ปีใหม่ของ Flex Message ปี 2019. Retrieved from

<https://medium.com/linedevth/%E0%B8%AD%E0%B8%B1%E0%B8%9E%E0%B9%80%E0%B8%94%E0%B8%97-11->

[%E0%B8%9F%E0%B8%B5%E0%B9%80%E0%B8%88%E0%B8%AD%E0%B8%A3%E0%B9%8C%E0%B9%83%E0%B8%AB%E0%B8%A1%E0%B9%88%E0%B8%82%E0%B8%AD%E0%B8%87-flex-message-%E0%B9%83%E0%B8%99-line-messaging-api-%E0%B8%9B%E0%B8%B5-2019-668e92721bc9](https://medium.com/linedevth/%E0%B8%AD%E0%B8%B1%E0%B8%9E%E0%B9%80%E0%B8%94%E0%B8%97-11-%E0%B8%9F%E0%B8%B5%E0%B9%80%E0%B8%88%E0%B8%AD%E0%B8%A3%E0%B9%8C%E0%B9%83%E0%B8%AB%E0%B8%A1%E0%B9%88%E0%B8%82%E0%B8%AD%E0%B8%87-flex-message-%E0%B9%83%E0%B8%99-line-messaging-api-%E0%B8%9B%E0%B8%B5-2019-668e92721bc9)

Kankate, K. (2562). ทำการตลาดออนไลน์ ให้ถูกใจลูกค้า ถูกที่ ถูกเวลา ต้องเริ่มวิเคราะห์ Customer & Channel. Retrieved from [https://stepstraining.co/strategy/digital-marketing-](https://stepstraining.co/strategy/digital-marketing-strategy-customer-channel)

[strategy-customer-channel](https://stepstraining.co/strategy/digital-marketing-strategy-customer-channel)

Keng Surapong. (2562). Activation Function คืออะไร ใน Artificial Neural Network. Retrieved

from <https://www.bualabs.com/archives/1261/what-is-activation-function-what-is-sigmoid-function-activation-function-ep-1/>

Kiglogistics. (2561). Shipping จีน และ Freight Forwarder สำคัญในการขนส่งอย่างไร. Retrieved from <https://kiglogistics.co.th/https-kiglogistics-co-th-shipping-%E0%B8%88%E0%B8%B5%E0%B8%99-%E0%B9%81%E0%B8%A5%E0%B8%B0-freight-forwarder-%E0%B8%AA%E0%B8%B3%E0%B8%84%E0%B8%B1%E0%B8%8D%E0%B9%83%E0%B8%99%E0%B8%81%E0%B8%B2%E0%B8%A3/>

Kongrath. (2560). 10 บริษัทโลจิสติกส์ (ขนส่งระหว่างประเทศ) บริการดีเยี่ยม ราคาไม่แพง 2020. Retrieved from <https://topbestbrand.com/%E0%B8%9A%E0%B8%A3%E0%B8%B4%E0%B8%A9%E0%B8%B1%E0%B8%97%E0%B9%82%E0%B8%A5%E0%B8%88%E0%B8%B4%E0%B8%AA%E0%B8%95%E0%B8%B4%E0%B8%81%E0%B8%AA%E0%B9%8C%E0%B8%97%E0%B8%B5%E0%B9%88%E0%B8%94%E0%B8%B5%E0%B8%97/>

Krungsri Plearn Plearn. (2560). ยกระดับธุรกิจให้ก้าวขึ้นด้วย “Chatbot” สุดยอดผู้ช่วยอัจฉริยะแห่งยุค 4.0. Retrieved from <https://www.krungsri.com/bank/th/plearn-plearn/chatbot-%E0%B8%AA%E0%B8%B8%E0%B8%94%E0%B8%A2%E0%B8%AD%E0%B8%94-%E0%B8%9C%E0%B8%B9%E0%B9%89%E0%B8%8A%E0%B9%88%E0%B8%A7%E0%B8%A2-%E0%B8%AD%E0%B8%B1%E0%B8%88%E0%B8%89%E0%B8%A3%E0%B8%B4%E0%B8%A2%E0%B8%B0.html>

Lowphansirikul, L. (2561). สอน Machine Learning — Linear Regression. Retrieved from <https://medium.com/@artificialcc/%E0%B8%AA%E0%B8%AD%E0%B8%99-machine-learning-linear-regression-bf07d230a7ee>

Mr. Automated. (2559). เหมืองข้อมูล (data mining). Retrieved from <http://www.autosoft.in.th/data-warehouse/%E0%B9%80%E0%B8%AB%E0%B8%A1%E0%B8%B7%E0%B8%AD%E0%B8%87%E0%B8%82%E0%B9%89%E0%B8%AD%E0%B8%A1%E0%B8%B9%E0%B8%A5-data-mining/>

Nalisa. (2562). ตลาด E-Commerce ยังหอมหวาน จากนี้ถึงปี 2565 โตเฉลี่ย 22%. Retrieved from <https://marketeeronline.co/archives/126314>

- Omise. (2562). WebHooks. Retrieved from <https://www.omise.co/th/api-webhooks/thailand>
- Pacharawongsakda, E. (2560). Introduction to Data Mining with RapidMiner Studio. Retrieved from http://dataminingtrend.com/2014/wp-content/uploads/2014/02/RM7_chapter1.pdf
- Pawoot. (2551). E-Marketing คืออะไร ? Retrieved from <http://www.pawoot.com/node/380>
- Petch Kruapanich. (2561). ลองทำแชทบอทง่ายๆด้วย Dialogflow. Retrieved from <https://medium.com/readmoreth/%E0%B8%A5%E0%B8%AD%E0%B8%87%E0%B8%97%E0%B8%B3%E0%B9%81%E0%B8%8A%E0%B8%97%E0%B8%9A%E0%B8%AD%E0%B8%97%E0%B8%A5%E0%B8%87%E0%B8%97%E0%B8%B0%E0%B9%80%E0%B8%9A%E0%B8%B5%E0%B8%A2%E0%B8%99%E0%B8%87%E0%B9%88%E0%B8%B2%E0%B8%A2%E0%B9%86%E0%B8%94%E0%B9%89%E0%B8%A7%E0%B8%A2-dialogflow-%E0%B8%81%E0%B8%B1%E0%B8%99%E0%B9%80%E0%B8%96%E0%B8%AD%E0%B8%B0-4bd3a8c550de>
- Pp. (2563). 5 เทรนด์เทคโนโลยีธุรกิจขนส่งและโลจิสติกส์ ในปี 2020. Retrieved from <https://www.brandbuffet.in.th/2020/02/5-trends-of-transportation-and-logistics/>
- Programmerbonn. (2562). DialogFlow – Chatbot. Retrieved from <https://programmerbonn.home.blog/2019/01/20/dialogflow-chatbot>
- Saetang, W. (2561). DATA ANALYTICS. Retrieved from http://bps.moph.go.th/new_bps/sites/default/files/03Data%20Analytics%2017-21Dec2018.pdf
- SAS Institute Inc. (2560). การวิเคราะห์เหมืองข้อมูลคืออะไร และสำคัญอย่างไร. Retrieved from https://www.sas.com/th_th/insights/analytics/data-mining.html
- Satangmongkol, K. (2561). สร้างโมเดล Tree Based ง่ายๆด้วย R. Retrieved from <https://datarockie.com/2018/11/13/tree-based-models-in-r/>
- Topten2U. (2556). โลจิสติกส์ คืออะไร. Retrieved from <http://xn--12cn3b8c0ac0ec4j.blogspot.com/2013/10/blog-post.html>
- WP. (2562). เปิดเบื้องหลัง “DHL Express” ทำไมเป็น “ผู้นำส่งด่วนระหว่างประเทศ”? ในยุคโลจิสติกส์แข่งเดือด. Retrieved from <https://www.marketingoops.com/news/biz-news/dhl-express-key-success-strategies/>
- โกเมศ อัมพวัน. (2561). การจำแนกประเภทและการทำนายข้อมูล (Classification and Prediction).

Retrieved from <https://staff.informatics.buu.ac.th/~komate/886464/%5B6%5D-Classification.pdf>

นงคราญ คำวิชัย. (2559). Practical Data Mining With RapidMiner Studio 7. Retrieved from http://bps.moph.go.th/new_bps/sites/default/files/Practical%20Data%20mining%20with%20Rapidminer%20Studio%207.pdf

นิภาพร ชนะมาร. (2557). การวิเคราะห์ปัจจัยการเรียนรู้ด้วยการคัดเลือกคุณสมบัติและการพยากรณ์. Retrieved from research.kpru.ac.th/conference5/form/EX2.pdf

รุจิรา ธรรมสมบัติ. (2554). รายงานการวิจัยเรื่อง ระบบสนับสนุนการตัดสินใจในการเลือกใช้แพคเกจอินเทอร์เน็ตมือถือโดยใช้ต้นไม้ตัดสินใจ. Retrieved from http://www.rpu.ac.th/Library_web/doc/RC_RR/2554_ComBus_Rujira.pdf

วีรศักดิ์ ช่อสูงเหลื่อม. (2555). การจัดกลุ่มข้อมูลด้วยเทคนิคกราฟเคมิตอยส์แบบขนานบนหน่วยประมวลผลกลางแบบหลายแกนหลัก. Retrieved from <http://sutir.sut.ac.th:8080/jspui/bitstream/123456789/4096/2/Fulltext.pdf>

ศจี วานิช. (2558). Data Mining (เหมืองข้อมูล). Retrieved from <http://sajeegm301.blogspot.com/2015/11/data-mining.html>

ศราวุธ เลิศพลังสันติ. (2553). การประยุกต์โครงข่ายประสาทเทียม เพื่อพยากรณ์ความพึงพอใจของผู้โดยสาร ในการออกแบบชิ้นส่วนยานยนต์. Retrieved from https://www2.mtec.or.th/th/e-magazine/admin/upload/252_17-25.pdf

สิริวิชญ์ ธนาวิชญ์. (2561). การตรวจสอบสถานะเครือข่ายด้วยแอปพลิเคชันไลน์. Retrieved from [http://msit.mut.ac.th/thesis/Thesis_2561/\(MISS\)%20NetworkMonitorLineApplication.pdf](http://msit.mut.ac.th/thesis/Thesis_2561/(MISS)%20NetworkMonitorLineApplication.pdf)

อัศม์เดช วานิชชินชัย. (2553). สแน็คพัพ ไม่สับสน ในแวดวงโลจิสติกส์. Retrieved from http://www.thailandindustry.com/indust_newweb/articles_preview.php?cid=10860

ประวัติผู้เขียน

ชื่อ-สกุล	นางสาว ทัดทิกา นันยากรสกุล
วัน เดือน ปี เกิด	23 มิถุนายน 2534
สถานที่เกิด	กรุงเทพมหานคร
วุฒิการศึกษา	คณะศิลปศาสตร์, เอกภาษาอังกฤษ มหาวิทยาลัยธรรมศาสตร์
ที่อยู่ปัจจุบัน	40/9 หมู่บ้านมหาดไทย 1 ถนนพุทธมณฑลสาย 1 แขวงบางระมาด เขตตลิ่งชัน กรุงเทพมหานคร 10170



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY