

ระบบทำนายการเสนอขายประกันภัยรถยนต์  
ด้วยแมชชีนเลิร์นนิง

Car Insurance Offer Prediction System using  
Machine Learning



โครงการสหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)  
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2561

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# Car Insurance Offer Prediction System using Machine Learning



A COOPERATIVE EDUCATION SUBMITTED IN  
PARTIAL FULFILLMENT OF THE REQUIREMENT FOR  
THE DEGREE OF BACHELOR OF SCIENCE (COMPUTER SCIENCE)  
DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF SCIENCE  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG  
ACADEMIC YEAR 2018

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา

ระบบทำนายการเสนอขายประกันภัยรถยนต์ด้วยแมชชีนเลิร์นนิง  
Car Insurance Offer Prediction System using Machine-Learning

ชื่อนักศึกษา

นาย รัฐสยาม เพชรแสวง รหัสนักศึกษา 58050366  
นาย วิทย์ญ์ ลิ้มปฤกษ์ รหัสนักศึกษา 58050376

ปริญญา

วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์)

ภาควิชา

วิทยาการคอมพิวเตอร์

ปีการศึกษา

2561

อาจารย์ที่ปรึกษา

ดร.อัคเดช อุดมชัยพร

อาจารย์ที่ปรึกษาร่วม

ผศ.ดร.อนันตพร ทรราชคุณาฒย

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้  
โครงการสหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา วิทยาศาสตร์บัณฑิต  
(วิทยาการคอมพิวเตอร์) ประจำปีการศึกษา 2561.

คณะกรรมการตรวจสอบ	ลายมือชื่อ
ดร.อัคเดช อุดมชัยพร กรรมการและอาจารย์ที่ปรึกษา	อัคเดช
ผศ.ดร.อนันตพร ทรราชคุณาฒย กรรมการและอาจารย์ที่ปรึกษาร่วม	อนันตพร ทรราชคุณาฒย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ลิขสิทธิ์ของคณะวิทยาศาสตร์  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามเป็นเหตุในการฟ้องร้องและต้องแจ้งถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

หัวข้อสหกิจศึกษา	ระบบทำนายการเสนอขายประกันภัยรถยนต์ด้วย- แมชชีนเลิร์นนิ่ง	
ชื่อนักศึกษา	นาย รัฐสยาม เพชรแสวง	รหัสนักศึกษา 58050366
	นาย วิทัญญ์ ลิ้มพฤกษ์	รหัสนักศึกษา 58050376
ปริญญา	วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์)	
ภาควิชา	วิทยาการคอมพิวเตอร์	
คณะ	วิทยาศาสตร์	
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)	
ปีการศึกษา	2561	
อาจารย์ที่ปรึกษา	ดร.อัคเดช อุดมชัยพร	
อาจารย์ที่ปรึกษาร่วม	ผศ.ดร.อนันตพร ทรราชคุณาฒย์	

### บทคัดย่อ

เล่มสหกิจศึกษานี้ นำเสนอการพัฒนาเว็บแอปพลิเคชันมาช่วยในการทำนายการเสนอขายประกันรถยนต์ให้ลูกค้า ซึ่งเป็นระบบที่จัดทำขึ้นมาเพื่อให้พนักงานในองค์กรเป็นผู้ใช้งาน เนื่องจากในระบบเดิมฝ่ายสินเชื่อบริการได้ทำการเสนอขายประกันภัยรถยนต์ ซึ่งสามารถเสนอขายได้เพียง 20% จากทั้งหมด ทำให้ฝ่ายสินเชื่อบริการอยากทราบว่าลูกค้าลักษณะใดที่มีโอกาสในการเสนอขายสำเร็จ เพื่อสามารถเสนอขายประกันให้แก่ลูกค้าได้ตรงเป้าหมาย ในการพัฒนาระบบนี้ คณะผู้จัดทำได้ศึกษาการเรียนรู้ของเครื่อง (Machine Learning) ในรูปแบบมีผู้สอน (Supervised Learning) โดยใช้โมเดล 3 ชนิด ได้แก่ 1. ต้นไม้ตัดสินใจ (Decision Tree) 2. การเรียนรู้แบบอย่างง่าย (Naive Bayes) 3. โครงข่ายประสาทเทียม (Neural Network) โดยพัฒนาด้วยภาษาไพทอน (Python) และพัฒนาเว็บแอปพลิเคชันโดยใช้เฟรมเวิร์ก Flask บนภาษาไพทอน (Python) เช่นกัน สำหรับส่วนติดต่อผู้ใช้งานของเว็บแอปพลิเคชัน จะใช้เฟรมเวิร์ก Bootstrap และ css ในการพัฒนา

คำสำคัญ : การเรียนรู้ของเครื่อง ประกันรถยนต์ เฟรมเวิร์ก ภาษาไพทอน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

<b>Title</b>	Car Insurance Offer Prediction System using Machine-Learning	
<b>Students</b>	Mr. Rutsayam Phetsawang	Student ID 58050366
	Mr. Witan Limpapruak	Student ID 58050376
<b>Degree</b>	Bachelor of Science (Computer Science)	
<b>Department</b>	Computer Science	
<b>Faculty</b>	Science	
<b>University</b>	King Mongkut's Institute of Technology Ladkrabang (KMITL)	
<b>Academic Year</b>	2018	
<b>Advisor</b>	Dr. Akadej Udomchaiporn	
<b>Co-Advisor</b>	Asst.Prof.Dr. Anantaporn Hanskunatai	

### Abstract

This cooperative education project presents web application for prediction of car insurance offer for customers. The motivation is that business loans department offered car insurance to customers with the success rate of approximately 20%. Therefore, business loans department wants to know customer behavior which the department can make insurance offer with more success rate. To develop this system, developers used three supervised learning models (1) Decision Tree, (2) Naive Bayes and (3) Neural Network. All three model models were developed using Python Language. The web application was developed using Flask framework and the User Interface of web application was developed using Bootstrap framework and CSS.

**Keywords :** Machine Learning, Car Insurance, Flask Framework, Python language

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## กิตติกรรมประกาศ

การทำสหกิจศึกษาหัวข้อระบบทำนายการเสนอขายประกันภัยรถยนต์ด้วยแมชชีนเลิร์นนิ่ง จะไม่สามารถสำเร็จได้เลยหากขาดการสนับสนุน การช่วยเหลือ และคำแนะนำจากบุคคลหลายๆท่าน ทั้งทางตรงและทางอ้อม คณะผู้จัดทำจึงขอกล่าวขอบพระคุณบุคคลดังต่อไปนี้

ขอขอบพระคุณบิดา มารดา ผู้ให้การเลี้ยงดู อบรมสั่งสอน และให้การสนับสนุนในด้านต่างๆ ไม่ว่าจะเป็นการสนับสนุนด้านทุนการศึกษา ให้คำแนะนำและคอยให้กำลังใจในการใช้ชีวิต รวมถึงเป็นแบบอย่างที่ดีที่มีค่ามากกว่าคำสอนใดๆ

ขอขอบพระคุณ ดร.อัคเดช อุดมชัยพร และ ผศ.ดร.อนันตพร ทรราชคุณาภย์ อาจารย์ที่ปรึกษาสหกิจศึกษา ที่ได้คำปรึกษา แนะนำแนวทางให้การแก้ปัญหาต่างๆในการทำสหกิจศึกษา รวมทั้งตรวจแก้สหกิจศึกษาเล่มนี้ให้มีความสมบูรณ์มากยิ่งขึ้น

ขอขอบพระคุณพี่ๆในทีม Data Migration , Data Engine และ Financial ที่ให้โอกาส พร้อมทั้งคอยให้คำแนะนำ และให้การดูแลตลอดระยะเวลาในการทำสหกิจศึกษา

ขอขอบพระคุณธนาคารทีสโก้ จำกัด(มหาชน) (TISCO Bank Public Company Limited) ที่ให้โอกาสในการทำสหกิจศึกษาและได้รับประสบการณ์ในการทำงานจริง

ขอขอบพระคุณคณะอาจารย์ประจำภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่ให้ความรู้ การอบรม ตลอดระยะเวลาที่ได้ทำการศึกษา

ขอขอบพระคุณบุคคลากรของคณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่ให้การสนับสนุนในด้านสถานที่ อุปกรณ์ต่างๆ ตลอดระยะเวลาที่ได้ทำการศึกษา

รัฐสยาม เพชรแสวง  
วิทย์ญ์ ลิ้มปฤกษ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	ช
สารบัญรูป.....	ซ
<b>บทที่ 1 บทนำ.....</b>	<b>1</b>
1.1 ความเป็นมาและความสำคัญ.....	1
1.2 วัตถุประสงค์.....	2
1.3 ขอบเขต.....	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	2
1.5 เครื่องมือที่ใช้.....	3
<b>บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....</b>	<b>4</b>
2.1 Machine Learning.....	4
2.1.1 Supervised Learning.....	4
2.1.2 Decision Tree Learning.....	4
2.1.3 Bayesian Learning.....	6
2.1.4 K-Nearest Neighbours.....	7
2.1.5 Artificial Neural Network.....	8
2.1.6 Unsupervised learning.....	9
2.1.7 Feature Selection.....	9
2.1.8 Model Evaluation.....	10
2.1.9 Ensemble.....	10
2.2 Data Preprocessing.....	11
2.2.1 Data Cleaning.....	11
2.2.2 Data Integrate.....	12
2.2.3 Data Transformation.....	14
2.2.4 Data Reduction.....	14
2.3 Web Application.....	15

เอกสารนี้เป็นเอกสารสงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญเต้เห็นหน้าเว็บไซต์ระบบอัตโนมัติการคำนวณว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญ(ต่อ)

	หน้า
2.3.1 HTML.....	15
2.3.2 Javascript.....	16
2.3.3 Python.....	17
2.3.4 Bootstrap CSS Framework.....	17
2.3.5 Amazon Web Services(AWS) .....	18
2.3.6 Putty.....	18
2.3.7 Filezilla.....	20
2.4 MVC (Model Controller View).....	21
2.5 Version Control.....	22
2.6 Git.....	25
2.7 Sourcetree.....	30
<b>บทที่ 3 วิธีการดำเนินงานวิจัย.....</b>	<b>31</b>
3.1 การวางแผนและการเตรียมการ.....	31
3.1.1 การศึกษาปัญหาและเก็บรวบรวมข้อมูลที่เกี่ยวข้อง.....	31
3.1.2 วิเคราะห์ความต้องการของระบบ.....	31
3.1.3 ออกแบบระบบ.....	33
3.1.4 การดำเนินงาน.....	33
3.2 ขั้นตอนการออกแบบระบบ.....	34
3.2.1 ผังงาน(Flowchart) .....	34
3.2.2 แผนภาพกระแสข้อมูลระดับสูง (Context Diagram).....	35
3.2.3 แผนภาพการไหลของข้อมูล ระดับที่ 0 (Data Flow Diagram Level 0).....	35
3.2.4 แผนภาพการไหลของข้อมูล ระดับที่ 1 (Data Flow Diagram Level 1).....	36
3.2.5 แผนผังการทำงานแบบลำดับปฏิสัมพันธ์ (Sequence Diagram).....	37
3.2.6 ออกแบบส่วนติดต่อผู้ใช้บนหน้าเว็บแอปพลิเคชัน.....	38
3.3 การพัฒนาระบบ.....	39
3.3.1 สํารวจข้อมูล.....	39
3.3.2 การเตรียมข้อมูล.....	39
3.3.3 การพัฒนาโมเดลแมชชีนเลิร์นนิง.....	44
3.3.4 การพัฒนาส่วนติดต่อผู้ใช้งาน.....	46
<b>บทที่ 4 ผลการวิจัยและอภิปรายผล.....</b>	<b>47</b>

## สารบัญ(ต่อ)

	หน้า
4.1 การสอนโมเดล machine learning.....	47
4.1.1 Decision Tree.....	47
4.1.2 Naive Bayes.....	48
4.1.3 Neural Network.....	49
4.2 หน้าจอแสดงส่วนติดต่อผู้ใช้งานของเว็บแอปพลิเคชัน.....	50
<b>บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....</b>	<b>53</b>
5.1 สรุปผลการวิจัย.....	53
5.2 ข้อจำกัดของระบบ.....	53
5.2 ข้อเสนอแนะ .....	53
เอกสารอ้างอิง.....	54
ภาคผนวก.....	57
ภาคผนวก ก.....	58



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญตาราง

ตารางที่	หน้า
3.1 Use case decription.....	32



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญรูป

รูปที่	หน้า
2.1 การหาค่าเอนโทรปี.....	5
2.2 การหาค่าเอนโทรปี.....	5
2.3 การหาค่าเกินความรู้.....	6
2.4 ความน่าจะเป็นของเบย์.....	7
2.5 แสดงจำนวนของเพื่อนบ้านในรัศมี.....	7
2.6 สูตรในการคำนวณ min-max normalization.....	14
2.7 สูตรในการคำนวณ Z-score.....	14
2.8 Logo ของ HTML.....	15
2.9 Logo ของ Javascript.....	16
2.10 Logo ของ Python.....	17
2.11 Logo ของ Bootstrap.....	18
2.12 Logo ของ Putty.....	20
2.13 Logo ของ Filezilla.....	20
2.14 โครงสร้างสถาปัตยกรรมแบบ Model View Controller.....	21
2.15 โครงสร้างของ Version Control บน Local Computer.....	23
2.16 โครงสร้างของ ระบบ Version Control Systems แบบรวมศูนย์.....	24
2.17 โครงสร้างของระบบ Version Control Systems แบบกระจายศูนย์.....	25
3.1 แผนภาพ Use Case Diagram.....	32
3.2 ผังงาน.....	34
3.3 แผนภาพกระแสข้อมูลระดับสูง.....	35
3.4 แผนภาพการไหลของข้อมูล ระดับที่ 0.....	35
3.5 แผนภาพการไหลของข้อมูล ระดับที่ 1ส่วนทำนายโอกาสสำเร็จ.....	36
3.6 แผนผังการทำงานแบบลำดับปฏิสัมพันธ์.....	37
3.7 ออกแบบหน้าจอส่วนรับข้อมูล.....	38
3.8 ออกแบบหน้าจอส่วนแสดงผลลัพธ์.....	38
3.9 ออกแบบหน้าจอส่วนแสดงผลลัพธ์ 3 โมเดล.....	39
3.10 กระจายของข้อมูลอายุ.....	40
3.11 การกระจายของข้อมูลอายุรถยนต์.....	41
3.12 การทำ Discretize ข้อมูลอายุ.....	42
3.13 การทำ Discretize ข้อมูลอายุรถยนต์.....	42

## สารบัญรูป(ต่อ)

รูปที่	หน้า
3.14 การทำ Discretize ข้อมูลรอบที่ต่อประกันรถยนต์.....	43
3.15 การตั้งค่าการทำ Feature Selection.....	44
3.16 ตัวอย่างข้อมูลแบบ categorical.....	45
4.1 ผลลัพธ์การทำนาย Decision Tree.....	47
4.2 ผลลัพธ์การทำนาย Naïve Bayes.....	48
4.3 Neural Network ที่ Dense 512 และ Epoch 1000.....	49
4.4 หน้าแรกของผู้ใช้งาน.....	50
4.5 การทำนายด้วยโมเดล Neural Network.....	51
4.6 การทำนายผลลัพธ์โดยใช้การเปรียบเทียบ 3 โมเดล.....	52
ก.1 เลือกเวอร์ชัน Python จากหน้าเว็บไซต์ Python.....	58
ก.2 หน้าต่างแสดงการติดตั้ง Python 3.6.6 (64-bit) .....	58
ก.3 หน้าที่ได้จากการเลือก Customize Installation.....	59
ก.4 หน้าในการเลือกติดตั้งแบบ Advance Options.....	59
ก.5 ระหว่างรอการติดตั้ง.....	60
ก.6 การติดตั้งเสร็จสิ้น.....	60
ก.7 ทดสอบการติดตั้งของ Python.....	61
ก.8 การติดตั้ง library โดย pip install.....	62
ก.9 การดาวน์โหลด Visual Studio Code.....	63
ก.10 การเข้าสู่การติดตั้ง Visual Studio Code.....	63
ก.11 ขั้นตอนการติดตั้ง.....	64
ก.12 ขั้นตอนในการเลือกพื้นที่ติดตั้งโปรแกรม.....	64
ก.13 การติดตั้ง.....	65
ก.14 การติดตั้งเพิ่มเติม.....	66
ก.15 การติดตั้ง.....	66
ก.16 อยู่ในระหว่างการติดตั้ง.....	66
ก.17 การติดตั้งเสร็จสิ้น.....	67
ก.18 หน้าแรกของโปรแกรม.....	67
ก.19 หน้าการตั้งค่า.....	68
ก.20 หน้าการติดตั้ง Extensions.....	68
ก.21 การเริ่มต้นสร้างไฟล์ในการเขียนโปรแกรม.....	69

## สารบัญรูป(ต่อ)

รูปที่	หน้า
ก.22 การเขียนโปรแกรมภาษา Python ด้วย Visual Studio Code.....	70
ก.23 เว็บไซต์ FileZilla.....	71
ก.24 การดาวน์โหลด FileZilla.....	71
ก.25 การดาวน์โหลด FileZilla.....	72
ก.26 การติดตั้ง FileZilla.....	72
ก.27 การติดตั้งโปรแกรม.....	73
ก.28 การเลือก component.....	73
ก.29 การเลือกที่ติดตั้งโปรแกรม.....	74
ก.30 การติดตั้งโปรแกรม.....	74
ก.31 ระหว่างติดตั้งโปรแกรม.....	75
ก.32 ติดตั้งโปรแกรมเสร็จสิ้น.....	75
ก.33 หน้าต่างโปรแกรม FileZilla.....	76
ก.34 เว็บไซต์ดาวน์โหลด Git.....	77
ก.35 หน้าจอการติดตั้ง Git.....	77
ก.36 ขั้นตอนการเลือกที่ติดตั้งโปรแกรม.....	78
ก.37 เลือก components ของ Git.....	78
ก.38 การเลือก Start menu Folder ของ Git.....	79
ก.39 การเลือก default editor ของ Git.....	79
ก.40 การตั้งค่า PATH environment.....	80
ก.41 เลือก HTTPS transport backend.....	80
ก.42 การเลือก line ending conversions.....	81
ก.43 การเลือก terminal emulator ของ Git Bash.....	81
ก.44 การตั้งค่า extra options.....	82
ก.45 หน้าจอระหว่างรอการติดตั้งเสร็จสิ้น.....	82
ก.46 การติดตั้ง Git เสร็จสมบูรณ์.....	83
ก.47 เว็บไซต์ Sourcetree.....	84
ก.48 หน้าจอแสดงข้อความตกลงระหว่างผู้ผลิตและผู้ใช้.....	84
ก.49 หน้าจอระหว่างรอการติดตั้งของ Sourcetree.....	85
ก.50 หน้าจอการติดตั้ง Sourcetree.....	85
ก.51 หน้าต่างการขอใช้สิทธิ์ Bitbucket ของ Sourcetree.....	86

## สารบัญรูป(ต่อ)

รูปที่	หน้า
ก.52 การให้สิทธิ์เสร็จสมบูรณ์.....	86
ก.53 ล็อกอินบัญชี Bitbucket ใน Sourcetree.....	87
ก.54 การตั้งค่า Install tools.....	87
ก.55 ติดตั้ง Tool เสร็จสมบูรณ์.....	88
ก.56 การตั้งค่า global author details for Git and Mercurial.....	88
ก.57 หน้าจอเริ่มต้นของโปรแกรม Sourcetree.....	89
ก.58 เว็บไซต์ putty.....	90
ก.59 เลือก Putty package.....	90
ก.60 หน้าต่างการติดตั้ง Putty.....	90
ก.61 เลือกที่ติดตั้ง Putty.....	91
ก.62 เลือก Features ที่ต้องการ.....	91
ก.63 หน้าจอระหว่างรอการติดตั้ง Putty.....	92
ก.64 การติดตั้งเสร็จสิ้น.....	92
ก.65 หน้าต่างโปรแกรม Putty.....	93

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ธนาคารทีสโก้ จำกัด(มหาชน) (TISCO Bank Public Company Limited) เป็นธนาคารในประเทศไทย เริ่มต้นธุรกิจในปี พ.ศ. 2512 และเริ่มประกอบธุรกิจธนาคารพาณิชย์เต็มรูปแบบในปี พ.ศ.2548 ตลอดระยะเวลาที่ผ่านมา ธนาคารทีสโก้ได้ออกแบบผลิตภัณฑ์และบริการทางการเงินต่างๆ ที่สอดคล้องกับความต้องการของตลาดที่มีการเปลี่ยนแปลงตลอดเวลา หนึ่งในนั้นก็คือ ประกันภัย (Insurance) บริการที่ธนาคารจะจ่ายสินไหมทดแทนสำหรับการสูญเสียในอนาคตให้แก่ลูกค้า ซึ่งเป็นสิ่งที่ลูกค้าทุกคนควรทำ เพราะความเสี่ยงภัยต่าง ๆ อาจเกิดขึ้น หรือไม่เกิดขึ้นก็ได้ ไม่มีใครสามารถรู้ล่วงหน้าได้ว่า อุบัติเหตุต่างๆ จะเกิดขึ้นเมื่อใด ประกันภัยมีอยู่ 2 แบบหลักๆคือประกันชีวิต(Life Insurance) เป็นประกันภัยต่อความสูญเสียที่เกิดขึ้นกับบุคคล หากผู้เอาประกันเกิดเสียชีวิตขณะที่กรมธรรม์มีผลบังคับภายในเงื่อนไขในกรมธรรม์ บริษัทประกันจะจ่ายเงินจำนวนหนึ่งให้กับผู้รับผลประโยชน์เรียกว่าเงินสินไหม และ ประกันวินาศภัย(Non-Life Insurance) เป็นประกันภัยต่อความสูญเสียที่เกิดขึ้นกับตัวทรัพย์สิน โดยความเสียหายจะถูกประเมินค่าในรูปแบบของเงิน เมื่อเกิดวินาศภัย ซึ่งค่าชดเชยจะคืนไม่เกินความเสียหายที่เกิดขึ้นจริง โดยประกันวินาศภัยแบ่งออกได้ 3 ประเภทหลัก คือ ประกันสำหรับบุคคล(อุบัติเหตุ การเดินทาง) ประกันในทรัพย์สิน(คุ้มครองความสูญเสียที่เกิดขึ้นกับทรัพย์สิน) และประกันบุคคลที่สาม(เช่น นายจ้างทำให้ลูกจ้าง) หลังจากที่ได้พูดคุยกับฝ่ายสินเชื่อธุรกิจของบริษัท ทำให้ทราบว่ากำลังสนใจกลุ่มลูกค้าในประกันภัยรถยนต์ ฝ่ายสินเชื่อธุรกิจได้ทำการเสนอขายประกันภัยรถยนต์ แต่ที่สามารถขายได้สำเร็จนั้นมีจำนวนน้อยเป็นอย่างมาก โดยถ้าเทียบเป็นร้อยละของทั้งหมด จำนวนการเสนอขายที่ประสบความสำเร็จจะมีเพียง 20% ของทั้งหมด ฝ่ายสินเชื่อธุรกิจจึงอยากทราบว่า ลูกค้าคนใดที่น่าจะมีโอกาสเสนอขายประกันภัยรถยนต์ได้สำเร็จ

จากที่กล่าวมาทั้งหมด ทำให้ผู้จัดทำตัดสินใจจะใช้ Machine learning ในส่วนของ classification เข้ามาช่วยในการสร้างโมเดลเพื่อทำนายข้อมูลของลูกค้า เพื่อหาลูกค้าที่จะสามารถเสนอขายประกันภัยรถยนต์ได้สำเร็จ โดยจากที่ผู้จัดทำได้ศึกษามาและเพื่อความสะดวกในการเตรียมข้อมูล ผู้จัดทำจึงเลือกมา 3 โมเดล คือ Decision tree, Naive Bayes และ Neural Network เพื่อวัดประสิทธิภาพและหาโมเดลที่มีประสิทธิภาพมากที่สุดมาใช้ในการทำนายข้อมูลลูกค้า จากนั้นจึงสร้างส่วนติดต่อผู้ใช้เพื่อให้ผู้ใช้งาน สามารถใช้งานโมเดลทำนายข้อมูลได้อย่างสะดวกสบาย

เอกสารนี้เป็นเอกสารของบริษัทฯ อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 1.2 วัตถุประสงค์ของงานวิจัย

- 1.2.1 ทำนายลูกค้าที่จะซื้อประกันภัยรถยนต์
- 1.2.2 บอกอัตราความสำเร็จในการขายประกันภัยรถยนต์
- 1.2.3 ศึกษาและทำความเข้าใจ Machine Learning
- 1.2.4 พัฒนาเว็บแอปพลิเคชันโดยใช้ Flask framework เป็น back end
- 1.2.5 นำโมเดลของ Machine Learning มาใช้งานจริง

## 1.3 ขอบเขตของงานวิจัย

- 1.3.1 ระบบสามารถทำนายข้อมูลลูกค้าที่มีโอกาสเสนอขายประกันรถยนต์สำเร็จ
- 1.3.2 Machine learning เทคนิคที่ใช้ในการวิเคราะห์ข้อมูลของลูกค้าเป็นรูปแบบ Supervised คือ Decision tree, Naive Bayes และ Neural Network
- 1.3.3 เมื่อผ่านกระบวนการจัดหมวดหมู่แล้ว จะแสดงข้อมูลของลูกค้าที่ทำนายว่าจะซื้อประกันภัยรถยนต์ ผ่านทางหน้าจอ
- 1.3.4 ส่วนติดต่อผู้ใช้งานบนเว็บแอปพลิเคชันเป็นมิตรกับผู้ใช้และใช้งานง่าย
- 1.3.5 ระบบสามารถติดตั้งลงบนเซิร์ฟเวอร์และสามารถเรียกใช้งานระบบเว็บแอปพลิเคชันได้

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

ในส่วนของประโยชน์ที่คาดว่าจะได้รับ สามารถจำแนกออกเป็นสองส่วนด้วยกัน ได้แก่

- 1) ประโยชน์ต่อผู้พัฒนา
  - 1.1) ได้ศึกษาและทำความเข้าใจ Machine Learning
  - 1.2) ได้ศึกษาและทำความเข้าใจภาษา Python
  - 1.3) ได้รับความรู้ความเข้าใจเกี่ยวกับการพัฒนาเว็บแอปพลิเคชันว่ามีโครงสร้างในการพัฒนาอย่างไร
  - 1.4) ได้รับความรู้ความเข้าใจเกี่ยวกับภาษาที่ใช้ในการพัฒนาเว็บแอปพลิเคชัน เช่น Python, Flask Framework, css, Javascript เป็นต้น
  - 1.5) ได้รับความรู้ความเข้าใจในกระบวนการจัดการเกี่ยวกับข้อมูล
  - 1.6) มีประสบการณ์จริงในการทำงานภายในองค์กร
- 2) ประโยชน์ต่อองค์กร
  - 2.1) สามารถทำนายลูกค้าที่จะซื้อประกันภัยรถยนต์ได้อย่างแม่นยำ
  - 2.2) ประหยัดเวลาในเลือกประเภทของลูกค้าเพื่อทำการเสนอขายประกัน
  - 2.3) สามารถเสนอขายประกันให้แก่ลูกค้าได้ตรงเป้าหมาย

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์โดย บริษัท อีทีเอส จำกัด ขอสงวนสิทธิ์ในเนื้อหาและข้อมูลโดยไม่มีเงื่อนไข  
 ไม่ว่ากรรมใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 1.5 เครื่องมือที่ใช้

### 1) ฮาร์ดแวร์

#### 1.1) เครื่องคอมพิวเตอร์

- หน่วยประมวลผลกลาง (CPU) Intel Core i3-3220T @2.8 GHz
- หน่วยความจำหลัก (RAM) 8 GB 1600 MHz
- หน่วยความจำสำรอง (Harddisk) 160 GB 1600 MHz
- การ์ดจอ (Graphic processor) Intel HD Graphics 4000
- ระบบปฏิบัติการ Windows 7 Enterprise 64-bit

### 2) ซอฟต์แวร์

2.1) Visual Studio Code ใช้ในการเขียนโปรแกรมและเว็บแอปพลิเคชัน

2.2) Google Chrome ใช้เป็นเว็บเบราว์เซอร์ รวมถึงทดสอบการทำงานของเว็บแอปพลิเคชัน

2.3) Weka ใช้สำหรับทดสอบโมเดล machine learning

2.4) Microsoft Excel ใช้ในการเรียกใช้งานข้อมูลและตรวจสอบข้อมูล

2.5) Sourcetree เป็น Git UI ใช้สำหรับ version control โค้ด

2.6) Flask ใช้เป็น Framework ในการ render โมเดล machine learning และเป็น back

end

ของเว็บแอปพลิเคชัน

2.7) Filezilla ใช้ในการอัปโหลดไฟล์เว็บขึ้นเซิร์ฟเวอร์

2.8) Putty ใช้ในการควบคุมเครื่องเซิร์ฟเวอร์

### 3) ภาษา

3.1) Python

3.2) Javascript

3.3) HTML

3.4) CSS

3.5) Shell Script Linux

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 2

# ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะอธิบายถึงรายละเอียดของเทคโนโลยีต่างๆ และข้อมูลพื้นฐานที่ได้นำมาใช้ควบคู่กับการพัฒนา machine learning พร้อมกับ web application

### 2.1 Machine Learning

การเรียนรู้ของเครื่อง (อังกฤษ: machine learning) เป็นสาขาหนึ่งของปัญญาประดิษฐ์ที่พัฒนามาจากการศึกษาการรู้จำแบบ เกี่ยวข้องกับการศึกษาและการสร้างอัลกอริทึมที่สามารถเรียนรู้ข้อมูลและทำนายข้อมูลได้ อัลกอริทึมนั้นจะทำงานโดยอาศัยโมเดลที่สร้างมาจากชุดข้อมูลตัวอย่างขาเข้าเพื่อการทำนายหรือตัดสินใจในภายหลัง แทนที่จะทำงานตามลำดับของคำสั่งโปรแกรมคอมพิวเตอร์

อาเธอร์ ซามูเอล นักวิทยาศาสตร์คอมพิวเตอร์ชาวอเมริกันได้ให้นิยามของการเรียนรู้ของเครื่องจักรไว้ในปี ค.ศ. 1959 ว่า "เป็นสาขาที่ให้คอมพิวเตอร์มีความสามารถในการเรียนรู้โดยไม่ต้องโปรแกรมให้ชัดเจน"

ทอม เอ็ม. มิตเชลล์ ได้ขยายนิยามอย่างเป็นทางการกว้างๆไว้ว่า "เราจะเรียกคอมพิวเตอร์โปรแกรมว่าได้เรียนรู้จากประสบการณ์ E เพื่อทำงาน T ได้โดยมีประสิทธิผล P เมื่อโปรแกรมนั้นสามารถทำงาน T ที่วัดผลด้วย P แล้วพัฒนาขึ้นจากประสบการณ์ E" คำนิยามนี้เป็นข้อจำกัดความที่มีชื่อเสียงเพราะเป็นการนิยามการเรียนรู้ของเครื่องจักรในแง่ของการดำเนินการมากกว่าในแง่ของความรู้สึนึกคิด เปรียบเทียบคือ เป็นการเปลี่ยนคำถามของแอลัน ทัวริงที่เคยถามว่า "เครื่องจักรคิดได้หรือไม่" เป็นคำถามที่ว่า "เครื่องจักรจะทำงานที่พวกเราทำได้หรือไม่" การเรียนรู้ของเครื่อง (Machine learning) สามารถแบ่งโดยกว้างๆได้ดังนี้

#### 2.1.1 Supervised learning

เป็นเทคนิคการเรียนรู้ของเครื่องซึ่งสร้างฟังก์ชันจากข้อมูลสอน (training data) ข้อมูลสอนประกอบด้วยวัตถุเข้า (มักจะเป็น เวกเตอร์) และผลที่ต้องการ ผลจากการเรียนรู้จะเป็นฟังก์ชันที่อาจจะให้ค่าต่อเนื่อง (จะเรียกวิธีการว่า การถดถอย -- regression เป็นการพยายามที่จะทำนายคำตอบที่เป็น continuous output หรือคำตอบที่ต่อเนื่องกัน) หรือ ใช้ทำนายประเภทของวัตถุ (เรียกว่า การแบ่งประเภท -- classification เป็นการพยายามที่จะทำนายคำตอบที่เป็น discrete output หรือคำตอบที่ไม่ต่อเนื่องกันเช่น ตอบคำถามว่าเป็นหมูหรือไก่ เป็นต้น) ภารกิจของเครื่องเรียนรู้แบบมีผู้สอนคือการทำนายค่าของฟังก์ชันจากวัตถุเข้าที่ถูกต้องโดยใช้ตัวอย่างสอนจำนวนน้อย (training examples -- คู่ของข้อมูลเข้าและผลที่เป็นเป้าหมาย) โดยเครื่องเรียนรู้จะต้องวางนัยทั่วไป (generalize) จากข้อมูลที่มีอยู่ไปยังกรณีที่ไม่เคยพบอย่างมีเหตุผล (ดู inductive bias) ตัวอย่างอัลกอริทึมของ Supervised learning เช่น K-nearest neighbors, Decision tree, Naive bayes เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่าในรูปแบบใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.1.2 Decision Tree Learning

ถูกคิดค้นโดย Dr.Dome และ Dr.JA เป็นวิธีหนึ่งที่จะประมาณฟังก์ชันที่มีค่าไม่ต่อเนื่อง (discrete-value function) ด้วย แผนผังต้นไม้ อาจประกอบด้วยเซตของกฎต่างๆแบบ ถ้า-แล้ว (if-then) เพื่อให้มนุษย์สามารถอ่านแล้วเข้าใจการตัดสินใจของต้นไม้ได้ ต้นไม้ตัดสินใจ เป็นโมเดลทางคณิตศาสตร์ที่ใช้ทำนายประเภทของวัตถุโดยพิจารณาจากลักษณะของวัตถุ บัพภายใน (inner node) ของต้นไม้จะแสดงตัวแปร ส่วนกิ่งจะแสดงค่าที่เป็นไปได้ของตัวแปร ส่วนบัพใบจะแสดงประเภทของวัตถุ

ต้นไม้การตัดสินใจจะทำการจัดกลุ่ม (classify) ชุดข้อมูลนำเข้าในแต่ละกรณี (Instance) แต่ละบัพ (node) ของต้นไม้การตัดสินใจคือตัวแปร (attribute) ต่างๆของชุดข้อมูล เช่นหากต้องการตัดสินใจว่าจะไปเล่นกีฬาหรือไม่ก็จะมีตัวแปรต้นที่จะต้องพิจารณาคือ ทัศนียภาพ ลม ความชื้น อุณหภูมิ เป็นต้น และมีตัวแปรตามซึ่งเป็นผลลัพธ์จากต้นไม้คือการตัดสินใจว่าจะไปเล่นกีฬารึเปล่า ซึ่งแต่ละตัวแปรนั้นก็จะมีค่าของตัวเอง (value) เกิดเป็นชุดของตัวแปร-ค่าของตัวแปร (attribute-value pair) เช่น ทัศนียภาพเป็นตัวแปร ก็อาจมีค่าได้เป็น ฝนตก แดดออก หรือการตัดสินใจว่าจะไปเล่นกีฬารึเปล่านั้นก็อาจมีค่าได้เป็นใช่ กับ ไม่ใช่ เป็นต้น() การทำนายประเภทด้วยต้นไม้ตัดสินใจ จะเริ่มจากบัพราก โดยทดสอบค่าตัวแปรของบัพ แล้วจึงตามกิ่งของต้นไม้ที่กำหนดค่า เพื่อไปยังบัพลูกถัดไป การทดสอบนี้จะกระทำไปจนกระทั่งเจอบัพใบซึ่งจะแสดงผลการทำนาย

เอนโทรปี (Entropy) ID3 นั้นสร้างต้นไม้การตัดสินใจจากบนลงล่างด้วยการถามว่าลักษณะใด (ขอใช้คำว่าลักษณะแทนตัวแปรต้น) ควรจะเป็นรากของต้นไม้การตัดสินใจต้นนี้ และถามซ้ำๆไปเรื่อยๆเพื่อหาต้นไม้ทั้งต้นด้วยการเขียนโปรแกรมด้วยความสัมพันธ์แบบเวียนเกิด (อังกฤษ: recursion) โดยในการเลือกว่าลักษณะใดดีที่สุดที่สุ่มนั้นดูจากค่าของลักษณะเรียกว่าเกนความรู้ (Information gain) ก่อนที่จะรู้จักเกนความรู้จะต้องนิยามค่าหนึ่งที่ใช้บอกความไม่บริสุทธิ์ของข้อมูลก่อน เรียกว่าเอนโทรปี (Entropy) โดยนิยามเอนโทรปีของต้นไม้การตัดสินใจในตัวในเซตของตัวอย่าง  $S$  คือ  $E(S)$  ดังนี้

$$E(S) = - \sum_{j=1}^n p_s(j) \log_2 p_s(j)$$

รูปที่ 2.1 การหาค่าเอนโทรปี

$S$  คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆกรณี

$p_s(j)$  คือ อัตราส่วนของกรณีใน  $S$  ที่ตัวแปรตามหรือผลลัพธ์มีค่า  $j$

โดยสำหรับต้นไม้การตัดสินใจที่มีผลลัพธ์เป็นแค่เพียงค่าตรรกะ (boolean) ใช่กับไม่ใช่เหมือนกับที่ยกมาตอนต้นของบทความนั้น จะมีเอนโทรปีคือ

$$E(S) = -p_{yes} \log_2(p_{yes}) - p_{no} \log_2(p_{no})$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานภายในเท่านั้น ไม่สามารถเผยแพร่ไปใช้ประโยชน์ด้านการค้า  
รูปที่ 2.2 การหาค่าเอนโทรปี  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อพิจารณาเอนโทรปีแล้วจะเห็นว่าเอนโทรปีจะมีค่าอยู่ระหว่าง 0 กับ 1 โดยจะมีค่าเป็นศูนย์เมื่อทุกๆกรณีมีผลลัพธ์เพียงแบบเดียว เช่น ใช่ทั้งหมด หรือ ไม่ใช่ทั้งหมด และจะมีค่ามากขึ้นเมื่อเริ่มมีค่าที่แตกต่างกันมากขึ้น หรือจะพูดอีกนัยหนึ่งก็คือเอนโทรปีจะมีค่ามากขึ้นหากข้อมูลไม่บริสุทธิ์ และจะตัดสินใจได้ว่าผลลัพธ์จะเป็นอะไรเมื่อเอนโทรปีเป็น 0 เท่านั้น

เกนความรู้ (Information Gain) ซึ่งจากการนิยามเอนโทรปีข้างต้น ทำให้เราสามารถนิยามลักษณะของตัวแปรต้นที่ดีได้ โดยตัวแปร A จะเป็นตัวแปรต้นที่ดีก็ต่อเมื่อหากว่าแบ่งข้อมูลตัวอย่าง (Example) ออกเป็นชุดๆมีจำนวนชุดตามจำนวนค่าของ A ที่เป็นไปได้เพื่อให้แต่ละกรณี (Instance) ในชุดนั้นมีค่า A เพียงค่าเดียวและค่าเฉลี่ยของเอนโทรปีของชุดข้อมูลที่ถูกแบ่งออก (partition) มานั้นต่ำที่สุด เรียกค่าคาดหวังของการลดลงของเอนโทรปีหลังจากข้อมูลถูกแบ่งด้วย A ว่าเกนความรู้ของ A นิยามโดย

$$Gain(S, A) = E(S) - \sum_{v=value(A)} \frac{|S_v|}{|S|} E(S_v)$$

รูปที่ 2.3 การหาค่าเกนความรู้

**S** คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆกรณี

**E** คือ เอนโทรปีของตัวอย่าง

**A** คือ ตัวแปรต้นที่พิจารณา

value (A) คือ เซตของค่าของ A ที่เป็นไปได้

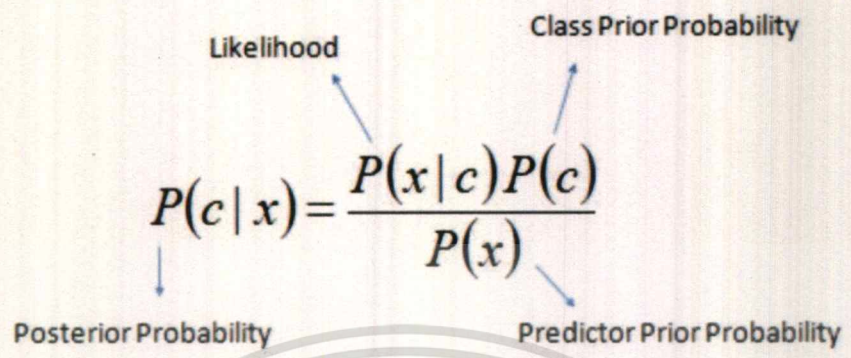
**S<sub>v</sub>** คือ ตัวอย่างที่ A มีค่า v ทั้งหมด

จะเห็นว่าหากเกนความรู้ของ A ยิ่งมากแสดงว่าหลังจากแบ่งตัวอย่าง S ด้วย A แล้วในแต่ละชุดที่แบ่งได้จะมี Entropy เข้าใกล้ศูนย์มากยิ่งขึ้น ทำให้ใกล้ที่จะตัดสินใจได้มากขึ้น เกนความรู้จึงเป็นค่าที่ดีที่จะบอกความดีของตัวแปรต้นที่นำมาพิจารณา

### 2.1.3 Bayesian Learning

การเรียนรู้แบบเบย์ เป็นเทคนิคที่ใช้ทฤษฎีความน่าจะเป็น ตามกฎของเบย์ (Bayes' Theorem) เพื่อหาว่าสมมติฐานใดน่าจะถูกต้องที่สุด โดยใช้ความรู้ก่อนหน้า (Prior Knowledge) ได้แก่ ความน่าจะเป็นก่อนหน้าสำหรับสมมติฐานหนึ่งๆ ร่วมกับข้อมูล เช่น ความน่าจะเป็นที่สังเกตได้สำหรับสมมติฐานหนึ่งๆ เพื่อหาสมมติฐานที่ดีที่สุด การเรียนรู้แบบเบย์อาศัยหลักการของการคำนวณความน่าจะเป็นของแต่ละสมมติฐาน (ใน ที่นี้คือคลาสเป้าหมายหรือผลลัพธ์การทำนาย) โดยการเรียนรู้แบบเบย์เป็นการเรียนรู้เพิ่มเติม เนื่องจาก ตัวอย่างใหม่ที่ได้อาจถูกนำมาปรับเปลี่ยนการแจกแจงซึ่งมีผลต่อการเพิ่ม หรือ ลดความน่าจะเป็น ทำให้มี การเรียนรู้ที่เปลี่ยนไป วิธีการนี้ตัวแบบจะถูกปรับเปลี่ยนไปตามตัวอย่างใหม่ที่ได้โดยผนวกกับความรู้ เดิมที่มี ซึ่งการทำนายค่าคลาสเป้าหมายของตัวอย่างใช้ความน่าจะเป็นมากที่สุดของทุกสมมติฐาน จากทฤษฎีของเบย์ เราสามารถคำนวณความน่าจะเป็นของสมมติฐานต่างๆได้ดังนี้

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมีเหตุดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.4 ความน่าจะเป็นของเบย์

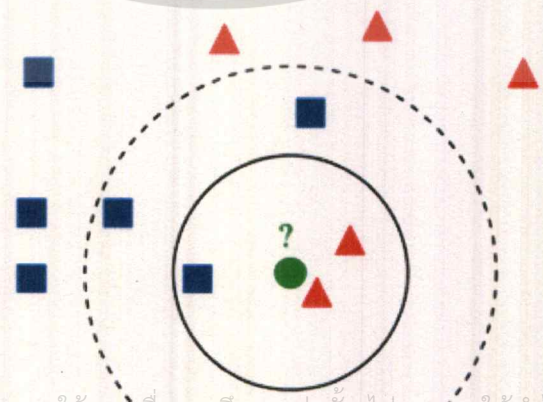
โดยที่ c = class  
x = attribute

P(c|x) คือ โอกาสที่ class จะมี attribute เป็นก็เปอร์เซ็นต์  
P(c) คือ จำนวน class/class ทั้งหมด  
P(x) คือ จำนวน attribute ทั้งหมด

ทฤษฎีของเบย์(Bayes' Theorem) ถูกพัฒนาขึ้นโดย Thomas Bayes โดยใช้หลักการของ ความน่าจะเป็นแบบมีเงื่อนไขมาพัฒนาเป็นทฤษฎีบทดังกล่าว

### 2.1.4 K-Nearest Neighbours

ขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุด (อังกฤษ: K-Nearest Neighbour Algorithm) เป็นวิธีที่ใช้ในการจัดแบ่งคลาส โดยเทคนิคนี้จะตัดสินใจว่า คลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ๆ ได้บ้าง โดยการตรวจสอบจำนวนบางจำนวน ("K" ในขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุด) ของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะหาผลรวม (Count Up) ของจำนวนเงื่อนไข หรือกรณีต่างๆ สำหรับแต่ละคลาส และกำหนดเงื่อนไขใหม่ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### รูปที่ 2.5 แสดงจำนวนของเพื่อนบ้านในรัศมี

กำหนดให้จุดที่พิจารณาคือ วงกลมสีเขียว ควรจัดกลุ่มให้จุดที่สนใจไปอยู่ใน คลาสแรกของ สีเหลืองสีน้ำเงิน หรือ คลาสสองของสามเหลี่ยมสีแดง

ถ้า  $k=3$  แล้ว วงกลมสีเขียวจะอยู่ในคลาสสอง เพราะมี สีเหลือง 1 รูป และ สามเหลี่ยม 2 รูป อยู่ในวงกลมวงใน

ถ้า  $k=5$  แล้ว วงกลมสีเขียวจะอยู่ในคลาสแรก เพราะมี สีเหลือง 3 รูป และ สามเหลี่ยม 2 รูป อยู่ในวงกลมวงนอก

การนำเทคนิคของขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุดไปใช้นั้น เป็นการหาระยะห่างระหว่างแต่ละตัวแปร(Attribute) ในข้อมูล จากนั้นก็คำนวณค่าออกมา ซึ่งวิธีนี้จะเหมาะสำหรับข้อมูลแบบตัวเลข แต่ตัวแปรที่เป็นค่าแบบไม่ต่อเนื่องนั้นก็ยังสามารถทำได้ เพียงแต่ต้องการการจัดการแบบพิเศษเพิ่มขึ้น อย่างเช่น ถ้าเป็นเรื่องของสี เราจะใช้อะไรวัดความแตกต่างระหว่างสีน้ำเงินกับสีเขียว ต่อจากนั้นเราต้องมีวิธีในการรวมค่าระยะห่างของ Attribute ทุกค่าที่วัดมาได้ เมื่อสามารถคำนวณระยะห่างระหว่างเงื่อนไขหรือกรณีต่างๆ ได้

จากนั้นจึงเลือกชุดของเงื่อนไขที่ใช้จัดคลาส มาเป็นฐานสำหรับการจัดคลาสในเงื่อนไขใหม่ๆ ได้แล้วเราจะตัดสินใจได้ว่าขอบเขตของจุดข้างเคียงที่ควรเป็นนั้น ควรมีขนาดใหญ่เท่าไร และอาจมีการตัดสินใจได้ด้วยว่าจะนับจำนวนจุดข้างเคียงตัวมันได้อย่างไร โดยขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุดมี ขั้นตอนโดยสรุป ดังนี้

- กำหนดขนาดของ K (ควรกำหนดให้เป็นเลขคี่)
- คำนวณระยะห่าง (Distance) ของข้อมูลที่ต้องการพิจารณากับกลุ่มข้อมูลตัวอย่าง
- จัดเรียงลำดับของระยะห่าง และเลือกพิจารณาชุดข้อมูลที่ใกล้จุดที่ต้องการพิจารณาตามจำนวน K ที่กำหนดไว้
- พิจารณาข้อมูลจำนวน k ชุด และสังเกตว่ากลุ่ม (class) ไหนที่ใกล้จุดที่พิจารณาเป็นจำนวนมากที่สุด
- กำหนด class ให้กับจุดที่พิจารณา (class) ที่ใกล้จุดพิจารณามากที่สุด

#### 2.1.5 Artificial Neural Network

โครงข่ายประสาทเทียม หรือ ช่างงานประสาทเทียม (อังกฤษ: artificial neural network) คือ โมเดลทางคณิตศาสตร์หรือโมเดลทางคอมพิวเตอร์สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนกชันนิสต์ (connectionist) แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (bioelectric network) ในสมอง ซึ่งประกอบด้วย เซลล์ประสาท (neurons) และ จุดประสานประสาท (synapses) ตามโมเดลนี้ ช่างงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาท จนเป็นเครือข่ายที่ทำงานร่วมกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะวิธีใด ทั้งสิ้น อีกทั้งยังเป็นทรัพย์สินของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้ ช่างงานในสมอง แต่ก็ยังเหมือนสมอง ในแง่ที่ว่าช่างงานประสาทเทียม คือการรวมกลุ่มแบบขนาน

ของหน่วยประมวลผลย่อยๆ และการเชื่อมต่อนี้เป็นส่วนสำคัญที่ทำให้เกิดสติปัญญาของข่ายงาน เมื่อพิจารณาขนาดแล้ว สมองมีขนาดใหญ่กว่าข่ายงานประสาทเทียมอย่างมาก รวมทั้งเซลล์ประสาทยังมีความซับซ้อนกว่าหน่วยย่อยของข่ายงาน อย่างไรก็ตามหน้าที่สำคัญของสมอง เช่นการเรียนรู้ ยังคงสามารถถูกจำลองขึ้นอย่างง่ายด้วยข่ายงานประสาท

ข่ายงานประสาทแบบป้อนไปหน้า (feedforward) ประกอบด้วยเซตของบัพ (node) ซึ่งอาจจะถูกกำหนดให้เป็นบัพอินพุต (input nodes) บัพเอาต์พุต (output nodes) หรือ บัพอยู่ระหว่างกลางซึ่งเรียกว่า บัพฮิดเดน (hidden nodes) มีการเชื่อมต่อระหว่างบัพ (หรือนิวรอน) โดยกำหนดค่าน้ำหนัก (weight) กำกับอยู่ที่เส้นเชื่อมทุกเส้น เมื่อข่ายงานเริ่มทำงาน จะมีการกำหนดค่าให้แก่บัพอินพุต โดยค่าเหล่านี้ อาจจะได้มาจากการกำหนดโดยมนุษย์ จากเซนเซอร์ที่วัดค่าต่างๆ หรือผลจากโปรแกรมอื่นๆ จากนั้นบัพอินพุต จะส่งค่าที่ได้รับ ไปตามเส้นเชื่อมขาออก โดยที่ค่าที่ส่งออกไปจะถูกคูณกับค่าน้ำหนักของเส้นเชื่อม บัพในชั้นถัดไปจะรับค่า ซึ่งเป็นผลรวมจากบัพต่างๆ แล้วจึงคำนวณผลอย่างง่าย โดยทั่วไปจะใช้ฟังก์ชันซิกมอยด์ (sigmoid function) แล้วส่งค่าไปยังชั้นถัดไป การคำนวณเช่นนี้จะเกิดขึ้นไปเรื่อยๆ ทีละชั้น จนถึงบัพเอาต์พุต โดยในยุคแรก (ราว ค.ศ. 1970) จำนวนชั้นจะถูกกำหนดไว้เป็นค่าคงที่ แต่ในปัจจุบันมีการนำขั้นตอนวิธีเชิงพันธุกรรม มาช่วยออกแบบโครงสร้างของข่ายงาน ดู นิวโรอีโวลูชัน (Neuroevolution)

### 2.1.6 Unsupervised learning

เป็นเทคนิคหนึ่งของการเรียนรู้ของเครื่อง โดยการสร้างโมเดลที่เหมาะสมกับข้อมูล การเรียนรู้แบบนี้แตกต่างจากการเรียนรู้แบบมีผู้สอน คือ จะไม่มีการระบุผลที่ต้องการหรือประเภทไว้ก่อน การเรียนรู้แบบนี้จะพิจารณาวัตถุเป็นเซตของตัวแปรสุ่ม แล้วจึงสร้างโมเดลความหนาแน่นร่วมของชุดข้อมูลการเรียนรู้แบบไม่มีผู้สอนสามารถนำไปใช้ร่วมกับการอนุมานแบบเบย์ เพื่อหาความน่าจะเป็นแบบมีเงื่อนไขของตัวแปรสุ่มโดยกำหนดตัวแปรที่เกี่ยวข้องให้ นอกจากนี้ยังสามารถนำไปใช้ในการบีบอัดข้อมูล ซึ่งโดยพื้นฐานแล้ว ขั้นตอนวิธีการบีบอัดข้อมูลจะขึ้นอยู่กับการแจกแจงความน่าจะเป็นของข้อมูลไม่อย่างชัดแจ้งก็โดยปริยายการเรียนรู้แบบไม่มีผู้สอนในอีกรูปแบบหนึ่งคือการแบ่งกลุ่มข้อมูล โดยจะไม่เกี่ยวข้องกับความน่าจะเป็น นอกจากนี้อาจจะดูได้จาก formal concept analysis

### 2.1.7 Feature Selection

ในการจำแนกประเภทข้อมูล (classification) จะพบว่าแอตทริบิวต์ (attribute) หรือ ฟีเจอร์ (feature) มีจำนวนมาก เช่น การจำแนกประเภทข้อความทัศนคติ (sentiment) ออกเป็นเชิงบวก (positive) หรือเชิงลบ (negative) นั้นจะมีจำนวนคำในข้อความต่างๆ ที่ใช้เป็นฟีเจอร์จำนวนมาก ฟีเจอร์เหล่านี้บางอันก็ไม่ได้มีความสำคัญในการแบ่งแยกคลาส (class) ออกเป็นเชิงบวกหรือเชิงลบได้

ดังนั้นจึงจำเป็นต้องทำการคัดเลือกฟีเจอร์ที่สำคัญมาใช้งาน ขั้นตอนนี้เรียกว่าการคัดเลือกฟีเจอร์ หรือ feature selection ซึ่งสามารถแบ่งได้เป็น 2 กลุ่มใหญ่ดังนี้

Filter approach เป็นการคัดเลือกฟีเจอร์โดยใช้การคำนวณค่าน้ำหนักซึ่งอาจจะเป็นค่าความสัมพันธ์ระหว่างแต่ละฟีเจอร์และคลาสต่างๆ และจะเลือกฟีเจอร์โดยเรียงลำดับตามค่าน้ำหนักที่คำนวณได้แล้วเลือกฟีเจอร์ที่มีค่าน้ำหนักมากกว่าที่ต้องการมาใช้งานต่อไป วิธีการนี้ต่างจากวิธีการ Wrapper ตรงที่วิธีการนี้จะไม่มีการสร้างโมเดลเพื่อคัดเลือกฟีเจอร์ เทคนิคในการคำนวณค่าน้ำหนักของฟีเจอร์ต่างๆ มีหลายวิธีครับ เช่น Information Gain, Chi-Square หรือ Correlation ครับ (ในบทความนี้จะแนะนำวิธีการคัดเลือกฟีเจอร์โดยการคำนวณ Information Gain ก่อนครับ)

Wrapper approach เป็นการคัดเลือกฟีเจอร์ด้วยการสร้างโมเดล (classification model) ขึ้นมาจากเซตของฟีเจอร์ที่กำหนดไว้และวัดประสิทธิภาพการทำงานของโมเดล และเลือกเซตของฟีเจอร์ที่ทำให้โมเดลมีประสิทธิภาพมากที่สุดมาใช้งาน เช่น โมเดลที่ให้ค่าความถูกต้อง (accuracy) มากที่สุด การคัดเลือกฟีเจอร์ด้วยวิธีการนี้แบ่งย่อยได้เป็น 2 แบบใหญ่ๆ คือ

Forward Selection เป็นการสร้างโมเดลโดยการเพิ่มฟีเจอร์ทีละ 1 ฟีเจอร์ ถ้าฟีเจอร์ที่ใส่เพิ่มให้ประสิทธิภาพที่ดีก็จะเก็บไว้และเลือกฟีเจอร์อื่นๆ มาเพิ่มต่อไปจนประสิทธิภาพของโมเดลไม่ได้ดีขึ้นก็จะหยุดทำงาน

Backward Elimination เป็นการสร้างโมเดลที่เริ่มจากการใช้ฟีเจอร์ทั้งหมดก่อนและตัด (eliminate) ฟีเจอร์ที่ไม่สำคัญทิ้งไปทีละฟีเจอร์ถ้าประสิทธิภาพดีขึ้นก็ตัดฟีเจอร์อื่นๆ ต่อไป

โดยในงานวิจัยครั้งนี้ได้เลือกใช้วิธีการ Wrapper approach เพื่อดำเนินงานวิจัยให้สำเร็จ

#### 2.1.8 Model Evaluation

การทำ Evaluation หรือ การประเมินผลว่า Model ของเรานั้นดีแค่ไหนนั้นเป็นสิ่งที่สำคัญมากๆ โดยเฉพาะอย่างยิ่งคนที่เตรียม model ไปใช้งานจริง โดยตัวหลักที่มักนำมาใช้กัน

Mean Square Error: หรือค่าผิดพลาดกำลังสองเฉลี่ย จะนิยมใช้กับงาน Regression

Accuracy: ค่าความแม่นยำ มักจะนิยมใช้กับงาน Classification

#### 2.1.9 Ensemble

เทคนิค Ensemble เป็นเทคนิคที่ใช้โมเดล classification หลายๆ โมเดล (model) มาช่วยในการหาคำตอบ เทคนิคนี้ได้เป็นเทคนิคที่มีประสิทธิภาพสูงซึ่งพบได้จากการแข่งขันต่างๆ สำหรับการทำนายด้วยเทคนิค Ensemble ซึ่งมีหลายๆ โมเดลนี้ แต่ละโมเดลก็จะให้คำตอบออกมา ในขั้นตอนสุดท้ายเราจะต้องนำคำตอบเหล่านี้มารวมกันเพื่อดูว่าคำตอบไหนเหมาะสมที่สุด โดยอาจจะใช้วิธีการโหวต (vote) เลือกคำตอบที่ตอบตรงกันมากที่สุดก็ได้ครับ หลักการสร้างโมเดล Ensemble คือโมเดลที่สร้างควรจะมีหลากหลายเพื่อให้ทำนายข้อมูลแบบต่างๆ กันได้มาก การสร้างโมเดลที่หลากหลายนี้ อาจจะทำได้โดยการใช้เทคนิค classification หลายๆ ประเภท หรือ การสร้างเทรนนิ่ง data ที่มีลักษณะต่างๆ กัน เช่น มีตัวอย่างต่างๆ กัน หรือมีแอตทริบิวต์ต่างๆ กัน โดยในงานวิจัยนี้ได้

เลือกใช้ Vote Ensemble วิธีการนี้เป็นการนำเทรนนิ่ง ดาต้าชุดเดียวกัน แต่เพื่อให้โมเดลมีความหลากหลายมากขึ้นจึงเลือกที่จะสร้างโมเดลด้วยเทคนิค classification ที่ต่างกัน หลังจากที่เราสร้างโมเดล Ensemble ด้วยหลากหลายเทคนิคได้แล้ว ขั้นตอนถัดไป คือ การนำโมเดลที่สร้างได้ไปทำนายข้อมูลใหม่ จากนั้นจึงดูผลลัพธ์คำตอบจากโมเดลเหล่านั้นเพื่อวัดว่าคำตอบควรจะเป็นอะไร

## 2.2 Data Preprocessing

จะเป็นขั้นตอนในการเตรียมข้อมูลที่จะนำไปประมวลผลขั้นตอนต่อไปซึ่ง ขั้นตอนนี้จะทำให้ข้อมูลดิบ หรือ Data set (มากกว่าหนึ่ง set ก็ได้) ที่เราได้มาตอนเริ่มต้น มีความถูกต้องแม่นยำ และมีขนาดตรงตามที่เรากำลังต้องการมากยิ่งขึ้นโดย กระบวนการเตรียมข้อมูลนั้น จะมีกระบวนการย่อยๆ ดังต่อไปนี้

### 2.2.1 Data Cleaning

การทำความสะอาดข้อมูล เป็นกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล เพราะหมายถึงความไม่สมบูรณ์ ความไม่ถูกต้อง ความไม่สัมพันธ์กับข้อมูลอื่นๆ เป็นต้น จึงต้องมีการแทนที่ การปรับปรุง หรือการลบข้อมูลที่ไม่ถูกต้องเหล่านั้นออกไป เพื่อให้ข้อมูลมีคุณภาพ

การทำความสะอาดข้อมูล เกิดขึ้นเนื่องจาก มีความไม่สอดคล้องของข้อมูล ซึ่งอาจเกิดจากข้อผิดพลาดของการบันทึกข้อมูล การส่งข้อมูล หรือการให้ความหมายของข้อมูลที่จัดเก็บแตกต่างกัน ยังต้องมีการบูรณาการกับฐานข้อมูลอื่นๆ เช่น คลังข้อมูล หรือหลายฐานข้อมูล จึงมีโอกาสมากที่จะเกิด "ข้อมูลที่ไม่สะอาด" ขึ้น

### วิธีการจัดการค่าที่ขาดหาย

1. ตัดทิ้งรายการที่มีข้อมูลสูญหาย นิยมใช้กับการทำเหมืองข้อมูลแบบจำแนกประเภท (Classification) ในกรณีที่ค่าคุณลักษณะขาดหายไปเป็นจำนวนมาก

2. เติมค่าที่ขาดหายด้วยมือ วิธีนี้ไม่เหมาะสมกรณีที่ชุดข้อมูลมีขนาดใหญ่ และมีข้อมูลขาดหายจำนวนมาก

3. เติมค่าคุณลักษณะของข้อมูลที่ขาดหายทุกค่า ด้วยค่าคงที่ค่าหนึ่งเช่น ไม่รู้ค่า หรือ unknown

4. ใช้ค่าเฉลี่ยของคุณลักษณะ เติมค่าข้อมูลที่ขาดหาย เช่น ถ้าทราบว่าคุณค่าที่รายได้เฉลี่ยเดือนละ 12,000 บาท จะใช้ค่านี้นั้นแทนค่ารายได้ของลูกค้าที่ขาดหาย

5. ใช้ค่าเฉลี่ยคุณลักษณะของตัวอย่างที่จัดอยู่ในประเภทเดียวกัน เพื่อเติมค่าข้อมูลที่ขาดหาย เช่น เติมค่ารายได้ของลูกค้าที่ขาดหาย ด้วยค่าเฉลี่ยของลูกค้าที่อยู่ในกลุ่มอาชีพ เดียวกัน

6. ใช้ค่าที่เป็นไปได้มากที่สุด เติมแทนค่าข้อมูลที่ขาดหาย เช่น ค่าที่ได้จากสมการความถดถอย (Regression) ค่าที่ได้จากการอนุมาน โดยใช้สูตรของเบย์ (Bayesian formula) หรือต้นไม้ตัดสินใจ

(Decision tree) เช่น ใช้ข้อมูลลูกค้า มาสร้างต้นไม้ตัดสินใจ เพื่อทำนายรายได้ของลูกค้า แล้วนำไปแทนค่าที่ขาดหาย วิธีนี้นิยมกันแพร่หลาย เนื่องจากทำนายค่าข้อมูลที่ขาดหาย โดยพิจารณาจากค่าของข้อมูลชุดปัจจุบัน และความสัมพันธ์ระหว่างคุณลักษณะในชุดข้อมูล

ข้อมูลรบกวน (Noisy data)

ข้อมูลรบกวน เป็นข้อมูลที่มีความผิดพลาดแบบสุ่ม หรือ คลาดเคลื่อนจากการวัด สาเหตุของความผิดพลาดอาจสืบเนื่องมาจาก

- อุปกรณ์เก็บรวบรวมข้อมูลท านหน้าที่ผิดพลาด
- ปัญหาการบันทึกหรือป้อนค่าข้อมูล
- ปัญหาการส่งข้อมูล (data transmission) ผิดพลาด
- ข้อจำกัดทางเทคโนโลยี เช่น ข้อจ ากัดของขนาดบัพเพอร์

การจัดการข้อมูลรบกวน

1. Binning Methods การปรับข้อมูลให้เรียงด้วยวิธีการแบบ binning ท านโดย เรียงล าดับข้อมูล แล้วใช้หลักการตัดแบ่ง (partition) แบ่ง ข้อมูลออกเป็นส่วนแต่ละส่วนเรียกว่า bin แล้วท านการปรับ เรียบข้อมูลในแต่ละ bin โดยใช้การปรับเรียบข้อมูลแบบ ท้องถิ่น (local smoothing) โดยใช้ค่าที่ได้จากเพื่อนบ้าน ใกล้เคียง (neighborhood) ใน bin หรือ bucket เดียวกัน เช่น ค่าเฉลี่ยของ bin (bin means) ค่ากลางของ bin (bin medians) หรือค่าขอบของ bin (bin boundaries)

2. Regression วิธีความถดถอย ใช้การทำนายค่าของตัวแปรจากสมการความถดถอยที่หาได้ ด้วยวิธีความผิดพลาดน้อยที่สุด (Least square error) จากชุดตัวอย่างตัวแปร สมการความถดถอย แบ่งเป็นความสัมพันธ์ระหว่าง 2 ตัวแปร หรือมากกว่า ได้แก่

- ความถดถอยเชิงเส้น (Linear Regression) เป็นวิธีการอธิบาย (explanatory) โดย มีตัวแปรอื่นที่สัมพันธ์กับตัวแปรที่เราต้องการพยากรณ์ และสามารถสร้าง model เพื่อแสดงความสัมพันธ์นั้นได้

- ความถดถอยพหุเชิงเส้น (Multiple Linear Regression)

3. Clustering การวิเคราะห์การจัดกลุ่ม จะช่วยตรวจหา และกำจัดค่าที่ผิดปกติเช่น ค่าที่สูงหรือต่ำกว่าปกติ จะไม่สามารถจัดอยู่ในกลุ่มใดๆและตกอยู่นอกกลุ่ม

### 2.2.2 Data Integrate

เมื่อข้อมูลเข้ามามีบทบาทในการทำงานขององค์กรมากขึ้น จึงมีความจำเป็นในการนำข้อมูลที่มีอยู่มาใช้ในหลายๆ ทาง เพื่อให้เกิดประโยชน์สูงสุด ปัญหาเกิดขึ้นเมื่อองค์กรใหญ่องค์กรหนึ่งๆ ซึ่งประกอบไปด้วยหลายๆ แผนกแต่ละแผนกเก็บข้อมูลที่เกี่ยวข้องกับแผนกของตนซึ่งอาจใช้ระบบจัดการฐานข้อมูลแตกต่างกันไปโครงสร้างของข้อมูลก็อาจแตกต่างกัน หรือแม้ กระทั่งใช้ชื่อเดียวกัน แต่หมายถึงของคนละอย่างหรือของอย่างเดียวกัน แต่ใช้ชื่อเรียกต่างกันในแต่ละแผนก สิ่งเหล่านี้ส่งผลให้การนำข้อมูลของบริษัทมาใช้โดยรวมต้องมีการจัดสร้างคลังข้อมูล (Data Warehousing) ขึ้นมา เพื่อที่จะทำสำเนา ข้อมูลจากฐานข้อมูลของแต่ละแผนก และมีการ

เปลี่ยนแปลงที่เหมาะสมและถูกต้องมาใช้พัฒนาสร้างฐานข้อมูลของส่วนกลางขึ้นซึ่งคลังข้อมูลนี้จะมีการปรับปรุงตามฐานข้อมูลย่อยของแต่ละแผนกแต่อาจไม่จำเป็นที่จะต้องปรับทันทีทันใดที่ข้อมูลถูกเปลี่ยนแปลง อาจจะแก้ไขข้อมูลทุกๆ คิน ซึ่งเป็นช่วงเวลาที่ไม่ค่อยมีการใช้งานฐานข้อมูล โดยคลังข้อมูลนี้ส่งผลให้ข้อมูลโดยรวมของบริษัท มีโครงสร้างเดียวกัน ทำให้นำมาใช้ในการโฆษณา ประชาสัมพันธ์ หรือประเมินผลงานของแต่ละแผนกได้อย่างมีประสิทธิภาพ และส่งผลให้การวางแผนและวิเคราะห์งานดีขึ้นอีกด้วย

ส่วนเหมืองข้อมูล (Data Mining) เป็นวิธีการค้นหารูปแบบหรือโครงสร้างของข้อมูล ที่ผู้ใช้ไม่สามารถเห็น หรือเข้าใจได้ในทันทีจากข้อมูลที่มีอยู่ หรือกล่าวได้ว่า เป็นเทคนิคเพื่อค้นหาความสัมพันธ์ในข้อมูลที่ใช้ไม่ทราบมาก่อน เช่น การค้นหารายชื่อกลุ่มลูกค้าที่สนใจสินค้าประเภทเดียวกัน จากฐานข้อมูลลูกค้าและการสั่งซื้อของเป็นต้น ทั้งนี้ เพื่อนำไปใช้ในการตัดสินใจ เพื่อให้เกิดประโยชน์ต่อการเพิ่มประสิทธิภาพการใช้งานข้อมูลกลุ่มนั้นๆ เทคโนโลยีที่นำมาใช้ในเหมืองข้อมูลมีมากมาย เช่น ระบบฐานข้อมูลการรู้จำรูปแบบข้อมูล โครงข่ายประสาทเทียม (Artificial Neural Networks) ฟัซซีลอจิก (Fuzzy Logic) ขั้นตอนวิธีทางยีน (Genetic Algorithms) เป็นต้น Data Integrate เป็นการรวบรวมข้อมูลจากแหล่งที่เก็บต่างๆ มาไว้ที่เดียวกัน การผสมผสานข้อมูลจากแหล่งต่างๆ เพื่อช่วย

1. ลดหรือหลีกเลี่ยงความซ้ำซ้อนของข้อมูล (Data Redundancies) ซึ่งจะนำไปสู่ปัญหาความไม่สอดคล้องกันของข้อมูล (Data inconsistencies)
2. เพิ่มความเร็วและคุณภาพในการทำเหมืองข้อมูล

#### Process of Data Integration

การผสมผสานโครงสร้างการเก็บข้อมูล (Schema Integration) โดยใช้ metadata ช่วยในการบ่งชี้ entities ในแหล่งเก็บข้อมูลต่างๆ เช่น คุณลักษณะ Cusid ในแหล่งข้อมูล A เป็นคุณลักษณะเดียวกับ CustNumber ในแหล่งข้อมูล B หรือไม่

ตรวจหาและแก้ไข ค่าข้อมูลที่ขัดแย้ง เช่น ค่าคุณลักษณะเดียวกัน แต่ใช้ หน่วยวัดต่างกัน ในแต่ละแหล่งข้อมูล

การกำจัดค่าข้อมูลซ้ำซ้อน เช่น ค่าคุณลักษณะในแหล่งข้อมูลหนึ่ง อาจเป็นค่าที่ได้จากการคำนวณในอีกแหล่งข้อมูลหนึ่ง อาทิ ค่ารายได้รายปี

ค่าสหสัมพันธ์ เป็น 0 หมายถึง ตัวแปรทั้งสองเป็นอิสระ ไม่มี ความสัมพันธ์ ระหว่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 2.2.3 Data Transformation

การแปลงข้อมูลที่พบบ่อยในการทำเหมืองข้อมูลคือ การทำนอร์มอลไลซ์(Normalization) โดยแปลงค่าข้อมูลให้อยู่ในช่วงสั้นๆ ที่อัลกอริทึมการทำเหมืองข้อมูลสามารถนำไปใช้ประมวลผลได้ วิธีการทำนอร์มอลไลซ์ข้อมูลได้แก่

1.Min-Max Normalization เป็นการแปลงข้อมูลเชิงเส้นจากช่วงที่เป็นไปได้เดิมของค่าอินพุต ให้เป็นช่วงข้อมูลใหม่ที่กำหนด ปกติ ช่วง 0-1 สมมติให้  $v$  คือค่าคุณลักษณะเดิม  $v'$  คือค่าคุณลักษณะใหม่  $min_A$  ,  $max_A$  คือค่าต่ำสุดและสูงสุดเดิมของคุณลักษณะ  $A$   $new\_min_A$  ,  $new\_max_A$  คือค่าต่ำสุดและสูงสุดใหม่ของคุณลักษณะ  $A$  ได้สูตร

$$v' = \frac{v - min_A}{max_A - min_A} (new\_max_A - new\_min_A) + new\_min_A.$$

รูปที่ 2.6 สูตรในการคำนวณ min-max normalization

Z-Score เป็นการปรับการกระจายของข้อมูลให้มีค่าเท่ากับ 0 และค่าเบี่ยงเบน มาตรฐานเท่ากับ 1 หาได้จากสูตร

$$v' = \frac{v - \bar{A}}{\sigma_A},$$

where  $\bar{A}$  and  $\sigma_A$  are the mean and standard deviation, respectively, of attribute  $A$ .

รูปที่ 2.7 สูตรในการคำนวณ Z-score

Decimal scaling เป็นการแปลงค่าข้อมูลเดิมให้เป็นเลขทศนิยม ตำแหน่งทศนิยมกำหนดโดยค่าสัมบูรณ์ที่มีค่ามากที่สุด เช่น ค่าที่เป็นไปได้ของคุณลักษณะ  $A$  อยู่ในช่วงระหว่าง -999 ถึง 997 จะได้ว่าค่าสัมบูรณ์ที่มีค่ามากที่สุด คือ  $|-999| = 999$  ดังนั้นเราจะหารข้อมูลแต่ละค่าด้วย 1000 ผลลัพธ์คือ ค่า -999 จะถูกแปลงเป็น -0.999

### 2.2.4 Data Reduction

คลังข้อมูลเก็บข้อมูลจำนวนมาก ขนาดเป็น เทราไบต์ ดังนั้น การวิเคราะห์ หรือการทำเหมืองข้อมูลที่ซับซ้อน ต้องใช้ เวลารนานในการประมวลผลข้อมูลปริมาณมากๆ

การลดรูป เป็นการกระทำเพื่อแทนข้อมูล ด้วยรูปแบบ กะทัดรัด กินเนื้อที่น้อยกว่า แต่สามารถนำไปใช้วิเคราะห์ แล้วเกิดผลลัพธ์ที่เท่ากันหรือเทียบเท่ากัน กับการวิเคราะห์จาก ข้อมูลทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.3 Web Application

การพัฒนาระบบงานบนเว็บ ซึ่งมีระบบมีการไหลเวียนในแบบ Online ทั้งแบบ Local ภายในวง LAN และ Global ออกไปยังเครือข่ายอินเทอร์เน็ต ทำให้เหมาะสำหรับงานที่ต้องการข้อมูลแบบ Real Time การทำงานของ Web Application นั้นโปรแกรมส่วนหนึ่งจะวางตัวอยู่บน Rendering Engine (เร็นเดอร์ริงเอนจิน) ซึ่งตัว Rendering Engine จะทำหน้าที่หลักๆ คือนำเอาชุดคำสั่งหรือรูปแบบโครงสร้างข้อมูลที่ใช้ในการแสดงผล นำมาแสดงผลบนพื้นที่ส่วนหนึ่งในจอภาพ โปรแกรมส่วนที่วางตัวอยู่บน Rendering Engine จะทำหน้าที่หลักๆ คือการเปลี่ยนแปลงแก้ไขสิ่งที่แสดงผล จัดการตรวจสอบข้อมูลที่รับเข้ามาเบื้องต้นและการประมวลผลบางส่วนแต่ส่วนการทำงานหลักๆ จะวางตัวอยู่บนเซิร์ฟเวอร์ ในลักษณะ Web Application แบบเบื้องต้น ฝั่งเซิร์ฟเวอร์จะประกอบไปด้วยเว็บเซิร์ฟเวอร์ซึ่งทำหน้าที่เชื่อมต่อกับไคลเอนต์ตามโปรโตคอล HTTP/ HTTPS (เอช ที ที พี / เอช ที ที พี เอส) โดยนอกจากเว็บเซิร์ฟเวอร์จะทำหน้าที่ส่งไฟล์ที่เกี่ยวข้องกับการแสดงผล ตามมาตรฐาน HTTP ตามปกติทั่วไปแล้ว เว็บเซิร์ฟเวอร์จะมีส่วนประมวลผลซึ่งอาจจะเป็นตัวแปลภาษา เช่น Script Engine ของภาษา PHP หรืออาจจะมีการติดตั้ง .NET Framework (ดอทเน็ต เฟรมเวิร์ก) ซึ่งมีส่วนแปลภาษา CLR (ซี แอล อาร์) ที่ใช้แปลภาษา intermediate (อินเทอร์เมดิเอท) จากโค้ดที่เขียนด้วย VB.NET (วี บี ดอทเน็ต) หรือ C#.NET (ซีฉับ ดอทเน็ต) หรืออาจจะเป็น J2EE (เจ ทู อี อี) ที่มีส่วนแปลไปดัดโค้ดของคลาสที่ได้จากโปรแกรมภาษาจาวา เป็นต้น

### 2.3.1 HTML

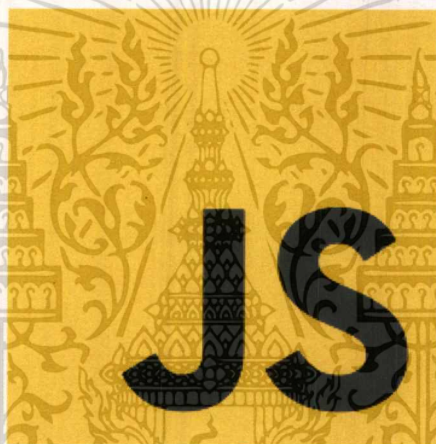


รูปที่ 2.8 Logo ของ HTML

ภาษาหลักที่ใช้ในการเขียนเว็บเพจ โดยใช้ Tag ในการกำหนดการแสดงผล HTML ย่อมาจากคำว่า Hypertext Markup Language โดย Hypertext หมายถึง ข้อความที่เชื่อมต่อกันผ่านลิงก์ (Hyperlink) Markup language หมายถึงภาษาที่ใช้ Tag ในการกำหนดการแสดงผลสิ่งต่างๆที่แสดงเอกสารอยู่บนเว็บเพจ ดังนั้น HTML จึงหมายถึง ภาษาที่ใช้ Tag ในการกำหนดการแสดงผลเว็บเพจที่ต่างก็การคำไม่ว่าเชื่อมถึงกันใน Hyperspace ผ่าน Hyperlink นั่นเอง

HTML เริ่มขึ้นเมื่อปี 1980 เมื่อ Tim Berners Lee เสนอต้นแบบสำหรับนักวิจัยใน CERN เพื่อแลกเปลี่ยนเอกสาร ข้อมูลด้านการวิจัย โดยใช้ชื่อว่า Enquire ในปี 1990 ได้เขียนโปรแกรมเบราเซอร์ และทดลองรันบนเซิร์ฟเวอร์ที่เค้าพัฒนาขึ้น HTML ได้รับการรู้จักจาก HTML Tag ซึ่งมีอยู่ 18 Tag ในปี 1991 HTML ถูกพัฒนาจาก SGML และ Tim ก็คิดเสมือนว่า HTML เป็นโปรแกรมย่อยของ SGML อยู่ในตอนนั้น ต่อมาในปี 1996 เพื่อกำหนดมาตรฐานให้ตรงกัน W3C World Wide Web Consortium จึงเป็นผู้กำหนดสเปกทั้งหมดของ HTML และปี 1999 HTML 4.01 ก็ถือกำเนิดขึ้น โดยมี HTML 5 ซึ่งเป็น Web Hypertext Application ถูกพัฒนาต่อมาในปี 2004 นอกจากนี้ยังมีการพัฒนาไปเป็น XHTML ซึ่ง คือ Extended HTMLซึ่งมีความสามารถและมาตรฐานที่รัดกุมกว่าอีกด้วย โดยอยู่ภายใต้การควบคุมของ W3C (World Wide Web Consortium)

### 2.3.2 Javascript



รูปที่ 2.9 Logo ของ Javascript

ภาษาคอมพิวเตอร์สำหรับการเขียนโปรแกรมบนระบบอินเทอร์เน็ตที่กำลังได้รับความนิยมอย่างสูง JavaScript เป็น ภาษาสคริปต์เชิงวัตถุ (ที่เรียกกันว่า "สคริปต์" (script) ซึ่งในการสร้างและพัฒนาเว็บไซต์ (ใช้ร่วมกับ HTML) เพื่อให้เว็บไซต์ของเราดูมีการเคลื่อนไหว สามารถตอบสนองผู้ใช้งานได้มากขึ้น ซึ่งมีวิธีการทำงานในลักษณะ "แปลความและดำเนินงานไปที่ละคำสั่ง" (interpret) หรือเรียกว่า อ็อบเจ็กต์โอเรียนเตด (Object Oriented Programming) ที่มีเป้าหมายในการ ออกแบบและพัฒนาโปรแกรมในระบบอินเทอร์เน็ต สำหรับผู้เขียนด้วยภาษา HTML สามารถทำงานข้ามแพลตฟอร์มได้ โดยทำงานร่วมกับ ภาษา HTML และภาษา Java ได้ทั้งทางฝั่งไคลเอนต์ (Client) และทางฝั่งเซิร์ฟเวอร์ (Server)

JavaScript ถูกพัฒนาขึ้นโดย เน็ตสเคปคอมมิวนิเคชันส์ (Netscape Communications Corporation) โดยใช้ชื่อว่า Live Script ออกมาพร้อมกับ Netscape Navigator2.0 เพื่อใช้สร้างเว็บเพจโดยติดต่อกับเซิร์ฟเวอร์แบบ Live Wire ต่อมาเน็ตสเคปจึงได้ร่วมมือกับ บริษัทซันไมโครซิสเต็มส์

ปรับปรุงระบบของบราวเซอร์เพื่อให้สามารถติดต่อกับภาษาจาวาได้ และได้ปรับปรุง LiveScript ใหม่เมื่อ ปี 2538 แล้วตั้งชื่อใหม่ว่า JavaScript JavaScript สามารถทำให้ การสร้างเว็บ

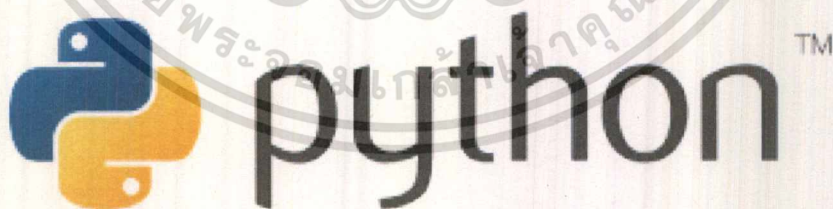
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปตีประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพจ มีลูกเล่น ต่าง ๆ มากมาย และยังสามารถโต้ตอบกับผู้ใช้ได้อย่างทันที เช่น การใช้เมาส์คลิก หรือ การกรอกข้อความในฟอร์ม เป็นต้น

เนื่องจาก JavaScript ช่วยให้ผู้พัฒนา สามารถสร้างเว็บเพจได้ตรงกับความต้องการ และมีความ น่าสนใจมากขึ้น ประกอบกับเป็นภาษาเปิด ที่ใครก็สามารถนำไปใช้ได้ ดังนั้นจึงได้รับความนิยมเป็น อย่างสูง มีการใช้งานอย่างกว้างขวาง รวมทั้งได้ถูกกำหนดให้เป็นมาตรฐานโดย ECMA การทำงานของ JavaScript จะต้องมีการแปลความคำสั่ง ซึ่งขั้นตอนนี้จะถูกจัดการโดยบราวเซอร์ (เรียกว่าเป็น client-side script) ดังนั้น JavaScript จึงสามารถทำงานได้ เฉพาะบนบราวเซอร์ที่สนับสนุน ซึ่ง ปัจจุบันบราวเซอร์เกือบทั้งหมดก็สนับสนุน JavaScript แล้ว อย่างไรก็ตาม สิ่งที่ต้องระวังคือ JavaScript มีการพัฒนาเป็นเวอร์ชันใหม่ๆออกมาด้วย (ปัจจุบันคือรุ่น 1.5) ดังนั้น ถ้านำโค้ดของเวอร์ชันใหม่ ไป รันบนบราวเซอร์รุ่นเก่าที่ยังไม่สนับสนุน ก็อาจจะทำให้เกิด error ได้

การทำงานของ JavaScript เกิดขึ้นบนบราวเซอร์ (เรียกว่าเป็น client-side script) ดังนั้น ไม่ว่าคุณจะใช้เซิร์ฟเวอร์อะไร หรือที่ไหน ก็ยังคงสามารถใช้ JavaScript ในเว็บเพจได้ ต่างกับภาษา สคริปต์อื่น เช่น Perl, PHP หรือ ASP ซึ่งต้องแปลความและทำงานที่ตัวเครื่องเซิร์ฟเวอร์ (เรียกว่า server-side script) ดังนั้นจึงต้องใช้บนเซิร์ฟเวอร์ ที่สนับสนุนภาษาเหล่านี้เท่านั้น อย่างไรก็ตาม จาก ลักษณะดังกล่าวก็ทำให้ JavaScript มีข้อจำกัด คือไม่สามารถรับและส่งข้อมูลต่างๆ กับเซิร์ฟเวอร์ โดยตรง เช่น การอ่านไฟล์จากเซิร์ฟเวอร์ เพื่อนำมาแสดงบนเว็บเพจ หรือรับข้อมูลจากผู้ชม เพื่อนำไป เก็บบนเซิร์ฟเวอร์ เป็นต้น ดังนั้นงานลักษณะนี้ จึงยังคงต้องอาศัยภาษา server-side script อยู่ (ความจริง JavaScript ที่ทำงานบนเซิร์ฟเวอร์เวอร์ก็มี ซึ่งต้องอาศัยเซิร์ฟเวอร์ที่สนับสนุนโดยเฉพาะ เช่นกัน แต่ไม่เป็นที่นิยมนัก)

### 2.3.3 Python



รูปที่ 2.10 Logo ของ Python

Python เป็นภาษาที่ใช้ในการเขียนโปรแกรมภาษาหนึ่งที่ได้รับความนิยมในการใช้งาน เนื่องจากมี library ต่างๆที่ครอบคลุมการใช้งาน โดยเฉพาะการใช้งานในเรื่อง Data Analysis และ machine learning

เอกสารนี้เป็นเอกสาร Python เป็นภาษาที่ถูกออกแบบมาให้ เป็นภาษาที่อ่านง่ายและเข้าใจได้ง่าย ออกแบบ การค้า ไม่ทำกำไร โครงสร้างไม่ซับซ้อน และเป็น Open source หมายความว่า ทุกคนสามารถนำภาษา Python ไปใช้

พัฒนาได้โดยไม่เสียค่าใช้จ่ายใดๆ และยังทำให้ทุกคนมีส่วนร่วมในการพัฒนาภาษา Python ให้ได้ขึ้นได้ด้วย

Python เป็นภาษากาว (Glue Language) ได้อย่างดีเนื่องจากสามารถเรียกใช้ภาษาโปรแกรมอื่น ๆ ได้หลายภาษา ทำให้เหมาะที่จะใช้เขียนเพื่อประสานงานโปรแกรมที่เขียนในภาษาต่างกันได้

### 2.3.4 Bootstrap CSS Framework



รูปที่ 2.11 Logo ของ Bootstrap

Bootstrap เป็น Front-end Framework ที่ประกอบด้วยโครงสร้าง CSS , HTML และ JavaScript ที่ช่วยให้เราสามารถสร้างหน้าจอ User Interface ได้ง่าย และ สวยงาม และรวดเร็ว ลดเวลาในการที่จะมานั่งออกแบบ Design หน้าจอ layout หรือรายการ Element อื่นๆ ที่เกี่ยวข้องกับหน้า Form ทั้งหมด ก็สามารถใช้ Bootstrap เข้ามาจัดการได้ทั้งหมด สำหรับ Bootstrap เองมีทั้ง CSS Component และ JavaScript Plugin ที่ทำงานร่วมกับ jQuery ที่สามารถเรียกใช้งานได้มากมาย และที่สำคัญคือ Bootstrap มีการแสดงผลในรูปแบบของ Responsive ซึ่งจะแสดงผลสภาพแวดล้อมที่เหมาะสมกับอุปกรณ์ที่ทำการเปิดอยู่ในขณะนั้น เช่น PC Desktop , Tablets , Mobile หรืออุปกรณ์อื่นๆ ที่เกี่ยวข้อง ซึ่งทำให้เรานั้นออกแบบเว็บและเขียนคำสั่งต่างๆ ก็สามารถที่จะรองรับอุปกรณ์ทั้งหมดได้เลย

ปัจจุบันมีการพัฒนาขึ้นด้วยกลุ่มนักพัฒนาจากทั่วทุกหนแห่งในโลก มีการอัปเดตแก้ไข bug อยู่ตลอดเวลา เพื่อรองรับการทำงานได้อย่างไม่มีปัญหาพร้อมๆ กับการพัฒนาเทคโนโลยีบน Web Browser เช่น CSS , HTML และในปัจจุบันเราจะเห็นว่า Bootstrap มีตัวอย่างธีมให้ดาวน์โหลดมากมาย มีทั้งที่สามารถดาวน์โหลดใช้ได้ฟรี และ แบบเสียเงินซื้อ และหลายๆ เว็บก็เลือกที่จะใช้ Bootstrap กันมากขึ้น สำหรับแอดมินเองก็มีการนำ Bootstrap มาใช้กับ Project ใหญ่ๆ หลายตัวแล้ว และที่ผ่านมาก็สามารถตอบใจของลูกค้าได้เป็นอย่างดี

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามทำให้อัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โครงสร้างของ Bootstrap Framework

- Scaffolding grid system จำนวน 12 คอลัมน์ สามารถเลือกใช้ได้ทั้งแบบ fixed และแบบ fluid เป็นโครงสร้างของ Layout ที่จะแสดงผลในหน้าจอ ซึ่งจำนวน Column นี้จะแสดงผลตามความกว้างของแต่ละอุปกรณ์ที่เรียกใช้งาน
- Base CSS style sheets สำหรับ html elements พื้นฐาน เช่น typography, tables, forms และ images เป็น Stylesheet พื้นฐานที่เราสามารถเรียกใช้งานได้เลย เช่น Button ที่อยู่ในรูปแบบของ สีต่างๆ การแสดงรูปภาพ ตาราง และอื่นๆ
- Components style sheets สำหรับสิ่งที่เราต้องใช้อย่างน้อย ไม่ว่าจะเป็น navigation, breadcrumbs รวมไปถึง pagination เป็นโครงสร้างพื้นฐานของ Bootstrap ที่ไว้จัดการ Menu , Navigation ซึ่งจะแปรผันกับขนาดของหน้าจอ ของอุปกรณ์ที่เรียกใช้งาน JavaScript jQuery plugins ต่างๆ ไม่ว่าจะเป็น modal, carousel หรือ tooltip ช่วยในการสร้าง Popup, Dialog , Tooltip ต่างๆ ซึ่งบอกได้เลยว่าเรียกใช้งานได้ง่ายมาก

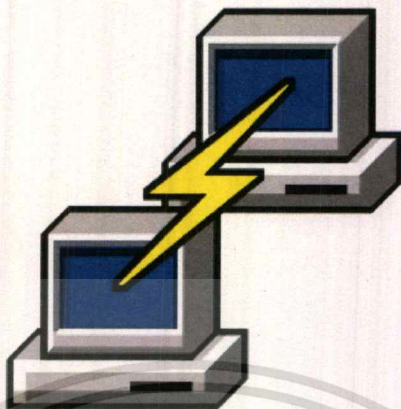
ถึงแม้ว่า Bootstrap จะมีโครงสร้างพื้นฐานที่บังคับให้การออกแบบเป็นไปตาม Framework ออกแบบและที่มีมาให้ แต่เราก็สามารถที่จะเขียนพวก CSS และ Stylesheet เพิ่มเติม เพื่อเข้าไปจัดการกับ UI ต่างๆ ที่ต้องการได้ แต่ทั้งนี้จะต้องให้เข้าใจโครงสร้างของมันซะก่อน มิฉะนั้นเมื่อนำไปใช้งานกับขนาดของอุปกรณ์ต่างๆ อาจจะมีปัญหาในการแสดงผลได้

### 2.3.5 Amazon Web Services(AWS)

บริการ AWS Cloud แบบครบวงจร ตั้งแต่ให้คำปรึกษาและวางแผนค่าใช้จ่าย รวมถึงบริการติดตั้งและดูแลระบบ โดยมีความชำนาญในการจัดทำโซลูชันให้ตอบโจทย์ และครอบคลุมทุกความต้องการ พร้อมวางแผนกลยุทธ์ ออกแบบ ติดตั้ง และดูแลระบบตลอด 24 ชม. เพื่อให้ทุกองค์กรสามารถใช้เทคโนโลยีคลาวด์ได้อย่างมีประสิทธิภาพและคุ้มค่า ที่การันตีด้วยใบรับรองสำหรับ Partner จาก AWS ถึง 3 ด้าน ได้แก่ Authorized Commercial Reseller, Consulting Partner และ Advanced Technology Partner

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.3.6 Putty



รูปที่ 2.12 Logo ของ PuTTY

PuTTY เป็นโปรแกรม Remote Server หรือ SSH ( Secure Shell ) พุดให้เข้าใจง่ายๆก็คือ เราสามารถใช้โปรแกรมนี้ในการ สั่งงาน Server ด้วย command line โดยส่วนใหญ่แล้วจะใช้เชื่อมต่อไปยัง server ที่เป็น linux รองรับการทำงานต่อหลากหลายรูปแบบดังนี้

- Raw
- Telnet
- Rlogin
- SSH
- Serial

## 2.3.7 Filezilla

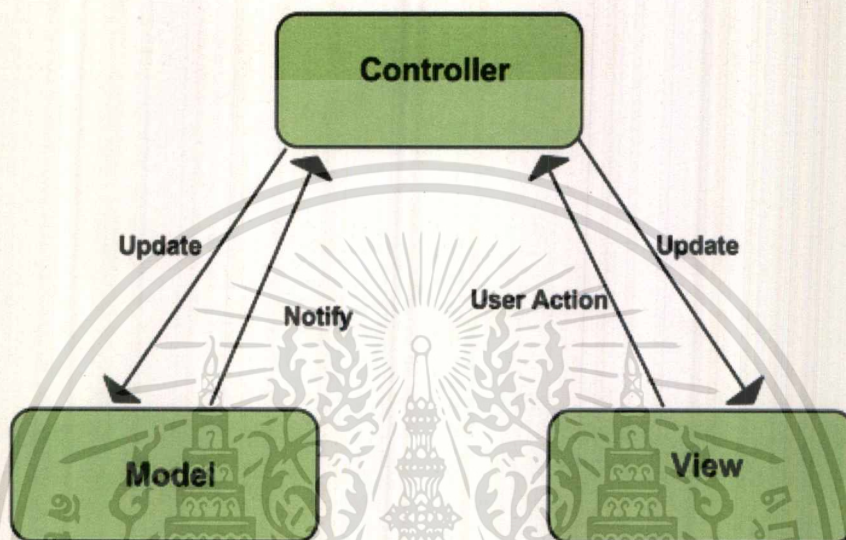


รูปที่ 2.13 Logo ของ Filezilla

โปรแกรมที่ใช้ในการถ่ายโอนไฟล์ระหว่างเครื่องคอมพิวเตอร์ของเรากับเว็บเซิร์ฟเวอร์ เป็นโปรแกรมที่สำคัญมาก ๆ ในการทำเว็บ เราเรียกกระบวนการนี้ว่า FTP ถ้าพูดถึงโปรแกรม FTP ก็จะ

หมายถึงโปรแกรมจำพวกนี้ ซึ่งมีหลายโปรแกรม แต่ในที่นี้เราจะแนะนำการใช้งาน FileZilla เนื่องจากเป็นที่นิยมของผู้ใช้งานและโฮสต์ไทยส่วนใหญ่ก็แนะนำให้ใช้โปรแกรมนี้

## 2.4 MVC (Model Controller View)



รูปที่ 2.14 โครงสร้างสถาปัตยกรรมแบบ Model View Controller

Model-View-Controller (MVC) คือ สถาปัตยกรรมซอฟต์แวร์ชนิดหนึ่ง ซึ่งในขณะนี้ถือว่าเป็นแบบแผนสถาปัตยกรรม (architectural pattern) ที่ใช้ในสาขาวิศวกรรมซอฟต์แวร์ รูปแบบ MVC ใช้เพื่อแยกส่วนซอฟต์แวร์ในส่วน ตรรกะเนื้อหา (domain logic) ได้แก่ความเข้าใจในระบบของผู้ใช้ และส่วนการป้อนข้อมูลและแสดงผล (GUI) ซึ่งช่วยให้การพัฒนา การทดสอบ และการดูแลรักษาซอฟต์แวร์ แยกออกจากกัน

โมเดล (Model) คือส่วนของการเก็บรวบรวมข้อมูล ไม่ว่าจะข้อมูลนั้น ๆ จะถูกจัดเก็บในรูปแบบใดก็ตาม ในฐานข้อมูลแบบเป็น Object Class หรือที่นิยมเรียกกันว่า VO ( Value Object ) หรือเก็บเป็นไฟล์ข้อมูลเลย เมื่อข้อมูลถูกโหลดเข้ามาจากที่ต่าง ๆ และเข้ามายังส่วนของโมเดล ตัวโมเดลจะทำการจัดการเตรียมข้อมูลให้เป็นรูปแบบที่เหมาะสม เพื่อรอการร้องขอข้อมูลจากส่วนของ Controller

ระบบซอฟต์แวร์หลายระบบใช้การเก็บข้อมูลถาวร เช่น ฐานข้อมูล เพื่อเก็บข้อมูลเหล่านี้ MVC ไม่ได้กำหนดถึงระดับการเข้าถึงข้อมูล เพราะเป็นที่เข้าใจกันว่าส่วนนี้จะอยู่ภายใต้ หรือถูกรอบคลุมด้วยโมเดล โมเดลไม่ได้เป็นเพียงอ็อบเจกต์ที่ใช้เข้าถึงข้อมูล แต่ในระบบซอฟต์แวร์เล็กๆ ซึ่งมีความซับซ้อนน้อยจะไม่เห็นความแตกต่างมากนัก

เอกสารนี้เป็น **เอวิว (View)** คือส่วนของการแสดงผล หรือส่วนที่จะปฏิสัมพันธ์กับผู้ใช้งาน ( User Interface ) การค้า  
ไม่ว่าหน้าที่ของ **view** ในการเขียนโปรแกรมแบบ MVC คือคอยรับคำสั่งจากส่วนของ Controller และไปใช้

End User เริ่มแรกเลยตัววิว อาจจะได้รับคำสั่งจาก Controller ให้แสดงผลหน้า Home และเมื่อผู้ใช้งานหน้าเว็บกดปุ่มสั่งซื้อ View จะส่งข้อมูลไปให้ Controller เพื่อประมวลผลและแสดงบางอย่างจาก Action นั้น

**คอนโทรลเลอร์ (Controller)** คือส่วนของการเริ่มทำงาน และรับคำสั่ง โดยที่คำสั่งนั้นจะเกิดขึ้นในส่วนการติดต่อกับผู้ใช้งานคือ view เมื่อผู้ใช้งานทำการ Interactive กับ UI view จะเกิดเหตุการณ์หรือข้อมูลบางอย่างขึ้น ตัววิวจะส่งข้อมูลนั้นมายัง controller ตัว controller จะทำการประมวลผลโดยบางคำสั่งอาจจะต้องไปติดต่อกับ model ก่อน เพื่อทำการประมวลผลข้อมูลอย่างถูกต้องเรียบร้อยแล้วก็จะส่งไปยัง view เพื่อแสดงผลตามคำสั่งที่ end user ร้องขอมา Controller จะทำหน้าที่เป็นตัวกลางระหว่าง Model และ View ให้ทำงานร่วมกันอย่างมีประสิทธิภาพและตรงกับความต้องการของ End User มากที่สุดแอปพลิเคชันที่ใช้ MVC อาจจะเป็นกลุ่มของ โมเดล/วิว/คอนโทรลเลอร์ โดยแต่ละกลุ่มใช้ในงานต่างกันไป

MVC มักจะพบได้ในเว็บแอปพลิเคชันโดย วิว จะเป็น HTML หรือ XHTML ที่สร้างโดยแอปพลิเคชันนั้น ส่วนคอนโทรลเลอร์รับค่า GET หรือ POST เข้ามา แล้วเลือกติดต่อกับโมเดลในส่วนที่เกี่ยวข้องเพื่อตอบสนอง โมเดลซึ่งมี business rules จะทำการจัดการตามคำร้องขอนั้นๆ

ในการพัฒนาโปรเจกต์นี้ได้สร้างเว็บแอปพลิเคชันโดยใช้สถาปัตยกรรมแบบ MVC มาเป็นต้นแบบในการสร้างเว็บแอปพลิเคชันนี้

## 2.5 Version Control

Version control คือ ระบบที่จัดการการเปลี่ยนแปลงที่เกิดขึ้นกับไฟล์หนึ่งหรือหลายไฟล์ เพื่อให้คุณสามารถเรียกเวอร์ชันใดเวอร์ชันหนึ่งกลับมาดูเมื่อไรก็ได้ จริง ๆ แล้วคุณสามารถใช้ version control กับไฟล์ชนิดใดก็ได้

ถ้าคุณเป็นนักออกแบบกราฟฟิกหรือเว็บดีไซน์เนอร์และต้องการเก็บทุกเวอร์ชันของรูปภาพหรือเลย์เอาต์ (ซึ่งคุณน่าจะอยากเก็บอยู่) การใช้ Version Control System (VCS) เป็นสิ่งที่ชาญฉลาดมาก เพราะมันช่วยให้คุณสามารถย้อนไฟล์บางไฟล์หรือแม้กระทั่งทั้งโปรเจกต์กลับไปเป็นเวอร์ชันเก่าได้นอกจากนั้นระบบ VCS ยังจะช่วยให้คุณเปรียบเทียบการแก้ไขที่เกิดขึ้นในอดีต ดูว่าใครเป็นคนแก้ไขคนสุดท้ายที่อาจทำให้เกิดปัญหา แก้ไขเมื่อไร ฯลฯ และยังช่วยให้คุณสามารถกู้คืนไฟล์ที่คุณลบหรือทำเสียโดยไม่ตั้งใจได้อย่างง่ายดาย

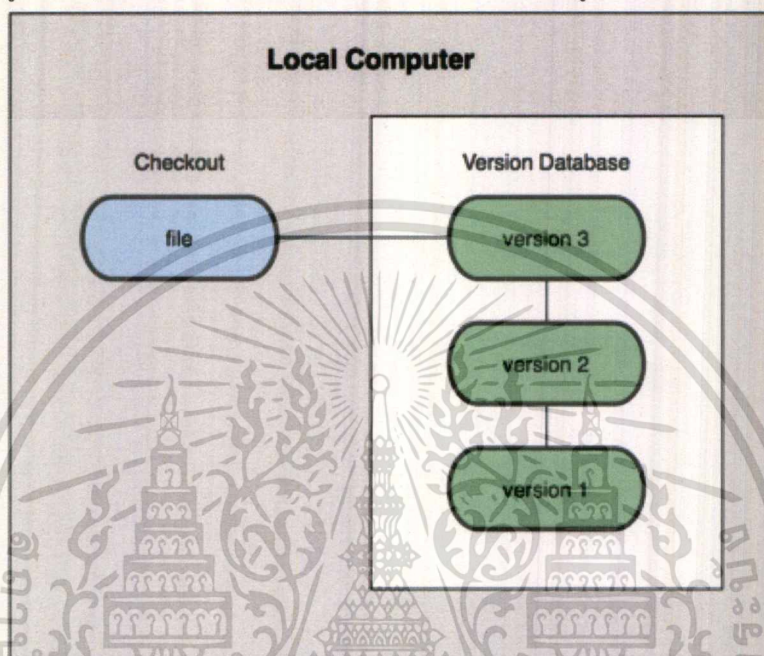
### Version Control Systems แบบ Local

หลาย ๆ คนจัดการประวัติการแก้ไขต่าง ๆ ด้วยมือโดยการคัดลอกไฟล์ไปไว้ในไดเรกทอรีใหม่ (อาจจะเป็นไดเรกทอรีที่มีชื่อเป็นวันเดือนปีและเวลาก็ได้) วิธีนี้เป็นวิธีที่ใช้กันโดยแพร่หลายเพราะว่าทำได้ง่าย แต่ในขณะที่เดียวกันก็เป็นวิธีที่เกิดข้อผิดพลาดได้ง่ายเช่นกัน ยกตัวอย่างเช่น คุณอาจไม่ทันดูว่าตอนนี้

เอกสารนี้เป็นเอกสารที่จัดทำขึ้นโดยศูนย์วิจัยและพัฒนาเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี หากมีข้อผิดพลาดประการใดขออภัยเป็นอย่างสูงและขอสงวนสิทธิ์ในเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คุณอยู่ในไดเรกทอรีไหนและเปลอเขียนทับไฟล์ที่คุณไม่น่าจะเขียนทับหรือทำการคัดลอกไฟล์ที่คุณไม่น่าจะคัดลอก

เพื่อที่จะลดปัญหาเหล่านี้ เมื่อนานมาแล้วโปรแกรมเมอร์ได้พัฒนาระบบ VCS ที่ใช้ในเครื่องของตัวเอง โดยใช้ฐานข้อมูลง่าย ๆ เพื่อเก็บการแก้ไขทั้งหมดที่เกิดขึ้นกับไฟล์ที่อยู่ภายใต้ revision control



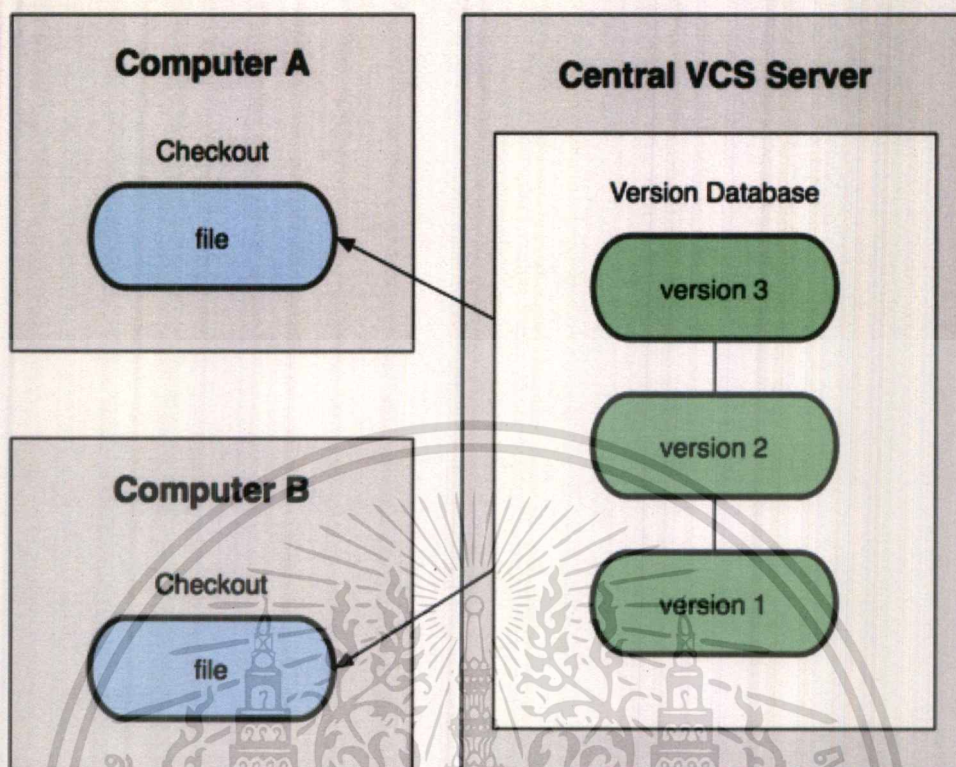
รูปที่ 2.15 โครงสร้างของ Version Control บน Local Computer

หนึ่งในเครื่องมือ VCS ที่ใช้กันมากทั้งในอดีตและปัจจุบันคือระบบที่เรียกว่า rcs แม้แต่ระบบปฏิบัติการ Mac OS X ก็ยังติดตั้ง rcs ให้เมื่อคุณติดตั้ง Developer Tools เครื่องมือนี้ทำงานโดยเก็บสิ่งที่เรียกว่า patch set (ซึ่งก็คือผลต่างของไฟล์แต่ละไฟล์) สำหรับการแก้ไขแต่ละครั้งในรูปแบบพิเศษในเครื่อง ทำให้มันสามารถเรียกคืนไฟล์ ณ ช่วงเวลาใดขึ้นมาดูก็ได้โดยการไล่เรียงไปตาม patch ที่มี

ระบบ Version Control Systems แบบรวมศูนย์

ปัญหาถัดไปที่คนใช้พบก็คือการร่วมมือกันกับนักพัฒนาคนอื่น ๆ เพื่อที่จะแก้ปัญหานี้เครื่องมือใหม่จึงได้ถูกพัฒนาขึ้นมา เรียกว่าระบบ Centralized Version Control Systems (CVCSs) หรือระบบ Version Control Systems แบบรวมศูนย์ ระบบเหล่านี้ เช่น CVS, Subversion และ Perforce มีเซิร์ฟเวอร์กลางที่เก็บไฟล์ทั้งหมดไว้ในที่เดียวและผู้ใช้หลาย ๆ คนสามารถต่อเข้ามาเพื่อดึงไฟล์จากศูนย์กลางนี้ไปแก้ไขได้ ระบบการทำงานแบบรวมศูนย์นี้ได้ถูกนำมาใช้เป็นเวลานานหลายปี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



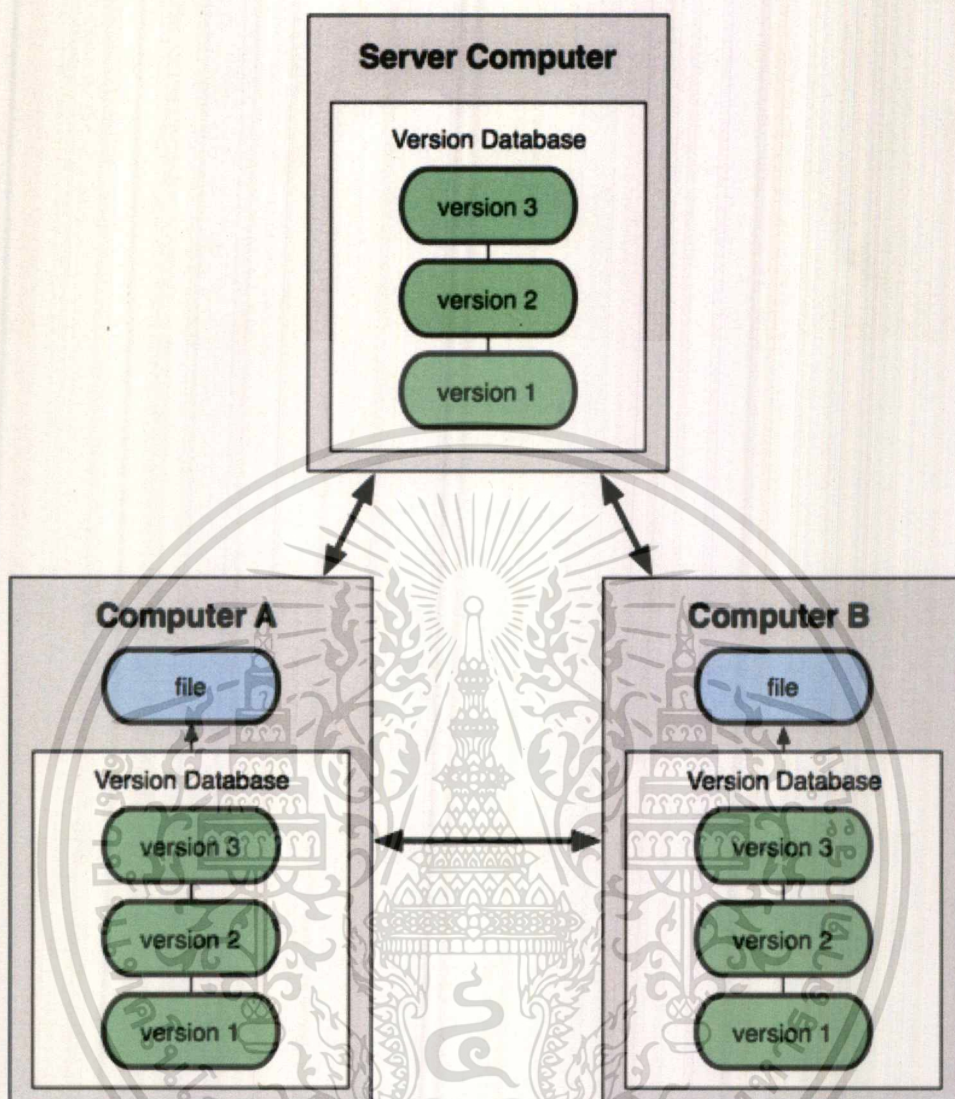
รูปที่ 2.16 โครงสร้างของ ระบบ Version Control Systems แบบรวมศูนย์

การทำงานแบบนี้มีประโยชน์เหนือ local VCS ในหลายด้าน เช่น ทุกคนสามารถรู้ได้ว่าคนอื่นในโปรเจกต์กำลังทำอะไร ผู้ควบคุมระบบสามารถควบคุมได้อย่างละเอียดว่าใครสามารถแก้ไขอะไรได้บ้าง การจัดการแบบรวมศูนย์นี้ที่เดียวทำได้ง่ายกว่าการจัดการฐานข้อมูลใน client แต่ละเครื่องเยอะ แต่ระบบแบบนี้ก็มีจุดอ่อนเหมือนกัน ตรงที่การรวมศูนย์ทำให้มันเป็นจุดอ่อนจุดเดียวที่จะล่มได้เหมือนกันเพราะทุกอย่างรวมกันอยู่ที่เซิร์ฟเวอร์ที่เดียว ถ้าเซิร์ฟเวอร์นั้นล่มซักชั่วโมงหนึ่ง หมายความว่าในช่วงเวลานั้นไม่มีใครสามารถทำงานร่วมกันหรือบันทึกการเปลี่ยนแปลงงานที่กำลังทำอยู่ไปที่เซิร์ฟเวอร์ได้เลย หรือถ้าฮาร์ดดิสก์ของเซิร์ฟเวอร์เกิดเสียขึ้นมาและไม่มีสำรองข้อมูลเอาไว้ คุณก็จะสูญเสียข้อมูลประวัติและทุกอย่างที่มี จะเหลือก็แค่ก๊อปปี้ของงานบนเครื่องแต่ละเครื่องเท่านั้นเอง

นี่คือที่มาของ Distributed Version Control Systems (DVCSs) หรือระบบ VCS แบบกระจายศูนย์ ในระบบแบบนี้ (เช่น Git, Mercurial, Bazaar หรือ Darcs) แต่ละคนไม่เพียงได้ก๊อปปี้ล่าสุดของไฟล์เท่านั้น แต่ได้ทั้งก๊อปปี้ของ repository เลย หมายความว่าถึงแม้ว่าเซิร์ฟเวอร์จะเสีย client ก็ยังสามารถทำงานร่วมกันได้ต่อไป และ repository เหล่านี้ของ client ยังสามารถถูกก๊อปปี้กลับไปเซิร์ฟเวอร์เพื่อถูกข้อมูลกลับคืนก็ได้ การ checkout แต่ละครั้งคือการทำสำรองข้อมูลทั้งหมดแบบเต็มๆ

นั่นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.17 โครงสร้างของระบบ Version Control Systems แบบกระจายศูนย์  
นอกจากนี้ระบบเหล่านี้ยังทำงานกับหลาย ๆ repository ได้อย่างดี ทำให้คุณสามารถทำงานกับคน  
หลายกลุ่มซึ่งทำงานในรูปแบบต่างกันโปรเจกต์เดียวกันได้อย่างง่ายดาย เนื่องจากระบบเหล่านี้  
สนับสนุนการทำงานได้หลากหลายรูปแบบ ซึ่งอาจทำได้ยากในระบบแบบรวมศูนย์

โดย Version Control ที่ใช้งานในปัจจุบันอย่างแพร่หลายคือ Git

## 2.6 Git

Git เริ่มต้นจากส่วนผสมของความคิดริเริ่มที่สั่นคลอนสถานะปัจจุบันเล็กน้อยแต่ทำให้เกิดผล  
กระทบในวงกว้าง โปรเจกต์ลินุกซ์เคอร์เนล (Linux kernel) ถือเป็นซอฟต์แวร์แบบโอเพนซอร์สที่มี  
ขนาดค่อนข้างใหญ่ ในช่วงปี 1991-2002 การพัฒนาเคอร์เนลถูกแชร์กันไปตามผ่าน patch และไฟล์  
ซอร์สโค้ดที่ถูกบีบอัด จากนั้นในปี 2002 ก็มีการเริ่มนำเครื่องมือ DVCS ที่ไม่ใช่โอเพนซอร์สชื่อ  
BitKeeper มาใช้งาน เอกสารที่เผยแพร่ไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

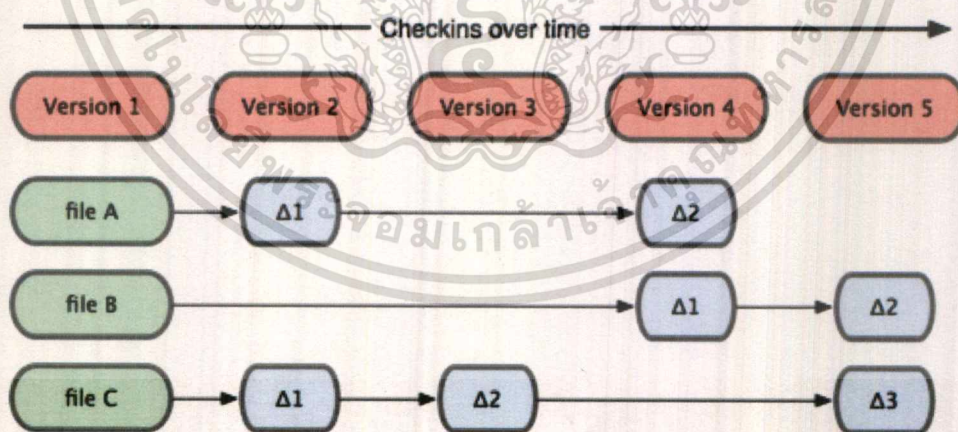
ในปี 2005 ความสัมพันธ์ระหว่าง community ที่พัฒนาลินุกซ์เคอร์เนลและบริษัทที่พัฒนา BitKeeper มีอันต้องจบลง และการใช้งานเครื่องมือนี้แบบฟรีก็ถูกยกเลิกไป ทำให้นักพัฒนาทั้งหลาย (โดยเฉพาะอย่างยิ่งไลน์ส ทอร์วอลด์ ผู้สร้างลินุกซ์) ต้องพัฒนาเครื่องมือของตัวเองขึ้นมาจากระสปรการณที่มีอยู่ระหว่างการใช้งาน BitKeeper โดยมีวัตถุประสงค์ดังนี้:

- ความเร็ว
- ดีไซน์ที่เรียบง่าย
- สนับสนุนการทำงานหลายทางพร้อม ๆ กัน (เช่นมี branch การพัฒนาเป็นพัน)
- แยกศูนย์
- สามารถรองรับโปรเจกต์ขนาดใหญ่อย่างลินุกซ์เคอร์เนลได้เป็นอย่างดี (ทั้งในแง่ความเร็วและขนาดของข้อมูล)

ตั้งแต่เริ่มต้นเมื่อปี 2005 Git ได้ถูกทำให้ใช้งานง่ายขึ้นแต่ก็ยังคงความสามารถตามวัตถุประสงค์เดิมเอาไว้ โดยสามารถทำงานได้เร็วเหลือเชื่อ ใช้พื้นที่น้อยสำหรับโปรเจกต์ใหญ่ ๆ และก็มีระบบการ branch ที่สนับสนุนการทำงานหลายทางพร้อม ๆ กัน

เก็บ Snapshot แทนผลต่าง

ความแตกต่างมากที่สุดระหว่าง Git และ VCS อื่น ๆ (เช่น Subversion และพองเพื่อน) คือ วิธีที่ Git มองข้อมูลต่าง ๆ โดยทั่วไประบบอื่นมักจะเก็บข้อมูลในรูปแบบของการแก้ไขที่เกิดขึ้นกับไฟล์ต่าง ๆ ระบบเหล่านี้ (เช่น CVS, Subversion, Perforce, Bazaar, ฯลฯ) จะมองข้อมูลในรูปแบบของไฟล์และการแก้ไขต่าง ๆ ที่เกิดขึ้นกับไฟล์แต่ละไฟล์

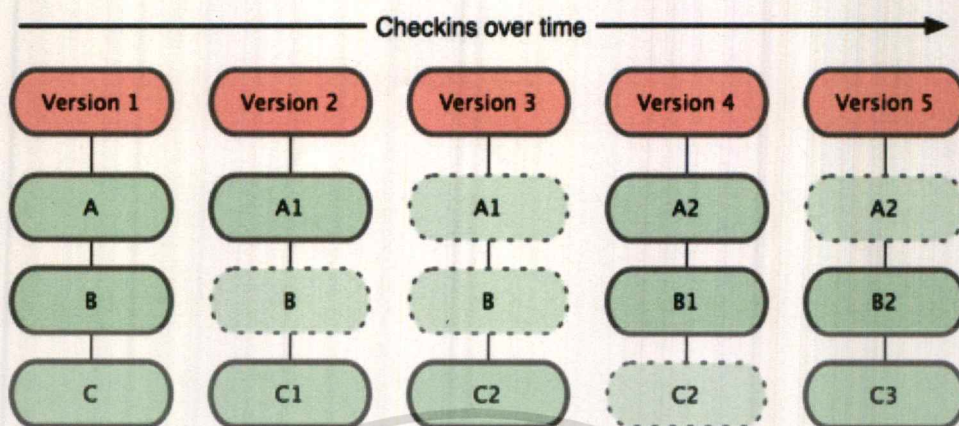


รูปที่ 2.18 การเก็บข้อมูลของระบบอื่นๆ

Git ไม่ได้มองและจัดเก็บข้อมูลในลักษณะนี้แต่จะคิดว่าข้อมูลของมันเป็นเสมือนภาพถ่าย (snapshot) ของระบบไฟล์ขนาดเล็กๆ ทุกครั้งที่มีการ commit หรือบันทึกสถานะของโปรเจกต์ลงใน Git มันจะทำการถ่ายภาพของไฟล์ทั้งหมดในตอนนั้นและบันทึกการอ้างอิงไปยัง snapshot นั้น

เพื่อให้การจัดเก็บนั้นมีประสิทธิภาพ ถ้าไฟล์ใดที่ไม่ได้มีการเปลี่ยนแปลง Git ก็จะไม่บันทึกไฟล์นั้นอีก ครั้ง เพียงแต่จะทำการเชื่อมโยงไปยังไฟล์เดิมที่เคยถูกบันทึกเอาไว้แล้ว Git จะมองข้อมูลดังรูปที่

ไม่ว่าการแก้ไขใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.19 Git เก็บข้อมูลเป็น snapshot ของโปรเจกต์

นี่คือความแตกต่างที่สำคัญระหว่าง Git กับ VCSs ตัวอื่นๆ มันทำให้ Git ทำได้เกือบทุกด้านของระบบ VCS อื่นๆ ซึ่งส่วนใหญ่ก็คัดลอกมาจากรุ่นก่อนๆ นี่ทำให้ Git เป็นเหมือนกับระบบไฟล์ขนาดเล็กที่มีเครื่องมือ(tools)อันทรงพลังอย่างน่าเหลือเชื่อครอบอยู่ แทนที่จะเป็น VCS แบบทั่วไป

การทำงานเกือบทุกอย่างเป็นการทำงานในเครื่องตัวเอง

การทำงานโดยส่วนใหญ่ของ Git จะใช้ไฟล์และทรัพยากรในเครื่องของเราเท่านั้น ปกติจะไม่มีข้อมูลใดๆ ที่จำเป็นต้องใช้จากคอมพิวเตอร์เครื่องอื่นๆ ในเน็ตเวิร์ก แต่หากเคยใช้ CVCS อื่นที่การทำงานส่วนใหญ่ต้องใช้ข้อมูลบนเน็ตเวิร์กจำนวนมาก ในแง่นี้ก็ทำให้คุณรู้สึกมีความสุขกับความเร็วในการทำงานของ Git เพราะว่าคุณจะมีประวัติการเปลี่ยนแปลงทั้งหมดของโปรเจกต์อยู่ในเครื่องของคุณอยู่แล้วและพร้อมที่จะทำงานได้ทันที

ตัวอย่างเช่น ถ้าจะดูประวัติย้อนหลังของโปรเจกต์ Git ไม่จำเป็นต้องจะไปดึงข้อมูลจากเซิร์ฟเวอร์แล้วจึงแสดงผลให้คุณได้ มันแค่อ่านโดยตรงจากฐานข้อมูลในเครื่องของคุณ หมายความว่า คุณสามารถที่จะดูประวัติของโปรเจกต์ได้ทันที หากจะดูความเปลี่ยนแปลงของไฟล์ในรุ่นปัจจุบันกับเมื่อหนึ่งเดือนที่แล้ว Git ก็สามารถค้นหาไฟล์เมื่อเดือนก่อนในเครื่องของเราแล้วทำการคำนวณความแตกต่างแทนที่จะถามเครื่องเซิร์ฟเวอร์ให้ดึงไฟล์เก่ามาให้

ซึ่งก็มีโอกาสน้อยมากที่คุณไม่สามารถทำงานได้ถ้าหากออฟไลน์อยู่ หากต้องเดินทางอยู่บนเครื่องหรือบนรถไฟและอยากจะทำงานสักหน่อย คุณก็สามารถ commit ได้อย่างมีความสุขจนกว่าจะเชื่อมต่อเน็ตเวิร์กได้แล้วออฟไลน์ หากที่บ้านของคุณมีปัญหาเรื่องเน็ตเวิร์กคุณก็ยังคงทำงานได้ ถ้าเป็นระบบอื่นทำแบบนี้ไม่ได้แน่ และจะทำอะไรแทบจะไม่ได้เลยหากเชื่อมต่อไปยังเซิร์ฟเวอร์ไม่ได้ ยิ่งในโปรแกรม Subversion และ CVS คุณสามารถแก้ไขไฟล์ต่างๆ ได้แต่จะไม่สามารถ commit ได้ เพราะฐานข้อมูลมันออฟไลน์อยู่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Git มีความเที่ยงตรง

ทุกอย่างที่ Git ทำการบันทึกเอาไว้จะถูกทำการ Checksum แล้วนำมาใช้เป็นตัวอ้างอิง นั่นทำให้ไม่มีทางที่เราจะแก้ไขข้อมูลของไฟล์และไดเรกทอรีใดโดยที่ Git จะไม่รู้ ซึ่งฟังก์ชันนี้จะอยู่ในระดับล่างและเป็นหลักการของ Git คุณจะไม่มีทางที่จะทำข้อมูลสูญหายระหว่างการโยกย้ายหรือรับไฟล์ที่เสียหายโดย Git จะสามารถตรวจพบได้

กลไกที่ Git ใช้ในการทำ Checksum คือการแฮช(hash)แบบ SHA-1 ซึ่งผลลัพธ์จะได้ออกมาเป็นตัวอักษร 40 ตัวที่แทนเลขฐานสิบหก(0-9 และ a-f)จากการคำนวณเนื้อหาในไฟล์หรือโครงสร้างของไดเรกทอรีของ Git ซึ่ง SHA-1 มีลักษณะดังนี้

24b9da6552252987aa493b52f8696cd6d3b00373

รูปที่ 2.20 checksum ของ Git ที่ได้จากการ hash แบบ SHA-1

จะเห็นว่าผลของการแฮช(hash)เหล่านี้อยู่ในทุกที่ใน Git เพราะจะถูกใช้บ่อยครั้ง ซึ่งจริงๆแล้ว Git ไม่ได้เก็บบันทึกข้อมูลทุกอย่างตามชื่อไฟล์แต่เก็บในฐานข้อมูลของ Git แล้วสามารถอ้างอิงด้วยค่าแฮช(hash)ของข้อมูลของไฟล์

Git เพียงแต่เพิ่มข้อมูล

เมื่อคุณกระทำอะไรสักอย่างใน Git เนื้อหาเกือบทั้งหมดนั้นก็จะถูกเพิ่มเข้าไปในฐานข้อมูลของ Git เท่านั้น มันเป็นเรื่องยากมากๆ ที่เราจะใช้ระบบที่ทำอะไรลงไปแล้วไม่สามารถย้อนคืนกลับมาได้หรือลบแล้วลบเลย เช่นเดียวกันกับ VCS ตัวอื่นๆ คุณสามารถสูญเสียข้อมูลหรือแก้ไขผิดพลาดได้โดยที่ยังไม่ทันได้ commit แต่ถ้าคุณ commit ลงใน snapshot ของ Git แล้วมันก็ยากที่จะสูญหายได้โดยเฉพาะอย่างยิ่งถ้าคุณทำการผลัก(push)ฐานข้อมูลของคุณไปไว้ที่อื่นๆ

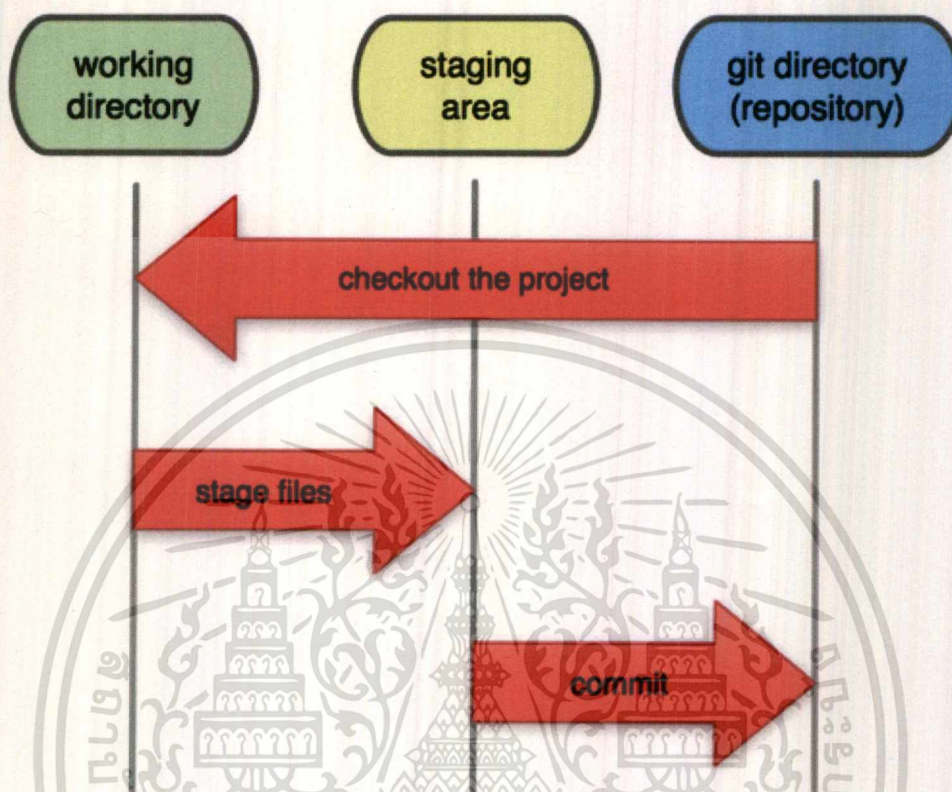
สามสถานะ

ไฟล์ของคุณใน Git จะมีอยู่ 3 สถานะ คือ ยืนยันแล้ว(committed), ถูกแก้ไข(modified) และ อยู่ในขั้นตอน(staged) ซึ่ง Committed หมายถึงข้อมูลที่ถูกบันทึกเรียบร้อยแล้วในฐานข้อมูลในเครื่องของคุณ Modified หมายถึงไฟล์ของคุณได้ถูกแก้ไขแล้วแต่ยังไม่ยืนยัน(commit)ลงในฐานข้อมูลของคุณ Staged หมายถึงคุณสามารถทำเครื่องหมายไว้ที่ไฟล์ที่ถูกแก้ไขในเวอร์ชันปัจจุบันเพื่อที่จะรอการ commit ใน snapshot ถัดไป

นี่ทำให้โปรเจกต์ที่ใช้ Git มี 3 ส่วน คือ the Git directory, working directory, และ staging area.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้拿去ใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## Local Operations



รูปที่ 2.21 Working directory, staging area, และ git directory

Git directory เป็นที่ที่ Git ใช้เก็บ metadata และออปเจ็คของฐานข้อมูลของโปรเจกต์ของคุณ นี่เป็นส่วนสำคัญที่สุดของ Git และเป็นส่วนที่จะถูกคัดลอกมาเมื่อคุณทำการโคลน (clone) คลังข้อมูลจากคอมพิวเตอร์เครื่องอื่น

working directory เป็นเวอร์ชันหนึ่งของไฟล์ในโปรเจกต์ที่ถูกดึงออกมาจากฐานข้อมูลที่ถูกบีบอัดไว้ใน Git directory แล้วเก็บไว้ในดิสก์เพื่อให้คุณนำไปใช้หรือเอามาแก้ไข

staging area เป็นไฟล์ธรรมดาไฟล์หนึ่ง โดยทั่วไปก็อยู่ใน Git directory ของคุณ ซึ่งเก็บข้อมูลส่วนที่คุณจะทำการ commit ในครั้งถัดไป บางครั้งก็เรียกว่าดัชนี(index) แต่ปกติก็จะเรียกว่า staging area

กระบวนการขั้นตอนพื้นฐานของ Git มีลักษณะดังนี้

1. คุณทำการแก้ไขไฟล์ใน working directory ของคุณ
2. แล้วทำการ stage ไฟล์เหล่านั้นเพื่อให้มีการใส่ snapshot ลงไปใน staging area ของคุณ
3. ทำการยืนยัน(commit)ซึ่งนำไฟล์ที่อยู่ใน staging area ไปเก็บอย่างถาวรใน Git

directory

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อไฟล์อยู่ใน Git directory มันจะถือว่าเป็นสถานะ committed ถ้าไฟล์ถูกแก้ไขแล้วถูกเพิ่มลงใน staging area สถานะจะเป็น staged และถ้าไฟล์ถูกแก้ไขแล้วหลังจากดึงออกมาแต่ไม่ได้เป็นสถานะ staged ก็จะเป็นสถานะ modified

## 2.7 Sourcetree



รูปที่ 2.22 Logo ของ Sourcetree

SourceTree คือโปรแกรมสำหรับจัดการ Source code ต่าง ๆ บน GIT โดยปกติเราสามารถใช้งานได้ผ่าน Command line ด้วยคำสั่งต่าง ๆ อย่างเช่น add commit push pull และโปรแกรมจำพวก Git UI จะเปลี่ยนคำสั่งต่าง ๆ ให้อยู่ในรูปแบบของหน้าจอโปรแกรม เพื่อให้ง่ายต่อการใช้งานนั่นเอง

จุดเด่นของ SourceTree

1. ง่ายต่อการใช้งาน ผ่านหน้าต่างโปรแกรมที่เป็นชัดเจน
2. ง่ายต่อการดู Git History เนื่องจากมีแผนผังแสดงประวัติให้เราดูได้อย่างชัดเจน
3. มีหน้าจอแสดงการเปลี่ยนแปลงของ Source code อย่างชัดเจน ทำให้รู้ว่า แต่ละ Commit เราทำอะไรลงไปบ้างแล้ว
4. สามารถใช้งานระดับ Advance ผ่านเมนู Terminal ที่มาพร้อมกับโปรแกรมได้เลย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 3

# วิธีการดำเนินงานวิจัย

### 3.1 การวางแผนและการเตรียมการ

การดำเนินการพัฒนาระบบทำนายการเสนอขายประกันภัยรถยนต์ด้วยแมชชีนเลิร์นนิงนี้ ได้มีการวางแผนและการเตรียมการเพื่อจัดการพัฒนาโดยแบ่งเป็นขั้นตอน ดังนี้

#### 3.1.1 การศึกษาปัญหาและเก็บรวบรวมข้อมูลที่เกี่ยวข้อง

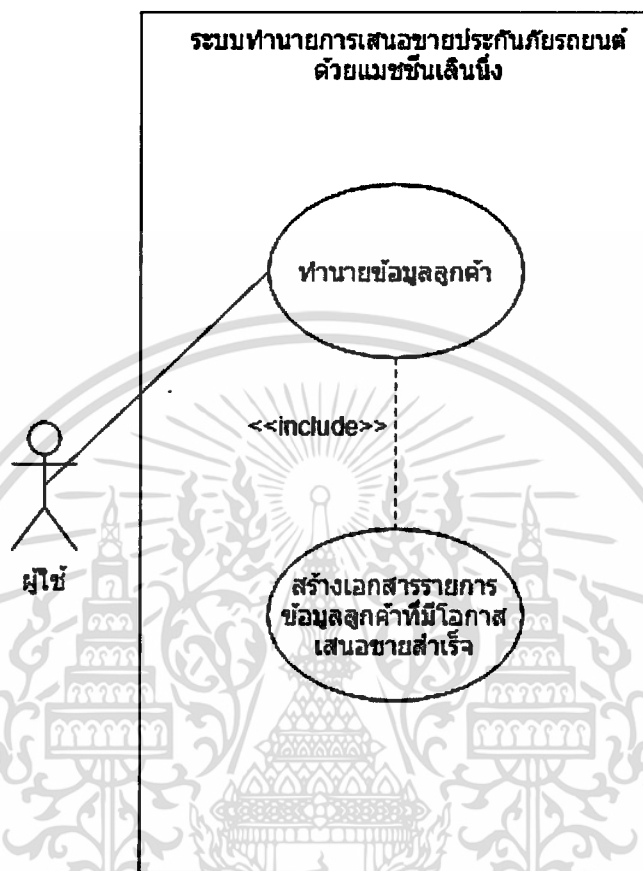
1. ประกันภัยรถยนต์ เป็นหนึ่งในผลิตภัณฑ์ของธนาคารทิสโก้ จำกัด(มหาชน) อยู่ในหมวดของประกันวินาศภัย ที่คุ้มครองได้สูงสุด 7 คน ในช่วงที่ผู้จัดทำได้เข้าไปพูดคุยกับฝ่ายสินเชื่อบริการของธนาคารเพื่อปรึกษาเรื่องปัญหาที่เกิดขึ้นกับโปรเจกต์ที่ทำให้ได้รับรู้ว่าฝ่ายสินเชื่อบริการมีความสนใจในกลุ่มลูกค้าที่เข้ามาทำประกันรถยนต์ และได้แนะนำให้ผู้จัดทำ ทำโปรเจกต์การทำนายข้อมูลลูกค้าเพื่อหาลูกค้าที่น่าจะประสบความสำเร็จในการเสนอขายประกันภัยรถยนต์ โดยทำให้ออกมาในรูปแบบของเว็บแอปพลิเคชันที่สามารถนำเข้าข้อมูลรูปแบบไฟล์นามสกุล CSV หรือ ใส่ข้อมูลที่แต่ละคนได้ อีกทั้งยังสามารถเลือกรูปแบบโมเดลที่จะใช้ทำนายข้อมูลนั้นๆได้ และใช้การเปรียบเทียบคำตอบการทำนาย
2. ศึกษาการใช้งานของเครื่องมือที่จำเป็น ดังต่อไปนี้
  - a. Python(library scikit-learn, keras, flask, pandas)
  - b. Weka
  - c. Excel
  - d. Visual Studio Code
  - e. Bootstrap
3. ศึกษากระบวนการทำงานของแมชชีนเลิร์นนิงที่ใช้

#### 3.1.2 วิเคราะห์ความต้องการของระบบ

1. ศึกษาเรียนรู้กระบวนการทำงานของแมชชีนเลิร์นนิงรูปแบบการจัดหมวดหมู่ข้อมูล เพื่อนำความรู้ที่ได้มาใช้ในการวางแผนการทำงานให้กับระบบทำนายการเสนอขายประกันภัยรถยนต์ด้วยแมชชีนเลิร์นนิง

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์และสงวนสิทธิ์ในเนื้อหา ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2. แผนภาพ Use Case



รูปที่ 3.1 แผนภาพ Use Case Diagram

## ตารางที่ 3.1 Use case decription

หัวข้อ	คำอธิบาย
Use Case Name:	ทำนายข้อมูลลูกค้า
Actor:	ผู้ใช้งานระบบ
Objective:	ทำนายข้อมูลของลูกค้าที่มีโอกาสเสนอขายประกันภัยรถยนต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Association:	เมื่อได้รับข้อมูลเข้ามาระบบจะทำการทำนายด้วยโมเดลแมชชีนเลิร์นนิงแล้วแสดงข้อมูลลูกค้าที่ระบบทำนายว่าซื้อผ่านทางหน้าจอ
Pre-condition:	ผู้ใช้ทำการอัปโหลดไฟล์ข้อมูลลูกค้าหรือกำหนดข้อมูลของลูกค้าแล้วกดปุ่มSubmit ระบบจะทำการทำนายข้อมูลต่อไป
Post-condition:	กดปุ่มDownloadเพื่อ โหลดเอกสารของผลลัพธ์ในการทำนายได้
Normal Flow of Events:	- ผู้ใช้ อัปโหลดไฟล์ หรือ กำหนดข้อมูลของลูกค้า - ระบบจะแสดงผลการทำนายที่ทำนายว่าสำเร็จ
Alternative/Sub flow of Events:	

### 3.1.3 ออกแบบระบบ

การออกแบบระบบ (Design) เป็นขั้นตอนการนำปัญหาและความต้องการต่าง ๆ ที่แบ่งแยกไว้นั้น เพื่อใช้ในการออกแบบระบบงาน แบ่งออกเป็นขั้นตอนย่อย ๆ ไว้ ดังนี้

1. ออกแบบผังงาน(Flowchart)
2. ออกแบบแผนภาพกระแสข้อมูลระดับสูง (Context Diagram)
3. ออกแบบแผนภาพการไหลของข้อมูล ระดับที่ 0 (Data Flow Diagram Level 0)
4. ออกแบบแผนภาพการไหลของข้อมูล ระดับที่ 1 (Data Flow Diagram Level 1)
5. ออกแบบแผนผังการทำงานแบบลำดับปฏิสัมพันธ์ Sequence Diagram
6. ออกแบบแบบฟอร์มสำหรับรับข้อมูลของผู้ใช้บนหน้าเว็บแอปพลิเคชัน

### 3.1.4 การดำเนินงาน

1. ทำการเขียนโปรแกรมเพื่อทำการสร้างโมเดลแมชชีนเลิร์นนิง เพื่อค้นหาโมเดลที่สามารถให้ผลลัพธ์คำตอบที่ดีที่สุดเพื่อนำมาใช้งานจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ทำการเขียนโปรแกรมเพื่อสร้างแบบฟอร์มรับข้อมูลของผู้ใช้และเรียกใช้โมเดลที่ได้สร้างไว้ในขั้นตอนก่อนหน้าเพื่อทำนายข้อมูลแล้วแสดงผลผ่านทางหน้าเว็บแอปพลิเคชัน

## 3.2 ชั้นการออกแบบระบบ

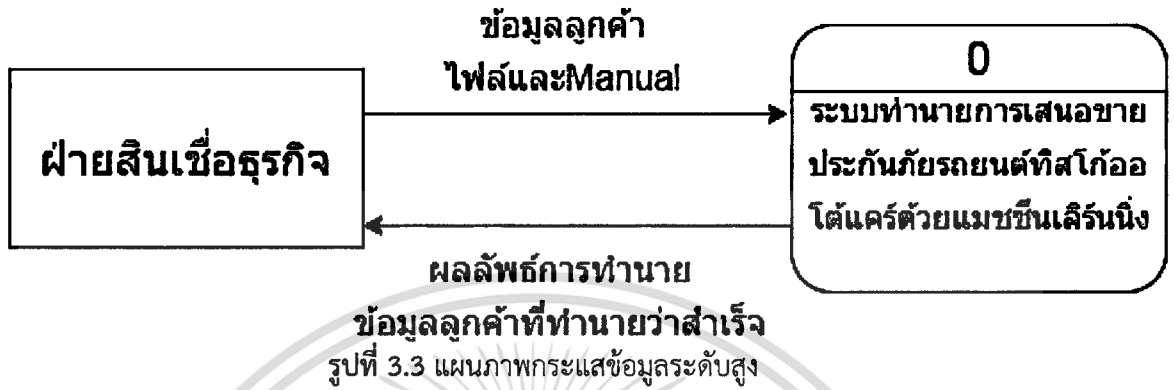
### 3.2.1 ผังงาน(Flowchart)



รูปที่ 3.2 ผังงาน

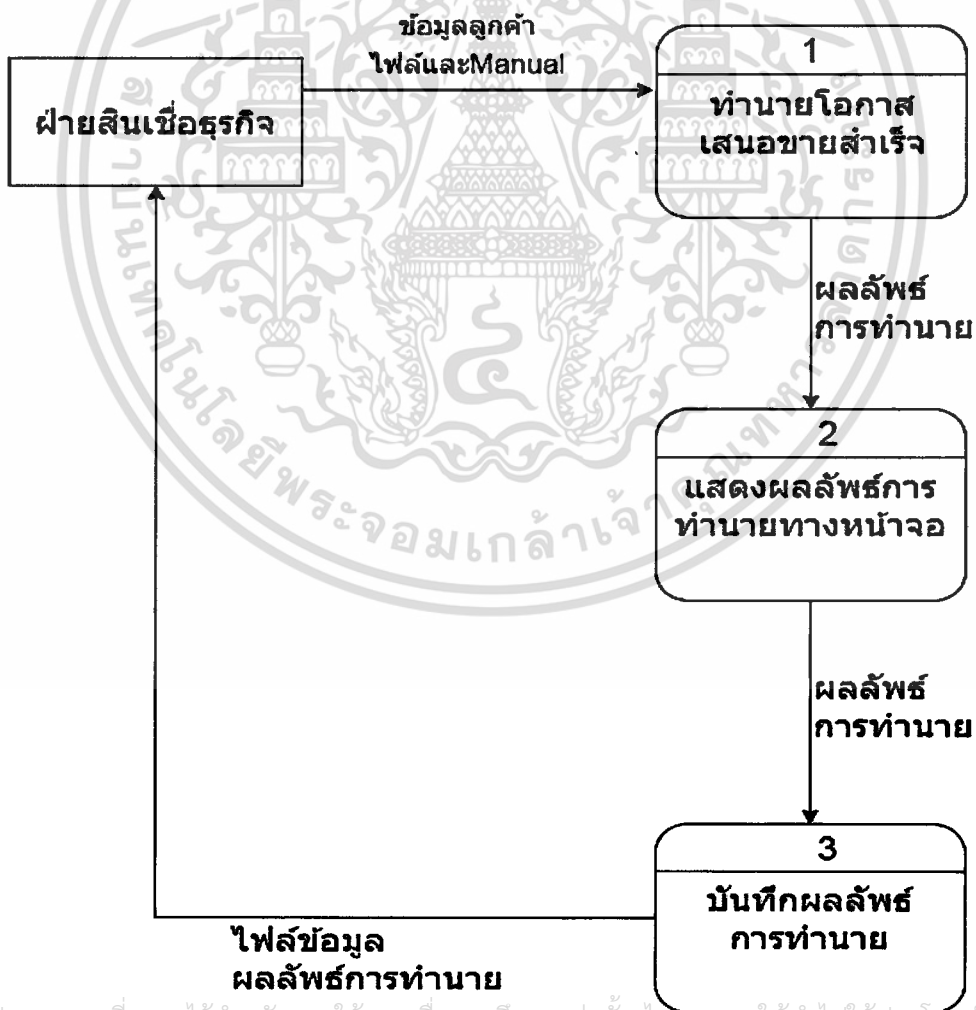
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2 แผนภาพกระแสข้อมูลระดับสูง (Context Diagram)



รูปที่ 3.3 แผนภาพกระแสข้อมูลระดับสูง

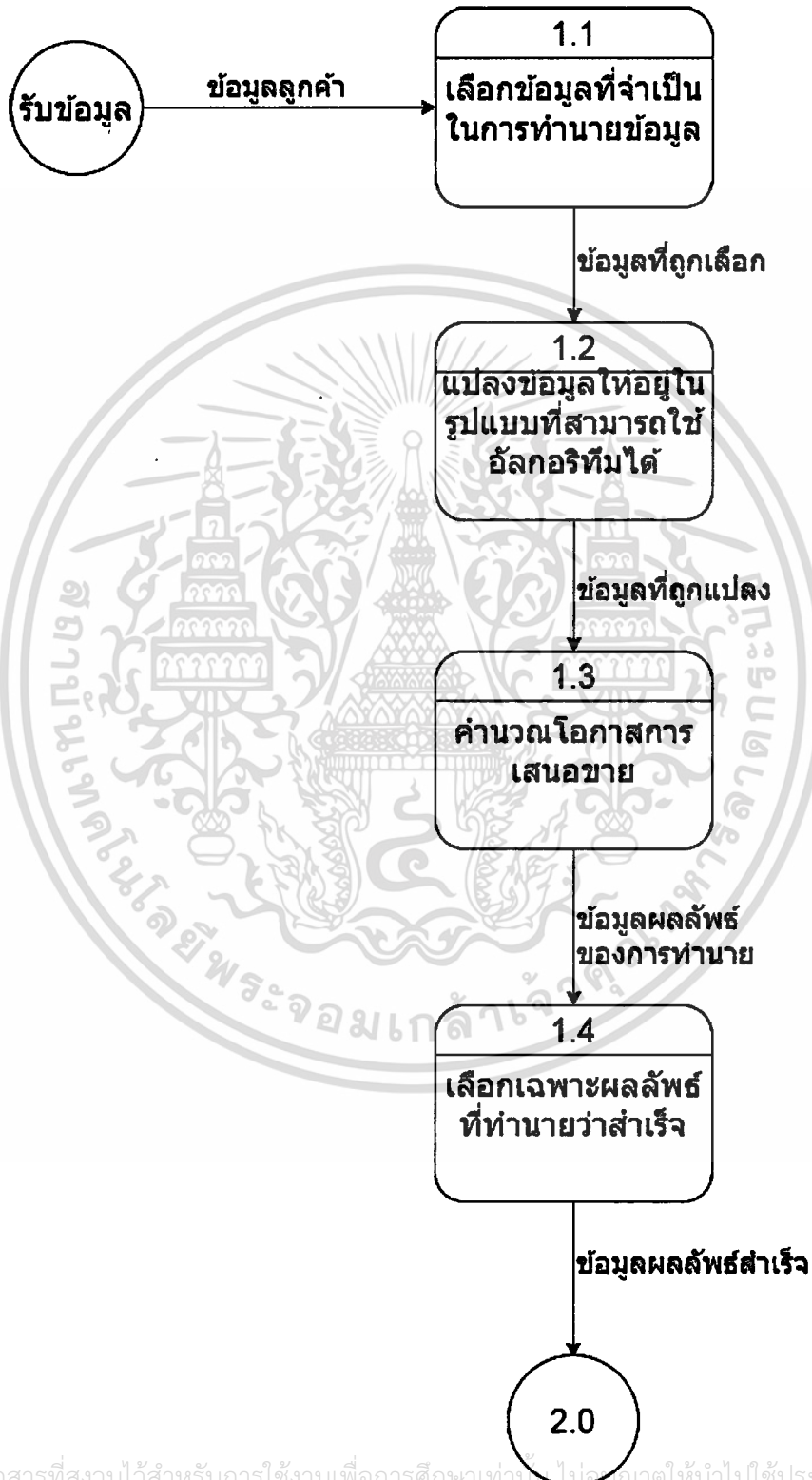
3.2.3 แผนภาพการไหลของข้อมูล ระดับที่ 0 (Data Flow Diagram Level 0)



รูปที่ 3.4 แผนภาพการไหลของข้อมูล ระดับที่ 0

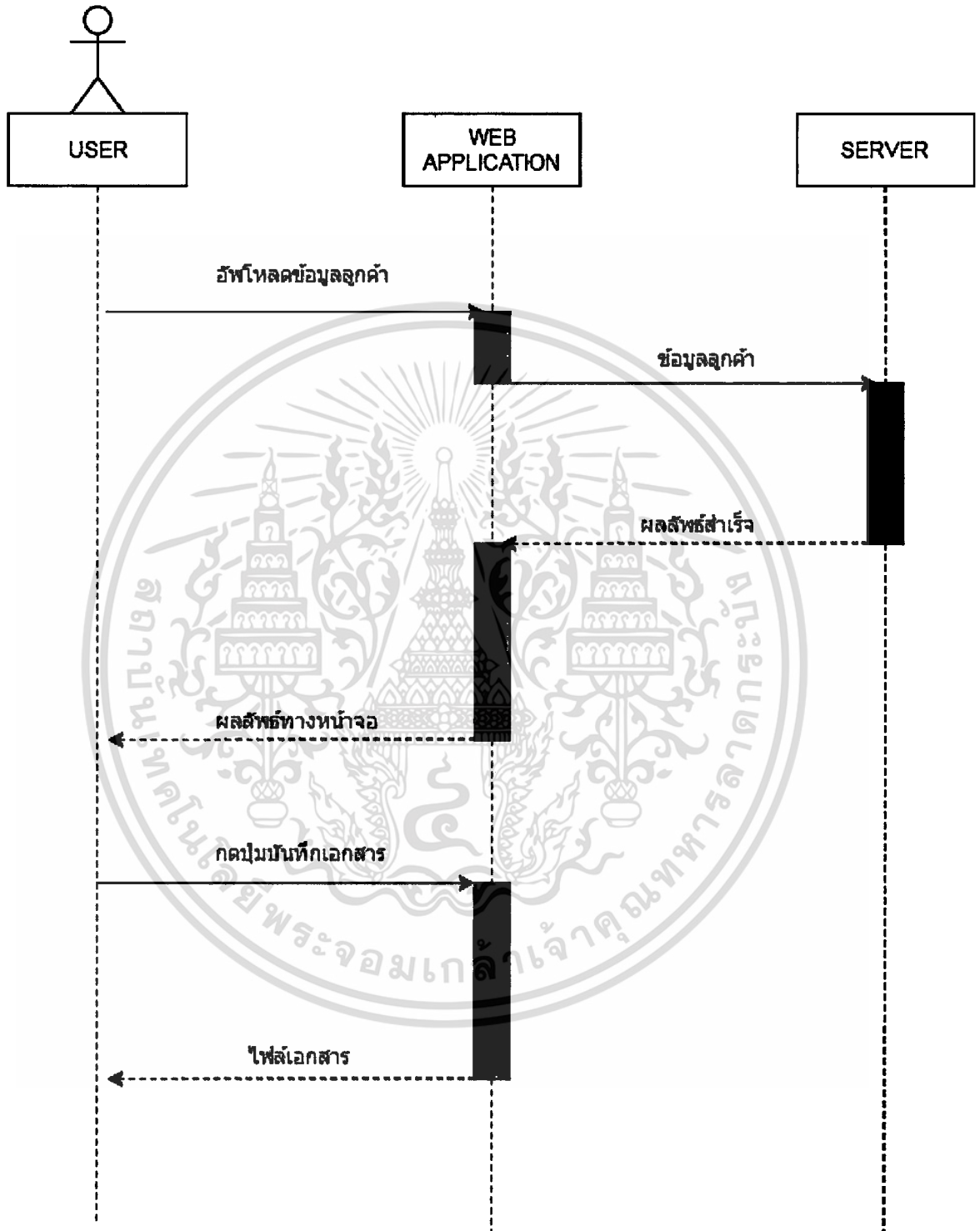
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต  
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.2.4 แผนภาพการไหลของข้อมูล ระดับที่ 1 (Data Flow Diagram Level 1)



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 รูปที่ 3.5 แผนภาพการไหลของข้อมูล ระดับที่ 1 ส่วนทำนายโอกาสสำเร็จ  
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.2.5 แผนผังการทำงานแบบลำดับปฏิสัมพันธ์ Sequence Diagram



รูปที่ 3.6 แผนผังการทำงานแบบลำดับปฏิสัมพันธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.2.6 ออกแบบส่วนติดต่อผู้ใช้บนหน้าเว็บแอปพลิเคชัน

#### 1. หน้าจอส่วนรับข้อมูล

The screenshot displays two input sections within a larger frame. The top section, titled 'FILE INPUT', contains a 'CHOOSE FILE' button followed by a text input field and a 'SUBMIT' button below it. The bottom section, titled 'MANUAL INPUT', features six data entry fields arranged in two rows: 'DATA1', 'DATA2', 'DATA3' in the first row, and 'DATA4', 'DATA5', 'DATA6' in the second row, with a 'SUBMIT' button centered below them.

รูปที่ 3.7 ออกแบบหน้าจอส่วนรับข้อมูล

#### 2. หน้าจอส่วนแสดงผลลัพธ์

The screenshot shows a 'RESULT' section containing a table with 5 rows and 3 columns. The columns are labeled 'ลำดับ' (Order), 'หมายเลขของจุดค่า' (Point Value Number), and 'โอกาสสำเร็จ' (Success Rate). Below the table is a 'Print Report' button.

ลำดับ	หมายเลขของจุดค่า	โอกาสสำเร็จ
1	HP110001	72%
2	HP110002	62%
3	HP110003	99%
4	HP110004	78%
5	HP110005	55%

รูปที่ 3.8 ออกแบบหน้าจอส่วนแสดงผลลัพธ์

#### 3. หน้าจอส่วนแสดงผลลัพธ์ เปรียบเทียบ 3 โมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

RESULT					
ลำดับ	หมายเลขลูกค้า	โมเดล1	โมเดล2	โมเดล3	โอกาส
1	HP00001	Y	Y	Y	56%
2	HP00002	Y	N	N	76%
3	HP00003	N	Y	N	59%
4	HP00004	N	N	Y	89%

**บันทึกเอกสาร**

รูปที่ 3.9 ออกแบบหน้าจอส่วนแสดงผลลัพธ์ 3 โมเดล

### 3.3 การพัฒนาระบบ

#### 3.3.1 สํารวจข้อมูล

จากการทำ สํารวจข้อมูล(Data Exploration) ทำให้ทราบว่า เป็นข้อมูลการเสนอขาย ประกันภัยรถยนต์ ปีพ.ศ.2561 ของลูกค้าจำนวน 188,894 คน โดยรวมมีเพียงแค่ 34,533 หรือ 18.76% ที่สามารถเสนอขายประกันได้สำเร็จ ข้อมูลมีแอตทริบิวต์ 26 ตัว มีข้อมูลขาดหายและข้อมูลผิดพลาด

#### 3.3.2 การเตรียมข้อมูล

เพื่อให้ข้อมูลสามารถทำงานร่วมกับโมเดลแมชชีนเลิร์นนิงเพื่อการจัดหมวดหมู่และเพื่อให้โมเดลมีประสิทธิภาพมากที่สุด ต้องทำการเตรียมข้อมูล(Data Preprocessing)ให้พร้อม เนื่องจากว่าข้อมูลจริงนั้นมีส่วนที่ไม่สมบูรณ์อยู่มาก ไม่ว่าจะเป็น มีข้อมูลที่ขาดหาย ข้อมูลผิดพลาด ข้อมูลไม่มีความสอดคล้องทั้งที่มีความหมายเดียวกัน เป็นต้น โดยขั้นตอนที่ทำประกอบไปด้วย

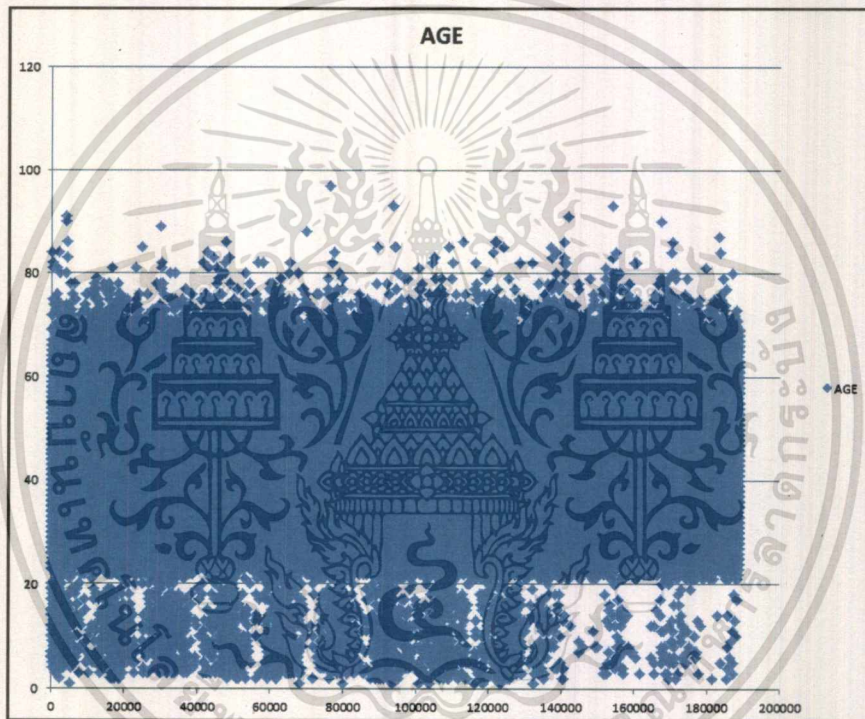
- Data Cleaning เพื่อจัดการกับข้อมูลที่สูญหาย และข้อมูลผิดพลาด
- Data Transformation เพื่อแปลงค่าข้อมูลให้อยู่ในช่วงสั้นๆ
- Data Integration เพื่อผสานข้อมูลจากหลายแห่งให้เป็นหนึ่งเดียว

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการเรียนการสอนเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลังจากนั้นจึงทำ Feature Selection เพื่อคัดเลือกเฉพาะแอตทริบิวต์ที่ส่งผลต่อผลลัพธ์แล้วนำข้อมูลที่ได้ ไปใช้ในการสร้างโมเดลแมชชีนเลิร์นนิงต่อไป

ในข้อมูลที่ได้รับมานั้น จะมีข้อมูลอยู่ 4 แอตทริบิวต์ ที่เป็นข้อมูลของลูกค้าที่เสนอขายประกันสำเร็จ ดังนั้นทางผู้ทำวิจัยจึงได้ทำการตัด 4 แอตทริบิวต์นี้ออกจากขั้นตอนการทำงาน ตัวอย่างข้อมูลอยู่ในรูปที่ 3.9 ตัวอย่างแอตทริบิวต์ที่มีข้อมูลเฉพาะลูกค้าที่เสนอขายสำเร็จ

ในข้อมูลที่ได้รับมานั้น จะมีข้อมูลที่ขาดหายในลูกค้าบางคน ซึ่งแอตทริบิวต์ที่มีข้อมูลขาดหาย ดังนั้นทางผู้ทำวิจัยจึงได้ดำเนินการจัดการกับข้อมูลขาดหายเหล่านี้ เช่น ข้อมูลอายุ ได้ทำการตรวจสอบการกระจายของข้อมูล เพื่อทำการตัดข้อมูลที่กระจายไกลข้อมูลตัวอื่นออก

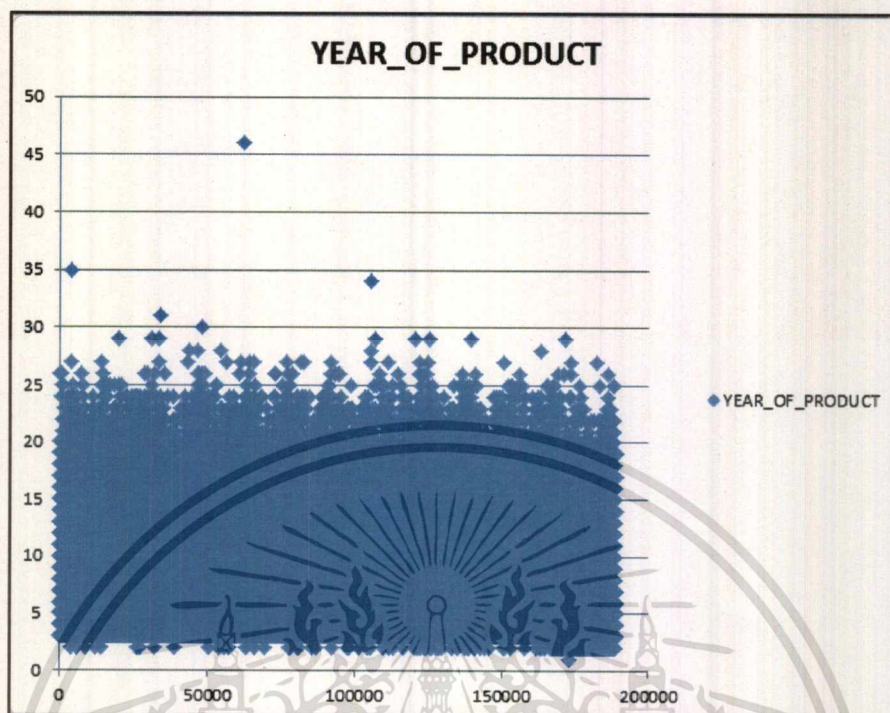


รูปที่ 3.10 การกระจายของข้อมูลอายุ

ในข้อมูลของแอตทริบิวต์ที่เกี่ยวกับรหัสไปรษณีย์ จะมีข้อมูลที่ผิดพลาด ซึ่งผู้ทำวิจัยคาดการณ์ว่าอาจเกิดจากการบันทึกข้อมูลที่ผิดพลาดในขั้นตอนการเก็บรวบรวมข้อมูล ดังนั้นผู้จัดทำจึงทำการลบรายการที่มีรหัสไปรษณีย์ผิดพลาด ออกจากข้อมูลที่จะใช้ทำโมเดลแมชชีนเลิร์นนิง

อีกหนึ่งแอตทริบิวต์ที่เจอข้อมูลผิดพลาดคือ แอตทริบิวต์ข้อมูลอายุรถ ซึ่งคืออายุของรถยนต์ลูกค้า โดยค่าที่เก็บจะเป็นค่าตัวเลขจำนวนอายุรถยนต์ แต่ในส่วนของข้อมูลที่ผิดพลาดนั้นจะมีทั้งตัวเลขติดลบ และ ข้อมูลที่เป็นปีคริสต์ศักราช ผู้ทำวิจัยได้ทำการตัดข้อมูลที่ผิดพลาดเหล่านี้ทิ้งไป แล้วทำการตรวจสอบการกระจายของข้อมูลก็พบว่า แอตทริบิวต์ข้อมูลอายุรถ มีการกระจายของข้อมูลลดลงอย่างเห็นได้ชัดเจน

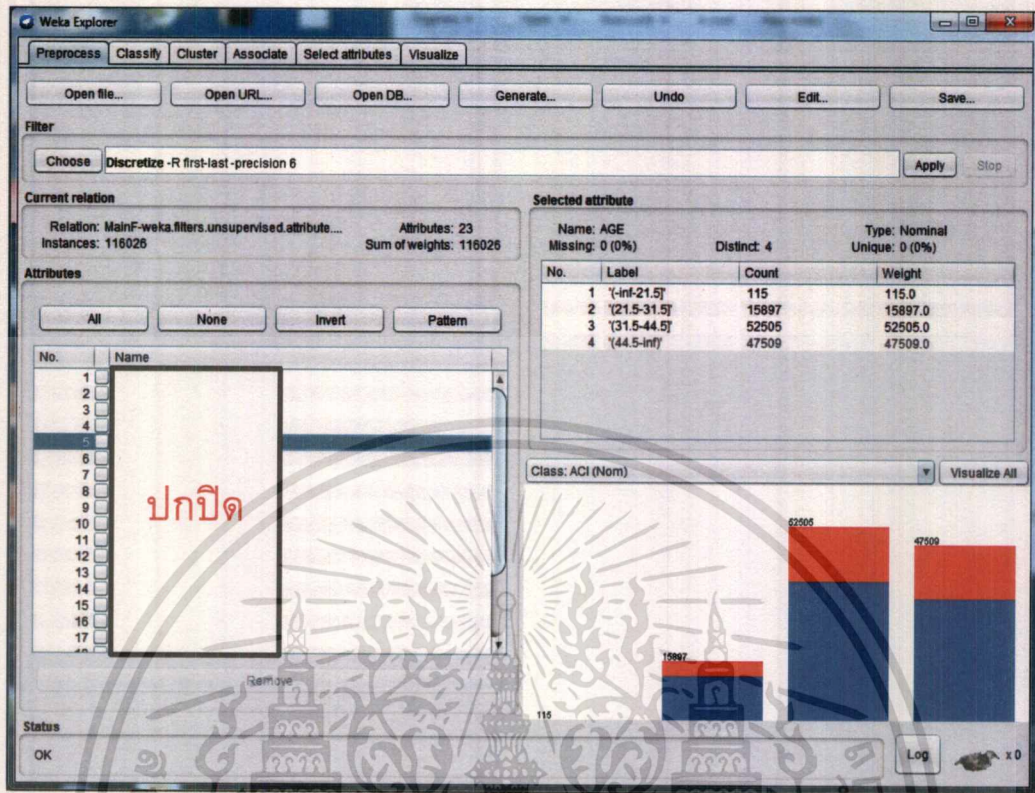
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



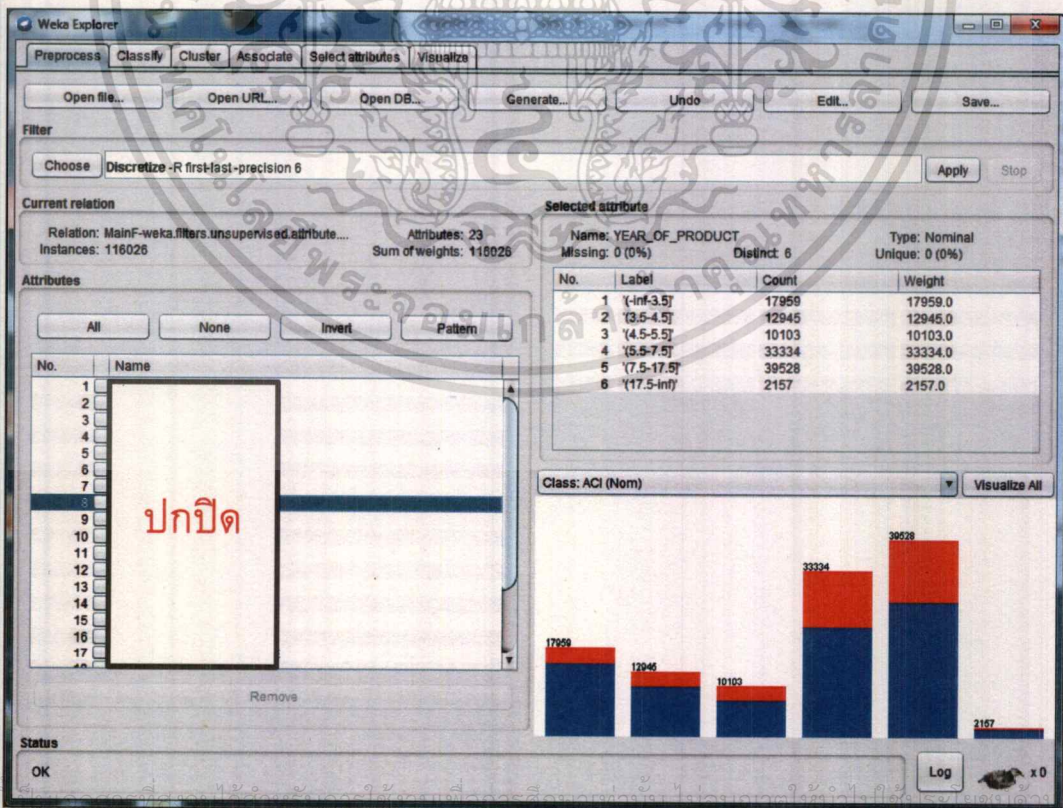
รูปที่ 3.11 การกระจายของข้อมูลอายุรถยนต์

ต่อไปจะเป็นในส่วนของการแบ่งช่วงข้อมูล(Discretize) โดยทางผู้ทำวิจัยได้เลือกใช้โปรแกรม Weka Explorer เข้ามาช่วยในขั้นตอนนี้ เพื่อเป็นการลดระยะเวลาในการปฏิบัติด้วย โดยในตอนนี้ทำการแปลงข้อมูลในแอตทริบิวต์ เพื่อลดความหลากหลายของข้อมูลในแอตทริบิวต์ รหัสไปรษณีย์ แต่ทางผู้ทำวิจัยรู้สึกว่าการแปลงข้อมูลของจังหวัดนั้นก็มีความหลากหลายมากเกินไปอยู่ที่ทางผู้ทำวิจัยได้เข้าขอคำปรึกษากับพี่เลี้ยงจึงได้คำตอบว่าให้เพิ่มข้อมูลที่เป็นตัวแทนความหลากหลายของข้อมูลนั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



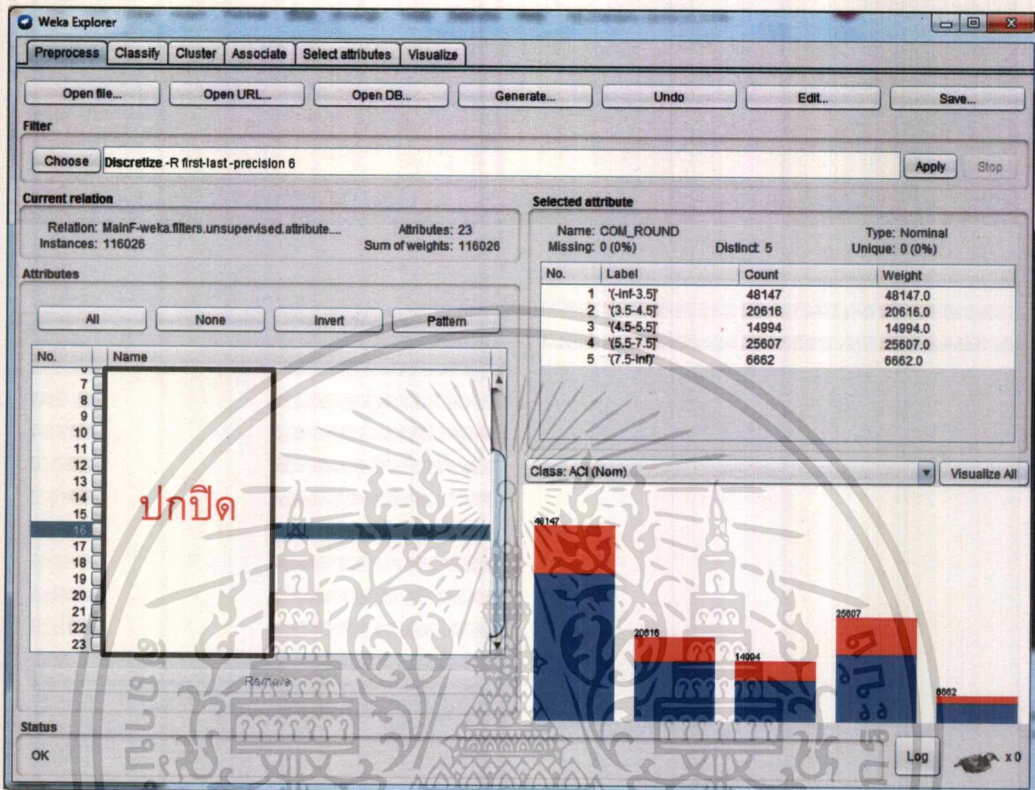
รูปที่ 3.12 การทำ Discretize ข้อมูลอายุ



รูปที่ 3.13 การทำ Discretize ข้อมูลอายุรถยนต์

เอกสารนี้จัดทำขึ้นเพื่อใช้ในการเรียนการสอนของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี การคัดลอกหรือเผยแพร่โดยไม่ได้รับอนุญาตถือว่าผิดกฎหมาย

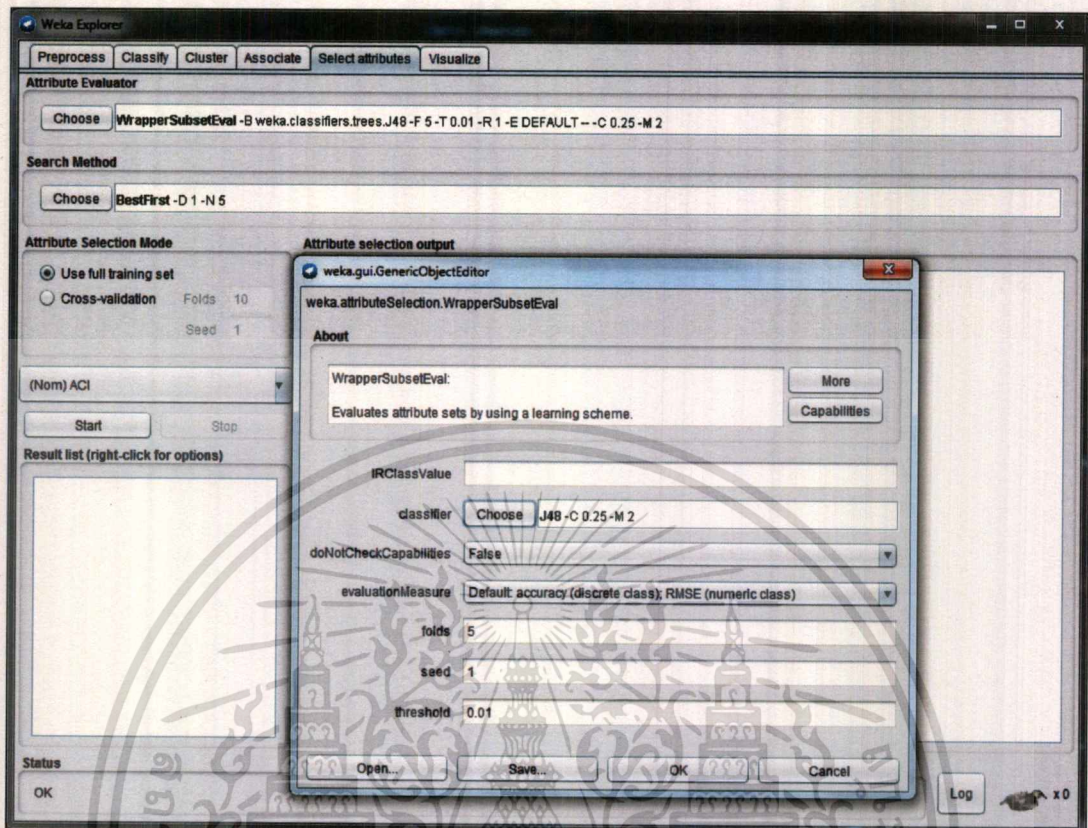
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้เด็ดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.14 การทำ Discretize ข้อมูลรอบที่ต่อประกันรถยนต์

ผลลัพธ์ที่ได้หลังจากการทำ Discretize ข้อมูลทำให้สามารถแบ่งข้อมูลออกเป็นขอบเขตและจำนวนที่ต้องการได้ เพื่อนำไปพัฒนาไปสู่โมเดล แมชชีนเลิร์นนิงต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.15 การตั้งค่าการทำ Feature Selection

ตั้งค่าการทำ Feature Selection ในโปรแกรม WEKA Explorer โดยเลือกใช้รูปแบบ WrapperSubsetEval ให้ทำการ classifier ด้วยโมเดล Decision Tree J48 จำนวน 5 Fold ผลลัพธ์ที่ได้ Feature ที่นำมาใช้ในการทำโมเดลแมชชีนเลิร์นนิงต่อไป

ซึ่งต้องทำความเข้าใจเกี่ยวกับข้อมูลและความเหมาะสมของโมเดลแมชชีนเลิร์นนิงที่เลือกใช้ด้วยเช่นกัน เพราะชนิดของข้อมูลจะมีผลอย่างมากต่อการเลือกโมเดลของแมชชีนเลิร์นนิงมาใช้งาน

### 3.3.3 การพัฒนาโมเดลแมชชีนเลิร์นนิง

โมเดลแมชชีนเลิร์นนิงที่เลือกนำมาใช้ในการพัฒนาสำหรับโครงการนี้นั้น ผู้จัดทำได้เลือกรูปแบบ การเรียนรู้แบบมีผู้สอน(Supervised Learning) ที่เป็นอัลกอริทึมสำหรับทำ การจัดหมวดหมู่ (Classification) เพื่อหาคำตอบสำหรับปัญหาซึ่งในที่นี้คือทำนายว่าลูกค้าคนใดที่จะสามารถเสนอขายประกันภัยรถยนต์ได้สำเร็จ โมเดลที่นำมาทดสอบใช้งานมี 3 ตัว คือ 1.Decision Tree 2.Naive Bayes 3.Neural Network โดยจะทำการวัดประสิทธิภาพของทั้ง 3 โมเดลแล้วจึงเลือกโมเดลที่ดีที่สุดมาใช้ จากนั้นจึงสร้างโมเดล Ensemble เพื่อให้สามารถทำการเปรียบเทียบคำตอบของทั้ง 3 โมเดล โดยมีประโยชน์ที่วิธีการนี้จะทำให้คำตอบที่ได้รับจากการทำนายนั้นมีความแม่นยำมากขึ้น โดยในขั้นตอนนี้จะทำโดยใช้งานภาษา Python และ Library ต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) ในขั้นตอนนี้จะทำการ one hot encoding ข้อมูลที่เป็น categorical ซึ่งเป็นการแตก feature เป็นหลายๆ column และมีค่าตอบเป็น 0 และ 1

	A	B	C	D	E
1	SEX	STATUS			
2	F	M			
3	M	M			
4	F	O			
5	F	M			
6	M	S			
7	F	S			
8					
9					
10					

รูปที่ 3.16 ตัวอย่างข้อมูลแบบ categorical

จึงทำการ one hot encoding ด้วย library pandas ในภาษา python

2) จากนั้นจึงนำไปเข้า machine learning ด้วย โมเดลที่เราเลือกไว้เช่น ต้นไม้ตัดสินใจ (Decision Tree) , การเรียนรู้แบบเบย์อย่างง่าย (Naive Bayes) ด้วย library scikit learn โดยจะเริ่มจาก Decision Tree ก่อน ซึ่งจะมีการแบ่งข้อมูลเป็น Train , Test set โดยใช้ Train 70% และ Test 30% และมีการหาค่าความแม่นยำ , confusion matrix ผ่าน scikit learn ด้วยเช่นกัน มีการตรวจสอบจำนวนข้อมูลที่ทำนายถูกต้อง และผิดเพื่อนำไปปรับปรุงโมเดลต่อไป

ต่อมาจึงทำการทำนายข้อมูลผ่าน Naive Bayes แต่เนื่องจาก Naive Bayes มีโมเดลอีก 3 ตัว คือ

1. Gaussian
2. Multinomial
3. Bernoulli

ได้ทำการทำนายโดยใช้ทั้ง 3 โมเดลและพบว่า Multinomial ให้ค่าความแม่นยำมากที่สุดสำหรับข้อมูลที่ต้องการ

3) เมื่อปรับแต่งโมเดล machine learning จาก scikit learn ได้ตามต้องการแล้ว จึงทำการ export โมเดลเพื่อนำมาใช้

4) โมเดลสุดท้ายที่เลือกมาใช้คือ Neural Network แต่เนื่องจาก scikit learn ไม่รองรับ Neural Network จึงใช้ library keras ที่มี back end เป็น tensorflow นั่นเอง

โดย tensorflow เป็น library ที่ได้รับความนิยมสูง แต่เนื่องจากไม่เป็นมิตรกับผู้ใช้งานใหม่ จึงมีการพัฒนา keras ขึ้นมาเป็น front end เพื่อให้ผู้ใช้งานใหม่สามารถใช้งานผ่าน keras ได้ไม่ยาก

จนเกินไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) เมื่อได้โมเดล Neural Network ที่พอใจแล้ว จึง export โมเดลเพื่อนำไปใช้ เนื่องจาก keras สามารถ export โมเดลได้ด้วย

6) เนื่องจากโมเดลเป็น python จึงต้องใช้ python web framework เพื่อทำการ rendering โมเดลและนำไปแสดงบนหน้าเว็บแอปพลิเคชัน โดยใช้ flask framework ร่วมกับ bootstrap

7) ทำการ export environment ทั้งหมดของโปรเจกต์นี้ เพื่อนำไป deploy บน server

8) จากนั้นจึงส่งไฟล์ทั้งหมดในโปรเจกต์เข้า server โดยผ่านโปรแกรม filezilla

9) ควบคุม server เพื่อให้ลง environment ตามที่เราต้องการ และสั่งให้รันสคริปต์เว็บแอปพลิเคชัน

### 3.3.4 การพัฒนาส่วนติดต่อผู้ใช้งาน

การพัฒนาส่วนติดต่อผู้ใช้งาน(User Interface) หลังจากได้โมเดลที่ดีที่สุด จึงมาทำในส่วนติดต่อผู้ใช้งานเพื่อให้ฝ่ายสินค้าธุรกิจสามารถใช้งานระบบได้อย่างสะดวก โดยจะใช้ HTML เพื่อสร้างหน้ารับข้อมูลจากผู้ใช้และหน้าแสดงผลลัพธ์ ตามที่ได้ออกแบบเอาไว้ในขั้นตอนก่อนหน้า แล้วจึงใช้ Bootstrap ซึ่งเป็น CSS Framework ในการตกแต่งหน้ารับข้อมูลให้ดูดีและน่าใช้งาน แล้วสุดท้ายจึงใช้ javascript ร่วมกับ Flask เพื่อรับ-ส่งข้อมูลจากผู้ใช้งานไปทำการประมวลผลลัพธ์ และเพื่อกระบวนการบันทึกเอกสารผลลัพธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 4

### ผลการวิจัยและอภิปรายผล

เนื้อหาในส่วนนี้จะกล่าวถึงผลจากการสร้างโมเดลการทำนายจาก machine learning ตามโมเดลที่ได้เลือกไว้ และส่วนของหน้าตาเว็บแอปพลิเคชันที่ติดต่อกับผู้ใช้ ซึ่งทั้งสองส่วนนั้นจะมีการนำมาใช้ร่วมกัน

โดยจะนำโมเดลมาวางบนระบบ โดยใช้ flask framework เป็นตัวกลางในการเรียกใช้โมเดลและส่งค่าไปยังหน้าเว็บแอป ซึ่งการสรุปผลการพัฒนาระบบจะถูกแบ่งเป็นหัวข้อต่างๆดังนี้

#### 4.1 การสอนโมเดล machine learning

นำข้อมูลการเสนอขายประกันที่ได้มาแปลงข้อมูลให้สามารถเข้าโมเดล machine learning ได้ โดยเลือกใช้ machine learning แบบ supervised learning และมีการใช้ 10k fold validation เพื่อให้ข้อมูลทั้งหมดได้ถูก train และ test ครบทุกส่วน โดยจะทำการ 10k fold validation ใน decision tree และ naive bayes

##### 4.1.1 Decision Tree

```
Confusion Matrix = [[8309 471]
 [2368 455]]
Confusion Matrix = [[7727 455]
 [2880 541]]
Confusion Matrix = [[7438 541]
 [2928 696]]
Confusion Matrix = [[7425 431]
 [3087 660]]
Confusion Matrix = [[7651 449]
 [2895 608]]
Confusion Matrix = [[7875 409]
 [2841 478]]
Confusion Matrix = [[8025 333]
 [2800 444]]
Confusion Matrix = [[8017 445]
 [2590 550]]
Confusion Matrix = [[8053 361]
 [2719 469]]
Confusion Matrix = [[7733 348]
 [3093 428]]
Accuracy Score For Each Round = Each Round
1 0.755322
2 0.712574
3 0.701026
4 0.696803
5 0.711799
6 0.719900
7 0.729960
8 0.738407
9 0.734529
10 0.703413
Accuracy Mean = 0.7203732337262521
TEST CLASSIFICATION RECORD
precision recall f1-score support
NO 0.71 0.96 0.82 8081
YES 0.55 0.12 0.20 3521
micro avg 0.70 0.70 0.70 11602
macro avg 0.63 0.54 0.51 11602
weighted avg 0.66 0.70 0.63 11602
---Runtime 13.2847318649292 seconds---
```

เอกสารนี้เป็นเอกสาร

ส่วนตัวด้านการค้า

รูปที่ 4.1 ผลลัพธ์การทำนาย Decision Tree

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมีเหตุดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใน decision tree มีค่า parameter ที่ต้องปรับเปลี่ยนหลายค่า จึงได้ทำการทดสอบโดยการปรับเปลี่ยนทีละ parameter และหาค่าความแม่นยำของโมเดล จนได้ค่าความแม่นยำที่น่าพอใจ

#### 4.1.2 Naive Bayes แบบ Multinomial

เนื่องจาก Naive bayes ใน scikit-learn มี 3 แบบ คือ Bernoulli , Gaussian และ Multinomial ซึ่งทางผู้จัดทำได้ทำการทดสอบทั้ง 3 โมเดลและพบว่า Multinomial ให้ค่าความแม่นยำมากที่สุดใน 3 โมเดลของ Naive bayes

```
Confusion Matrix = [[8267 513]
 [2436 387]]
Confusion Matrix = [[7672 510]
 [3013 408]]
Confusion Matrix = [[7390 589]
 [3132 492]]
Confusion Matrix = [[7363 493]
 [3287 460]]
Confusion Matrix = [[7624 476]
 [3053 450]]
Confusion Matrix = [[7779 505]
 [2953 366]]
Confusion Matrix = [[8082 276]
 [2993 251]]
Confusion Matrix = [[8260 202]
 [2963 177]]
Confusion Matrix = [[8271 143]
 [3037 151]]
Confusion Matrix = [[7965 116]
 [3372 149]]
Accuracy Score For Each Round = Each Round
1 0.745842
2 0.696372
3 0.679307
4 0.674222
5 0.695855
6 0.701974
7 0.718238
8 0.727202
9 0.725909
10 0.699362
Accuracy Mean = 0.7064282574501958
TEST CLASSIFICATION RECORD
```

	precision	recall	f1-score	support
NO	0.70	0.99	0.82	8081
YES	0.56	0.04	0.08	3521
micro avg	0.70	0.70	0.70	11602
macro avg	0.63	0.51	0.45	11602
weighted avg	0.66	0.70	0.60	11602

```
---Runtime 8.02885103225708 seconds ---
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับรูปที่ 4.2 ผลลัพธ์การทำนาย Naïve Bayes แต่ให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 4.1.3 Neural Network

```

92820/92820 [=====] - 15s 158us/step - loss: 0.3557 - acc: 0.8252
Epoch 985/1000
92820/92820 [=====] - 15s 159us/step - loss: 0.3573 - acc: 0.8282
Epoch 986/1000
92820/92820 [=====] - 15s 157us/step - loss: 0.3587 - acc: 0.8254
Epoch 987/1000
92820/92820 [=====] - 14s 156us/step - loss: 0.3567 - acc: 0.8275
Epoch 988/1000
92820/92820 [=====] - 15s 156us/step - loss: 0.3562 - acc: 0.8276
Epoch 989/1000
92820/92820 [=====] - 15s 159us/step - loss: 0.3559 - acc: 0.8266
Epoch 990/1000
92820/92820 [=====] - 15s 157us/step - loss: 0.3575 - acc: 0.8268
Epoch 991/1000
92820/92820 [=====] - 15s 159us/step - loss: 0.3564 - acc: 0.8259
Epoch 992/1000
92820/92820 [=====] - 15s 161us/step - loss: 0.3564 - acc: 0.8279
Epoch 993/1000
92820/92820 [=====] - 15s 160us/step - loss: 0.3579 - acc: 0.8267
Epoch 994/1000
92820/92820 [=====] - 15s 157us/step - loss: 0.3566 - acc: 0.8258
Epoch 995/1000
92820/92820 [=====] - 15s 156us/step - loss: 0.3557 - acc: 0.8278
Epoch 996/1000
92820/92820 [=====] - 15s 157us/step - loss: 0.3540 - acc: 0.8281
Epoch 997/1000
92820/92820 [=====] - 15s 158us/step - loss: 0.3572 - acc: 0.8278
Epoch 998/1000
92820/92820 [=====] - 15s 156us/step - loss: 0.3572 - acc: 0.8255
Epoch 999/1000
92820/92820 [=====] - 15s 157us/step - loss: 0.3589 - acc: 0.8256
Epoch 1000/1000
92820/92820 [=====] - 15s 160us/step - loss: 0.3571 - acc: 0.8270
Accuracy Score For Each Round = Each Round
1 0.688055
Accuracy Mean = 0.6880548134103249
Confusion Matrix = [[14225 2346]
 [4893 1742]]
      precision    recall  f1-score   support

     0       0.74       0.86       0.80      16571
     1       0.43       0.26       0.32       6635

 micro avg       0.69       0.69       0.69      23206
 macro avg       0.59       0.56       0.56      23206
weighted avg       0.65       0.69       0.66      23206

---Runtime 14817.473422050476 seconds ---

```

รูปที่ 4.3 Neural Network ที่ Epoch 1000

จากการใช้โมเดล Neural Network ที่ Dense ที่ต้องการ และ Epoch 1000 โดยใช้ Activation Function RELU พบว่า ค่าความแม่นยำของ Neural Network ให้ความแม่นยำมากที่สุดจากทุกโมเดล โดยรอบที่ Epoch 1000 มีค่าความแม่นยำ 0.8270 แต่โมเดลนี้จะใช้เวลาในการเทรนมากที่สุด เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

เช่นกัน  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยเมื่อได้โมเดลที่ต้องการแล้ว จะทำการบันทึกโมเดลและนำไปรวมเข้ากับส่วน back end ของ flask framework เพื่อทำเว็บแอปพลิเคชันต่อไป

#### 4.2 หน้าจอแสดงส่วนติดต่อผู้ใช้งานของเว็บแอปพลิเคชัน

ในส่วนหน้าแรกของเว็บจะให้ผู้ใช้ทำการป้อน input โดยใช้ไฟล์นามสกุล .csv และจะให้ผู้ใช้เลือกโมเดลที่ต้องการใช้งาน โดยจะมี 4 รูปแบบดังนี้

1. Neural Network ใช้โมเดล Neural Network ในการทำนายผลลัพธ์
2. Naive Bayes ใช้โมเดล Naive Bayes ในการทำนายผลลัพธ์
3. Decision Tree ใช้โมเดล Decision Tree ในการทำนายผลลัพธ์
4. Compare 3 Model คือการเรียกใช้โมเดลทั้ง 3 โมเดล ได้แก่ Decision Tree , Naive Bayes และ Neural Network มาใช้ในการทำนายผลลัพธ์

และหากผู้ใช้งานไม่ต้องการ input ด้วยไฟล์ .csv ก็สามารถป้อนค่า input ด้วยตนเองได้โดยการเลือกใช้งานในส่วน Manual Input

The image shows two screenshots of a web application interface. The top screenshot is titled "FILE INPUT" and features a "Choose File" button with "testDataPart1.csv" selected. Below it is a dropdown menu for selecting a model, with "NEURAL NETWORK" selected. A red text overlay reads "ข้อควรระวังก่อนใช้งาน กรุณาเตรียมไฟล์ข้อมูลดังนี้" (Caution before use, please prepare the data file as follows). Below this are two bullet points: "ข้อมูลที่จำเป็น: click" and "การเตรียมข้อมูล: click". The bottom screenshot is titled "MANUAL INPUT" and shows a form with various input fields. At the bottom of this form is a red "OK" button. The text "version 1.3" is visible at the very bottom of the page.

#### รูปที่ 4.4 หน้าแรกของผู้ใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้


ตัวอย่างการทำนายผลด้วยโมเดล Neural Network โดยเว็บจะทำนายโอกาสที่จะขายประกันสำเร็จเป็นเปอร์เซ็นต์ (แสดงเฉพาะลูกค้าที่ทำนายว่าขายสำเร็จ)

BACK

## RESULT

### Neural Network

รายชื่อลูกค้าที่ระบบทำนายว่าจะเสนอขายสำเร็จ



No	ACCOUNT	(%)
1	HN0000000015	61.254%
2	HP0000000023	66.133%
3	HP0000000028	76.880%
4	HN0000000030	99.989%
5	HN0000000031	98.469%
6	HN0000000032	81.820%
7	HP0000000046	86.557%
8	HP0000000064	73.729%
9	HP0000000065	72.508%
10	HN0000000085	99.290%
11	HN0000000115	81.618%
12	HN0000000116	85.389%
13	HN0000000121	56.267%
14	HP0000000131	55.904%
15	HN0000000140	100.000%
16	HN0000000145	99.946%

รูปที่ 4.5 การทำนายด้วยโมเดล Neural Network

หากผู้ใช้งานเลือก Compare 3 Model ระบบก็จะนำโมเดล Neural Network , Naive Bayes และ Decision Tree มาทำนายร่วมกัน โดยจะหารเฉลี่ยโอกาสในการขายสำเร็จและแสดงผลให้แก่ผู้ใช้งาน หากผู้ใช้งานอยากทราบผลลัพธ์ทั้งหมด ก็สามารถกดที่ปุ่ม csv มุมขวาบนของเว็บ เพื่อดาวน์โหลดไฟล์ผลลัพธ์ทั้งหมดในรูปแบบ .csv เพื่อนำไปใช้งานต่อไป


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

BACK

## RESULT

### Compare 3 Model

รายชื่อลูกค้าที่ระบบทำนายว่าจะเสนอขายสำเร็จ



NO.	ACCOUNT	NEURAL NETWORK	DECISION TREE	NAIVE BAYES	%
1	HP0000000028	YES	YES	NO	68.06
2	HP0000000046	YES	YES	NO	63.40
3	HN0000000140	YES	YES	NO	64.58
4	HP0000000218	YES	YES	NO	59.88
5	HP0000000252	YES	YES	YES	70.95
6	HP0000000253	YES	YES	YES	56.34
7	HP0000000264	YES	YES	NO	55.08
8	HP0000000281	YES	YES	NO	56.62
9	HP0000000296	YES	YES	NO	73.20
10	HP0000000337	YES	YES	NO	62.87
11	HP0000000396	YES	YES	NO	56.60
12	HP0000000397	YES	YES	NO	63.26
13	HP0000000399	NO	YES	YES	53.42
14	HP0000000404	YES	NO	YES	64.64
15	HP0000000434	YES	YES	NO	49.67
16	HP0000000472	YES	YES	NO	62.19

รูปที่ 4.6 การทำนายผลลัพธ์โดยใช้การเปรียบเทียบ 3 โมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 5

# สรุปผลการวิจัยและข้อเสนอแนะ

### 5.1 สรุปผลการวิจัย

จากการพัฒนาโมเดล machine learning และเว็บแอปพลิเคชันของระบบทำนายการเสนอขายประกันภัยรถยนต์ด้วยแมชชีนเลิร์นนิง สามารถสรุปผลการดำเนินงานได้ดังนี้

- เว็บแอปพลิเคชันและส่วนเรียกใช้งานโมเดล machine learning สามารถทำงานร่วมกันได้เมื่อนำไปติดตั้งบนเซิร์ฟเวอร์
- เว็บแอปพลิเคชันถูกออกแบบมาสอดคล้องกับความต้องการในการใช้งานของผู้ใช้งาน
- การใช้งานข้อมูลเป็นไปตามที่กำหนดขององค์กร

### 5.2 ข้อจำกัดของระบบ

จากความสามารถของเว็บแอปพลิเคชันที่ได้กล่าวมาข้างต้นนั้นมีความสามารถที่ยังเป็นข้อจำกัดดังนี้

- เนื่องจากมีการใช้ one hot encoding ซึ่งเป็นการแยก features ของข้อมูลออกเป็นคอลัมน์ตามจำนวน feature ทั้งหมด ทำให้เกิดปัญหาคือ ถ้า feature ของเว็บแอปพลิเคชันและของข้อมูลไม่ตรงกัน จะทำให้ระบบไม่สามารถทำนายได้ เนื่องจากจำนวนคอลัมน์ไม่ตรงกัน

### 5.2 ข้อเสนอแนะ

เนื่องจากระบบนี้เป็นระบบที่ถูกพัฒนาขึ้นมาใหม่โดยไม่มีข้อมูลอ้างอิง จึงอาจทำให้เกิดปัญหาได้ดังนี้

- ในส่วนของ features มีปัญหาในการเก็บข้อมูล เนื่องจากมีข้อมูลซ้ำซ้อนกันจากการพิมพ์ผิดได้เช่น Toyota Vios และ Toyota Vioss ซึ่งทั้งสองนี้เป็นข้อมูลเดียวกัน
- เนื่องจากไม่มีการทำ data engineering มาจากต้นทาง อาจทำให้ข้อมูลคลาดเคลื่อนได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



## เอกสารอ้างอิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Potiwarakorn B. 2015. Machine Learning เรียนอะไร, รู้ไปทำไม . [Online]. Available :  
<https://medium.com/o-v-e-r-f-i-t-t-e-d/machine-learning-เรียนอะไร-รู้ไปทำไม-4a1dcae85b96>

จิราวรรณ รอนราญ. 2016. การทำงานในรูปแบบ Data Mining. [Online]. Available :  
<https://erp.mju.ac.th/articleDetail.aspx?qid=502>

Eakasit P. 2014. การสร้างโมเดล classification ด้วย Weka Explorer. [Online]. Available :  
<http://dataminingtrend.com/2014/building-weka-classification-model/>

Eakasit P. 2014. การสร้างโมเดล Ensemble แบบต่างๆ. [Online]. Available :  
<http://dataminingtrend.com/2014/data-mining-techniques/ensemble-model/>

Eakasit P. 2014. การแบ่งข้อมูลเพื่อนำทดสอบประสิทธิภาพของโมเดล. [Online]. Available :  
<http://dataminingtrend.com/2014/data-mining-techniques/cross-validation/>

Eakasit P. 2014. โมเดล Naive Bayes และการแปลความหมาย. [Online]. Available :  
<http://dataminingtrend.com/2014/naive-bayes/>

Eakasit P. 2014. ขั้นตอนการสร้างโมเดล Decision Tree. [Online]. Available :  
<http://dataminingtrend.com/2014/decision-tree-model/>

Suphan F. 2018. Machine Learning ตอนที่ 2: Neural Network แบบจำลองของสมอง?. [Online].

Available : <http://codeonthehill.com/machine-learning-2-neural-network/>

Suphan F. 2018. Machine Learning ตอนที่ 4: เตรียมข้อมูลเพื่อสร้าง model. [Online]. Available :

<http://codeonthehill.com/machine-learning-4-data-set/>

ผศ.วิภาวรรณ บัวทอง. 2014. Data Preprocessing. [Online]. Available :

<https://wipawanblog.files.wordpress.com/2014/06/chapter-3-data-preprocessing.pdf>

CORALINE CO. LTD. 2017. สถิติเบื้องต้นง่ายๆ ที่จะช่วยให้คุณเข้าใจการวิเคราะห์มากขึ้น (ตอนที่ 2).

[Online]. Available : [https://medium.com/@info\\_46914/สถิติเบื้องต้นง่ายๆ-ที่จะช่วยให้คุณเข้าใจการวิเคราะห์มากขึ้น-ตอนที่-2-1f94b6664ede](https://medium.com/@info_46914/สถิติเบื้องต้นง่ายๆ-ที่จะช่วยให้คุณเข้าใจการวิเคราะห์มากขึ้น-ตอนที่-2-1f94b6664ede)

Boyd BigData RPG. 2018. Confusion Matrix|Recall|Precision|Accuracy เขี้ยมันคืออะไร?

[Online]. Available : <https://medium.com/bigdataeng/confusion-matrix-recall-precision-accuracy-เขี้ยมันคืออะไร-aa-fc3e1fcfc579>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Nuttavut T. 2017. [Machine Learning#2] รู้จักการจำแนกประเภทข้อมูลด้วย k-Nearest Neighbors.

[Online]. Available : <https://www.babelcoder.com/blog/posts/k-nearest-neighbors>

MounirMesselmeni. 2018. Reading CSV file with Javascript and HTML5 File API.

[Online]. Available : <https://github.com/MounirMesselmeni/html-fileapi>

Codebee. 2016. MVC คืออะไร ทำความเข้าใจรูปแบบการเขียนโปรแกรม

[Online]. Available : <https://www.codebee.co.th/labs/mvc-คืออะไร-ทำความเข้าใจ/>

Git-scm. 2014. เริ่มต้นใช้งาน - เกี่ยวกับ Version Control

[Online]. Available : <https://git-scm.com/book/th/v1/เริ่มต้นใช้งาน-เกี่ยวกับ-Version-Control>

Git-scm. 2014. เริ่มต้นใช้งาน - ประวัติย่อของ Git

[Online]. Available : <https://git-scm.com/book/th/v1/เริ่มต้นใช้งาน-ประวัติย่อของ-Git>

Git-scm. 2014. เริ่มต้นใช้งาน - Git ขั้นพื้นฐาน

[Online]. Available : <https://git-scm.com/book/th/v1/เริ่มต้นใช้งาน-Git-ขั้นพื้นฐาน>

Onlymycode. 2017. Git version control UI SourceTree

[Online]. Available : <http://onlymycode.blogspot.com/2017/05/git-version-control-ui-sourcetree.html>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



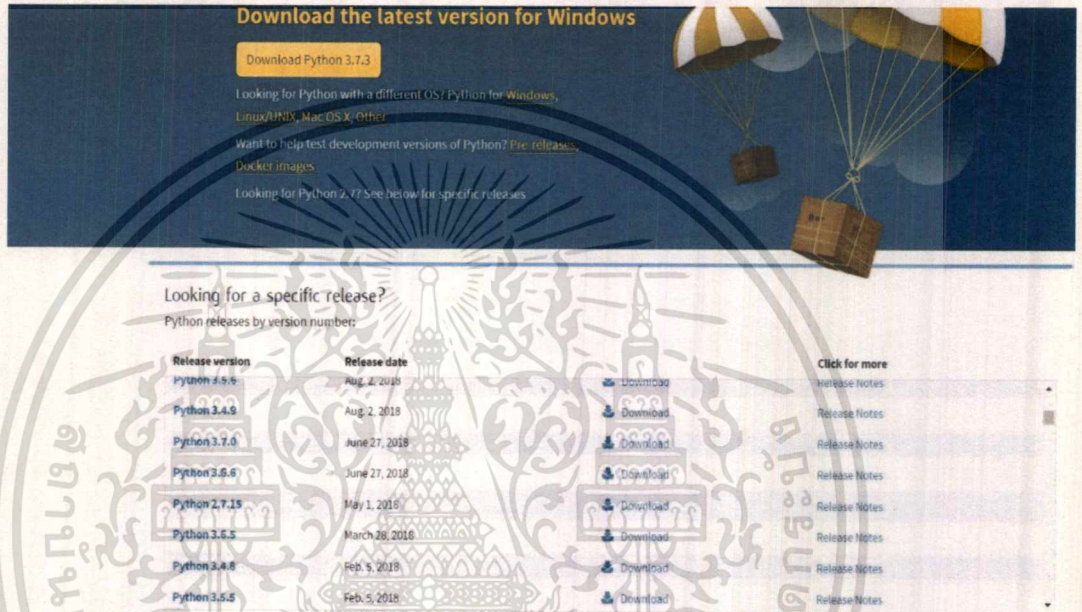
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ภาคผนวก ก

# การเตรียมเครื่องมือเพื่อใช้ในการพัฒนาระบบ

### ก.1 ติดตั้ง Python

- 1) ดาวน์โหลดตัวติดตั้ง Python จาก <https://www.python.org/downloads/> โดยสามารถเลือกเวอร์ชันที่ต้องการได้ (ในที่นี้จะเลือกใช้เวอร์ชัน 3.6.6 64 bits)



รูปที่ ก.1 เลือกเวอร์ชัน Python จากหน้าเว็บไซต์ Python

- 2) เลือก Customize installation

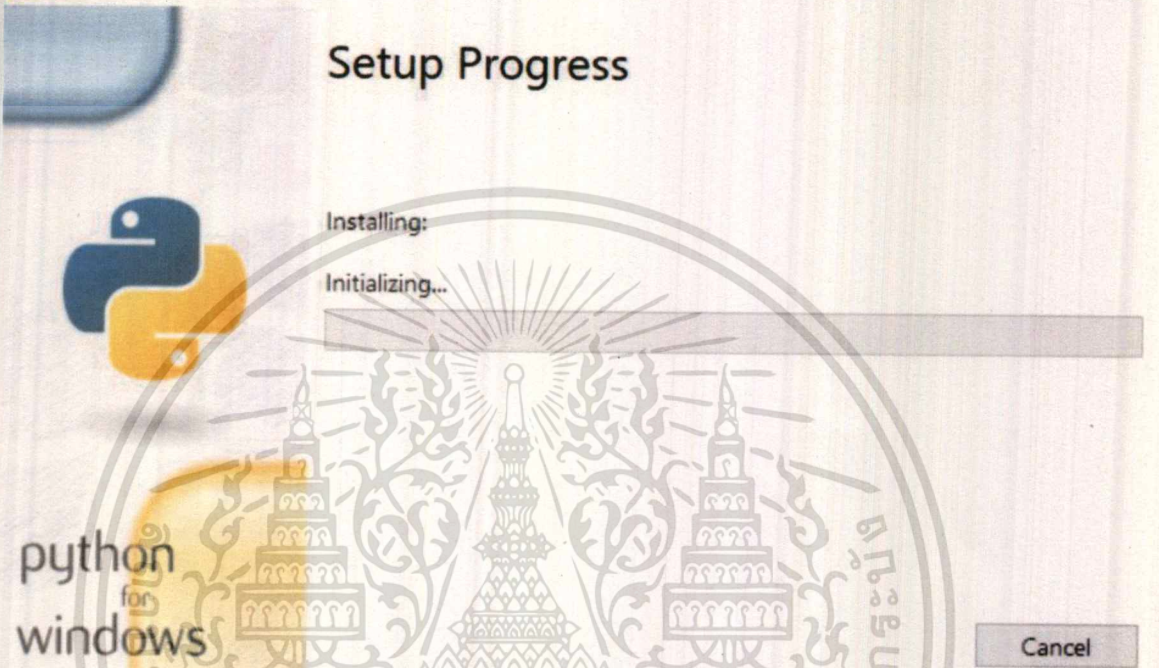


รูปที่ ก.2 หน้าต่างแสดงการติดตั้ง Python 3.6.6 (64-bit)

เอกสารนี้เป็นเอกสารที่ส่งมอบให้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านธุรกิจ  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมีเหตุดัดแปลงเนื้อหาและต้องอ้างอิงถึงชื่อของเอกสารทุกครั้งที่มีการนำไปใช้

5) รอติดตั้งจนเสร็จ

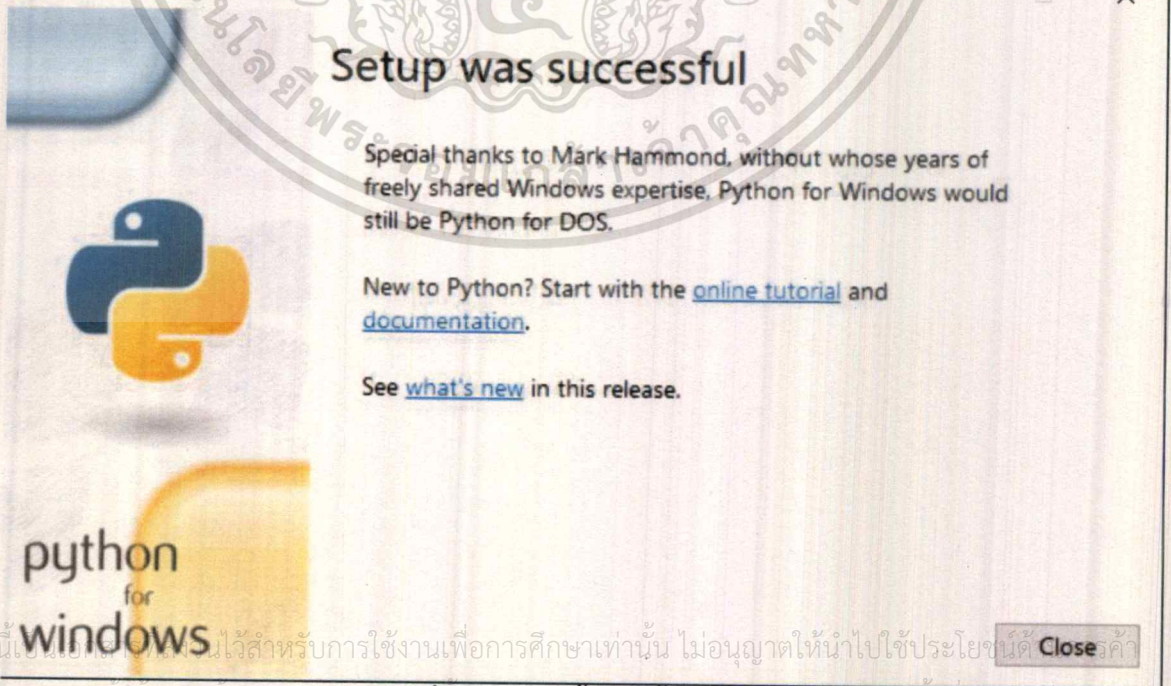
Python 3.6.6 (64-bit) Setup



รูปที่ ก.5 ระหว่างรอการติดตั้ง

6) การติดตั้งเสร็จสิ้น

Python 3.6.6 (64-bit) Setup



เอกสารนี้เป็นเอกสารที่จัดทำขึ้นไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์อื่นใด

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงหรือทำซ้ำโดยไม่ได้รับอนุญาตจากเจ้าของเอกสารทุกครั้งหากมีการนำไปใช้

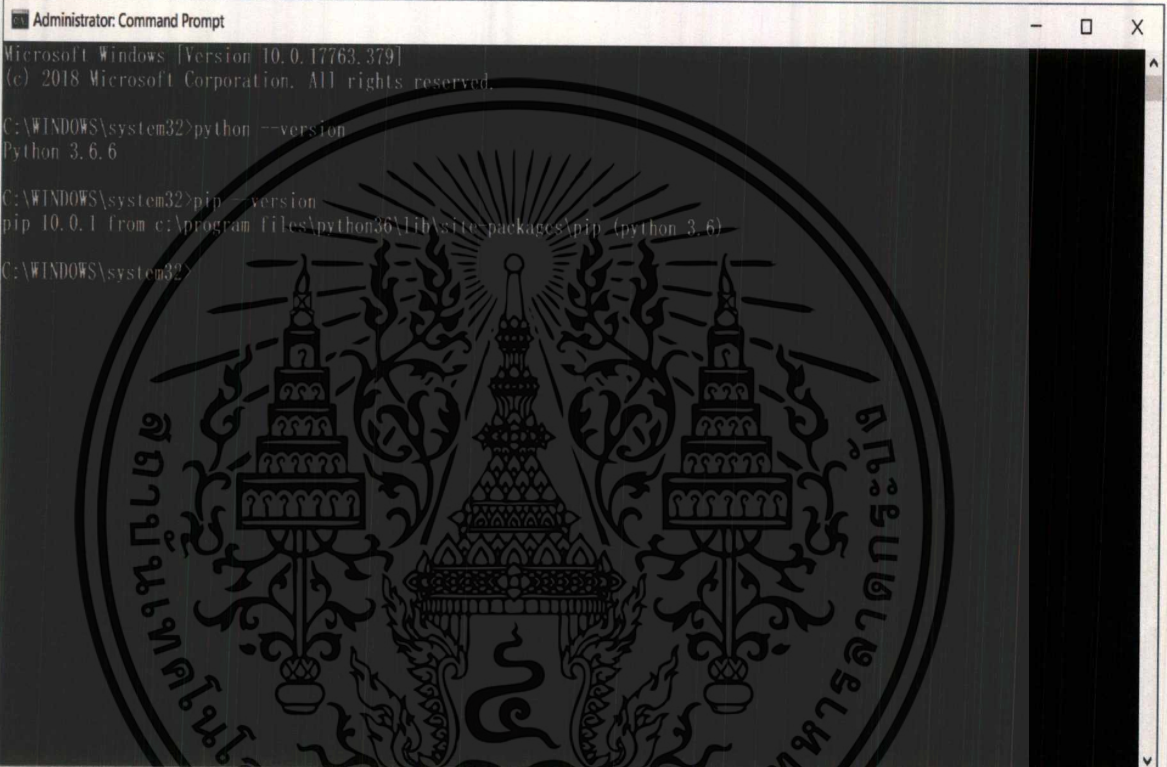
รูปที่ ก.6 การติดตั้งเสร็จสิ้น

7) ทดสอบการติดตั้ง Python โดยเปิดโปรแกรม Command Prompt

8) พิมพ์คำสั่ง python --version เพื่อทดสอบว่า python ได้ถูกติดตั้งตรงตามเวอร์ชันที่ต้องการ  
พิมพ์คำสั่ง pip --version เพื่อตรวจสอบเวอร์ชันของ pip (Python Install Package) ที่ใช้

ในการ

ดาวน์โหลด library ต่างๆ ของ python



```

Administrator: Command Prompt
Microsoft Windows [Version 10.0.17763.379]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\WINDOWS\system32>python --version
Python 3.6.6

C:\WINDOWS\system32>pip --version
pip 10.0.1 from c:\program files\python36\lib\site-packages\pip (python 3.6)

C:\WINDOWS\system32>
  
```

รูปที่ ก.7 ทดสอบการติดตั้งของ Python

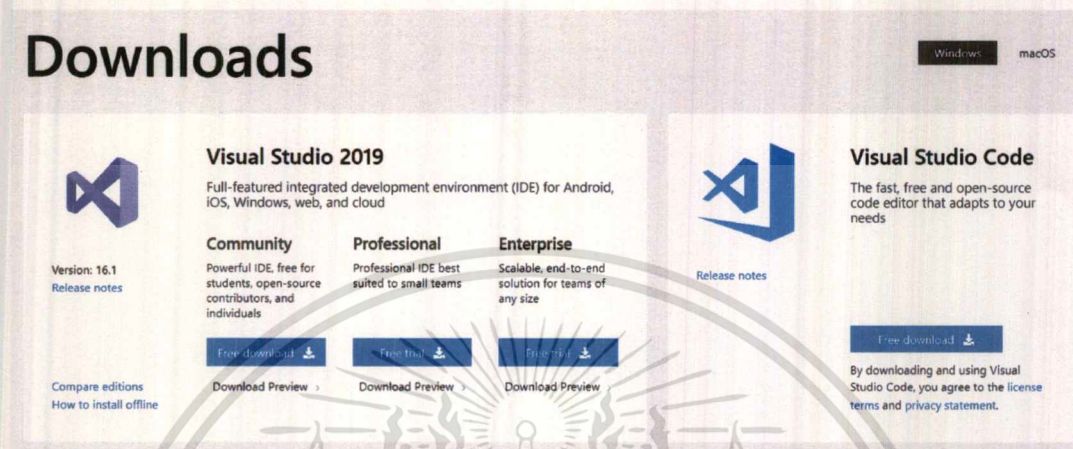
9) ตัวอย่างการติดตั้ง library ต่างๆของ Python โดยผ่านคำสั่ง pip install (ในตัวอย่างคือ library pandas)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



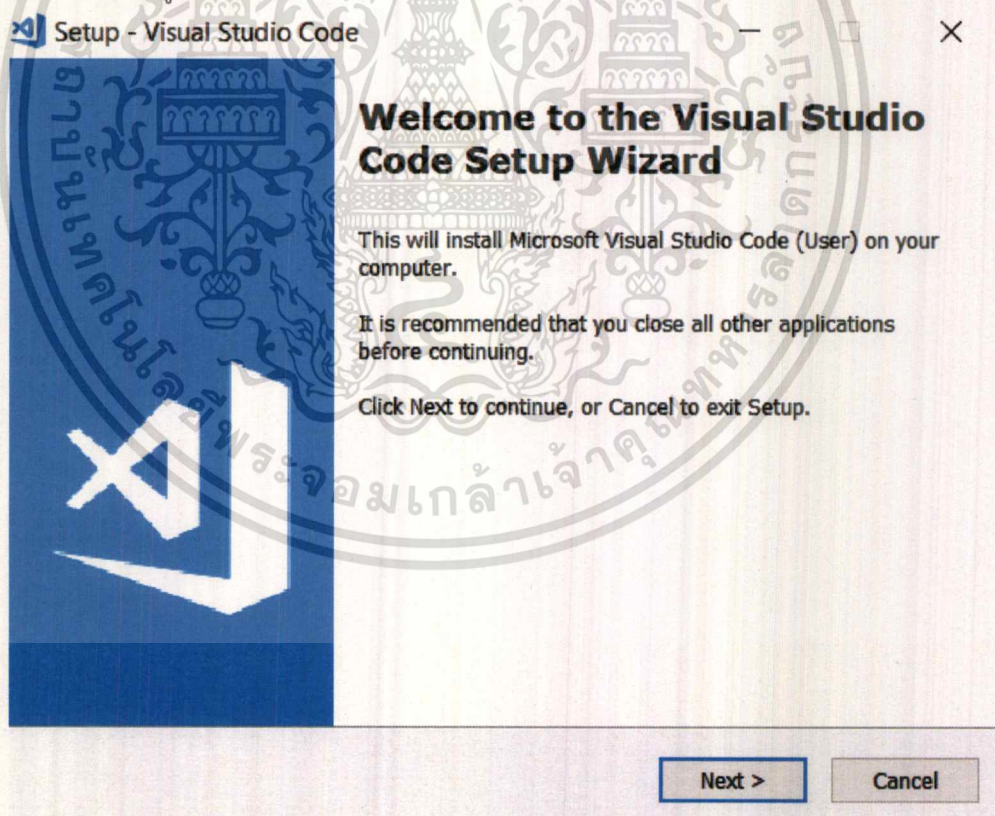
## ก.2 ติดตั้ง Visual Studio Code

1) เข้าไปที่เว็บไซต์ <https://visualstudio.microsoft.com/downloads/> โดยเลือก Visual Studio Code และกด Free Download



รูปที่ ก.9 การดาวน์โหลด Visual Studio Code

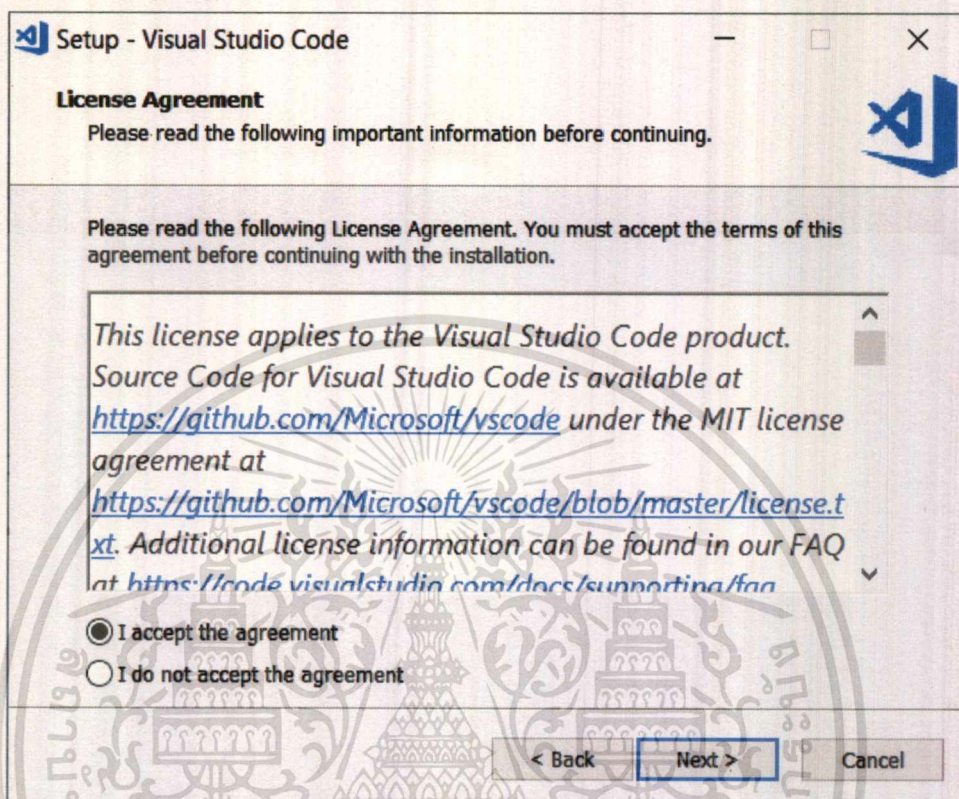
2) หน้าจอเข้าสู่การติดตั้ง



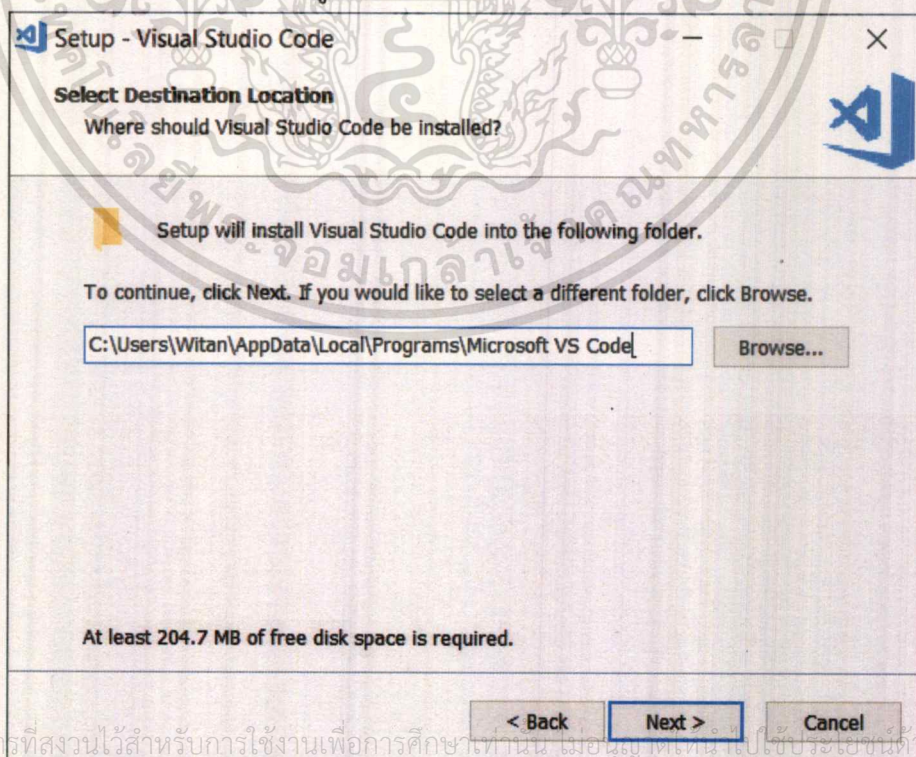
รูปที่ ก.10 การเข้าสู่การติดตั้ง Visual Studio Code

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 3) คลิก Next เพื่อเข้าสู่หน้าถัดไป  
 4) ตี๊ก I accept agreement และกด Next

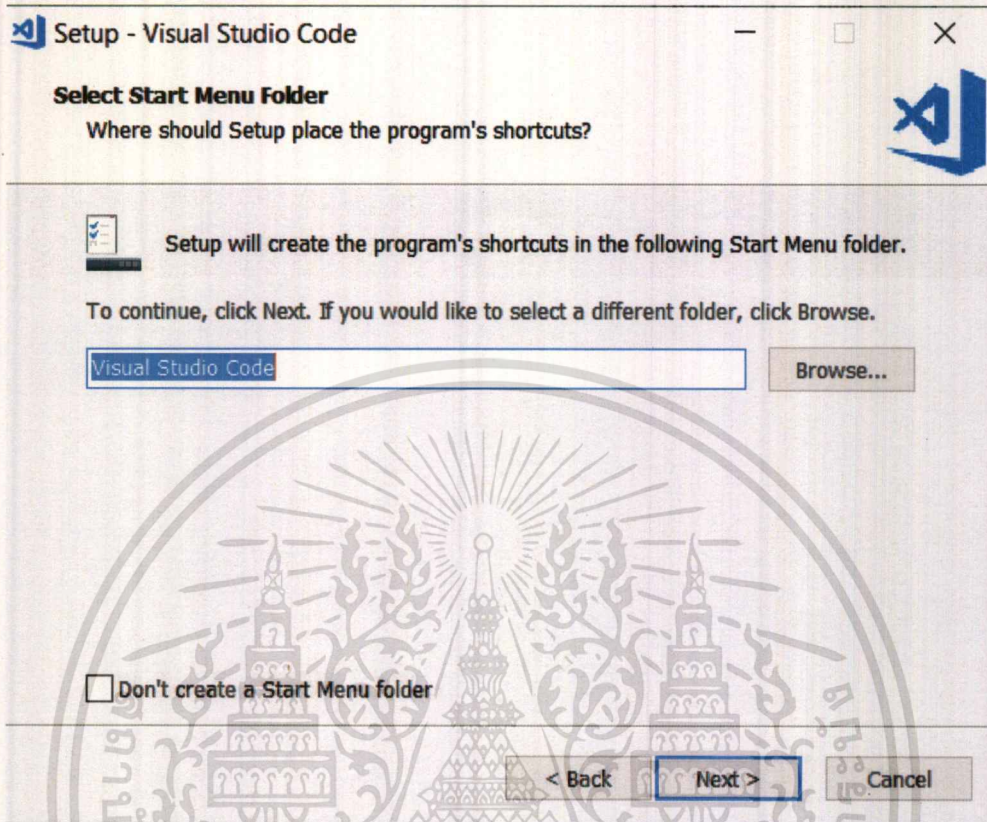


รูปที่ ก.11 ขั้นตอนการติดตั้ง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออยู่ในที่เห็นไปใช้ประโยชน์ด้านการค้า  
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิรูปที่ ก.12 ขั้นตอนในการเลือกพื้นที่ติดตั้งโปรแกรมทุกครั้งที่มีการนำไปใช้

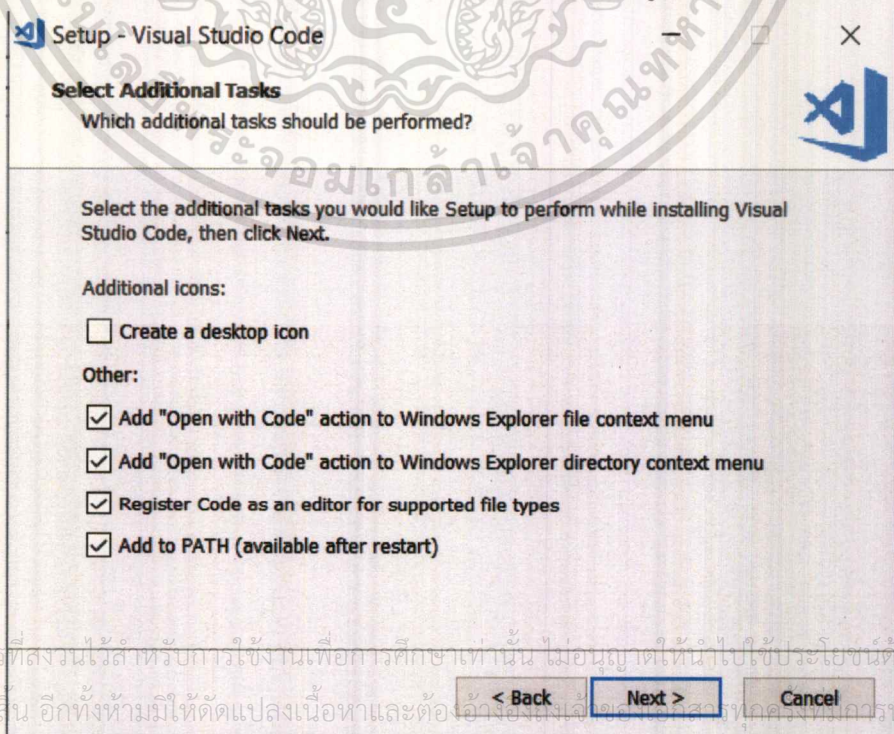
5) เมื่อเลือกพื้นที่ในการติดตั้งแล้วจึงกด Next



รูปที่ ก.13 การติดตั้ง

6) เลือก Next

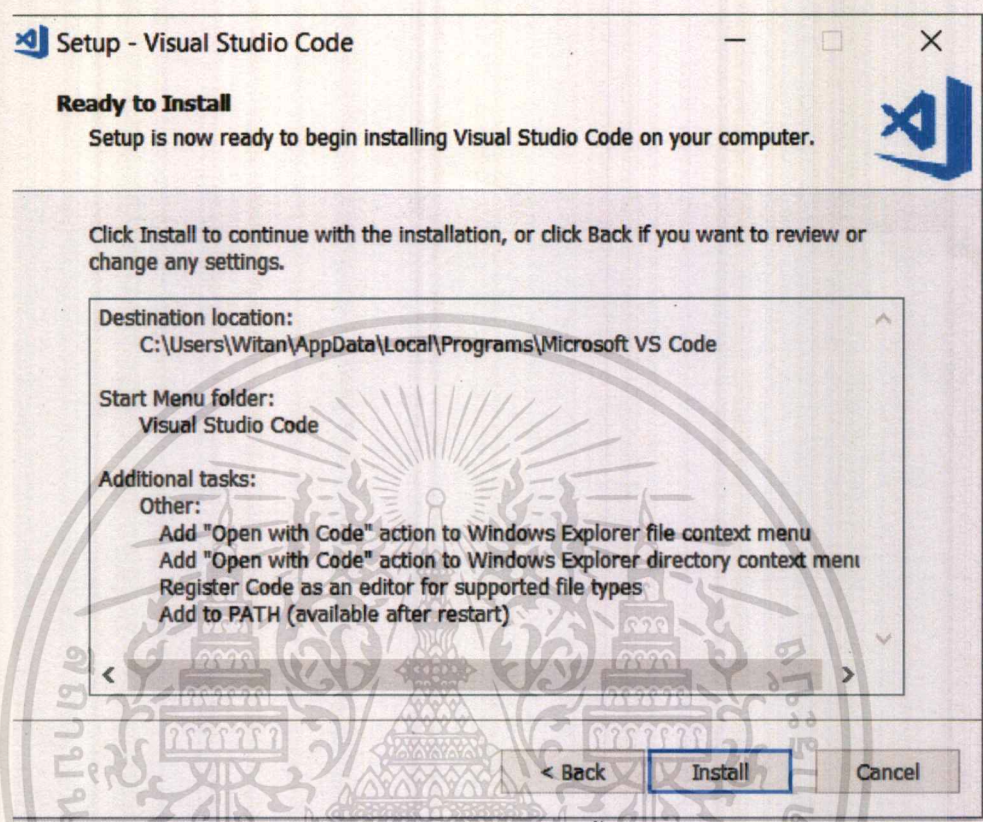
7) เลือกการติดตั้งเพิ่มเติมตามที่ต้องการ (Add to PATH) สำคัญมาก



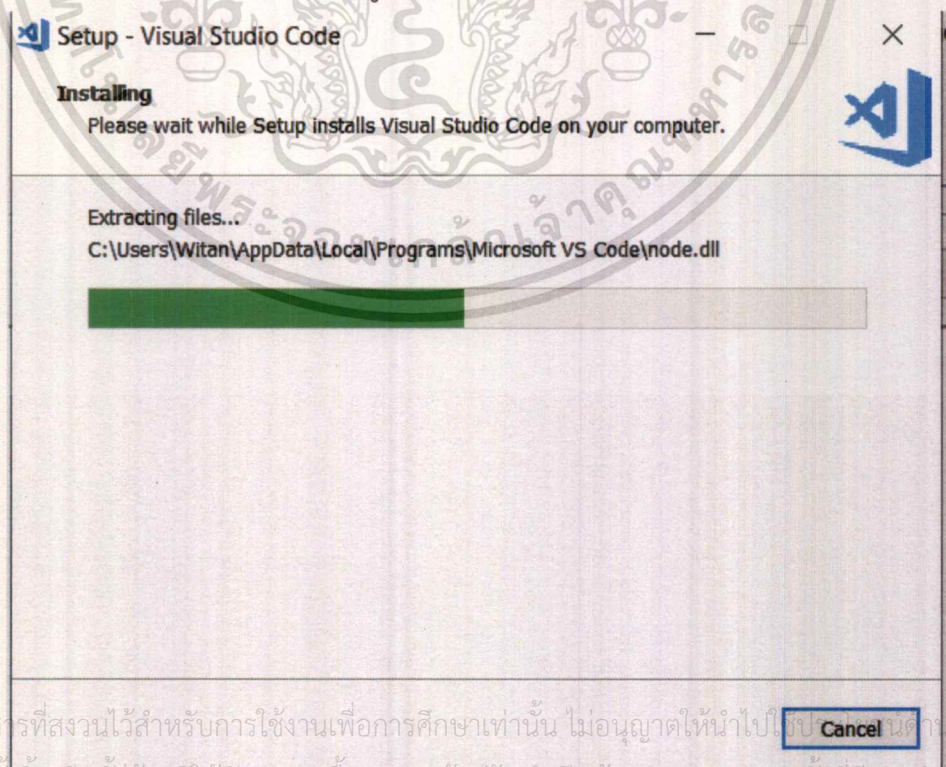
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงแหล่งที่มาของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ ก.14 การติดตั้งเพิ่มเติม

8) กด Install



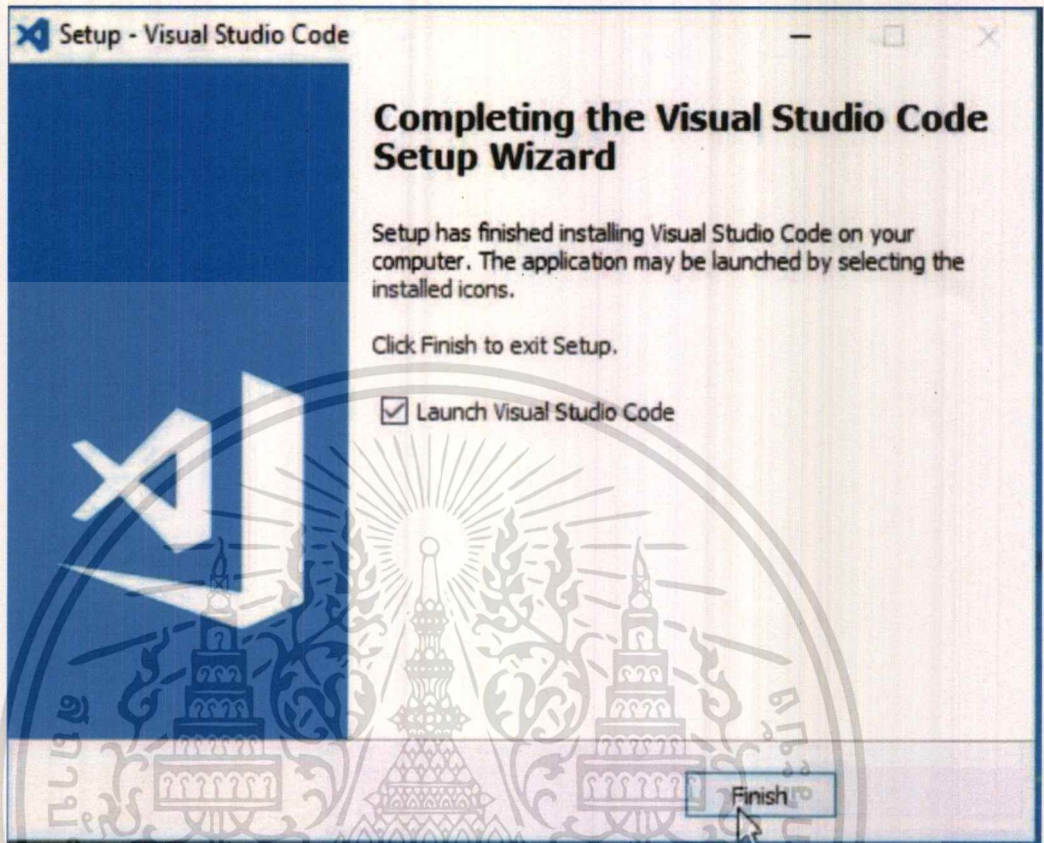
รูปที่ ก.15 การติดตั้ง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ในทางการค้า  
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและแจ้งอ้างถึงถึงผู้ซึ่งเอกสารทุกครั้งที่มีการนำไปใช้

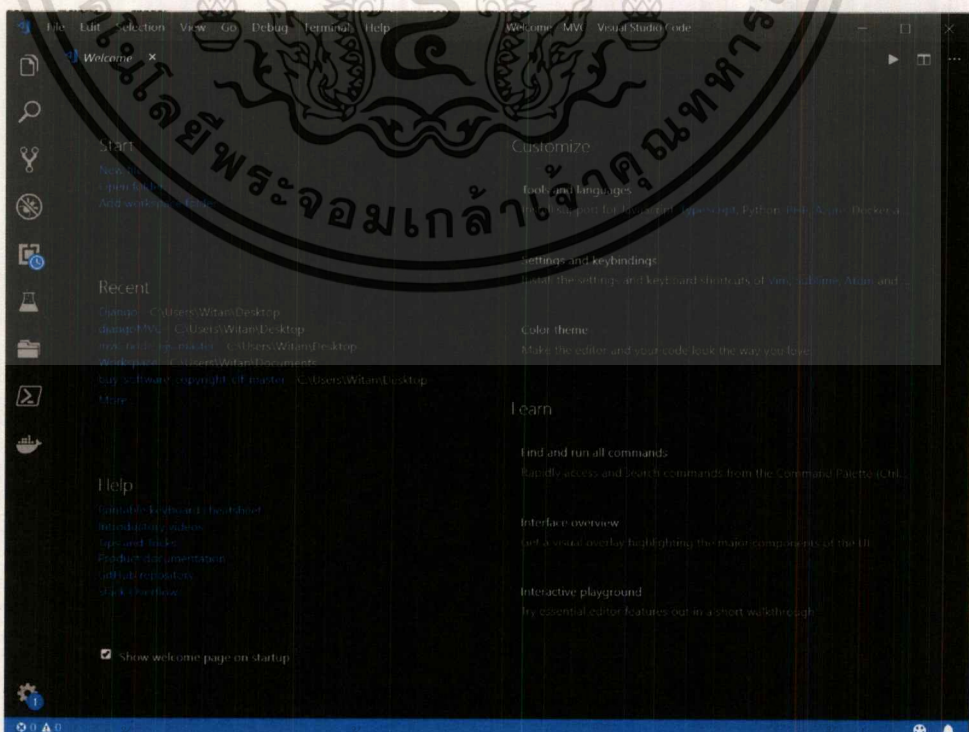
รูปที่ ก.16 อยู่ในระหว่างการติดตั้ง

9) กด Finish



รูปที่ ก.17 การติดตั้งเสร็จสิ้น

10) หน้าแรกของโปรแกรม



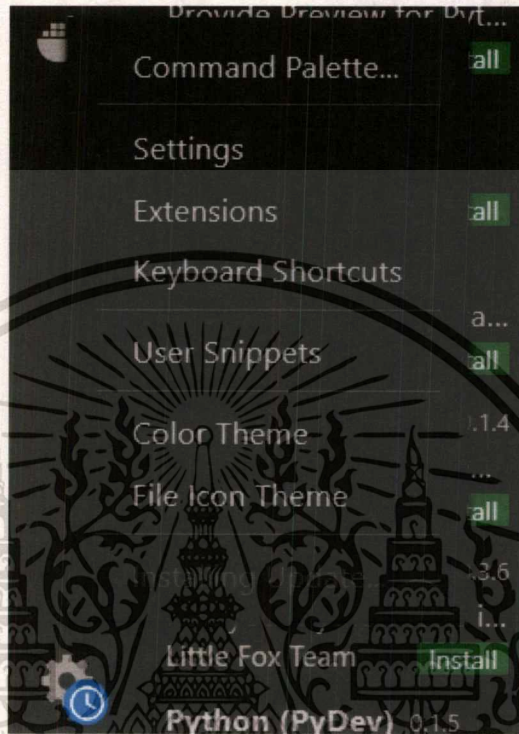
รูปที่ ก.18 หน้าแรกของโปรแกรม

เอกสารนี้เป็นเอก

การคำ

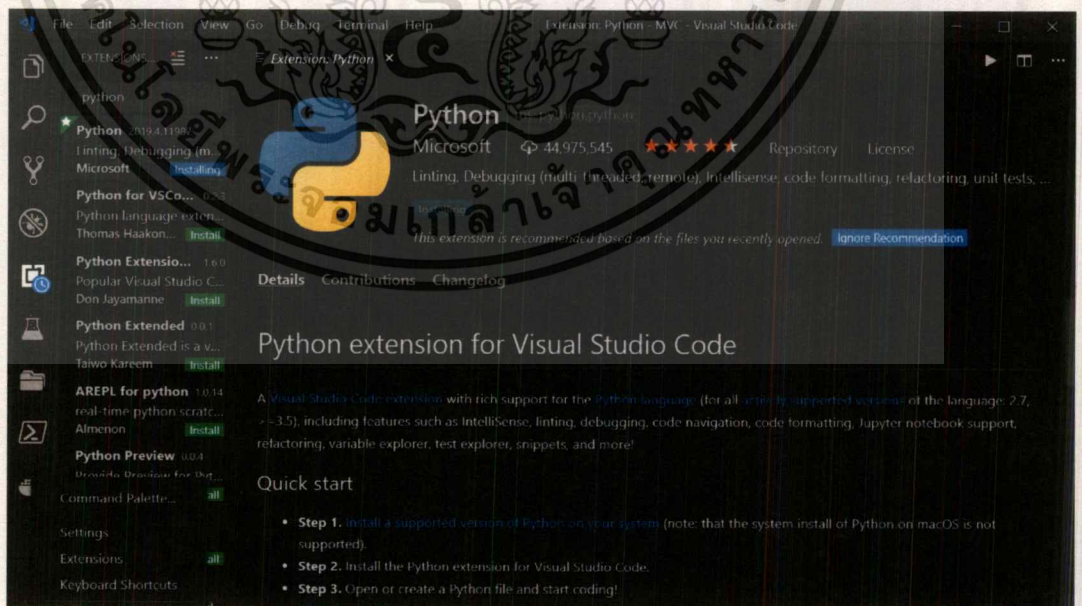
ไม่ว่ากรณีใดๆ หงสน ออกกฎหมายเขตดแปลงนทการลางคองรณลางคองเอกสารทุกครั้งทมการนาไปใช้

11) โดยต้องทำการติดตั้ง Extensions เพิ่มเติมเพื่อให้ Visual Studio Code สามารถใช้งานได้อย่างถูกต้อง เช่น ลง Extension ของ Python เพื่อให้ Visual Studio Code สามารถเขียนภาษา Python ได้



รูปที่ ก.19 หน้าการตั้งค่า

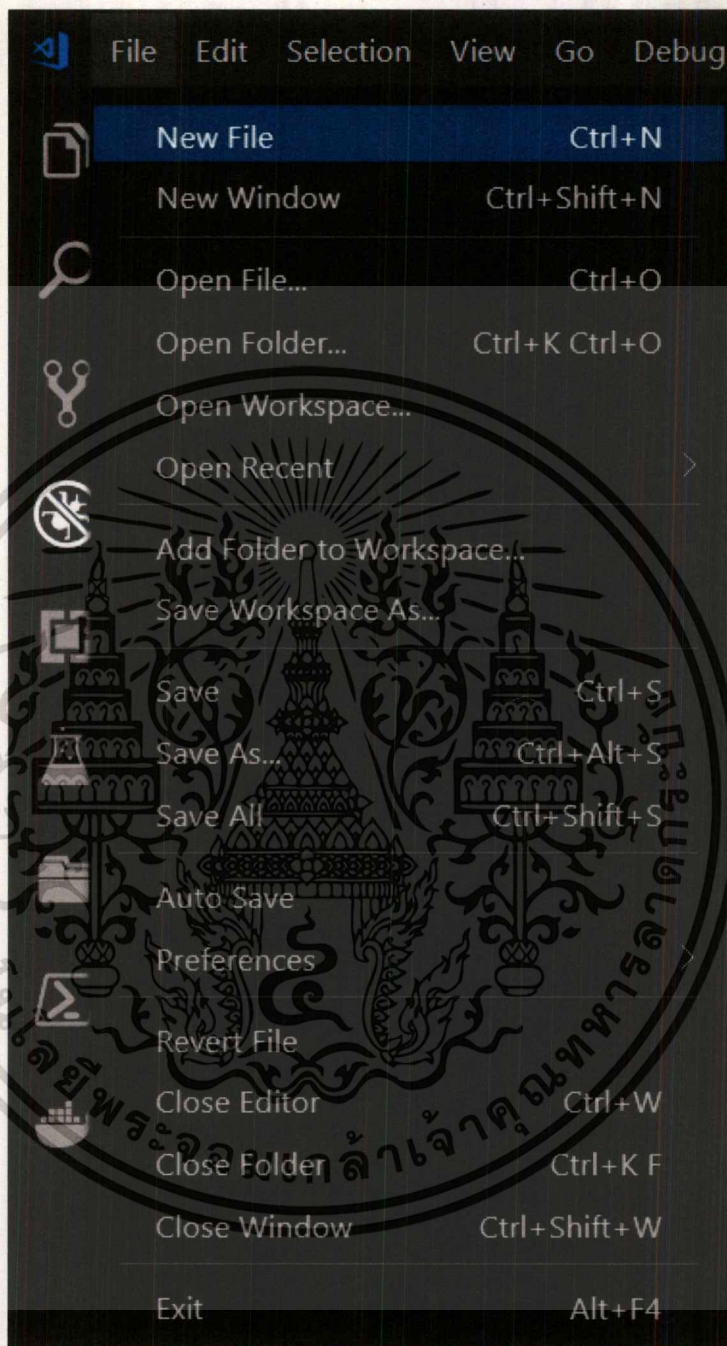
12) เสิร์ช python ในช่องการค้นหา และกด Install จากนั้นจะขึ้นสถานะ Installing



รูปที่ ก.20 หน้าการติดตั้ง Extensions

เอกสารนี้เป็นเอกสารที่ 13) หลังจากนี้ยังสามารถติดตั้ง Extensions อื่นๆเพิ่มได้ตามความต้องการใช้งาน โยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

14) เริ่มต้นเขียนโปรแกรมด้วย Visual Studio Code โดยการกดที่ File > New File

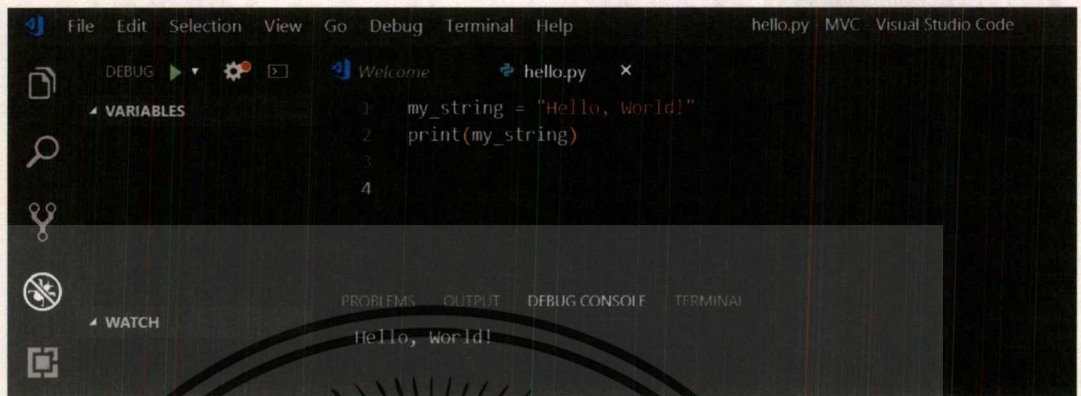


รูปที่ ก.21 การเริ่มต้นสร้างไฟล์ในการเขียนโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

15) เมื่อเขียนโปรแกรมแล้วต้องการรัน ให้กด Debug > Start Debugging หรือกด F5

16) ได้ผลลัพธ์ดังรูป



```

File Edit Selection View Go Debug Terminal Help hello.py MVC Visual Studio Code
DEBUG
VARIABLES
1 my_string = "Hello, World!"
2 print(my_string)
3
4
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL
Hello, world!

```

รูปที่ ก.22 การเขียนโปรแกรมภาษา Python ด้วย Visual Studio Code



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### ก.3 การติดตั้งโปรแกรม Filezilla

1) เข้าไปที่เว็บไซต์ <https://filezilla-project.org/> และเลือก Download FileZilla Client

**FileZilla** The free FTP solution

Home

**FileZilla**  
Features  
Screenshots  
Download  
Documentation  
FileZilla Pro

**FileZilla Server**  
Download

**Community**  
Forum  
Project page  
Wiki

**General**  
FAQ  
Support  
Contact  
License  
Privacy Policy  
Trademark Policy

**Development**  
Source code  
Nightly builds  
Translations  
Version history  
Changelog

Promotion:  
**FileZilla Pro**  
S3, Google Drive - Cloud, OneDrive

## Overview

Welcome to the homepage of FileZilla®, the free FTP solution. The *FileZilla Client* not or free of charge under the terms of the GNU General Public License.

We are also offering *FileZilla Pro*, with additional protocol support for WebDAV, Amazon Storage, and Google Cloud Storage.

Last but not least, *FileZilla Server* is a free open source FTP and FTPS Server.

Support is available through our [forums](#), the [wiki](#) and the [bug and feature request track](#)

In addition, you will find documentation on how to compile FileZilla and nightly builds for

◆ **Quick download links**

**Download FileZilla Client**  
All platforms

**Download FileZilla Server**  
Windows only

Pick the client if you want to transfer files. Get the server if you want to make file

รูปที่ ก.23 เว็บไซต์ FileZilla

2) กด Download FileZilla Client

## Download FileZilla Client for Windows (64bit)

The latest stable version of FileZilla Client is 3.42.1

Please select the file appropriate for your platform below.


◆ **Windows (64bit)** 🇺🇸

**Download FileZilla Client**

This installer may include bundled offers. Check below for more options.

The 64bit versions of Windows 7, 8, 8.1 and 10 are supported.

◆ **More download options**

Other platforms: 

Not what you are looking for?

เอกสารนี้เป็น  [Show additional download options](#) ารศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดรูปที่ ก.24 การดาวน์โหลด FileZilla ของเอกสารทุกครั้งที่มีการนำไปใช้

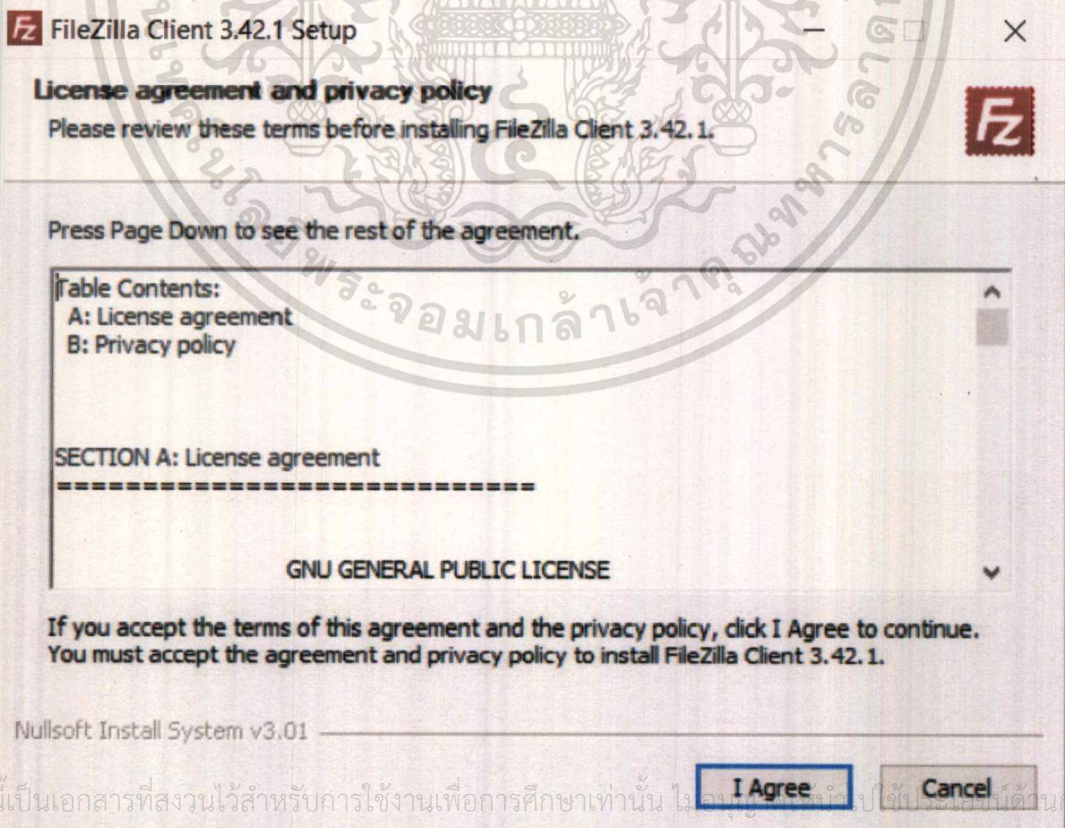
3) กด Download ตรง FileZilla

## Please select your edition of FileZilla Client

	FileZilla Pro	FileZilla
Standard FTP	Yes	Yes
FTP over TLS	Yes	Yes
SFTP	Yes	Yes
Amazon S3	Yes	-
Backblaze B2	Yes	-
Dropbox	Yes	-
Microsoft OneDrive	Yes	-
Google Drive	Yes	-
Google Cloud Storage	Yes	-
Microsoft Azure Blob and File Storage	Yes	-
WebDAV	Yes	-
OpenStack Swift	Yes	-

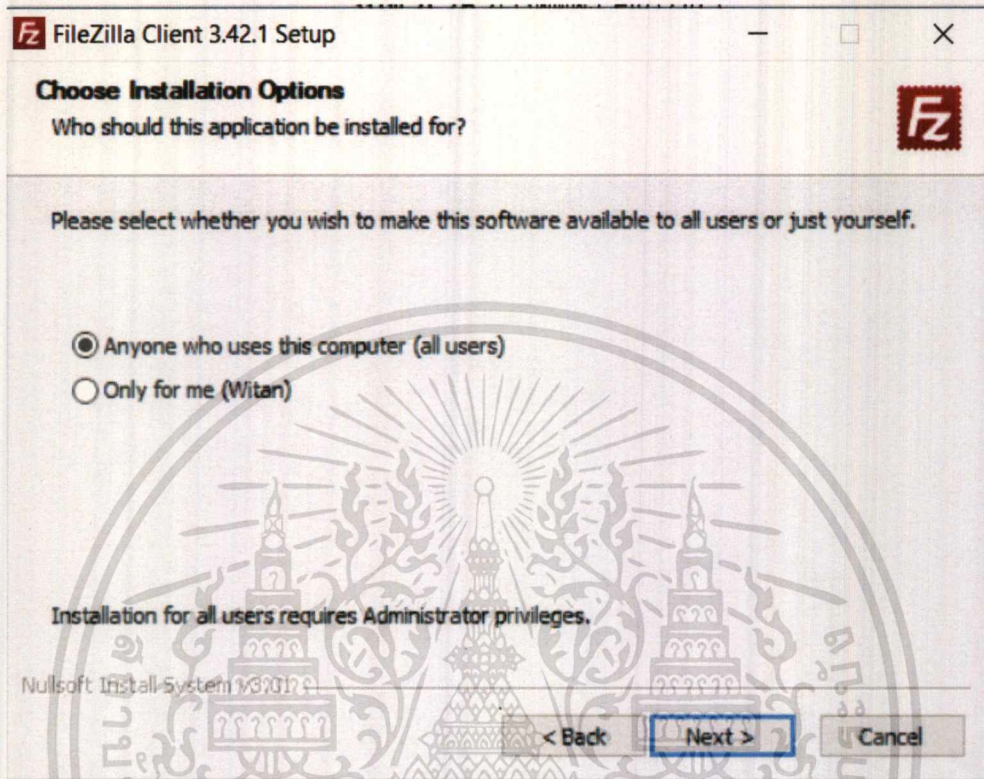
รูปที่ ก.25 การดาวน์โหลด FileZilla

4) เมื่อเข้าสู่ขั้นตอนการติดตั้งให้กด I agree



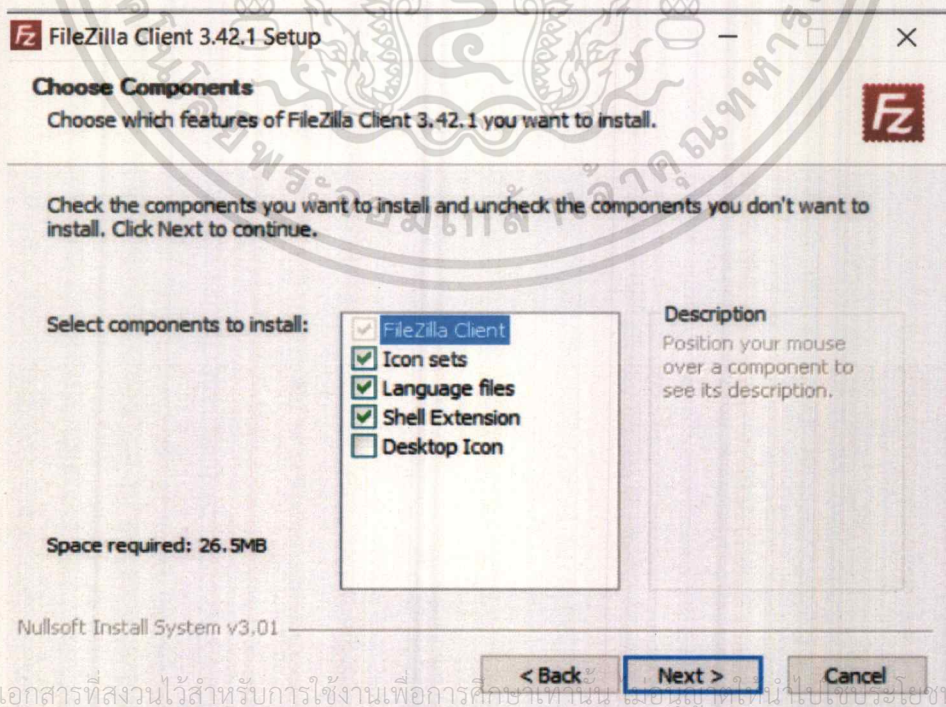
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่สามารถนำออกจำหน่ายหรือใช้เพื่อการค้า  
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเอกสารนี้ไปเผยแพร่โดยไม่ได้รับอนุญาต  
 รูปที่ ก.26 การติดตั้ง FileZilla

5) เลือกประเภทการ Install ว่าต้องการแบบไหน ในที่นี้เลือก Anyone who uses this computer (all users) และกด Next



รูปที่ ก.27 การติดตั้งโปรแกรม

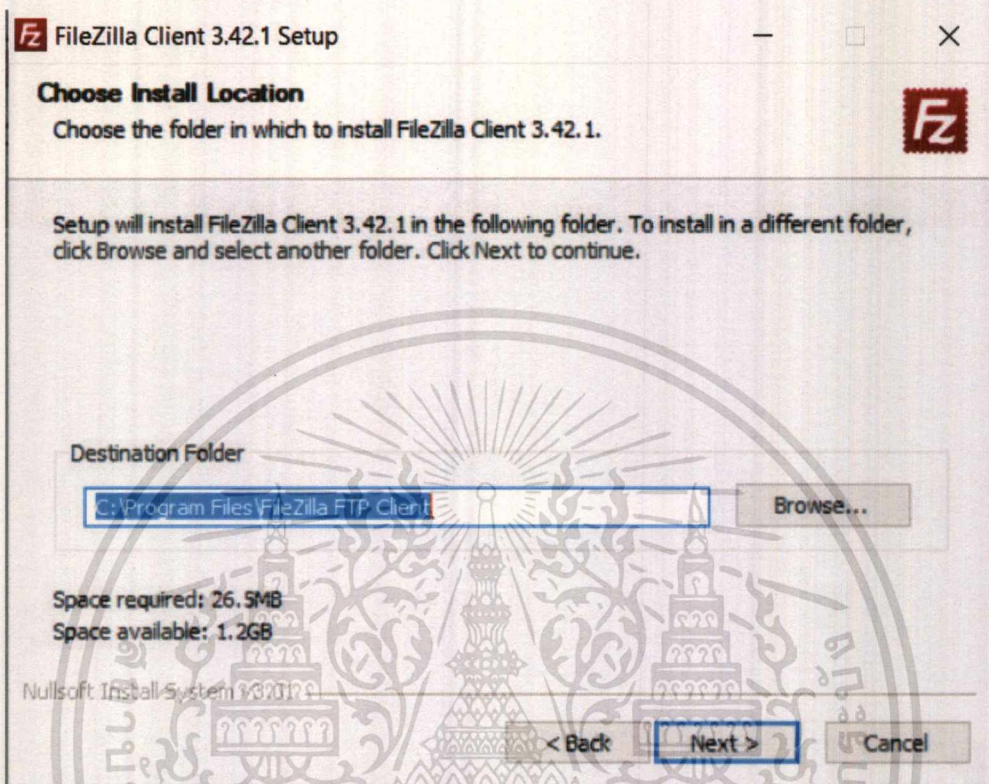
6) เลือก Component ที่ต้องการ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อผู้จัดทำนำไปเผยแพร่โดยไม่ขอขออนุญาต  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดรูปที่ ก.28 การเลือก component ของเอกสารทุกครั้งที่มีการนำไปใช้

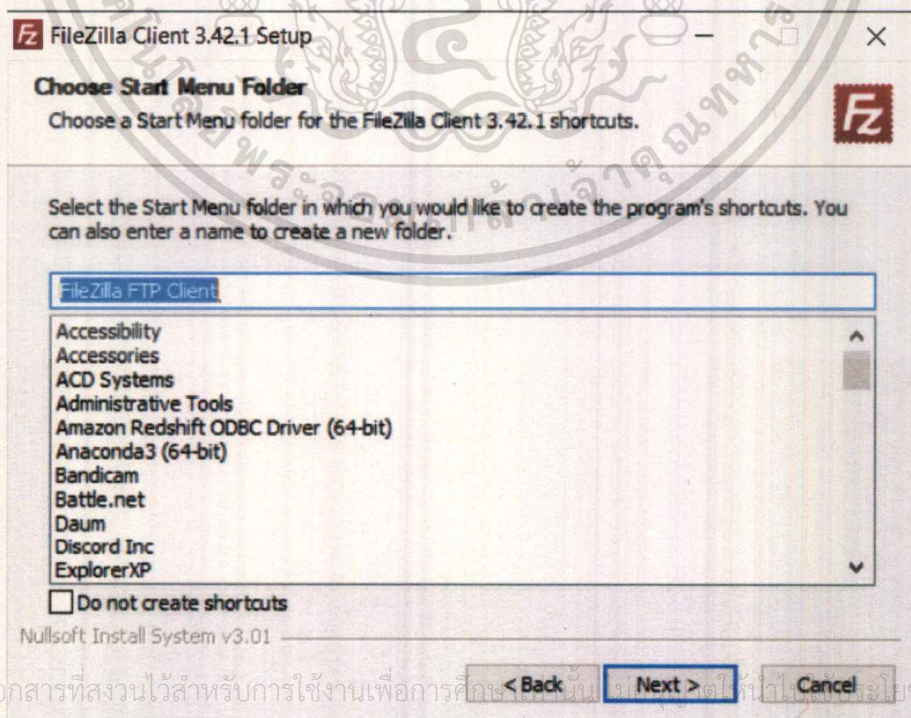
7) กด Next

8) เลือกที่ติดตั้งโปรแกรม



รูปที่ ก.29 การเลือกที่ติดตั้งโปรแกรม

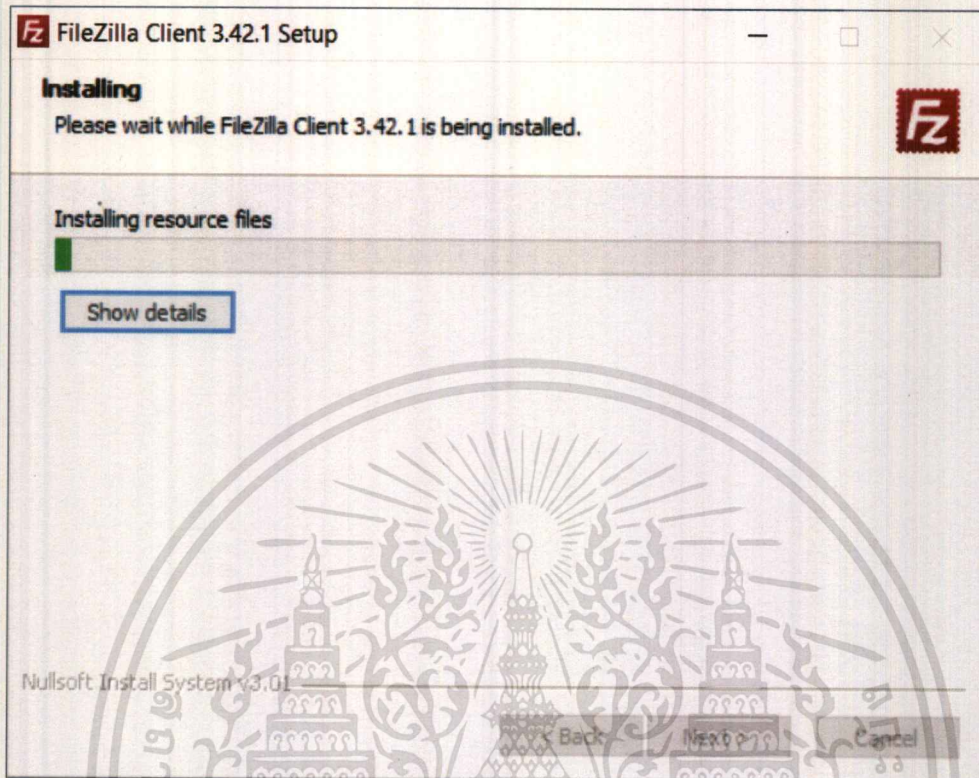
9) กด Next



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่สามารถนำออกเผยแพร่โดยไม่ได้รับอนุญาตจากเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

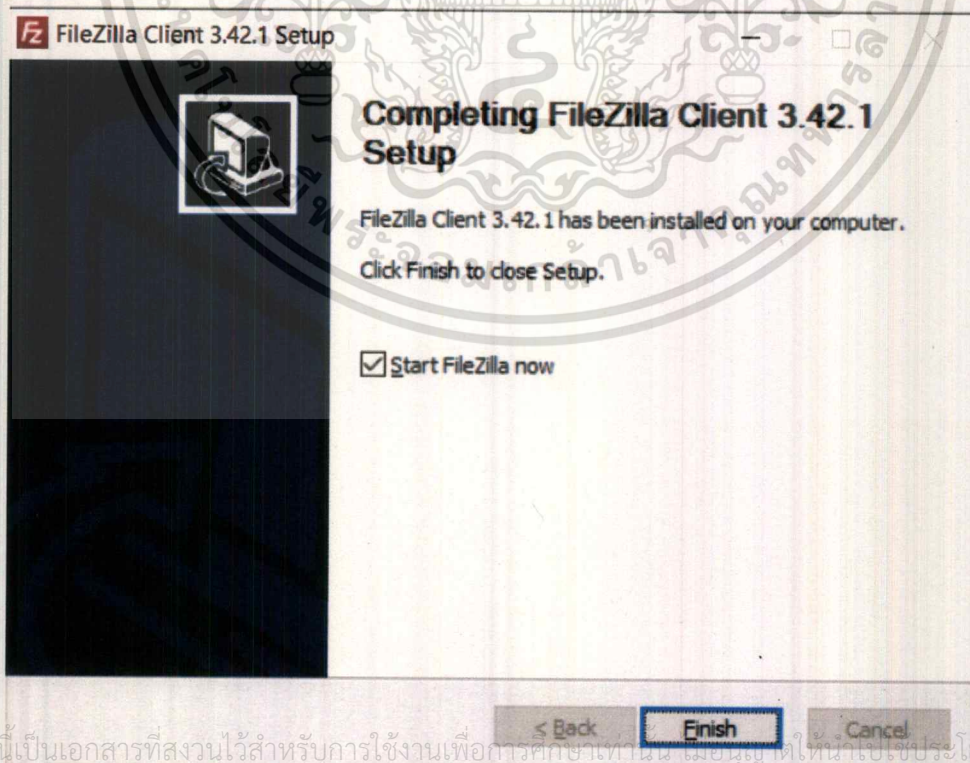
รูปที่ ก.30 การติดตั้งโปรแกรม

## 10) ระหว่างรอการติดตั้งโปรแกรม



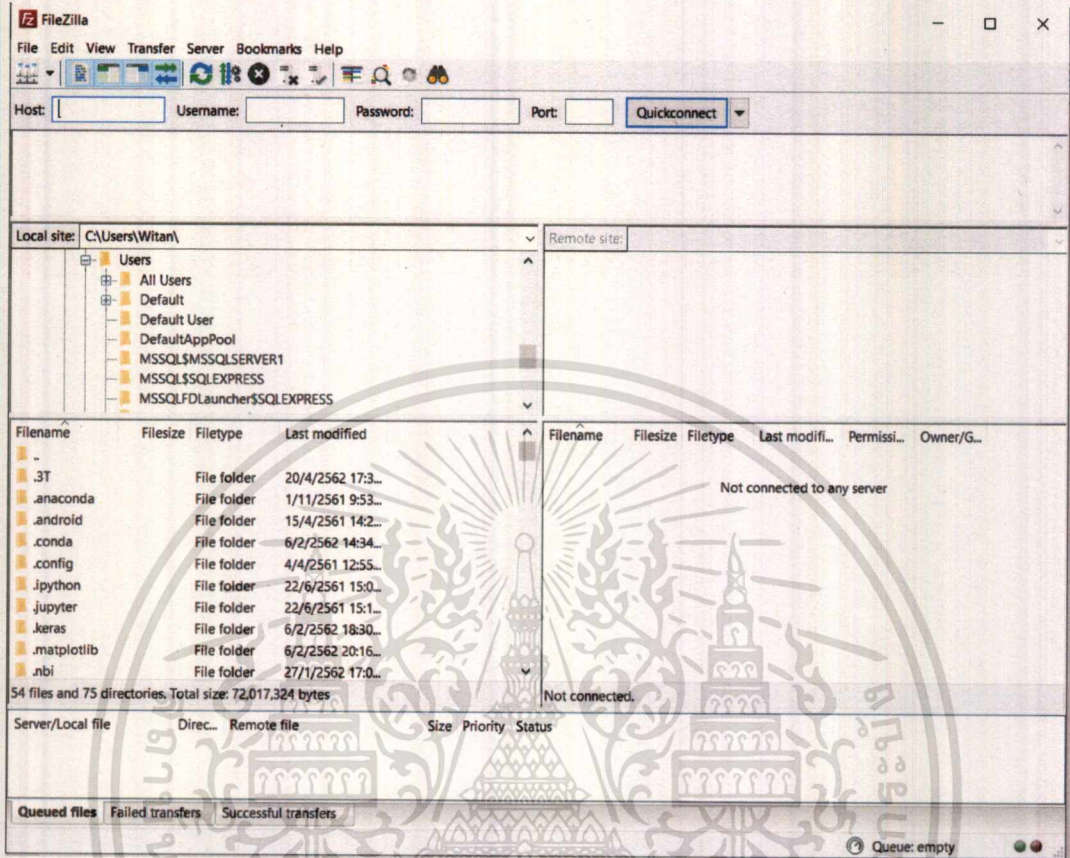
รูปที่ ก.31 ระหว่างติดตั้งโปรแกรม

## 11) กด Finish



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อผู้ผู้ใดให้นักเบีเซประโชยชนด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดรูปที่ ก.32 ติดตั้งโปรแกรมเสร็จสิ้นของเอกสารทุกครั้งที่มีการนำไปใช้

## 12) หน้าต่างโปรแกรม FileZilla

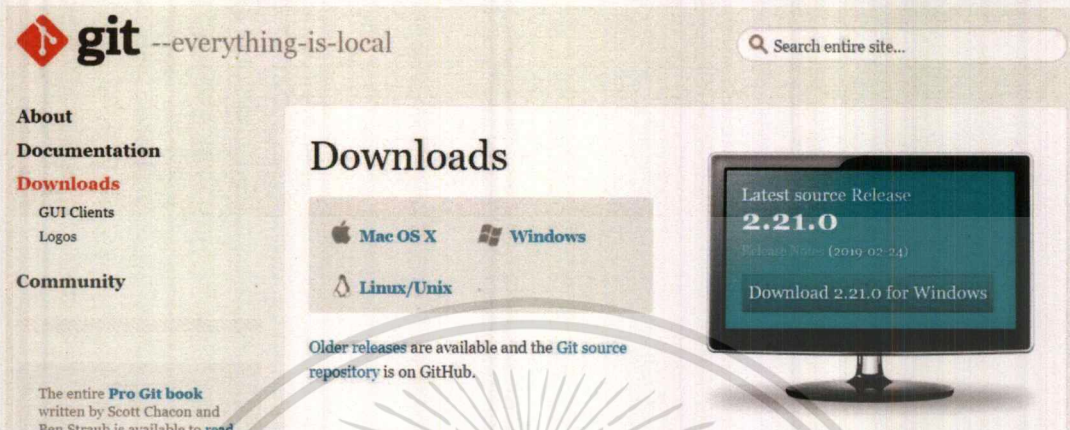


รูปที่ ก.33 หน้าต่างโปรแกรม FileZilla

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

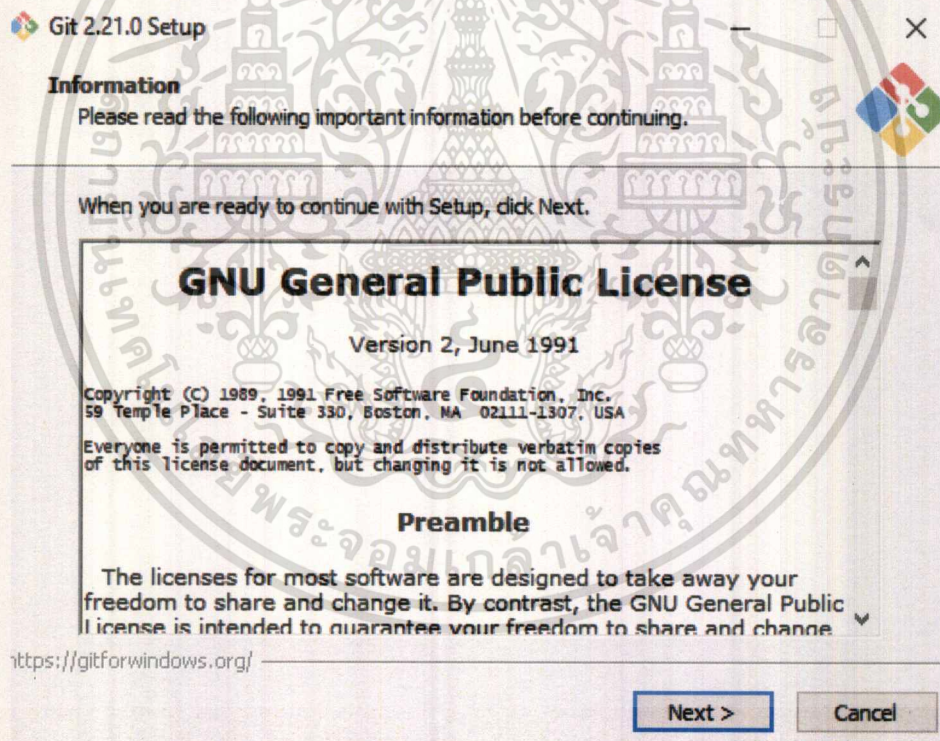
## ก.4 การติดตั้ง Git

1) ไปที่เว็บไซต์และเลือกระบบปฏิบัติการที่ต้องการ <https://git-scm.com/downloads>



รูปที่ ก.34 เว็บไซต์ดาวน์โหลด Git

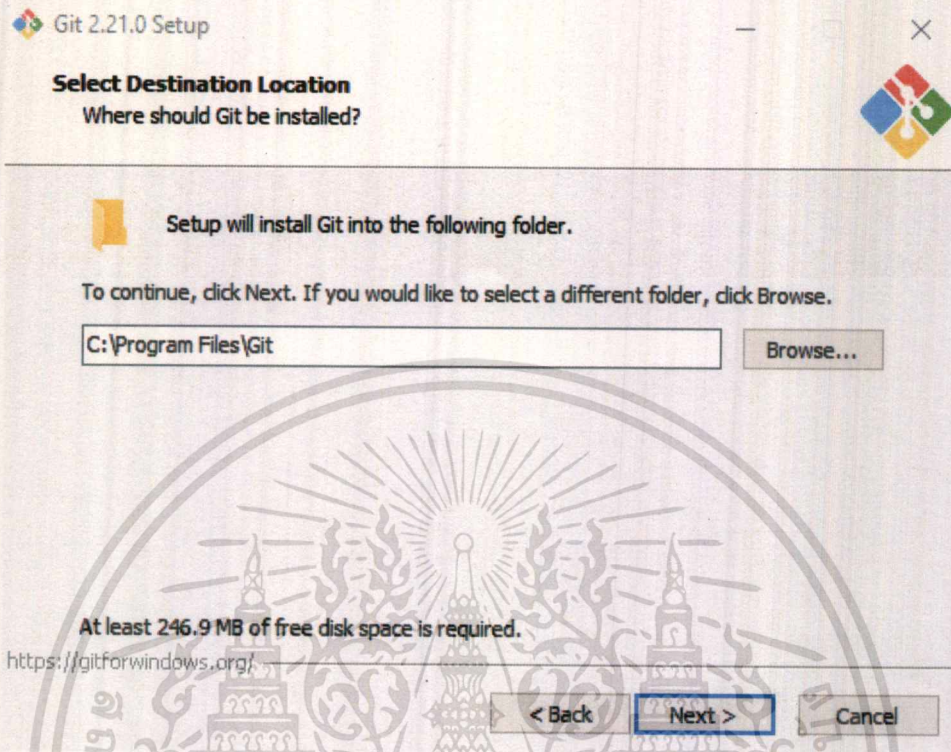
2) หน้าขั้นตอนการติดตั้งให้กด Next



รูปที่ ก.35 หน้าจอการติดตั้ง Git

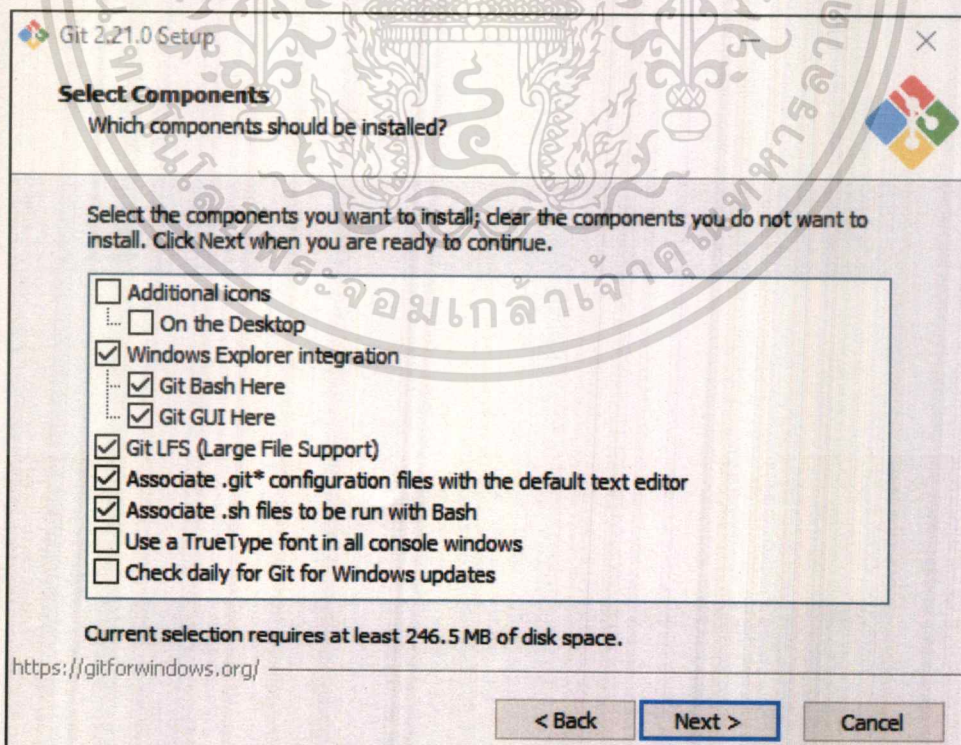
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 3) เลือกที่ติดตั้ง Git และกด Next



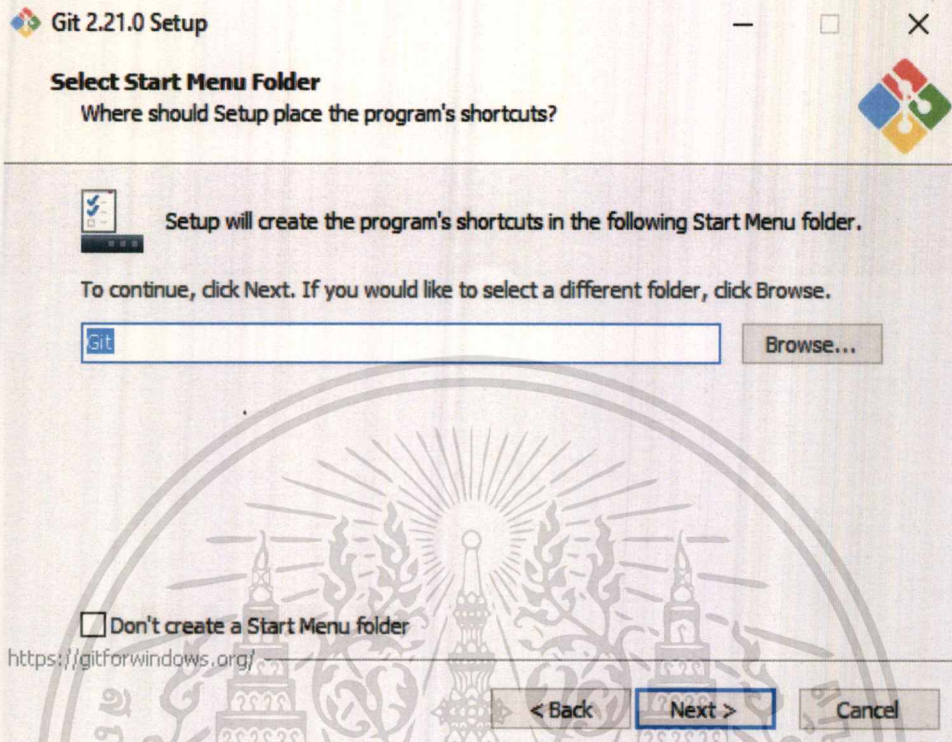
รูปที่ ก.36 ขั้นตอนการเลือกที่ติดตั้งโปรแกรม

## 4) เลือก components ของ Git ที่ต้องการและกด Next



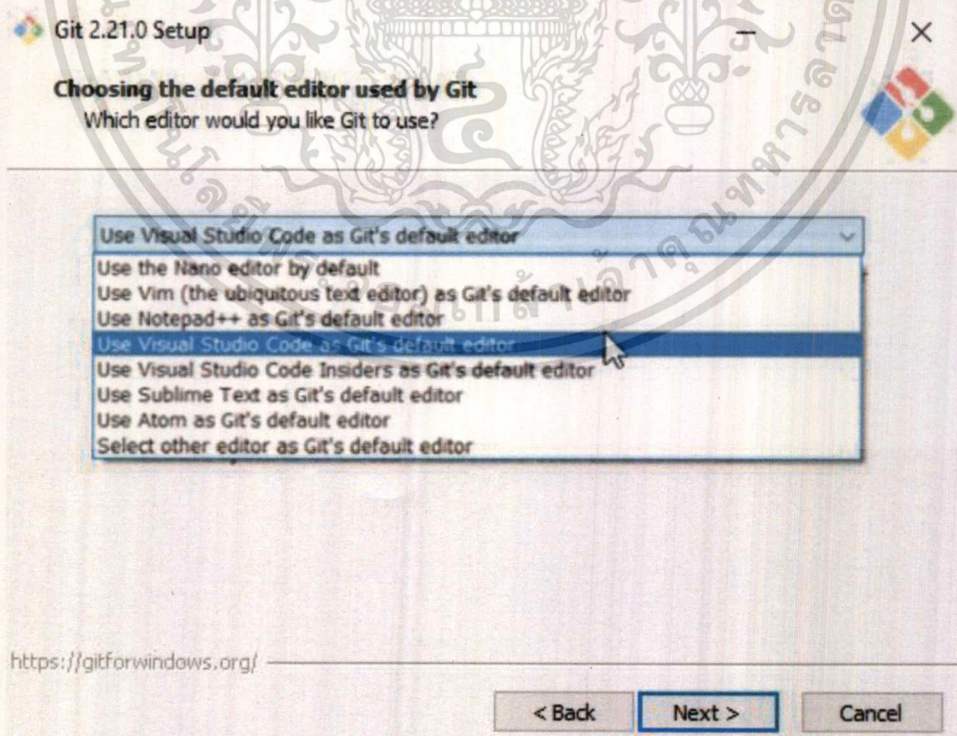
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับรูปที่ ก.37 เลือก components ของ Git ภายใต้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) กด Next



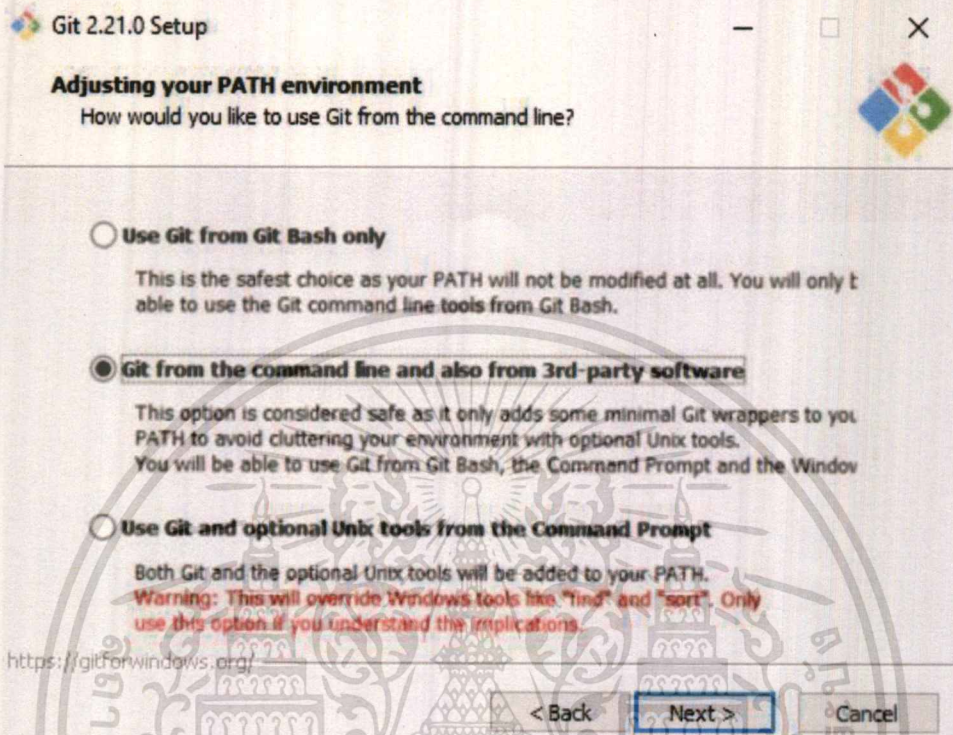
รูปที่ ก.38 การเลือก Start menu Folder ของ Git

6) การเลือก default editor ของ Git ในที่นี้จะเลือก Visual Studio Code และกด Next



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับรูปที่ ก.39 การเลือก default editor ของ Git ให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

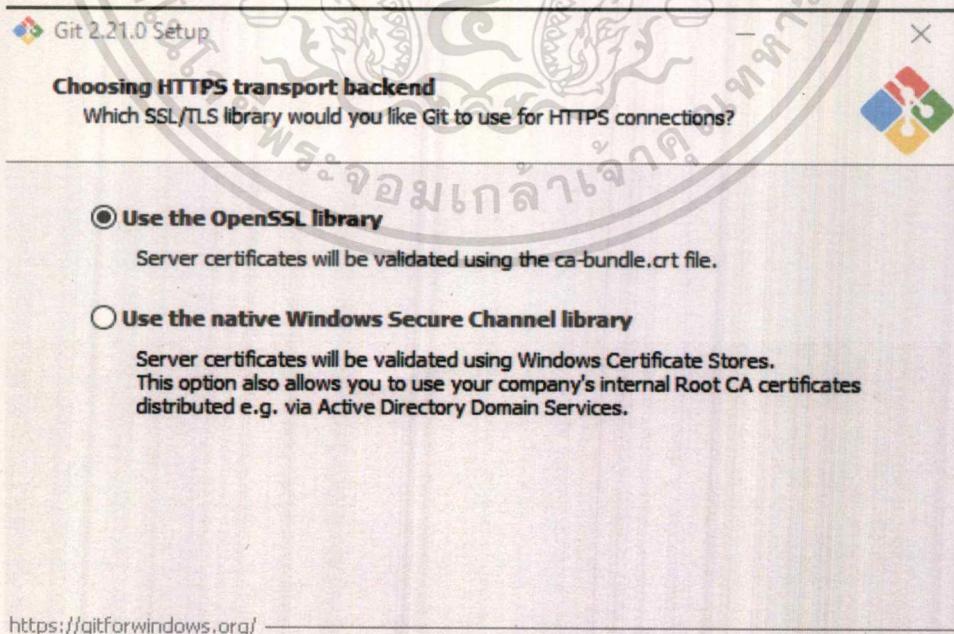
7) กดเลือก Git from the command line and also from 3rd-party software เพื่อให้สามารถใช้ git from command line จาก Editor อื่นๆได้ เช่น Visual Studio Code



รูปที่ ก.40 การตั้งค่า PATH environment

8) เลือก HTTPS transport backend ในที่นี้จะเลือก Use the OpenSSL library และกด

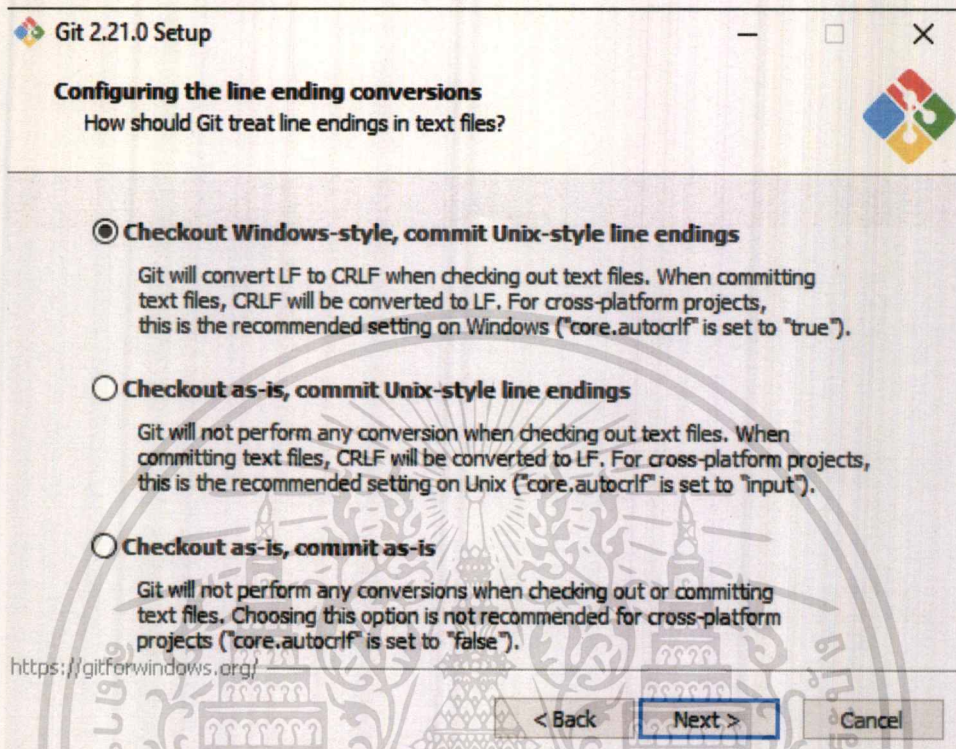
Next



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่ควรนำไปเผยแพร่โดยไม่ได้รับอนุญาต

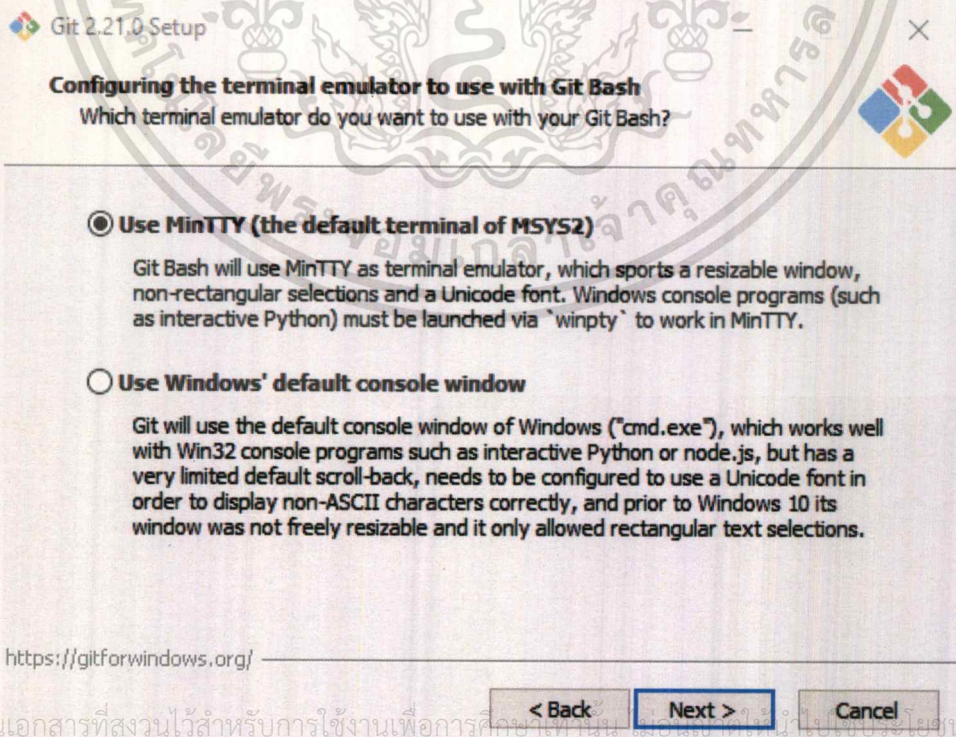
รูปที่ ก.41 เลือก HTTPS transport backend

9) การเลือก line ending conversions ในที่นี้จะเลือก Checkout Windows-style, commit Unix-style line endings ตามค่าพื้นฐานของ Git



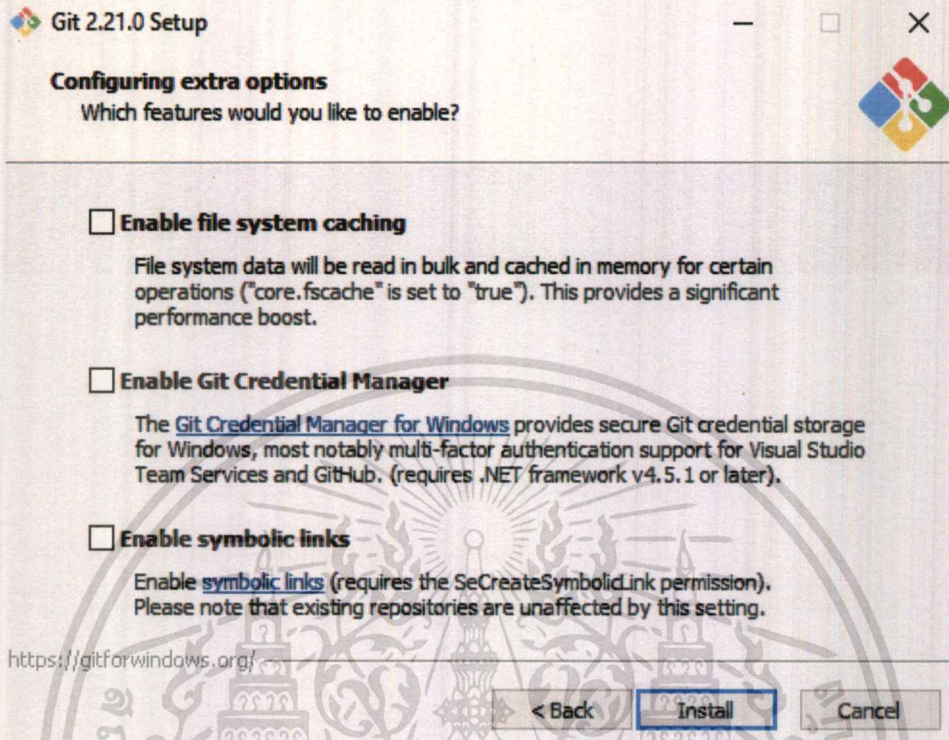
รูปที่ ก.42 การเลือก line ending conversions

10) เลือก Use MinTTY ที่เป็นค่าพื้นฐานของ Git และกด Next



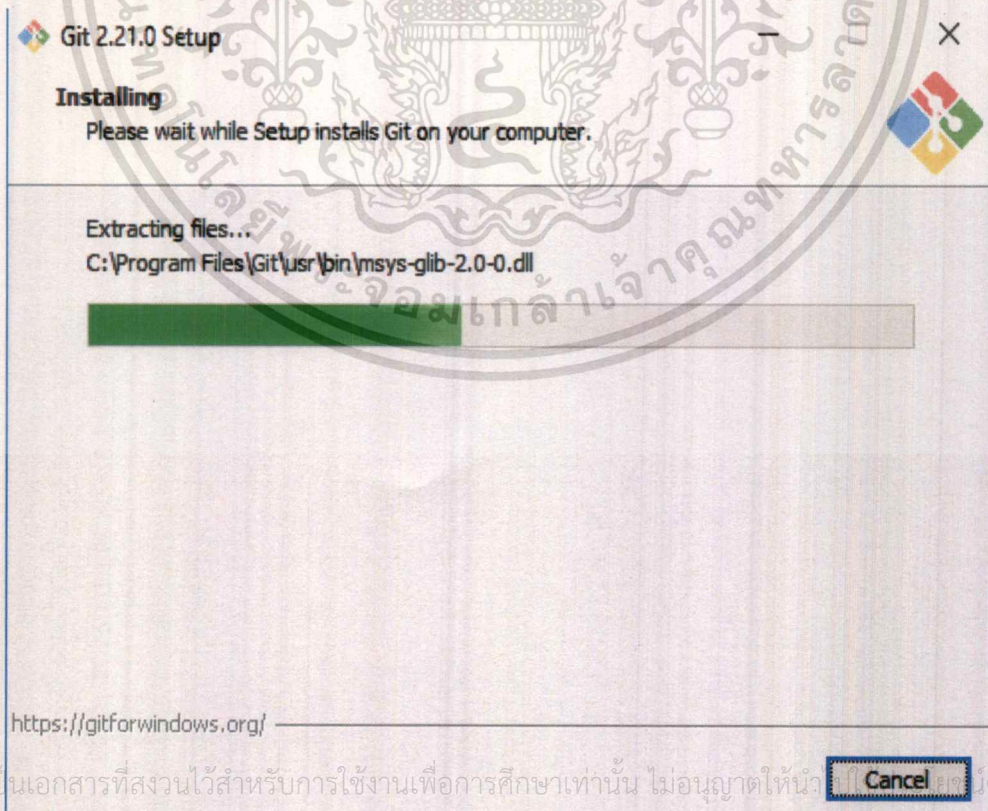
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่ควรนำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกที รูปที่ ก.43 การเลือก terminal emulator ของ Git Bash

11) ตีถูกออกทั้งหมด (ไม่เลือก) extra options และกด Install



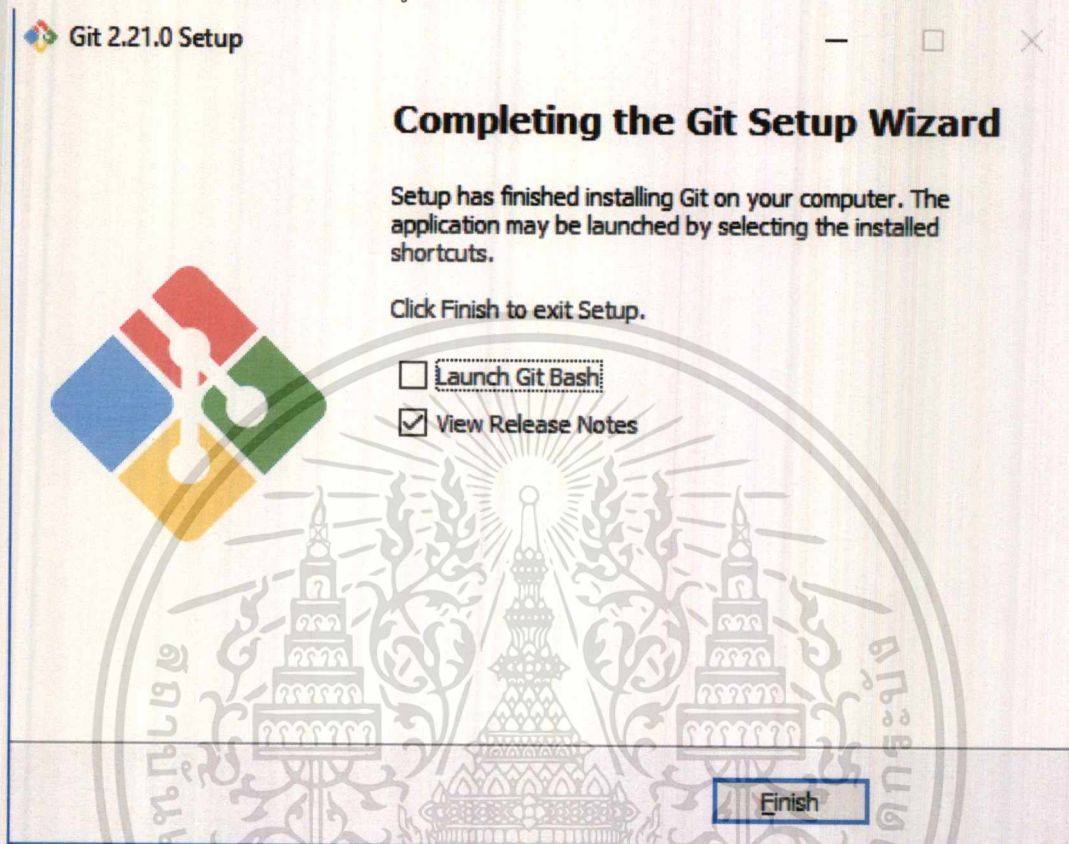
รูปที่ ก.44 การตั้งค่า extra options

12) ระหว่างรอติดตั้งโปรแกรมจนเสร็จ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปทำนการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามรูปที่ ก.45 หน้าจอระหว่างรอการติดตั้งเสร็จสิ้น เอกสารทุกครั้งที่มีการนำไปใช้

## 13) การติดตั้ง Git เสร็จสมบูรณ์ และกด Finish




รูปที่ ก.46 การติดตั้ง Git เสร็จสมบูรณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ก.5 การติดตั้ง Sourcetree

1) ไปที่เว็บไซต์ <https://www.sourcetreeapp.com/> และเลือกตามระบบปฏิบัติการที่ต้องการ

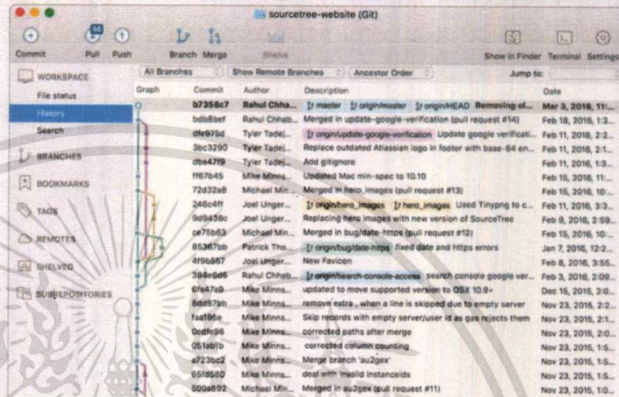
 Sourcetree

Download free

Simplicity and power in  
a beautiful Git GUI

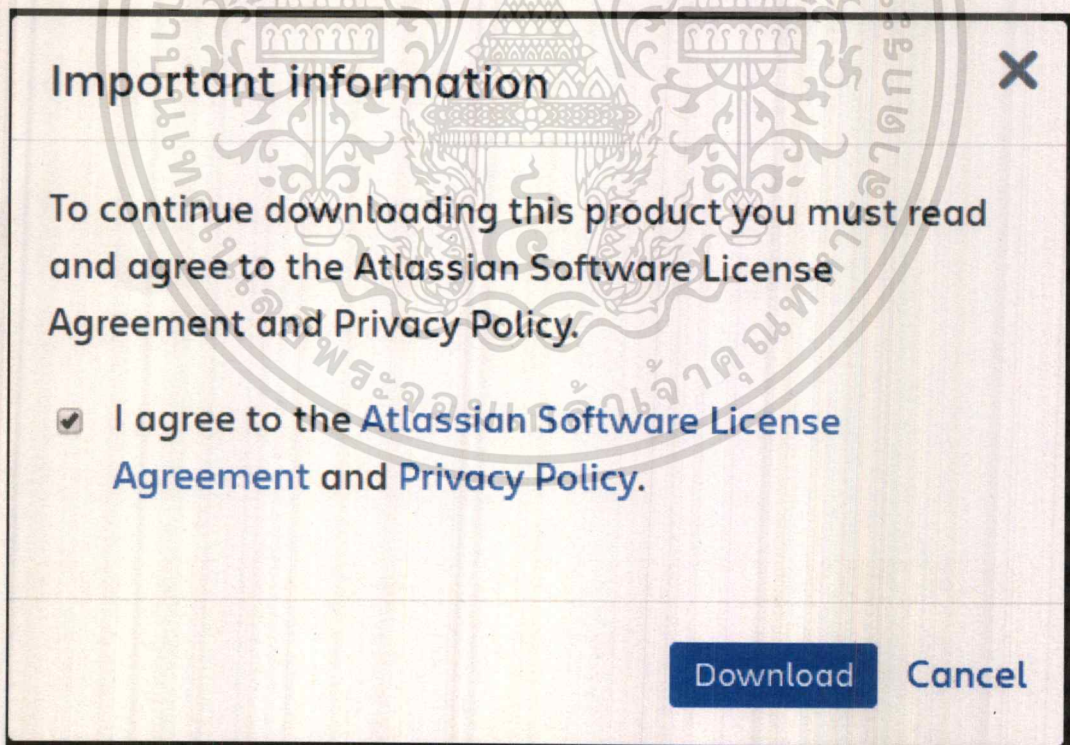
Download for Windows

Also available for Mac OS X



รูปที่ ก.47 เว็บไซต์ Sourcetree

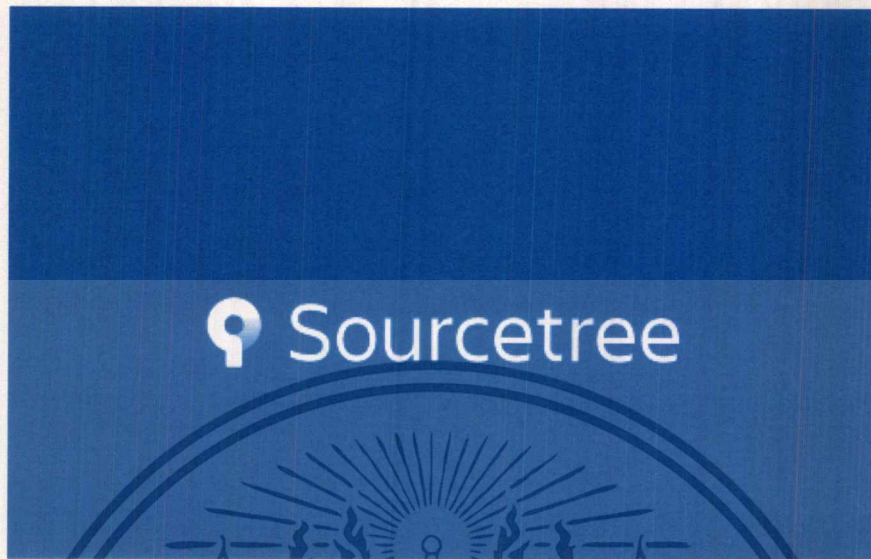
2) กด I agree และกด Download



รูปที่ ก.48 หน้าจอแสดงข้อความตกลงระหว่างผู้ผลิตและผู้ใช้

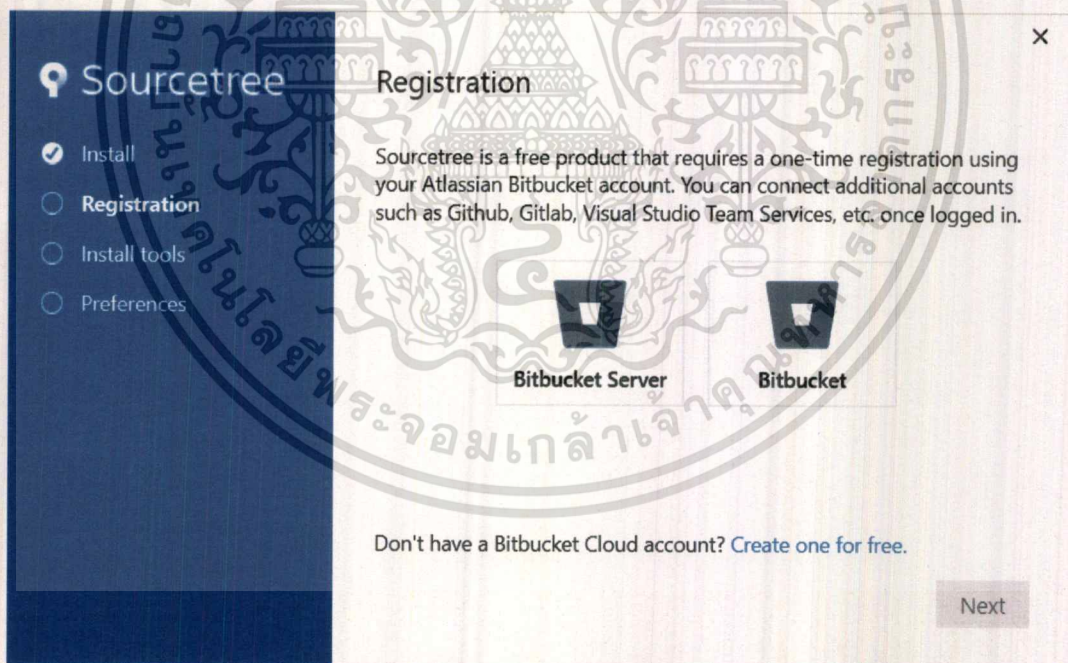
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) หน้าจอระหว่างรอการติดตั้งของ Sourcetree



รูปที่ ก.49 หน้าจอระหว่างรอการติดตั้งของ Sourcetree

4) เมื่อเข้าหน้าจอการติดตั้งแล้วจะมีการให้สมัคร Bitbucket ซึ่งเป็นผู้ให้บริการ Git รายหนึ่ง และกด Bitbucket












รูปที่ ก.50 หน้าจอการติดตั้ง Sourcetree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) เมื่อล็อกอินด้วย Bitbucket account แล้ว Bitbucket จะถามว่าเราจะให้สิทธิ์แก่ Sourcetree ในการจัดการ repository ต่างๆของบัญชีเราไหม

## Confirm access to your account

**Sourcetree for Windows** is requesting access to the following:

-  Read and modify your account information
-  Read and modify your repositories' issues
-  Read and modify your team's project settings, and read and transfer repositories within your team's projects
-  Read and modify your repositories and their pull requests
-  Administer your repositories
-  Read and modify your snippets
-  Read and modify your team membership information
-  Read and modify your repositories' webhooks
-  Read and modify your repositories' wikis

By installing the App you agree to the [privacy policy](#) and [terms of use](#) provided by Sourcetree for Windows.

Grant access Cancel

รูปที่ ก.51 หน้าต่างการขอใช้สิทธิ์ Bitbucket ของ Sourcetree

6) เมื่อการให้สิทธิ์ในบัญชี bitbucket แก่ Sourcetree สำเร็จจะขึ้นหน้านี้

## Authentication Successful

SourceTree has been successfully authenticated. You may now close this page.

 ATlassian

รูปที่ ก.52 การให้สิทธิ์เสร็จสมบูรณ์

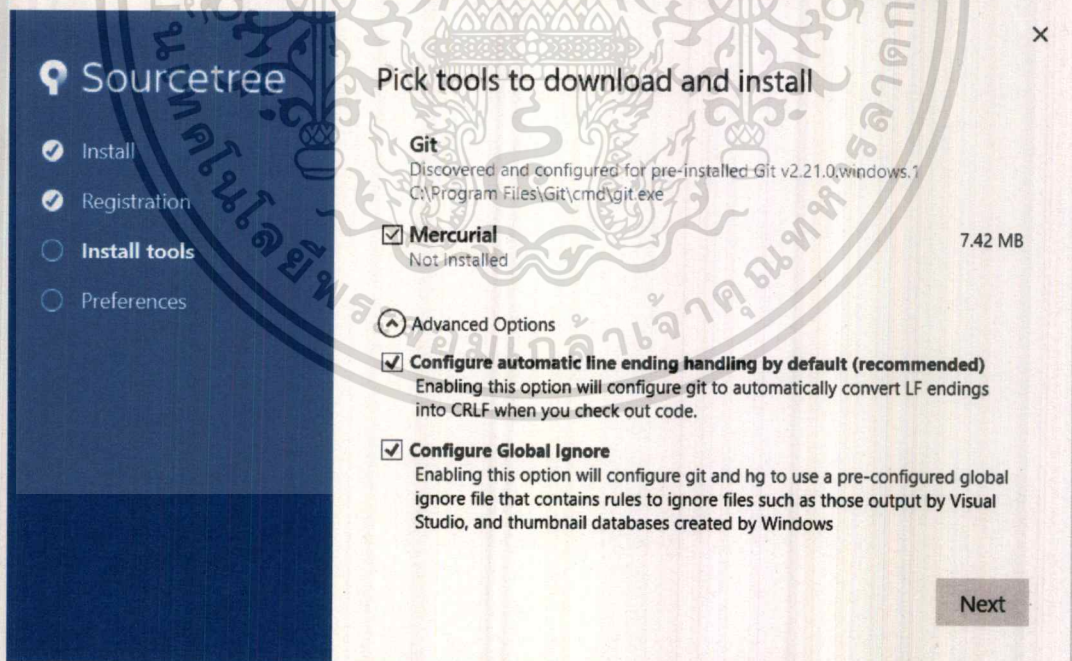
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7) เมื่อการให้สิทธิ์ในบัญชี Bitbucket แก่ Sourcetree เสร็จสิ้น ในโปรแกรม Sourcetree จะขึ้นหน้านี้และกด Next



รูปที่ ก.53 ล็อกอินบัญชี Bitbucket ใน Sourcetree

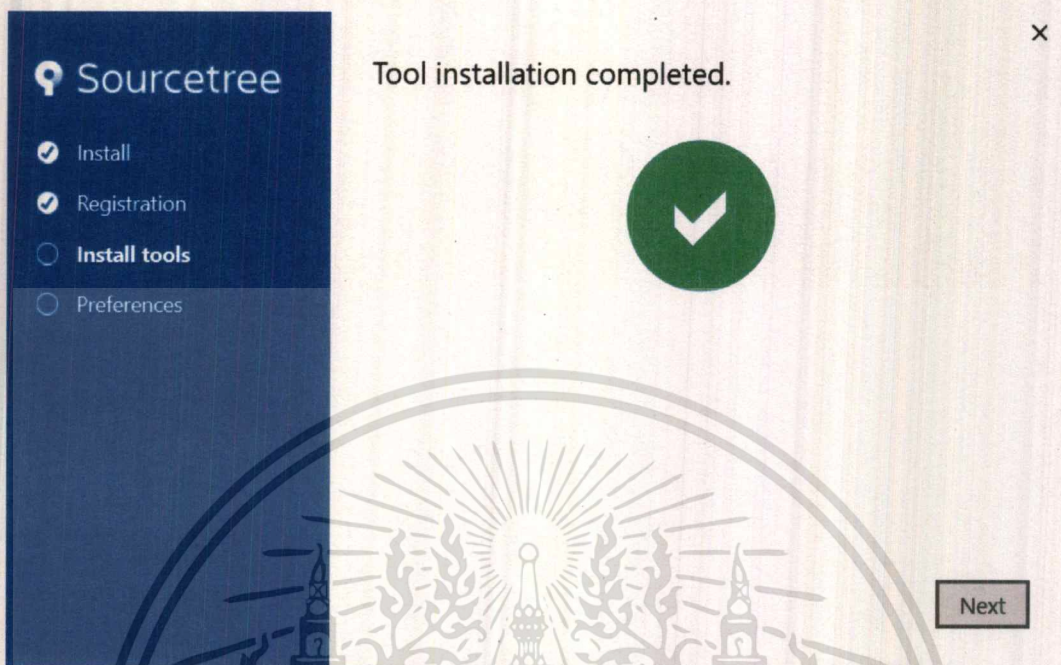
8) หน้าจอให้เลือกติดตั้งเครื่องมือเพิ่มเติม เมื่อเลือกตามต้องการแล้วให้กด Next



รูปที่ ก.54 การตั้งค่า Install tools

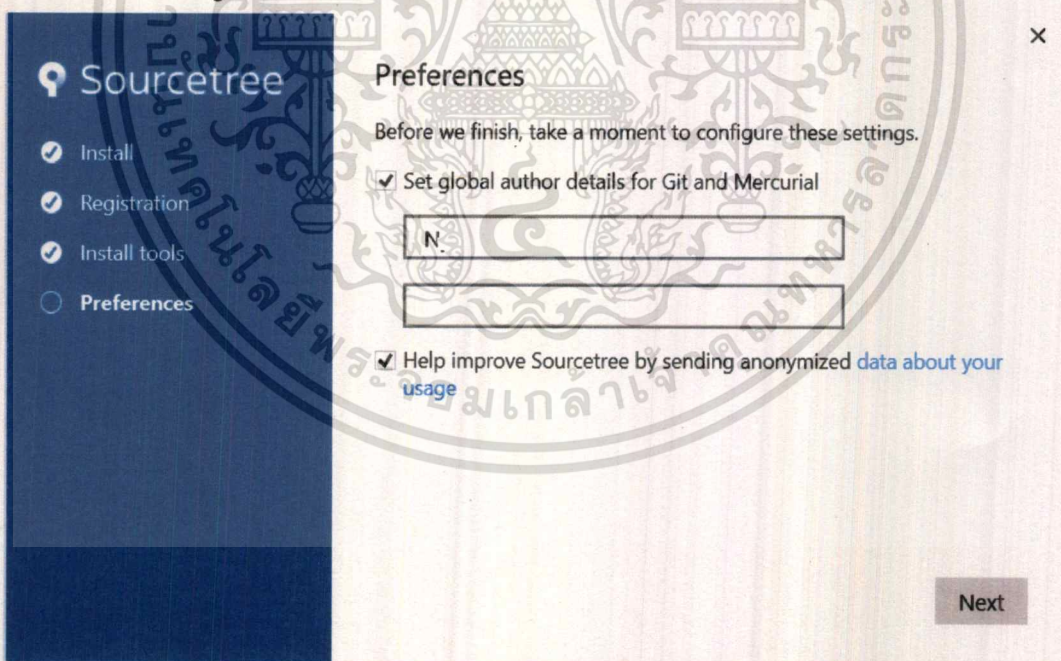
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9) ติดตั้ง Tools เสร็จแล้วให้กด Next



รูปที่ ก.55 ติดตั้ง Tool เสร็จสมบูรณ์

10) ตั้งค่า global author details for Git and Mercurial และกด Next



รูปที่ ก.56 การตั้งค่า global author details for Git and Mercurial

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

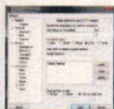
## 11) หน้าจอเริ่มต้นของโปรแกรม Sourcetree เพื่อใช้จัดการ Git site จากเว็บต่างๆ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ก.6 การติดตั้ง Putty

1) ดาวน์โหลด Putty จากเว็บไซต์ <https://www.putty.org/> กดที่ [here](#)



### Download PuTTY

PuTTY is an SSH and telnet client, developed originally by Simon Tatham for the Windows platform. PuTTY is open source software that is available with source code and is developed and supported by a group of volunteers.

You can download PuTTY [here](#).

Below suggestions are independent of the authors of PuTTY. They are not to be seen as endorsements by the PuTTY project.

รูปที่ ก.58 เว็บไซต์ putty

2) เลือก package ที่ต้องการ

### Package files

You probably want one of these. They include versions of all the PuTTY utilities.  
(Not sure whether you want the 32-bit or the 64-bit version? Read the [FAQ entry](#).)

**MSI ("Windows Installer")**

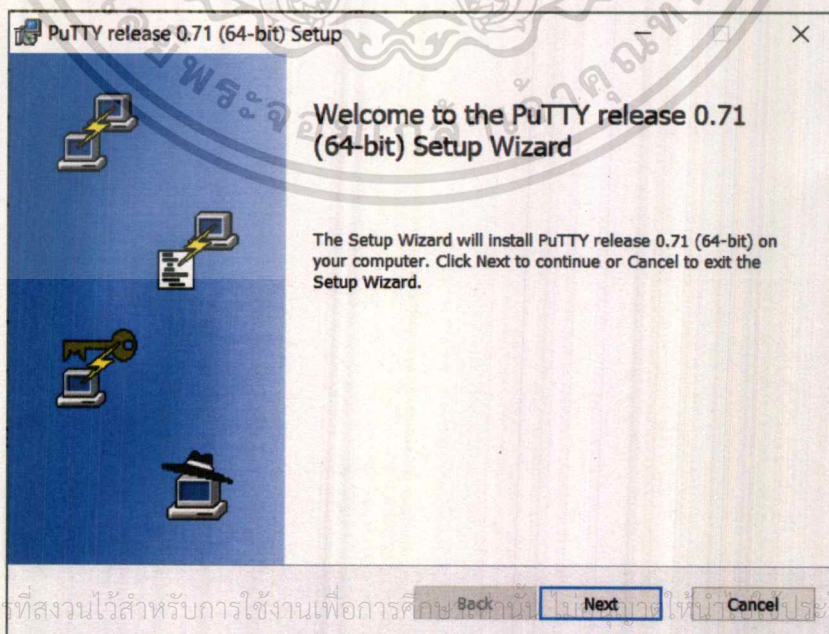
32-bit:	<a href="#">putty-0.71-installer.msi</a>	<a href="#">(or by FTP)</a>	<a href="#">(signature)</a>
64-bit:	<a href="#">putty-64bit-0.71-installer.msi</a>	<a href="#">(or by FTP)</a>	<a href="#">(signature)</a>

**Unix source archive**

.tar.gz:	<a href="#">putty-0.71.tar.gz</a>	<a href="#">(or by FTP)</a>	<a href="#">(signature)</a>
----------	-----------------------------------	-----------------------------	-----------------------------

รูปที่ ก.59 เลือก Putty package

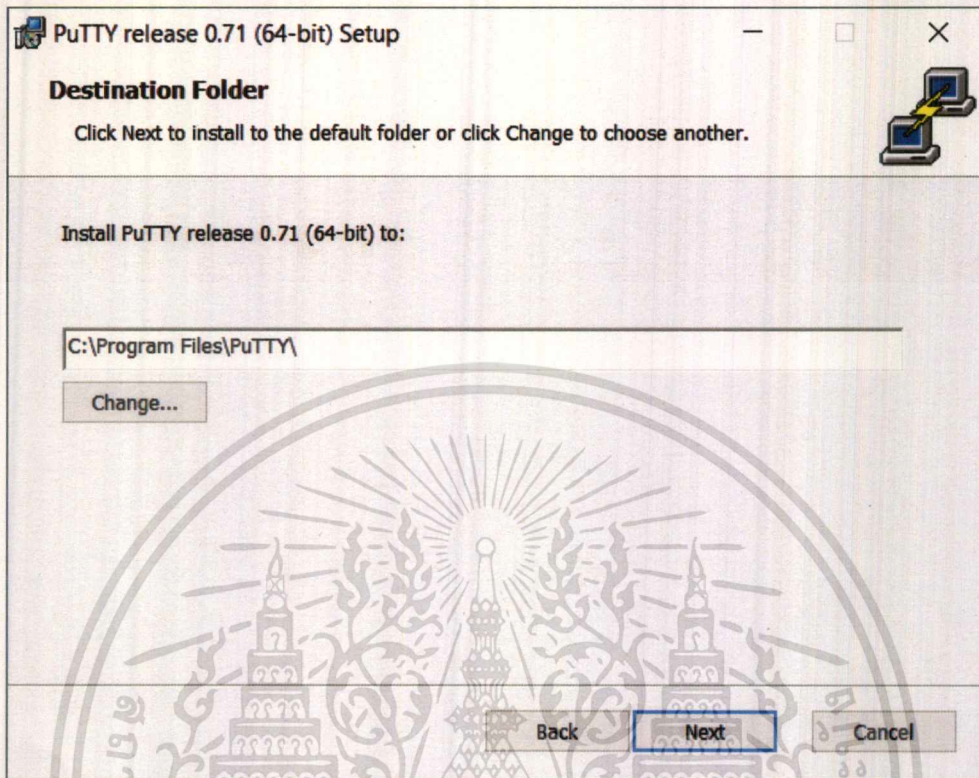
3) หน้าต่างการติดตั้ง Putty



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษานี้ ไม่สามารถเผยแพร่โดยไม่ได้รับอนุญาตจากเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้  
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงหรือแก้ไขเนื้อหาใดๆ ของเอกสาร

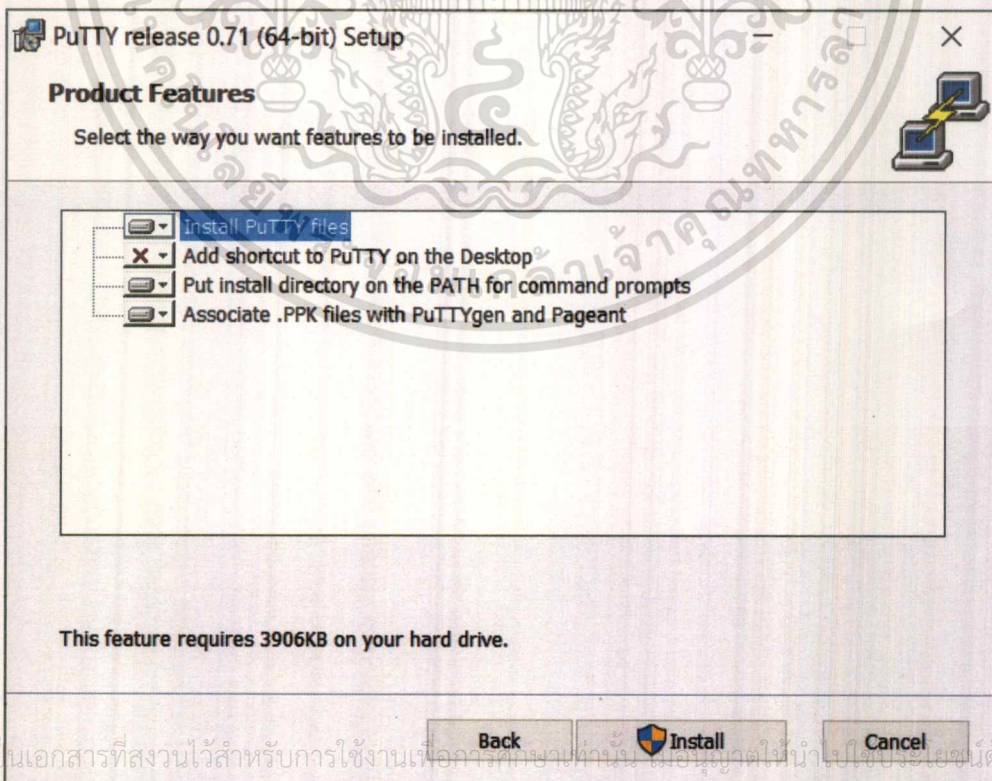
รูปที่ ก.60 หน้าต่างการติดตั้ง Putty

## 4) เลือกที่ติดตั้ง Putty



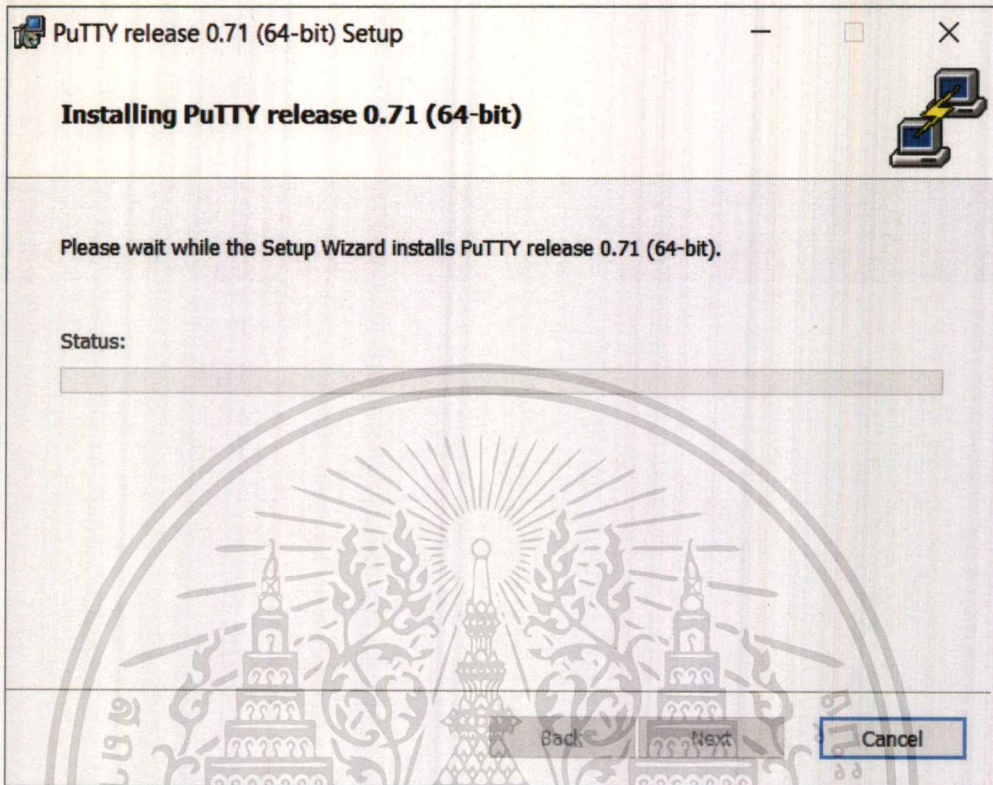
รูปที่ ก.61 เลือกที่ติดตั้ง Putty

## 5) เลือก Features ของ Putty ที่ต้องการ และกด Install



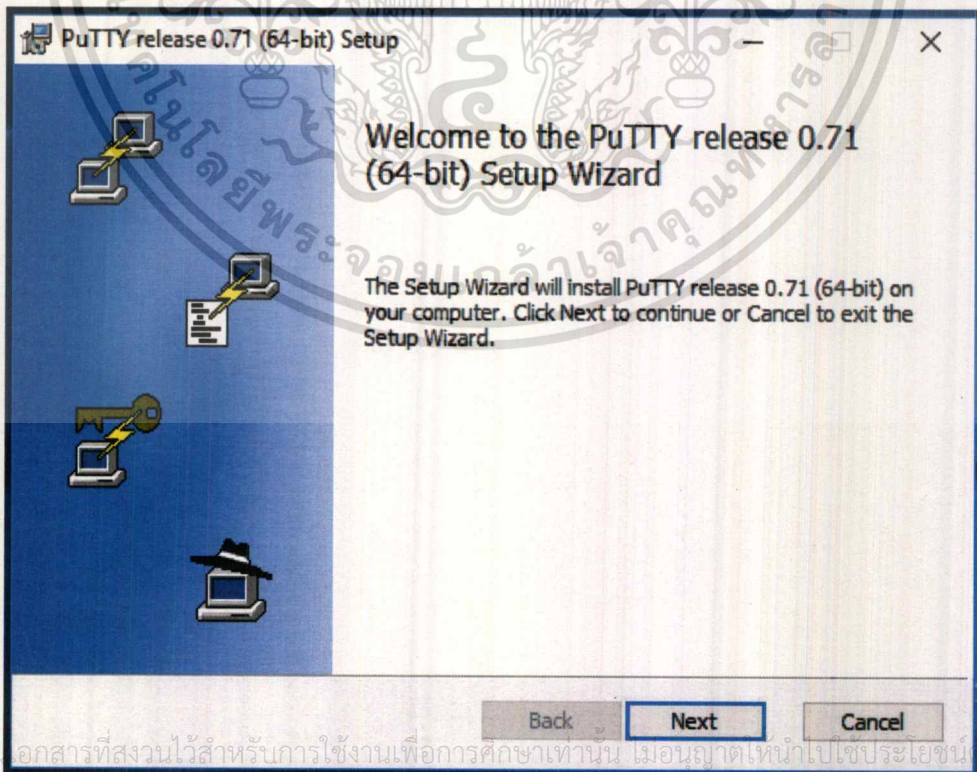
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษานานาชาติเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้รูปที่ ก.62 เลือก Features ที่ต้องการ ของเอกสารทุกครั้งที่มีการนำไปใช้

## 6) ระหว่างรอการติดตั้ง PuTTY



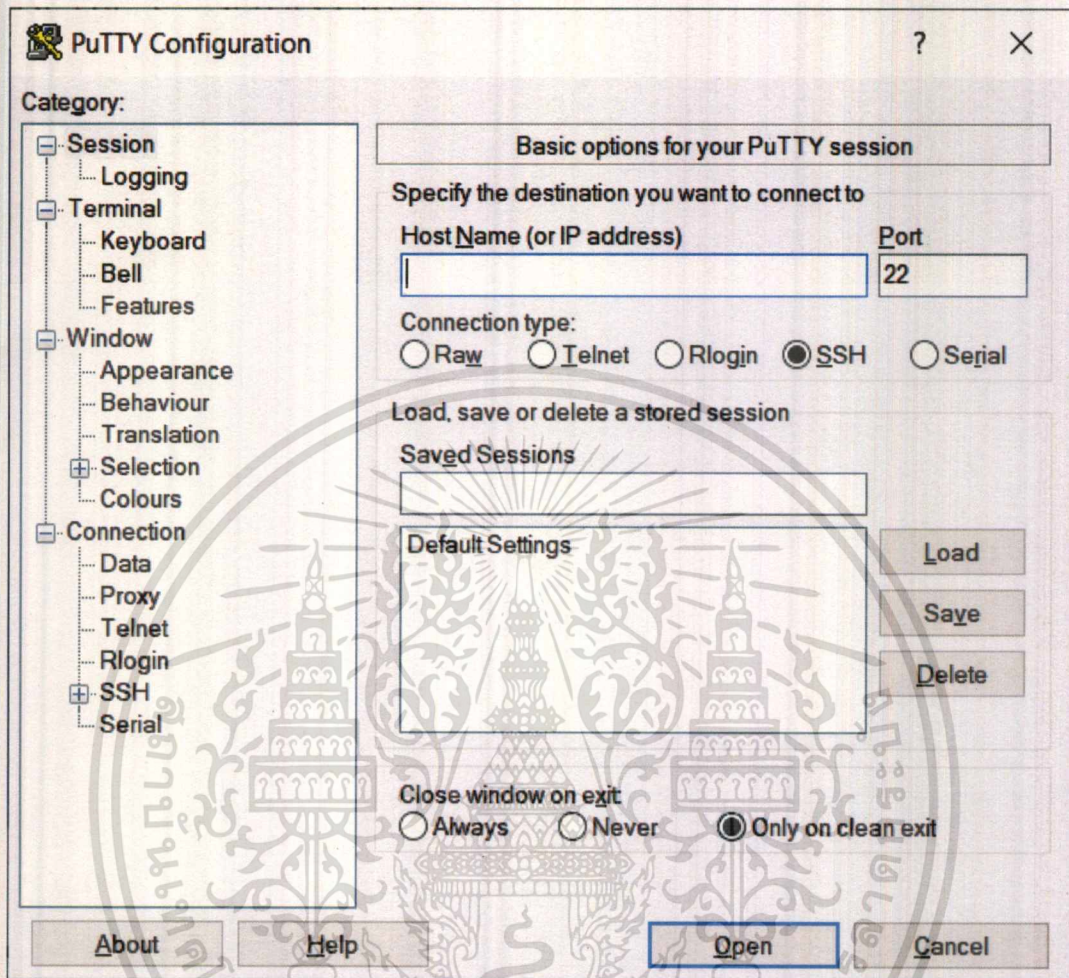
รูปที่ ก.63 หน้าจอระหว่างรอการติดตั้ง PuTTY

## 7) การติดตั้งเสร็จสิ้น



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนำไปใช้ประโยชน์ การค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดรูปที่ ก.64 การติดตั้งเสร็จสิ้นเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 8) หน้าต่างโปรแกรม Putty



รูปที่ ก.65 หน้าต่างโปรแกรม Putty

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้