ระบบแนะนำสถานที่ท่องเที่ยวในประเทศไทย

TOURIST DESTINATION RECOMMENDATION SYSTEM IN THAILAND

สุมิตรา กองเป็ง

SUMITRA KONGPENG

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2567

KMITL-2024-SC-M-002-023

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TOURIST DESTINATION RECOMMENDATION SYSTEM IN THAILAND

SUMITRA KONGPENG

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE
DEGREE OF MASTER OF SCIENCE IN COMPUTER SCIENCE
DEPARTMENT OF COMPUTER SCIENCE SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2024

KMITL-2024-SC-M-002-023

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2024

SCHOOL OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	ระบบแนะนำสถานที่ท่องเที่ยวในประเทศไทย
ชื่อนักศึกษา	นางสาวสุมิตรา กองเป็ง
รหัสประจำตัว	61605038
ปริญญา	วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชา	วิทยาการคอมพิวเตอร์
พ.ศ.	2567
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผู้ช่วยศาสตราจารย์ ดร. อนันตพร หรรษคุณาฒัย

บทคัดย่อ

สถานที่ท่องเที่ยวในประเทศไทยมีความหลากหลายส่งผลให้เป็นการยากต่อนักท่องเที่ยวสำหรับการเลือกจุดหมายปลายทางในการท่องเที่ยว การสร้างระบบแนะนำสถานที่ท่องเที่ยวที่เหมาะสมรายบุคคลจึงเป็นความท้าทายอย่างมาก ดังนั้นจุดมุ่งหมายหลักของการวิจัยนี้จึงครอบคลุมวัตถุประสงค์สองประการ ประการแรกเพื่อสร้างระบบการแนะนำสถานที่ท่องเที่ยวในประเทศไทยโดยใช้อัลกอริทึมการเรียนรู้ของเครื่อง และประการที่สองเพื่อวิเคราะห์ปัจจัยที่มีอิทธิพลต่อการเลือกจุดหมายปลายทางของนักท่องเที่ยว โดยข้อมูลที่นำมาใช้ในการสร้างระบบแนะนำเป็นการนำข้อมูลมาจากการเก็บแบบสอบถามออนไลน์ และมีการเตรียมข้อมูลโดยนำข้อมูลมาเลือกคุณลักษณะที่แตกต่างกันสี่ประการ ในส่วนของกระบวนการสร้างแบบจำลองผู้วิจัยใช้อัลกริทึมการเรียนรู้ของเครื่องสี่วิธี วิธีเหล่านี้ถูกนำมาใช้ในการสร้างแบบจำลองและเปรียบเทียบประสิทธิภาพการทำนายของระบบแนะนำ เมื่อเปรียบเทียบด้วยค่า Hit Rate และ NDCG พบว่าภาคเหนือมีค่าสูงกว่าในทุกภูมิภาค และค่าเฉลี่ยของ Hit Rate และ NDCG สำหรับทั้งห้าภูมิภาคคือ 0.8 และ 0.59 ตามลำดับ และนอกจากนี้ยังมีการวิเคราะห์หาความสำคัญของปัจจัยที่ส่งผลต่อการเลือกสถานที่ท่องเที่ยว เช่น กิจกรรม เดือนที่เดินทาง งบประมาณในการเดินทาง และอายุของนักท่องเที่ยว เพื่อให้เข้าใจถึงอิทธิพลต่อการเลือกเดินทางท่องเที่ยวในแต่ละภูมิภาคของประเทศไทย

คำสำคัญ : การทำเหมืองข้อมูล การเรียนรู้ของเครื่อง เมืองรองในประเทศไทย ระบบแนะนำนักท่องเที่ยว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	Tourist Destination Recommendation System in Thailand
Student Name	Sumitra Kongpeng
Student ID	61605038
Degree	Master of Science (Computer Science)
Department	Computer Science
Year	2024
Thesis Advisor	Asst. Prof. Dr. Anantaporn Hanskunatai

Abstract

Thailand's diverse attractions make choosing a vacation destination difficult for tourists. The tourist destination recommendation system is a challenge for creating. Therefore, the principal aims of this research encompass two distinct objectives: firstly, to create a recommendation system for tourist destinations in Thailand by applying machine learning algorithms; and secondly, to analyze factors influencing tourists' choices of destinations. The dataset was gathered from an online survey conducted. In the experiments, four different types of feature selection methods were applied in a data preprocessing step. In the modeling process, four machine learning algorithms were used to construct the model and compare the predictive performance of the recommendation system based on hit rate and NDCG. The experimental results showed that suggesting tourist destinations in the Central region was the most effective, with the highest Hit rate and NDCG compared to other regions. The average Hit rate and NDCG for the five regions were 0.8 and 0.59, respectively. In addition, there has been an analysis of key factors influencing destination selection, such as activity, month travel, travel budget, and the age of tourists, to understand their impact on travel choices in each region of Thailand.

Keywords : data mining, machine learning, less visited area, tourist recommendation system

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดี ด้วยได้รับความอนุเคราะห์การสนับสนุนด้วยดีจากอาจารย์ที่ปรึกษา ผศ.ดร. อนันตพร หรรษคุณาฒัย ที่สละเวลาในการให้ความรู้ในการทำวิจัยตลอดจนคำแนะนำในทุกๆ ขั้นตอนการทำวิทยานิพนธ์ รวมถึงการเรียบเรียงการเขียนงานวิจัยภาษาอังกฤษ อีกทั้งยังช่วยตรวจสอบแก้ไขปัญหาและข้อบกพร่องต่างๆ ในวิทยานิพนธ์ให้สำเร็จลุล่วงด้วยดี

ขอขอบพระคุณ รศ.ดร. อนุชิต จิตพัฒนกุล อาจารย์ผู้ทรงคุณวุฒิจากภายนอกสถาบันฯ ประธานกรรมการสอบ และ ผศ.ดร. ปัทมา เจริญพร อาจารย์บัณฑิตประจำภาควิชา กรรมการสอบวิทยานิพนธ์ ที่ให้คำแนะนำและชี้แนะอีกทั้งยังช่วยแก้ไขข้อบกพร่องจนวิทยานิพนธ์เล่มนี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบพระคุณอาจารย์ภาควิชาวิทยาการคอมพิวเตอร์ และอาจารย์ท่านอื่นๆ ตลอดจนบุคคลากรภายในสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุกท่านที่ให้ความรู้ตลอดระยะเวลาหลักสูตรและมีส่วนช่วยผลักดันให้วิทยานิพนธ์เล่มนี้สำเร็จไปได้ด้วยดี

ขอขอบพระคุณ คุณพ่อ คุณแม่ สามี ครอบครัว และเพื่อนๆ ที่ให้ความช่วยเหลือสนับสนุนและเป็นกำลังใจในการเรียนและการทำวิจัยจนสำเร็จการศึกษาในระดับชั้นปริญญาโท

ขอขอบพระคุณท่านอื่นๆ ที่มิได้กล่าวไว้ ณ ที่นี้ ที่มีส่วนร่วมในการให้ความช่วยเหลือในการทำวิทยานิพนธ์ครั้งนี้

สุมิตรา กองเป็ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	ฉ
สารบัญรูป	ช
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของงานวิจัย	1
1.2 วัตถุประสงค์ของงานวิจัย	3
1.3 ขอบเขตของงานวิจัย	3
1.4 ขั้นตอนการศึกษาและดำเนินการวิจัย	4
1.4 ประโยชน์ที่คาดว่าจะได้รับ	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	5
2.1 เมืองรองในประเทศไทย	5
2.2 การเรียนรู้ของเครื่อง (Machine Learning Algorithm)	5
2.2.1 ต้นไม้ตัดสินใจ (Decision Tree)	7
2.2.2 การสุ่มแบบป่าไม้ (Random Forest)	10
2.2.3 การหาเพื่อนบ้านใกล้เคียง (k-Nearest Neighbors: k-NN)	13
2.2.4 เพอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron: MLP)	14
2.3 การเตรียมข้อมูล	17
2.3.1 การเลือกคุณลักษณะ	17
2.3.2 การเข้ารหัสวันฮอต	19
2.4 ระบบแนะนำ (Recommendation System)	20
2.4.1 ระบบแนะนำที่ใช้ข้อมูลเชิงเนื้อหา	20
2.4.2 ระบบแนะนำการกรองร่วมกัน	21
2.4.3 ระบบแนะนำผสมผสาน	23

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
2.5 Leave-One-Out Cross Validation	23
2.6 การวัดประสิทธิภาพ	24
2.6.1 ตัววัดอัตราคำแนะนำที่ถูกต้อง (Hit Rate)	24
2.6.2 ตัววัดอัตราการจัดลำดับคำแนะนำที่ถูกต้อง (NDCG)	26
2.7 งานวิจัยที่เกี่ยวข้อง	28
2.7.1 ระบบแนะนำที่ใช้ข้อมูลเชิงเนื้อหา	28
2.7.2 ระบบแนะนำการกรองร่วมกัน	29
2.7.3 ระบบแนะนำผสมผสาน	31
2.7.4 การวัดประสิทธิภาพด้วย Hit Rate และ NDCG	32
บทที่ 3 วิธีดำเนินงานวิจัย	33
3.1 การเก็บรวบรวมชุดข้อมูล	34
3.2 การเตรียมข้อมูล	40
3.2.1 การเลือกคุณลักษณะ	40
3.2.2 การแปลงข้อมูลโดยใช้เทคนิคการเข้ารหัสวันฮอต	41
3.3 การพัฒนาแบบจำลอง	41
3.4 การออกแบบส่วนประสานงานผู้ใช้	44
บทที่ 4 ผลการวิจัยและอภิปรายผล	46
4.1 ผลการวิจัย	46
4.2 อภิปรายผล	56
4.2.1 การวิเคราะห์ปัจจัยที่มีผลต่อการท่องเที่ยว	56
4.2.2 ความรู้ที่ได้จากแบบจำลอง	61
4.2.3 การนำแบบจำลองไปพัฒนาระบบ	62
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	67
5.1 สรุปผลการวิจัย	67
5.2 ข้อเสนอแนะ	68
เอกสารอ้างอิง	69
ประวัติผู้เขียน	72

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 Peudo code การทำงานของขั้นตอนวิธีต้นไม้ตัดสินใจ (Decision Tree)	9
2.2 Peudo code การทำงานของขั้นตอนวิธีเทคนิคการสุ่มแบบป่าไม้ (Random Forest)	11
2.3 Peudo code การทำงานของขั้นตอนเพอร์เซ็ปตรอนหลายชั้น (MultiLayer Perceptron: MLP)	15
3.1 จำนวนข้อมูลหน่อยตัวอย่างในแต่ละภูมิภาคของประเทศไทย	34
3.2 รายละเอียดในแต่ละข้อคำถามที่ใช้ในแบบสอบถาม	37
3.3 รายละเอียดของอัลกอริทึมที่ใช้ในการวิจัย	42
4.1 รายละเอียดของชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบในแต่ละภูมิภาค	46
4.2 ผลการทดลองในแต่ละแบบจำลองโดยแบ่งตามภูมิภาค	47
4.3 ประสิทธิภาพการทำงานของแบบจำลองในแต่ละภูมิภาค	55
4.4 คุณลักษณะที่ถูกเลือกมาใช้ในแบบจำลองโดยแบ่งตามภูมิภาค	56
4.5 คะแนนความพึงพอใจต่อโปรแกรมแนะนำสถานที่ท่องเที่ยว	66

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 แผนภาพต้นไม้ตัดสินใจ (Decision Tree)	7
2.2 แผนภาพการสุ่มแบบป่า (Random Forest)	11
2.3 แผนภาพเทคนิคการหาเพื่อนบ้านใกล้เคียง (k-Nearest Neighbors: k-NN)	14
2.4 แผนภาพโครงข่ายประสาทเทียม (Multilayer Perceptron: MLP)	15
2.5 ตัวอย่างการทำเทคนิคการเข้ารหัสวันฮอต (One-hot encoding)	20
2.6 ตัวอย่างของขั้นตอนการทำ Leave-One-Out Cross Validation	24
2.7 ตัวอย่างผลลัพธ์ที่ได้จากแบบจำลองเพื่อยกตัวอย่างการคำนวณค่า Hit Rate	25
2.8 ตัวอย่างผลลัพธ์ที่ได้จากแบบจำลองเพื่อยกตัวอย่างการคำนวณค่า NDCG	27
3.1 โครงสร้างการทำดำเนินงานวิจัยของระบบแนะนำสถานที่ท่องเที่ยวในเมืองรอง	33
3.2 ข้อคำถามใน Google Forms	34
3.3 ข้อคำถามใน Google Forms	35
3.4 ข้อคำถามใน Google Forms	36
3.5 ข้อคำถามใน Google Forms	37
3.6 ตัวอย่างการแปลงคุณลักษณะอายุเป็นประเภทหมวดหมู่โดยใช้การเข้ารหัสวันฮอด	41
3.7 ภาพรวมของกระบวนการทดลอง	42
3.8 แผนผังและโครงสร้าง (Wire Frame) ก่อนนำไปพัฒนาเว็บแอปพลิเคชัน	44
4.1 ค่าสัดส่วนของคุณลักษณะในแต่ละภูมิภาค	58
4.2 แผนผังต้นไม้ตัดสินใจของภาคกลาง	61
4.3 ตัวอย่างเว็บแอปพลิเคชันของโปรแกรมแนะนำสถานที่ท่องเที่ยวในภาคเหนือ	63
4.4 แบบสอบถามที่ใช้ในการสำรวจความพึงพอใจต่อโปรแกรมแนะนำ	65
4.5 กราฟเปรียบเทียบความพึงพอใจตามช่วงอายุ	66

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของงานวิจัย

ประเทศไทยมีชื่อเสียงด้านแหล่งท่องเที่ยวที่น่าอัศจรรย์ใจและน่าตื่นตาตื่นใจที่โด่งดังระดับโลกหลายแห่งตั้งแต่อดีตจนถึงปัจจุบัน เนื่องด้วยลักษณะภูมิประเทศที่หลากหลายทำให้แหล่งท่องเที่ยวในประเทศไทยมีความหลากหลายอย่างมาก อาทิ เช่น สถานที่ท่องเที่ยวเชิงภูเขา สถานที่ท่องเที่ยวเชิงทะเล สถานที่ท่องเที่ยวเชิงชุมชน หรือสถานที่ท่องเที่ยวเชิงประวัติศาสตร์ เป็นต้น ด้วยเหตุนี้ทำให้อุตสาหกรรมการท่องเที่ยวจึงถูกจัดเป็นรายได้อันดับต้นๆ ของประเทศไทยมาช้านาน ไม่ว่าจะสร้างรายได้ให้แก่ผู้ประกอบการธุรกิจโดยตรง เช่น ธุรกิจโรงแรมและที่พัก ธุรกิจอาหารและเครื่องดื่ม และธุรกิจนำเที่ยว แต่ยังมีส่วนสำคัญในการสร้างรายได้ให้แก่คนในท้องถิ่นอีกด้วย จึงส่งผลให้การท่องเที่ยวเป็นปัจจัยสำคัญอย่างหนึ่งในการทำให้เศรษฐกิจของประเทศขับเคลื่อนไปได้อย่างดี ผลสำรวจจากกระทรวงการท่องเที่ยวและกีฬา พบว่าพฤติกรรมการเดินทางท่องเที่ยวภายในประเทศของคนมีอัตราเพิ่มขึ้นจากปี พ.ศ. 2566 ร้อยละ 23 ซึ่งจังหวัดในประเทศไทยหากแบ่งในเชิงการท่องเที่ยวสามารถแบ่งได้ตามจำนวนนักท่องเที่ยวที่เดินทางไปท่องเที่ยวในจังหวัดนั้นๆ ซึ่งสามารถแบ่งเป็น 2 แบบ ได้แก่ เมืองหลัก และเมืองรอง เมืองหลัก คือ เมืองดังกล่าวมีนักท่องเที่ยวเข้าไปเที่ยวเกิน 6 ล้านคนต่อปี หรือเรียกได้ว่าเป็นเมืองที่ได้รับความนิยมอย่างมากจากนักท่องเที่ยว ทำให้มีการกระจายรายได้ภายในจังหวัดได้เป็นอย่างดี ส่วนเมืองรอง คือ เมืองดังกล่าวมีนักท่องเที่ยวเข้าไปเที่ยวไม่เกิน 6 ล้านคนต่อปี ทำให้การกระจายรายได้มีความแตกต่างจากเมืองหลัก

นักท่องเที่ยวส่วนมากเลือกวางแผนการเดินทางไปยังสถานที่ท่องเที่ยวที่ได้รับความนิยมต่างๆ ซึ่งสถานที่ท่องเที่ยวที่ได้รับความนิยมนั้นส่วนใหญ่ล้วนอยู่ในจังหวัดเมืองหลักของประเทศไทย ทำให้เกิดความแออัด และเกิดการกระจุกตัวของนักท่องเที่ยวในสถานที่ท่องเที่ยวยอดนิยมเป็นจำนวนมาก ด้วยเหตุนี้ทำให้เมืองรองได้รับความสนใจจากนักท่องเที่ยวมากยิ่งขึ้น เนื่องจากเมืองรองสามารถได้รับความสงบความเป็นส่วนตัว และสามารถหนีความวุ่นวายจากจำนวนนักท่องเที่ยวที่มีจำนวนมากในแต่ละสถานที่ท่องเที่ยวยอดนิยมได้เป็นอย่างดี และผลสำรวจจาก การท่องเที่ยวแห่งประเทศไทย (ททท.) พบว่ามุมมองของนักท่องเที่ยวที่มีต่อการท่องเที่ยวหลัง Covid-19 พบว่า ร้อยละ 68 เลือกที่จะหลีกเลี่ยงเมืองใหญ่ที่มีความหนาแน่นของนักท่องเที่ยว ทำให้เมืองรองเป็นที่จับตามองในการท่องเที่ยวเป็นอย่างมาก เนื่องจาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมืองรองเป็นเมืองที่มีนักท่องเที่ยวเดินทางมาท่องเที่ยวไม่มากนักจึงสามารถช่วยลดการพบปะกับนักท่องเที่ยวจำนวนมากได้เป็นอย่างดี และยังมีผลสำรวจจาก จอห์น บราวน์ ประธานเจ้าหน้าที่บริหาร อโกด้า (Agoda) ประเทศไทย เปิดข้อมูลเกี่ยวกับสถานการณ์การท่องเที่ยวของประเทศไทย เกี่ยวกับพฤติกรรมท่องเที่ยวของนักท่องเที่ยวหลังผ่านสถานการณ์ Covid-19 พบว่านักท่องเที่ยวชาวไทยหันมาสนใจเมืองรองมากขึ้น เนื่องจากมีสถานที่ท่องเที่ยวในเมืองรองเริ่มเข้ามาติด Top 10 การค้นหาจุดหมายปลายทางมากที่สุดในอโกด้า และไม่เพียงแต่สถานการณ์ Covid-19 ที่ช่วยกระตุ้นการท่องเที่ยวในเมืองรองโดยทางอ้อมแล้ว ในส่วนของภาครัฐก็ได้เล็งเห็นความสำคัญในการกระตุ้นการท่องเที่ยวในเมืองรอง โดยรัฐบาลมีนโยบายชัดเจนให้ความสำคัญด้านการท่องเที่ยวด้วยการผลักดันให้นักท่องเที่ยวเดินทางเชื่อมโยงจากเมืองหลักสู่เมืองรองมากขึ้น เช่น ในปี 2562 ภาครัฐได้ออกกิจกรรม “เที่ยวเมืองรองลดหย่อนภาษี” โดยใช้ใบเสร็จรับเงินจากค่าแพคเก็จทัวร์ หรือค่าตั๋วเครื่องบินมาใช้ลดหย่อนภาษี ซึ่งไม่แม้แต่รัฐบาลที่มีกิจกรรมส่งเสริมการท่องเที่ยวเมืองรอง ยังรวมไปถึงภาคเอกชนหลายๆ แห่งที่ร่วมมือกับการท่องเที่ยวแห่งประเทศไทย (ททท.) เพื่อออกกิจกรรมส่งเสริมการท่องเที่ยวเมืองรอง เช่น บริษัท ปตท. น้ำมันและการค้าปลีก จำกัด (มหาชน) ออกกิจกรรมที่ชื่อว่า “เที่ยวเมืองรองความสุขยกกำลัง 2 กับ พีทีที บลูการ์ด” เพื่อสะสมคะแนนในขณะท่องเที่ยวเมืองรอง และสามารถนำคะแนนไปแลกของรางวัลมากมาย และอีกกิจกรรมสำหรับในช่วงฤดูกาลท่องเที่ยวของปี 2563 จนถึงสิ้นปี 2564 การท่องเที่ยวแห่งประเทศไทย(ททท) ร่วมกับ จส.100 ได้จัดทำโครงการ “เที่ยวเมืองรอง ช่วง LONG WEEKEND” เพื่ออำนวยความสะดวกแก่นักท่องเที่ยวในเรื่องของการเดินทาง เช่น ข้อมูลการจราจร เส้นทางถนน และทางเลี่ยงต่างๆ เพื่อสร้างความเชื่อมั่นและความปลอดภัยให้แก่นักท่องเที่ยวในการท่องเที่ยวเมืองรอง และปัจจุบันในปี 2567 รัฐบาลได้ออกมาตรการเที่ยวเมืองรองลดหย่อนภาษีสูงสุด 15,000 บาท ในช่วง low season เพื่อหวังกระตุ้นให้เงินสะพัดและกระตุ้นการท่องเที่ยวเมืองรองอีกด้วย

ด้วยเหตุผลดังที่กล่าวจึงส่งผลให้เมืองรองได้รับความสนใจ และถูกจับตามองจากนักท่องเที่ยวชาวไทยเพิ่มมากขึ้น แต่เนื่องจากสถานที่ท่องเที่ยวในเมืองรองยังเป็นสถานที่แปลกใหม่ และไม่คุ้นชินมากนักสำหรับนักท่องเที่ยวส่วนมาก อาจทำให้ยากลำบากในการวางแผนการท่องเที่ยวไปยังเมืองรอง และอีกหนึ่งปัจจัยที่ผู้เขียนเล็งเห็นถึงความสำคัญในการส่งเสริมการท่องเที่ยวเมืองรองคือการกระจายรายได้จากเมืองหลักเข้าสู่เมืองรองเพื่อยกระดับคุณภาพชีวิตของคนในท้องถิ่น ผู้วิจัยจึงได้คิดค้นระบบแนะนำสถานที่ท่องเที่ยวในเมืองรอง โดยประยุกต์ใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) และเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) เพื่อเลือกใช้เทคนิคที่เหมาะสมที่สุดในการทำระบบแนะนำ และเพื่อตอบโจทย์กับการส่งเสริมการท่องเที่ยวเมืองรอง ซึ่งระบบนี้ถูกจัดทำขึ้นเพื่อพัฒนาและส่งเสริม 2 ด้าน คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ด้านระบบแนะนำ และด้านศึกษาปัจจัยที่ส่งผลต่อการท่องเที่ยวเมืองรอง ในด้านระบบแนะนำนั้น ประยุกต์ใช้เทคนิคการจำแนกประเภท (Classification) ที่หลากหลายวิธี เพื่อเป็นการเปรียบเทียบและหาอัลกอริทึม ที่เหมาะสมต่อการสร้างระบบ ซึ่งทำให้ระบบมีความถูกต้องมากที่สุด โดยใช้ข้อมูลจากพฤติกรรมการท่องเที่ยว และข้อมูลประชากรของนักท่องเที่ยว เพื่อคาดการณ์สถานที่ท่องเที่ยวให้ตรงกับรสนิยม และพฤติกรรมการท่องเที่ยวของผู้ใช้มากที่สุด โดยระบบแนะนำนอกจากจะช่วยอำนวยความสะดวกแก่ผู้ใช้แล้ว ยังสามารถช่วยประหยัดเวลาในการค้นหาสถานที่ท่องเที่ยวในเมืองรองได้อีกด้วย และในส่วนของด้านศึกษาปัจจัยที่ส่งผลการท่องเที่ยวเมืองรอง ประยุกต์ใช้วิธีต้นไม้ตัดสินใจเพื่อศึกษาปัจจัยที่ส่งผลให้นักท่องเที่ยวเลือกเดินทางมายังสถานที่ท่องเที่ยวเมืองรอง ไม่ว่าจะเป็นด้านวิธีการเดินทาง ด้านข้อมูลประชากร หรือด้านงบประมาณที่จำกัด และยังสามารถนำความรู้ที่ได้จากการศึกษาปัจจัยเหล่านี้ไปช่วยใช้ในการพัฒนาเมืองรอง ช่วยจัดการวางแผนการเดินทาง หรือช่วยในด้านการออกแบบแพคเก็จที่เหมาะสมกับสถานที่ท่องเที่ยว กระตุ้นให้นักท่องเที่ยวเดินทางเข้าสู่เมืองรอง เพื่อให้เม็ดเงินมาสะพัดที่เมืองรองและกระจายลงสู่ท้องถิ่น หนุนให้เกิดการกระจายรายได้อย่างรวดเร็ว เพื่อลดความเหลื่อมล้ำภายในสังคม ยกระดับรายได้ในชุมชนและท้องถิ่นให้ดีขึ้น ซึ่งเป็นรากฐานสำคัญของการพัฒนาสังคมและเศรษฐกิจไทยอย่างยั่งยืน

1.2 วัตถุประสงค์ของงานวิจัย

1) เพื่อสำรวจประสบการณ์ และปัจจัยที่มีผลต่อการท่องเที่ยวของนักท่องเที่ยวชาวไทย และต่อสถานที่ท่องเที่ยวในเมืองรองของประเทศไทย

2) เพื่อออกแบบ และพัฒนาแบบจำลองในการสร้างระบบแนะนำสถานที่ท่องเที่ยวเมืองรองตามรสนิยมของแต่ละบุคคล

1.3 ขอบเขตของงานวิจัย

1) งานวิจัยนี้เป็นการศึกษาชุดข้อมูลจากการเก็บแบบสอบถามเกี่ยวกับพฤติกรรม และประสบการณ์การท่องเที่ยวเมืองรองภายในประเทศไทยผ่าน Google Forms ทั้งหมด 429 ตัวอย่าง

2) งานวิจัยนี้มุ่งเน้นศึกษาพฤติกรรม และประสบการณ์การท่องเที่ยวของนักท่องเที่ยวชาวไทยเท่านั้น

3) งานวิจัยนี้ประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ได้แก่ ต้นไม้ตัดสินใจ (Decision Tree) การสุ่มแบบป่า (Random Forest) เพื่อนบ้านใกล้เคียงที่สุด (k-Nearest Neighbor: k-

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

NN) เพอร์เซ็ปตรอนหลายชั้น (Multilayer perceptron: MLP) เพื่อหาวิธีที่เหมาะสมที่สุดในแต่ละแบบจำลอง

1.4 ขั้นตอนการศึกษาและดำเนินการวิจัย

1) ศึกษาเกี่ยวกับปัญหา และทฤษฎีที่นำมาวิจัย ต้นไม้ตัดสินใจ (Decision Tree) การสุ่มแบบป่า (Random Forest) เพื่อนบ้านใกล้เคียงที่สุด (k-Nearest Neighbor: k-NN) เพอร์เซ็ปตรอนหลายชั้น (Multilayer perceptron: MLP) การเลือกคุณลักษณะที่เหมาะสม (Feature Selection) การหาความสัมพันธ์ของตัวแปร (Chi-Square: $\boldsymbol{\chi}^2$)

2) ออกแบบข้อคำถามเกี่ยวกับพฤติกรรมการท่องเที่ยว และเก็บประชากรตัวอย่างจากแบบสอบถาม

3) จัดเก็บและรวบรวมข้อมูลที่ได้จากการเก็บแบบสอบถาม นำข้อมูลมาวิเคราะห์ด้วย การเลือกคุณลักษณะที่เหมาะสม (Feature Selection) การหาความสัมพันธ์ของตัวแปร (Chi-Square: $\boldsymbol{\chi}^2$) เลือกข้อคำถามส่งผลต่อการตัดสินใจเลือกสถานที่ท่องเที่ยวของชาวไทย

4) ออกแบบและสร้างแบบจำลองในงานวิจัยโดยประยุกต์ใช้เทคนิค การเรียนรู้ของเครื่อง (Machine Learning)

5) ทดสอบแบบจำลองและวัดประสิทธิภาพของแบบจำลอง

6) วิเคราะห์ผลการทดลอง

7) พัฒนาเว็บแอปพลิเคชันสำหรับระบบแนะนำการท่องเที่ยวในเมืองรอง

8) สรุปผลการทดลอง

9) เรียบเรียงวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1) ได้แบบจำลองสำหรับการแนะนำสถานที่ท่องเที่ยวเมืองรองที่สามารถใช้งานได้อย่างถูกต้อง มีประสิทธิภาพ

2) เพื่อส่งเสริมและกระตุ้นการท่องเที่ยวเมืองรองภายในประเทศไทย ให้มีรายได้จากการท่องเที่ยวมากยิ่งขึ้น เพื่อกระจายรายได้สู่เมืองรอง ส่งผลต่อการกระตุ้นเศรษฐกิจของประเทศ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 เมืองรองในประเทศไทย

เมืองรอง หมายถึง เมืองที่ไม่ได้เป็นเมืองท่องเที่ยวหลัก หรือยังมีนักท่องเที่ยวเข้าไปไม่มากนัก การกระจายรายได้ถึงชุมชนท้องถิ่นจึงแตกต่างกับเมืองท่องเที่ยวหลัก เกณฑ์การพิจารณาความเป็นเมืองรองคือ เมืองดังกล่าวมีนักท่องเที่ยวเข้าไปเที่ยวไม่ถึง 6 ล้านคนต่อปี โดยเมืองรองในประเทศไทยมีทั้งหมด 55 จังหวัด กระจายอยู่ในทุกภาคของประเทศ ดังนี้

1. ภาคเหนือ 16 จังหวัด ได้แก่ จังหวัดเชียงราย จังหวัดพิษณุโลก จังหวัดตาก จังหวัดเพชรบูรณ์ จังหวัดนครสวรรค์ จังหวัดสุโขทัย จังหวัดลำพูน จังหวัดอุตรดิตถ์ จังหวัดลำปาง จังหวัดแม่ฮ่องสอน จังหวัดพิจิตร จังหวัดแพร่ จังหวัดน่าน จังหวัดกำแพงเพชร จังหวัดอุทัยธานี และจังหวัดพะเยา

2. ภาคตะวันออกเฉียงเหนือ 18 จังหวัด ได้แก่ จังหวัดอุดรธานี จังหวัดอุบลราชธานี จังหวัดหนองคาย จังหวัดเลย จังหวัดมุดาหาร จังหวัดบุรีรัมย์ จังหวัดชัยภูมิ จังหวัดศรีสะเกษ จังหวัดสุรินทร์ จังหวัดสกลนคร จังหวัดนครพนม จังหวัดร้อยเอ็ด จังหวัดมหาสารคาม จังหวัดบึงกาฬ จังหวัดกาฬสินธุ์ จังหวัดยโสธร จังหวัดหนองบัวลำภู และ จังหวัดอำนาจเจริญ

3. ภาคกลาง 7 จังหวัด ได้แก่ จังหวัดลพบุรี จังหวัดสุพรรณบุรี จังหวัดราชบุรี จังหวัดสมุทรสงคราม จังหวัดชัยนาท จังหวัดอ่างทอง และจังหวัดสิงห์บุรี

4. ภาคตะวันออก 5 จังหวัด ได้แก่ จังหวัดสระแก้ว จังหวัดตราด จังหวัดจันทบุรี จังหวัดปราจีนบุรี และจังหวัด นครนายก

5. ภาคใต้ 9 จังหวัด ได้แก่ จังหวัดนครศรีธรรมราช จังหวัดพัทลุง จังหวัดตรัง จังหวัดสตูล จังหวัดชุมพร จังหวัดระนอง จังหวัดนราธิวาส จังหวัดยะลา และจังหวัดปัตตานี

2.2 การเรียนรู้ของเครื่อง (Machine Learning Algorithm)

Machine learning เป็นสาขาหนึ่งของปัญญาประดิษฐ์ที่มุ่งเน้นการพัฒนาและการใช้งานแบบจำลองทางคณิตศาสตร์และวิศวกรรมคอมพิวเตอร์ เพื่อให้คอมพิวเตอร์สามารถเรียนรู้และปรับปรุงพฤติกรรมได้ด้วยตัวเองจากข้อมูล โดยไม่ต้องมีการโปรแกรมโดยตรง โดยเรามักจะใช้ข้อมูลขนาดใหญ่ (Big Data) และเทคนิคทางคณิตศาสตร์ เช่น การจำแนกประเภท (Classification) การทำนายค่าต่อเนื่อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Regression) การจัดกลุ่ม (Clustering) และการเรียนรู้แบบมีกำหนดระบบรากฐาน (Reinforcemet Learning) เพื่อทำนายผลลัพธ์หรือตัดสินใจต่างๆ อัตโนมัติและมีประสิทธิภาพมากยิ่งขึ้นตามข้อมูลที่มีให้มากขึ้น การเรียนรู้เชิงเครื่องมักนำไปใช้ในหลายแนวทาง เช่น การแยกแยะอีเมลล์ขยะ การวิเคราะห์ภาพทางการแพทย์ การตรวจจับโจมตีไซเบอร์ การพยากรณ์ทางการเงิน และอื่นๆ อีกมากมาย โดยมีการใช้งานกว้างขวางในภาคธุรกิจและภาครัฐทั้งหมด โดยหลักการของ Machine Learning ที่นิยมนำมาประยุกต์ใช้ในการการทำเหมืองข้อมูล (data mining) จะแบ่งออกเป็น 2 ประเภทตามรูปแบบการเรียนรู้ด้วยกัน นั่นคือ supervised learning และ unsupervised learning ดังนี้

1. Supervised Learning (การเรียนรู้แบบมีการควบคุม)

   Supervised learning เป็นกระบวนการที่ใช้ข้อมูลที่มีป้ายกำกับ (labeled data) เพื่อการเรียนรู้แบบจำลอง เมื่อต้องการจำแนกประเภทข้อมูลหรือทำนายผลลัพธ์ที่ถูกต้อง โดยในขั้นตอนการฝึก (training) ข้อมูลจะถูกนำเข้าแบบจำลองพร้อมกับป้ายกำกับเพื่อให้แบบจำลองเรียนรู้จากข้อมูลที่มีการตอบสนองที่ถูกต้อง หลังจากฝึกเสร็จสิ้น แบบจำลองนั้นสามารถทำนายผลลัพธ์ของข้อมูลใหม่ที่ไม่มีป้ายกำกับได้

2. Unsupervised Learning (การเรียนรู้แบบไม่มีการควบคุม):

   Unsupervised learning เป็นกระบวนการที่ใช้ข้อมูลที่ไม่มีป้ายกำกับ (unlabeled data) เพื่อการเรียนรู้แบบจำลอง โดยแบบจำลองจะพยายามหาโครงสร้างภายในข้อมูลโดยไม่มีคำแนะนำจากป้ายกำกับ ส่วนใหญ่จะเป็นการจัดกลุ่ม (clustering) ข้อมูลหรือการลดมิติ (dimensionality reduction) เพื่อให้เราเข้าใจโครงสร้างหรือลักษณะเฉพาะของข้อมูลได้ดีขึ้น

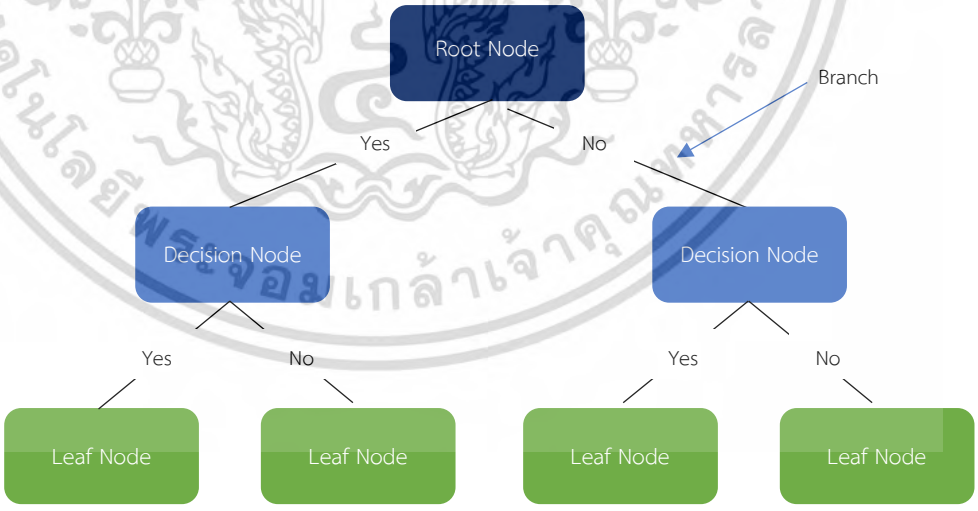
โดยการจำแนกประเภทข้อมูลหรือ Classification เป็นกระบวนการทางคณิตศาสตร์ที่ใช้การเรียนรู้จากข้อมูลที่มีอยู่เพื่อสร้างแบบจำลองที่สามารถแบ่งแยกข้อมูลออกเป็นกลุ่มหรือประเภทต่างๆ ตามลักษณะและคุณลักษณะของข้อมูล โดยแบบจำลอง ที่สร้างขึ้นจะใช้กฎหรือลักษณะเฉพาะที่ได้เรียนรู้จากข้อมูลในขั้นตอนการฝึกจากข้อมูลเรียนรู้ (Training Data) โดยใช้ข้อมูลที่มีป้ายกำกับ (Label) จากนั้นแบบจำลองที่เรียนรู้มานั้นสามารถใช้ในการจำแนกประเภทข้อมูลใหม่ที่ไม่มีป้ายกำกับหรือต่อการจำแนกประเภทข้อมูลในอนาคต ในทฤษฎีการจำแนกประเภทข้อมูลนั้น มักจะใช้วิธีการแบ่งกลุ่มหรือจัดการแบบเฉพาะเจาะจง เช่น ในการจำแนกประเภทลูกค้าในธุรกิจค้าปลีก หรือการจำแนกประเภทโรคของผู้ป่วย เป็นต้น ซึ่งการจำแนกประเภทข้อมูลมีการประยุกต์ใช้ในหลายงาน เช่น การจำแนกตามลักษณะเฉพาะของบุคคล (Characters classification) การจำแนกประเภทของภาพถ่าย (Image classification) การจำแนกประเภทของข้อความ (Text classification) หรือการจำแนกประเภทของสัญญาณ (Signal

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

classification) และอื่นๆ อีกมากมาย โดยมีการใช้เทคนิคและวิธีการต่างๆ เช่น เรียนรู้เชิงลึก (Deep learning) การใช้เทคนิคการเรียนรู้เชิงสถิติ (Statistical learning) หรือการใช้เทคนิคการเรียนรู้เชิงกฎ (Rule-based learning) ในการพัฒนาแบบจำลองสำหรับการจำแนกประเภทข้อมูลต่างๆ ออกไป กล่าวได้ว่าการจำแนกประเภทเป็นเครื่องมือที่มีประโยชน์อย่างมากในหลายด้านของชีวิตประจำวันและธุรกิจ เพื่อช่วยในการตัดสินใจ การวินิจฉัย หรือการทำนายในอนาคตโดยใช้ข้อมูลที่มี ขั้นตอนวิธีที่ใช้ในงานวิจัยนี้ (Algorithms) ได้แก่

2.2.1 ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจเป็นวิธีการสร้างแบบจำลองทางคณิตศาสตร์ที่ใช้ในการตัดสินใจแบบชั้นเชิงลึก (hierarchical) ซึ่งสามารถใช้ในการจำแนกประเภทข้อมูลและทำนายผลลัพธ์ของข้อมูลได้ โดยใช้กฎตัดสินใจ (decision rules) ซึ่งอยู่ในรูปแบบของโครงสร้างต้นไม้ โครงสร้างของต้นไม้ตัดสินใจประกอบด้วยโหนดตัดสินใจ (Decision Node) และเส้นสาย (Branch) โหนดแรกที่อยู่บนสุดเรียกว่า โหนดราก (Root Node) โดยสามารถแบ่งข้อมูลออกเป็นกลุ่มย่อยๆ ตามลักษณะต่างๆ ของข้อมูลที่กำหนดโดยกฎตัดสินใจที่อยู่ในแต่ละโหนด และโหนดสุดท้ายที่ไม่มีการแบ่งข้อมูลแล้วเรียกว่า โหนดใบ (Leaf Node) โดยในโหนดใบจะมีการกำหนดคลาสหรือค่าทำนายสำหรับข้อมูลในกลุ่มนั้น ดังรูปที่ 2.1



รูปที่ 2.1 แผนภาพต้นไม้ตัดสินใจ (Decision Tree)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การใช้ต้นไม้ตัดสินใจ (Decision Tree) ที่นิยมใช้มี 2 ประเภทคือ Gini และ Entropy เป็นวิธีการเลือกโหนดแรก (root node) ในต้นไม้ตัดสินใจ (Decision Tree) มีวัตถุประสงค์เพื่อหากฎตัดสินใจที่เหมาะสมที่สุดในการแบ่งข้อมูล เพื่อสร้างแบบจำลองที่มีประสิทธิภาพสูงสุด โดยมีรายละเอียดดังนี้

1. ดัชนี Gini (Gini Index)

การใช้ Gini Index เป็นวิธีการวัดความไม่บริสุทธิ์ (impurity) ของข้อมูลในโหนด เนื่องจาก Gini Index นั้นคำนึงถึงความน่าจะเป็นที่ข้อมูลจะถูกจำแนกผิดคลาส ค่า Gini Index จะอยู่ในช่วงระหว่าง 0 ถึง 0.5 โดยที่ค่า 0 หมายถึงข้อมูลไม่มีความไม่บริสุทธิ์เลย และค่า 0.5 หมายถึงข้อมูลมีความไม่บริสุทธิ์สูงสุด โดยการใช้ Gini Index นั้นช่วยให้สามารถปรับปรุงแบบจำลอง Decision Tree เพื่อให้มีความแม่นยำและประสิทธิภาพในการทำนายที่ดีขึ้นได้ คำนวณได้จากสมการที่ 2.1

$$Gini(p) = 1 - \sum_{i=1}^{K} p_i^2 \quad (2.1)$$

โดยที่ p_i คือความน่าจะเป็นของข้อมูลในกลุ่มหรือคลาส i และ K คือจำนวนกลุ่มหรือคลาสทั้งหมดที่เป็นไปได้ในข้อมูล

2. เอ็นโทรปี (Entropy)

เป็นการวัดความสับสนในข้อมูล โดยค่า Entropy จะอยู่ในช่วงระหว่าง 0 ถึง 1 โดยที่ค่า 0 หมายถึงข้อมูลไม่มีความสับสนเลย และค่า 1 หมายถึงข้อมูลมีความสับสนสูงสุด คำนวณได้จากสมการที่ 2.2

$$H(p) = -\sum_{i=1}^{K} p_i log_2(p_i) \quad (2.2)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.1 Peudo code การทำงานของขั้นตอนวิธีต้นไม้ตัดสินใจ (Decision Tree)

Decision Tree Algorithms
Function BuildDecisionTree(Data, Target, Features): Create a root node for the tree **If** all examples in Data are in the same class: **Return** a leaf node with class label **If** Features is empty: **Return** a leaf node with the most common class label in Data **Else:** BestFeature <- Select the best feature to split on from Features using a criterion (e.g., information gain) Split Data into subsets using BestFeature, creating a branch for each unique value of BestFeature **For** each branch: SubsetData <- The subset of Data corresponding to the branch **If** SubsetData is empty: Attach a leaf node with the most common class label in Data to the branch **Else:** Attach the node returned by BuildDecisionTree(SubsetData, Target, Features - {BestFeature}) to the branch **Return** root node

จากตารางที่ 2.1 แสดงขั้นตอนการทำงานของต้นไม้ตัดสินใจ (Decision Tree) สามารถสรุปขั้นตอนได้ดังนี้

1. สร้างต้นไม้ตัดสินใจด้วยการรับข้อมูลได้แก่ ข้อมูล (Data), เป้าหมาย (Target), และคุณลักษณะ (Feature)
2. ถ้าหน่วยตัวอย่างทั้งหมดในข้อมูลอยู่ในคลาสเดียวกัน ให้คืนโหนดใบ (Leaf Node) พร้อมป้ายกำกับคลาส (Class Label)

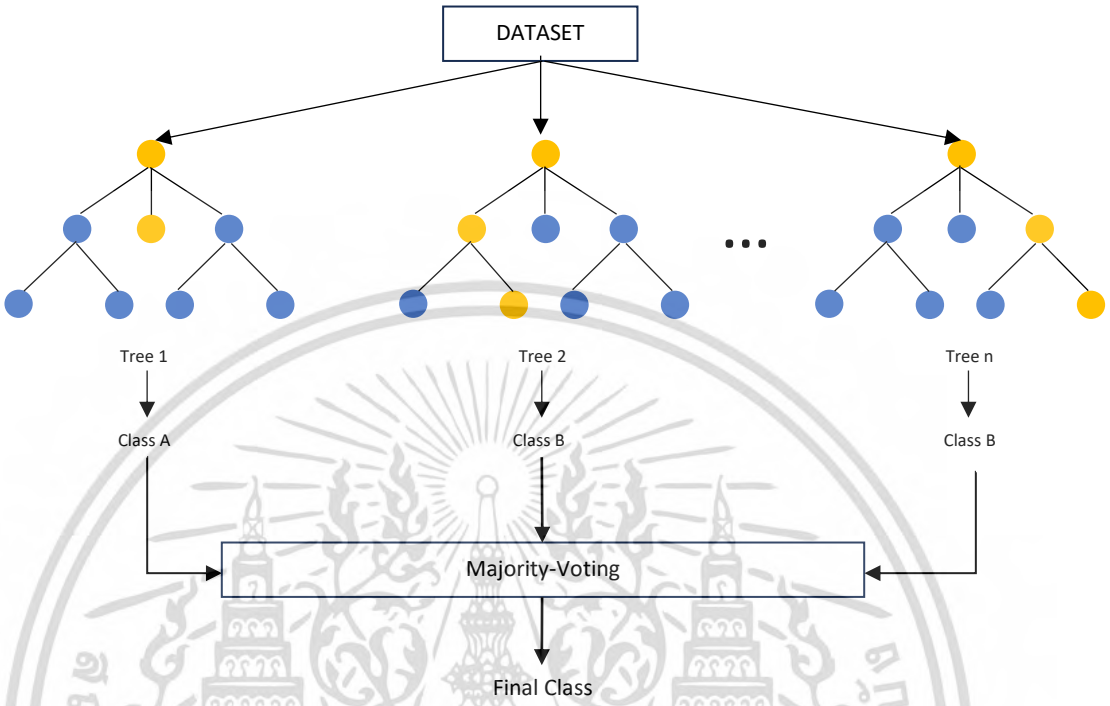
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ถ้าไม่มีคุณลักษณะ (Feature) ให้คืนโหนดใบ (Leaf Node) พร้อมป้ายกำกับคลาส (Class Label) ที่พบบ่อยที่สุดในข้อมูล
4. ถ้ามีคุณลักษณะ (Feature) ให้ลือกคุณลักษณะที่ดีที่สุดในการแบ่งข้อมูลจากคุณลักษณะด้วยเกณฑ์ที่กำหนด และเก็บในตัวแปร 'BestFeature'
5. แบ่งข้อมูลออกเป็นกลุ่มย่อยตาม 'BestFeature' สร้างเส้นสาขา (Branch) สำหรับแต่ละค่าของคุณลักษณะนั้น
6. ในส่วนของแต่ละเส้นสาขา (Branch) ให้เก็บรายละเอียดขอข้อมูลที่มีความเกี่ยวข้องกับเส้นสาขาไว้ในตัวแปร 'SubsetData'
7. ถ้าไม่มีข้อมูลใน 'SubsetData' ให้ใส่โหนดใบพร้อมป้ายกำกับคลาสที่พบบ่อยที่สุดในข้อมูลไปยังสาขา
8. ถ้ามีข้อมูลใน 'SubsetData' ให้ใส่โหนดที่ได้จากฟังก์ชันสร้างต้นไม้ตัดสินใจ (SubsetData, Target, Features - {BestFeature} ไปยังสาขา
9. คืนโหนดราก (Root Node)

2.2.2 การสุ่มแบบป่าไม้ (Random Forest)

Random forest เป็นหนึ่งในกลุ่มของแบบจำลองที่เรียกว่า Ensemble learning ซึ่งมีหลักการคือการฝึกแบบจำลอง (train Model) ที่เหมือนกันหลายครั้งบนข้อมูลชุดเดียวกัน โดยแต่ละครั้งของการฝึกฝนจะเลือกส่วนของข้อมูลที่ฝึกฝนไม่เหมือนกัน จากนั้นจะนำการตัดสินใจของแบบจำลองเหล่านั้นมาโหวตกันเพื่อตัดสินใจในการแยกแยะของข้อมูลและสร้างแบบจำลองที่มีประสิทธิภาพมากยิ่งขึ้น ด้วยกระบวนการที่มีการสร้างต้นไม้แบบสุ่มและกระบวนการโหวตผลลัพธ์ของต้นไม้ทั้งหมดดังรูปที่ 2.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.2 แผนภาพการสุ่มแบบป่า (Random Forest)

เทคนิค Random Forest สามารถลดโอกาสที่เกิด overfitting และมีประสิทธิภาพในการทำนายที่มีความแม่นยำและเชื่อถือได้สูง และมักถูกนำมาใช้ในงานที่ต้องการการจำแนกประเภทหรือการทำนายที่มีความแม่นยำและเชื่อถือได้สูง เช่น การวิเคราะห์ข้อมูลทางการแพทย์หรือการทำนายทางการเงิน และอื่นๆ อีกมากมาย

ตารางที่ 2.2 Peudo code การทำงานของขั้นตอนวิธีเทคนิคการสุ่มแบบป่าไม้ (Random Forest)

Random Forest Algorithms
Initialize: Determine the number of trees, N. Determine the sample size for each tree. **For** each tree from 1 to N: Create a bootstrap sample of the original dataset.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Random Forest Algorithms

Build a decision tree with the following modifications:

At each node:

If the node is pure (all samples belong to one class) or a stopping criterion is met (e.g., maximum depth, minimum node size):

Mark the node as a leaf and assign the most common class.

Else:

Randomly select features without replacement to consider for splitting.

Choose the best feature and split value to maximize the separation (e.g., based on Gini impurity or information gain).

Split the node into two child nodes.

Prediction:

For classification tasks:

For a new sample, make each tree in the forest predict the class and take a majority vote as the final prediction.

For regression tasks:

For a new sample, make each tree predict the value and take the average of these predictions as the final output.

จากตารางที่ 2.2 แสดงขั้นตอนการทำงานของเทคนิคการสุ่มแบบป่าไม้ (Random Forest) สามารถสรุปขั้นตอนได้ดังนี้

1. กำหนดค่าเริ่มต้น โดยกำหนดจำนวนต้นไม้ (N) และกำหนดขนาดตัวอย่างสำหรับแต่ละต้นไม้
2. สร้าง bootstrap sample จากชุดข้อมูลต้นฉบับในแต่ละต้นไม้
3. **สร้างต้นไม้แต่ละโหนดด้วยมีเงื่อนไขคือ ถ้าโหนดนั้นๆ บริสุทธิ์หรือเงื่อนไขถูกหยุดให้สร้างโหนดใบ (Leaf Node) และกำหนดคลาส (Class Label) ที่พบบ่อยที่สุด**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. ถ้าโหนดนั้นๆ ไม่บริสุทธิ์ ให้ทำ 3 ขั้นตอนคือ 1) เลือกคุณลักษณะอย่างสุ่มโดยไม่มีการเลือกซ้ำเพื่อพิจารณาการแบ่ง 2) เลือกคุณลักษณะและค่าการแบ่งที่ดีที่สุดเพื่อเพิ่มการแยก 3) แบ่งโหนดออกเป็นโหนดลูกสองโหนด.
5. ในส่วนของการทำนาย หากต้องการจำแนกประเภท ให้แต่ละต้นไม้ทำนายคลาสและใช้การโหวตเสียงข้างมากเป็นผลลัพธ์ หากต้องการถดถอย (Regression) ให้ต้นไม้ทำนาย และใช้ค่าเฉลี่ยของการทำนายเป็นผลลัพธ์

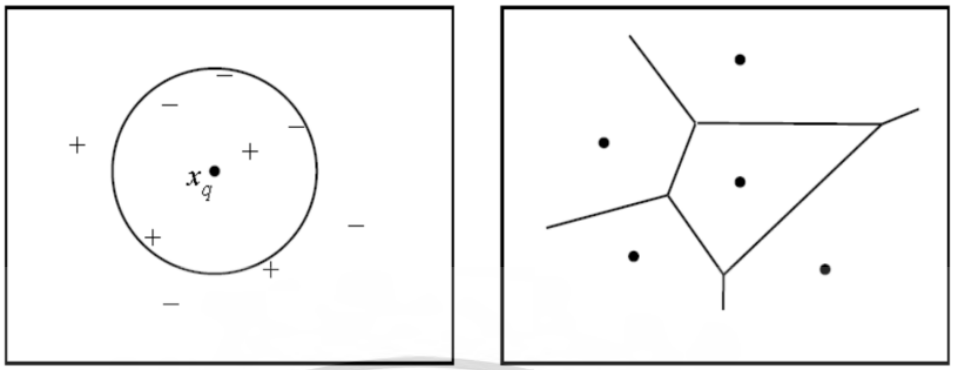
2.2.3 เพื่อนบ้านใกล้เคียง (k-Nearest Neighbors: k-NN)

การใช้วิธีของ k-Nearest Neighbors (kNN) เป็นอัลกอริทึมในการจำแนกประเภท (Classification) นำเสนอแนวคิดที่ใช้ในการจัดกลุ่มข้อมูลโดยพิจารณาความคล้ายคลึงระหว่างข้อมูลแต่ละตัว โดยการกำหนดค่าของพารามิเตอร์ k ซึ่งเป็นจำนวนของเพื่อนบ้านที่ใกล้ที่สุดที่จะใช้ในการตัดสินใจว่าข้อมูลที่กำลังจะจำแนกควรจะอยู่ในกลุ่มใด โดยการทำงานของ k-NN มีขั้นตอนดังนี้

1. การคำนวณระยะหรือความคล้ายคลึง ในขั้นตอนนี้จะใช้วิธีการวัดระยะหรือความคล้ายคลึงระหว่างข้อมูลที่ต้องการจำแนกกับข้อมูลที่มีอยู่แล้วในชุดข้อมูล ซึ่งส่วนใหญ่จะใช้ระยะหรือความคล้ายคลึงเชิงระยะทาง (Euclidean distance) หรือระยะห่างมิติ (Manhattan distance) เป็นต้น
2. การค้นหาเพื่อนบ้าน หลังจากคำนวณระยะหรือความคล้ายคลึงแล้ว จะเลือกข้อมูล k ตัวที่มีค่าระยะหรือความคล้ายคลึงน้อยที่สุดมาเป็นเพื่อนบ้าน
3. การหาความถูกต้อง ในขั้นตอนนี้จะใช้โหวตโดยเลือกคลาสที่มีจำนวนมากที่สุดในเพื่อนบ้าน k ตัว และกำหนดข้อมูลที่กำลังจะจำแนกให้เป็นคลาสนั้น
4. การทำนาย หลังจากที่กำหนดคลาสให้กับข้อมูลที่กำลังจะจำแนกแล้ว สามารถใช้แบบจำลองที่ได้มาใช้ทำนายกับข้อมูลใหม่ๆ

ขั้นตอนดังกล่าวแสดงเป็นภาพดังรูปที่ 2.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



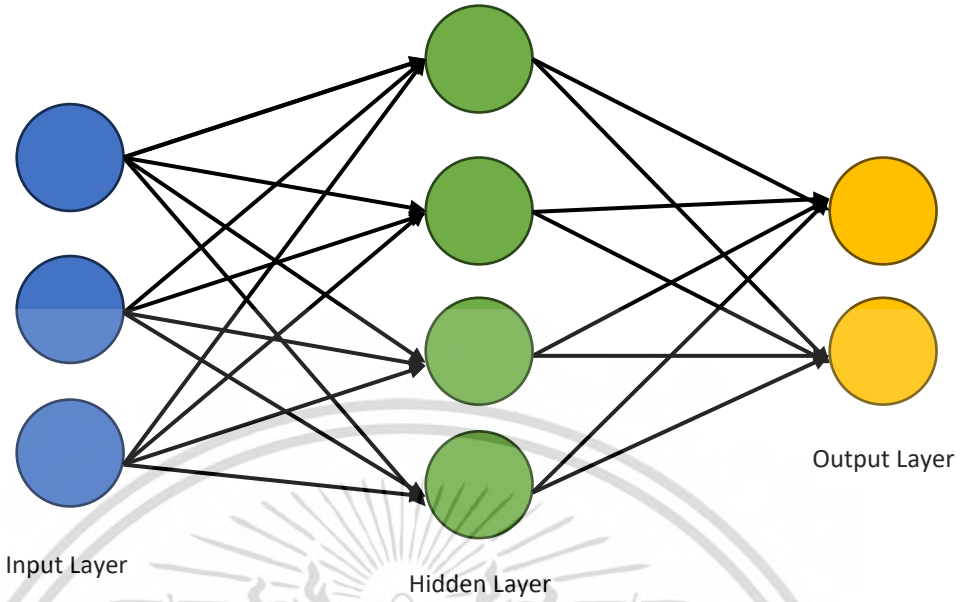
รูปที่ 2.3 แผนภาพเทคนิคการหาเพื่อนบ้านใกล้เคียง (k-Nearest Neighbors: k-NN)

2.2.4 เพอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron: MLP)

Multilayer Perceptron (MLP) เป็นแบบจำลองของโครงข่ายประสาทเทียม (Artificial Neural Network) ที่มีโครงสร้างที่มีหลายชั้น (multilayer) ประกอบด้วยชั้นของโหนด (nodes) หรือเรียกว่าเซลล์ประสาท (neurons) ในแต่ละชั้น ซึ่งแต่ละโหนดในชั้นก่อนหน้าจะเชื่อมต่อกับโหนดในชั้นถัดไป และโหนดในชั้นสุดท้ายจะเป็นชั้นเอาต์พุต (output layer) ซึ่งจะให้ผลลัพธ์ของการทำนายหรือการจำแนกประเภทของข้อมูลดังรูปที่ 2.4 โดย MLP มีสองชั้นหลักคือ

1. ชั้นซ่อน (hidden layer): เป็นชั้นที่ตัวแบบจำลองจะเรียนรู้และทำคำนวณในระหว่างกระบวนการฝึก (training) โดยมีหน้าที่เรียนรู้ลักษณะหรือลักษณะเชิงลึกของข้อมูล เช่น ลักษณะที่ซับซ้อนหรือความสัมพันธ์ที่ซับซ้อนระหว่างคุณลักษณะของข้อมูล
2. ชั้นเอาต์พุต (output layer): เป็นชั้นที่ให้ผลลัพธ์สุดท้ายของการทำนายหรือการจำแนกประเภทของข้อมูล โดยมักจะมีจำนวนโหนดเท่ากับจำนวนคลาสที่ต้องการจำแนกหรือพยากรณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.4 แผนภาพโครงข่ายประสาทเทียม (Multilayer Perceptron: MLP)

MLP มีการเชื่อมต่อของโหนดระหว่างชั้นที่ทำให้มีความยืดหยุ่นในการเรียนรู้และสามารถจำแนกหรือพยากรณ์ข้อมูลที่ซับซ้อนได้ดี เนื่องจากความสามารถในการเรียนรู้ลักษณะหรือลักษณะเชิงลึกของข้อมูลที่ซับซ้อน โดยสามารถใช้การคำนวณแบบถอดรหัส (backpropagation) เพื่อปรับค่าพารามิเตอร์ในแบบจำลองเพื่อให้มีการประมาณค่าที่ดีที่สุดของผลลัพธ์ที่ต้องการ โดยใช้วิธีการหาค่าความคลาดเคลื่อน (loss function) และเทคนิคการปรับค่าพารามิเตอร์เพื่อลดค่าคลาดเคลื่อนในการทำนายหรือจำแนกประเภทข้อมูล ซึ่งมักใช้เทคนิคการปรับค่าน้ำหนัก (weight adjustment) และการใช้ค่าความล่าช้า (momentum) เพื่อช่วยในกระบวนการฝึกแบบจำลองให้มีประสิทธิภาพมากขึ้น

ตารางที่ 2.3 Peudo code การทำงานของขั้นตอนเพอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron: MLP)

Multi-Layer Perceptron Algorithms
Initialize the weights and biases for all layers randomly. **For** each epoch (iteration over the entire dataset): **For** each input in the dataset: a. Forward Pass:

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Multi-Layer Perceptron Algorithms
Compute the input to each neuron in the hidden layers: **For** each hidden layer: Calculate neuron_input = weights * inputs + bias Apply ReLU activation function: **If** neuron_input > 0: neuron_output = neuron_input **Else:** neuron_output = 0 Use neuron_output as input for the next layer. Compute the output layer (similar to hidden layers, but may use a different activation function for tasks like classification or regression) b. Calculate Loss: Compare the output layer results with the actual target values to compute the loss. c. Backward Pass (Backpropagation): Update weights and biases to minimize the loss, using a gradient descent or any other optimization algorithm. Repeat until the algorithm converges or a maximum number of epochs is reached.

จากตารางที่ 2.3 แสดงขั้นตอนการทำงานของเทคนิคเพอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron: MLP)

1. กำหนดค่าน้ำหนักและค่าเอนเอียง (Bias)
2. การวนนำข้อมูลเข้าในแต่ละชุดข้อมูลมีดังนี้

   2.1 การส่งต่อข้อมูล (Forward Pass) เป็นคำนวณข้อมูลนำเข้าในแต่ละเซลล์ (neuron) ในชั้นซ่อน (hidden layers) โดยคำนวณตัวแปร neuron_input และนำไปใช้เป็นข้อมูลนำเข้าในชั้นถัดไป

   2.2 คำนวณความสูญเสีย (Calculate Loss) เป็นการเปรียบเทียบผลลัพธ์ของชั้นผลลัพธ์กับค่าเป้าหมายจริงเพื่อคำนวณความสูญเสีย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 การส่งย้อนกลับข้อมูล (Backpropagation) เป็นการอัปเดตน้ำหนักและค่าเอนเอียงเพื่อลดความสูญเสีย โดยใช้อัลกอริทึมหาค่าเหมาะที่สุด

3. ทำซ้ำจนกว่าอัลกอริทึมจะครบจำนวนการทำซ้ำที่กำหนด

2.3 การเตรียมข้อมูล (Data Preprocessing)

2.3.1 การเลือกคุณลักษณะ (Feature Selection)

การทำ Feature Selection เป็นกระบวนการที่ใช้เพื่อเลือกคุณลักษณะหรือตัวแปรที่มีผลต่อการทำนายหรือจำแนกประเภทของข้อมูล โดยการลดจำนวนของคุณลักษณะในชุดข้อมูลเพื่อลดความซับซ้อนของแบบจำลองและเพิ่มประสิทธิภาพในการฝึกแบบจำลอง หรือช่วยลดปัญหาการเรียนรู้ที่ไม่จำเป็น (overfitting) ทำให้แบบจำลองมีความสามารถในการทำนายหรือจำแนกประเภทข้อมูลที่ดีขึ้น การทำ Feature Selection เป็นขั้นตอนสำคัญในการกำหนดคุณลักษณะที่สำคัญและลดความซับซ้อนของแบบจำลองเพื่อให้ได้ผลลัพธ์ที่แม่นยำและมีประสิทธิภาพในการใช้งานจริง โดยอาศัยวิธีการที่เหมาะสมกับลักษณะของข้อมูลและการใช้งานของแบบจำลอง ข้อดีของการทำ Feature Selection คือ

1. ลดความซับซ้อน การลดจำนวนของคุณลักษณะช่วยลดความซับซ้อนของแบบจำลอง ทำให้แบบจำลองมีประสิทธิภาพและเร็วขึ้นในการฝึกและการทำนาย
2. ลดเวลาและทรัพยากรในการประมวลผล การลดขนาดของชุดข้อมูลทำให้กระบวนการประมวลผลเร็วขึ้น โดยไม่ต้องใช้ทรัพยากรในการจัดการกับคุณลักษณะที่ไม่จำเป็น
3. เพิ่มความแม่นยำ การเลือกคุณลักษณะที่สำคัญสามารถช่วยลด noise ในข้อมูล ทำให้แบบจำลองมีประสิทธิภาพในการทำนายมากขึ้น
4. ลดโอกาสเกิด Overfitting การลดจำนวนคุณลักษณะที่ไม่จำเป็นช่วยลดโอกาสในการเกิด overfitting ซึ่งเป็นปัญหาที่สำคัญในการฝึกแบบจำลอง

1) การเลือกคุณลักษณะที่สำคัญ (Feature Importance)

เป็นการหาความสำคัญของคุณลักษณะจากต้นไม้ (tree-based models) เช่น Decision Trees, Random Forests, Gradient Boosting Machines เป็นต้น โดยวิธีนี้จะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใช้วิธีการสร้างแบบจำลองในแต่ละโหนดของต้นไม้เพื่อหาค่าความสำคัญของแต่ละคุณลักษณะ โดยใช้การลดค่าความไม่แน่นอน (uncertainty) ในการแยกกลุ่มข้อมูล

เมื่อทำการฝึกแบบจำลองจะสามารถดึงค่าความสำคัญของแต่ละคุณลักษณะออกมาได้ ซึ่งมักจะแสดงในรูปแบบของลิสต์หรืออาเรย์ของค่าที่แสดงความสำคัญของแต่ละคุณลักษณะ โดยมักจะเรียงลำดับตามลำดับความสำคัญของคุณลักษณะในชุดข้อมูล ในการแยกกลุ่มข้อมูลในต้นไม้ หรือความสัมพันธ์กับผลลัพธ์ โดยค่าที่ได้จะอยู่ในช่วงของค่าที่แสดงถึงความสำคัญของคุณลักษณะในการทำนายผลลัพธ์ ซึ่งมักจะอยู่ระหว่าง 0 ถึง 1 โดยที่ค่ามากที่สุดจะบ่งบอกถึงคุณลักษณะที่มีผลต่อผลลัพธ์มากที่สุด และค่าที่น้อยที่สุดจะบ่งบอกถึงคุณลักษณะที่มีผลต่อผลลัพธ์น้อยที่สุด

2) ไคสแควร์ (Chi-Square)

การเลือกคุณลักษณะสำคัญโดยใช้ค่า p-value ของไคสแควร์ (Chi-square) เป็นวิธีที่นิยมใช้ในการทดสอบความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตามในการแยกข้อมูลเป็นกลุ่มหรือทำนายผลลัพธ์ วิธีนี้เป็นการทำ Feature Selection ในกระบวนการสร้างแบบจำลองที่มีประสิทธิภาพมากที่สุด การใช้ทดสอบสถิติเป็นขั้นตอนหนึ่งในกระบวนการนี้ช่วยให้สามารถคัดเลือกคุณลักษณะที่สำคัญที่สุดได้อย่างมีประสิทธิภาพในการสร้างแบบจำลองของระบบและการวิเคราะห์ข้อมูลในงานทางสถิติและวิทยาการคอมพิวเตอร์อย่างมีประสิทธิภาพ การประยุกต์ใช้เทคนิคนี้ช่วยเพิ่มความเป็นมาตรฐานในการวิเคราะห์ข้อมูลและการสร้างแบบจำลองให้มีประสิทธิภาพมากที่สุดในงานวิจัยและพัฒนาระบบต่างๆ กระบวนการทำงานของเทคนิคนี้สามารถแบ่งออกเป็นขั้นตอนต่างๆ ได้ดังนี้

1. กำหนดสมมติฐาน (Hypothesis Setting) เริ่มต้นด้วยการกำหนดสมมติฐานที่ต้องการทดสอบ เช่น สมมติฐานว่าความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตามนั้นมีความสำคัญทางสถิติ
2. สร้างตาราง Contingency คือการสร้างตารางที่แสดงความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตาม โดยจะมีเซลล์ของตารางเป็นความถี่ของการเกิดขึ้นระหว่างค่าของตัวแปรต้นและตัวแปรตาม
3. คำนวณค่าสถิติ Chi-square โดยใช้ตาราง Contingency ที่ได้สร้างขึ้นมา จะทำการคำนวณค่าสถิติ Chi-square ซึ่งเป็นตัววัดของความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตาม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. คำนวณ p-value จากค่าสถิติ Chi-square ที่ได้ จะทำการคำนวณค่า p-value ซึ่งเป็นความน่าจะเป็นที่ค่าสถิติ Chi-square ที่ได้มากกว่าหรือเท่ากับค่าที่สัมพันธ์กัน จะเป็นผลจากการสุ่มเท่ากับหรือน้อยกว่าค่าที่สัมพันธ์กัน
5. ตัดสินใจจากค่า p-value ที่ได้ จะใช้ในการตัดสินใจว่าตัวแปรต้นแต่ละตัวมีความสำคัญทางสถิติหรือไม่ โดยทั่วไปถ้าค่า p-value น้อยกว่าค่าที่กำหนด (เช่น 0.05) จะถือว่ามีความสำคัญทางสถิติและตัวแปรนั้นจะถูกเลือกไปใช้ในแบบจำลอง ในทางตรงกลับกันถ้าค่า p-value มากกว่าหรือเท่ากับค่าที่กำหนด จะถือว่าไม่มีความสำคัญทางสถิติและตัวแปรนั้นจะถูกตัดออก
6. การประเมินแบบจำลองหลังจากเลือกคุณลักษณะที่มีความสำคัญออกมาแล้ว จะนำคุณลักษณะเหล่านั้นมาใช้ในการฝึกและประเมินแบบจำลอง และทำการปรับแก้แบบจำลองต่อไปเพื่อให้ได้ผลลัพธ์ที่ดีที่สุด

การใช้เทคนิคการเลือกคุณลักษณะสำคัญจากค่า p-value ของ Chi-square มีความสำคัญในการลดขนาดของชุดข้อมูล และในการคัดเลือกคุณลักษณะที่มีผลต่อผลลัพธ์อย่างมีนัยสำคัญ

2.3.2 การเข้ารหัสวันฮอต (One-hot encoding)

การทำ One-hot Encoding เป็นกระบวนการที่สำคัญในการแปลงข้อมูลที่มีลักษณะเป็นประเภทหนึ่งให้กลายเป็นเวกเตอร์ที่มีความยาวเท่ากับจำนวนของค่าที่เป็นไปได้ในข้อมูลนั้น ๆ โดยที่แต่ละค่าจะมีค่า 1 อยู่ในตำแหน่งของมันและค่า 0 อยู่ในตำแหน่งอื่น ๆ ในเวกเตอร์ กระบวนการนี้มักนำมาใช้งานอย่างแพร่หลายในการจัดการข้อมูลที่มีลักษณะเป็นประเภทหมวดหมู่ (categorical data) เช่น สี, ผลไม้, หรือประเภทของสินค้า เป็นต้นดังรูปที่ 2.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Color
Red
Orange
Yellow
Orange

Red	Orange	Yellow
1	0	0
0	1	0
0	0	1
0	1	0

รูปที่ 2.5 ตัวอย่างการทำเทคนิคการเข้ารหัสวันฮอต (One-hot encoding)

จากรูปที่ 2.5 สามารถอธิบายวิธีการทำ One Hot Encoding เป็นขั้นตอนได้ดังนี้

1. ระบุค่าที่เป็นไปได้ทั้งหมดในข้อมูลหมวดหมู่: ก่อนที่จะทำ One Hot Encoding ต้องระบุค่าที่เป็นไปได้ทั้งหมดในข้อมูลหมวดหมู่นั้น ๆ เพื่อให้สามารถสร้างเวกเตอร์ของ One Hot Encoding ได้ถูกต้อง
2. สร้างเวกเตอร์ One Hot Encoding: หลังจากที่ระบุค่าที่เป็นไปได้ทั้งหมดแล้ว จะสร้างเวกเตอร์ที่มีความยาวเท่ากับจำนวนของค่าที่เป็นไปได้ และกำหนดให้มีค่า 1 ที่ตำแหน่งของค่านั้น ๆ และค่า 0 ที่ตำแหน่งอื่น ๆ

เมื่อได้เวกเตอร์ One Hot Encoding เรียบร้อยแล้ว สามารถนำมันไปใช้ในการประมวลผลข้อมูลต่อไป เช่น การฝึกแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่อง (machine learning) หรือ การเรียนรู้เชิงลึก (deep learning) โดยที่เวกเตอร์ One Hot Encoding จะถูกนำมาเป็นข้อมูลนำเข้า (input) ของแบบจำลองสำหรับการจำแนกประเภทต่าง ๆ

2.4 ระบบแนะนำ (Recommendation System)

ระบบแนะนำ (Recommendation System) เป็นเครื่องมือหรือโปรแกรมที่ออกแบบขึ้นเพื่อช่วยแนะนำสิ่งต่าง ๆ ให้กับผู้ใช้ โดยใช้ข้อมูลเกี่ยวกับความชอบ ประวัติการใช้งาน หรือลักษณะทางสังคมของผู้ใช้ เพื่อสร้างประสบการณ์ที่ดีและความพึงพอใจในการค้นหาหรือเลือกสิ่งที่เหมาะสมกับความต้องการของผู้ใช้งาน ซึ่งมีหลายวิธีในการสร้างระบบแนะนำ สามารถแบ่งออกเป็น 3 ประเภทหลักได้แก่

2.4.1 ระบบแนะนำที่ใช้ข้อมูลเชิงเนื้อหา (Content-Based Recommendation Systems)

ระบบแนะนำที่ใช้ข้อมูลเชิงเนื้อหา (Content-Based Recommendation Systems) คือ ระบบที่ใช้ข้อมูลเกี่ยวกับลักษณะหรือคุณลักษณะของสิ่งที่จะแนะนำ เช่น คำอธิบาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลักษณะเฉพาะ หรือคุณลักษณะอื่น ๆ เพื่อทำนายสิ่งที่ผู้ใช้จะชอบหรือสนใจ โดยไม่จำเป็นต้องใช้ข้อมูลเกี่ยวกับพฤติกรรมการใช้งานของผู้ใช้งาน โดยจะแนะนำสิ่งของที่สอดคล้องกับความชอบหรือคุณลักษณะที่ผู้ใช้ระบุไว้ เช่น แนะนำหนังสือที่มีเนื้อหาคล้ายกับหนังสือที่ผู้ใช้เคยอ่านและชื่นชอบมาก่อนหน้านี้ หรือแนะนำเพลงที่มีลักษณะเสียงหรือแนวเพลงคล้ายกับเพลงที่ผู้ใช้ชื่นชอบเป็นต้น

ในการสร้างระบบแนะนำเชิงเนื้อหา จะใช้เทคนิคการวิเคราะห์ข้อมูลเชิงเนื้อหา เช่น การใช้เทคนิคการคัดเลือกคำหรือคำสำคัญจากเนื้อหา เพื่อจัดหมวดหมู่หรือคุณลักษณะของสิ่งที่จะแนะนำ และใช้เทคนิคการค้นหาหรือการวิเคราะห์เชิงสถิติเพื่อหาสิ่งที่มีความคล้ายคลึงกันในมุมมองของเนื้อหา ทำให้สามารถสร้างรายการแนะนำที่มีความเหมาะสมและถูกต้องตามความชอบหรือคุณลักษณะที่ผู้ใช้ระบุไว้ได้อย่างมีประสิทธิภาพและแม่นยำ ตัวอย่างของแอปพลิเคชันที่ใช้ Content-Based Recommendation Systems เช่น

1. Netflix (และแอปพลิเคชันอื่น ๆ ที่ให้บริการวิดีโอสตรีมมิง): Netflix ใช้ Content-Based Recommendation Systems เพื่อแนะนำภาพยนตร์และรายการทีวีที่เป็นไปได้ที่ผู้ใช้จะชอบ โดยการตรวจสอบลักษณะของภาพยนตร์และรายการทีวีที่ผู้ใช้ชอบเคยดูแล้ว และค้นหาเนื้อหาที่มีลักษณะคล้ายกันเพื่อแนะนำให้กับผู้ใช้
2. Spotify (และแอปพลิเคชันอื่น ๆ ที่ให้บริการสตรีมเพลง): Spotify ใช้ Content-Based Recommendation Systems เพื่อแนะนำเพลงและเพลย์ลิสต์ที่เป็นไปได้ที่ผู้ใช้จะชอบ โดยการวิเคราะห์ลักษณะของเพลงที่ผู้ใช้ชื่นชอบ เช่น แนวเพลง, ศิลปิน, หรือเนื้อเพลง เพื่อค้นหาเพลงที่มีลักษณะคล้ายกันเพื่อแนะนำให้กับผู้ใช้

2.4.2 ระบบแนะนำการกรองร่วมกัน (Collaborative Filtering Recommendation Systems)

ระบบแนะนำประเภทนี้ใช้ข้อมูลเกี่ยวกับพฤติกรรมการใช้งานหรือข้อมูลของผู้ใช้หลาย ๆ คนในการทำการแนะนำ โดยการหาความคล้ายคลึงในพฤติกรรมหรือความชอบของผู้ใช้ หากมีผู้ใช้ที่มีลักษณะการใช้งานคล้ายกัน ระบบจะแนะนำสิ่งที่ผู้ใช้คนนั้นชอบให้กับผู้ใช้คนอื่นที่มีลักษณะคล้ายกัน เทคนิคหลักในการสร้างระบบแนะนำแบบ Collaborative Filtering ได้แก่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) Memory-Based Collaborative Filtering

Memory-Based ใช้ข้อมูลเก่าเพื่อคำนวณความคล้ายคลึงระหว่างผู้ใช้หรือระหว่างสิ่งของ เพื่อทำนายการชอบหรือความพอใจของผู้ใช้งานใหม่ ๆ โดยใช้ความคล้ายคลึงในพฤติกรรมการใช้งานระหว่างผู้ใช้ หรือคะแนนที่ให้กับสิ่งของ ทำให้สามารถแนะนำสิ่งที่เป็นไปได้ที่ผู้ใช้จะชอบ โดยพิจารณาจากการกระทำของผู้ใช้งานท่านอื่น ๆ วิธีการสำคัญที่ใช้ใน Memory-Based ได้แก่ Nearest Neighbor Approaches เช่น User-Based และ Item-Based ซึ่งใช้เทคนิคการคำนวณความคล้ายคลึง เช่น Cosine Similarity, Pearson Correlation, หรือ Jaccard Coefficient เพื่อหาผู้ใช้หรือสิ่งของที่คล้ายกันมากที่สุดเพื่อทำนายความชอบหรือความพอใจ

2) Model-Based Collaborative Filtering

Model-Based ใช้ข้อมูลการใช้งานเพื่อสร้างแบบจำลองทางสถิติหรือแบบจำลองเชิงลึก เพื่อทำนายความชอบหรือความพอใจของผู้ใช้งานใหม่ ๆ โดยใช้การเรียนรู้จากข้อมูลการใช้งานในอดีต และเมื่อมีการใช้งานใหม่ๆ เกิดขึ้น ระบบสามารถทำนายผลลัพธ์ที่เป็นไปได้ที่มีความแม่นยำมากขึ้น วิธีการสำคัญใน Model-Based ได้แก่ Matrix Factorization, Singular Value Decomposition (SVD), Principal Component Analysis (PCA), และ Deep Learning ซึ่งสามารถใช้โครงสร้างข้อมูลขนาดใหญ่และความซับซ้อนได้ดี เพื่อสร้างแบบจำลองที่มีประสิทธิภาพในการทำนายความชอบของผู้ใช้งานใหม่ ๆ

มีหลายตัวอย่างของแอปพลิเคชันที่ใช้ระบบแนะนำแบบ Collaborative Filtering ในการแนะนำสิ่งต่าง ๆ ให้กับผู้ใช้ ตัวอย่างที่มีความนิยมและใช้เทคโนโลยี Collaborative Filtering เช่น

1. Amazon ใช้ระบบแนะนำเพื่อแนะนำผลิตภัณฑ์ที่เกี่ยวข้องหรือเสมือนกันให้กับผู้ใช้ เช่น หนังสือ, เครื่องใช้ไฟฟ้า, เสื้อผ้า, และสินค้าอื่น ๆ โดยอาศัยการซื้อสินค้าก่อนหน้าของผู้ใช้ และพฤติกรรมการค้นหาเพื่อแนะนำสินค้าที่เป็นไปได้ที่จะมีความสนใจ
2. LinkedIn ใช้ระบบแนะนำเพื่อแนะนำบุคคลที่เกี่ยวข้องหรือมีความสนใจในการเชื่อมโยง เช่น บุคคลในวงการอาชีพ, บุคคลในบริษัทเดียวกัน, หรือบุคคลที่มีความสนใจและพฤติกรรมการใช้งานที่คล้ายคลึงกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.3 ระบบแนะนำผสมผสาน (Hybrid Recommendation System)

ระบบแนะนำแบบ hybrid คือ ระบบที่ใช้การผสมผสานหรือการรวมกันระหว่างวิธีการทำนายและแนะนำที่มีลักษณะแตกต่างกัน เช่น Collaborative Filtering และ Content-Based Filtering เพื่อสร้างระบบแนะนำที่มีประสิทธิภาพและความแม่นยำมากขึ้น เนื่องจากแต่ละวิธีการนั้นมีข้อดีและข้อเสียที่แตกต่างกัน เมื่อผสมกันอย่างเหมาะสมสามารถลดความเสี่ยงหรือข้อจำกัดของแต่ละวิธีได้ เช่น

1. ความแม่นยำที่มีประสิทธิภาพ โดยการผสมผสานข้อมูลจากทั้ง Collaborative Filtering และ Content-Based Filtering สามารถทำให้ระบบแนะนำมีความแม่นยำมากขึ้น เนื่องจากสามารถดึงประโยชน์จากข้อได้เปรียบของแต่ละวิธีและเพิ่มประสิทธิภาพในการทำนายและการแนะนำ
2. ความเหมาะสมกับความต้องการของผู้ใช้ การผสมผสานระหว่างข้อมูลเชิงเนื้อหาและข้อมูลการกระทำของผู้ใช้สามารถช่วยให้ระบบแนะนำเสนอสิ่งที่เหมาะสมและน่าสนใจตามความชอบและความต้องการของผู้ใช้
3. การรับมือกับปัญหาความขัดแย้ง (Cold Start) ระบบแนะนำแบบ hybrid สามารถรับมือกับปัญหาความขัดแย้งในระบบแนะนำที่ไม่มีข้อมูลการใช้งานเพียงพอหรือไม่มีข้อมูลเชิงเนื้อหาในระยะเริ่มต้น โดยสามารถใช้ Content-Based Filtering ในการทำนายในระยะเริ่มต้นและเพิ่มข้อมูลการใช้งานเข้าไปในภายหลังเพื่อปรับปรุงความแม่นยำของระบบแนะนำในระยะยาว

ระบบแนะนำแบบ hybrid มักถูกนำมาใช้ในหลายด้าน เช่น การค้าออนไลน์, การสตรีมมิ่ง, การบริการสื่อสาร, และสถานการณ์การใช้งานอื่น ๆ ที่ต้องการระบบแนะนำที่มีประสิทธิภาพและตอบสนองต่อความต้องการของผู้ใช้งานในระดับสูง

2.5 Leave-One-Out Cross-Validation

ทฤษฎีการใช้ leave-one-out cross-validation เป็นวิธีการในการประเมินประสิทธิภาพของแบบจำลองทางสถิติหรือแบบจำลองที่สร้างขึ้น โดยการใช้ข้อมูลที่มีอยู่ให้เป็นการทดสอบแต่ละรายการหนึ่งในแต่ละครั้ง โดยใช้ข้อมูลที่เหลือในการฝึกแบบจำลอง กระบวนการนี้ทำซ้ำกับข้อมูลทั้งหมดทุกครั้งโดยใช้แต่ละรายการในชุดข้อมูลเป็นตัวทดสอบอย่างเดียว จากนั้นเก็บค่าประสิทธิภาพของแบบจำลองที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ได้จากการทดลองทุกครั้ง โดยมักจะนำค่าเฉลี่ยหรือค่าเฉลี่ยของค่าประสิทธิภาพเหล่านี้เพื่อประเมินประสิทธิภาพของแบบจำลองโดยรวม

test	train	train	train	train	train	train
train	test	train	train	train	train	train
train	train	test	train	train	train	train
.						
.						
.						
train	train	train	train	train	train	test

รูปที่ 2.6 ตัวอย่างของขั้นตอนการทำ Leave-One-Out Cross Validation

การใช้ leave-one-out cross-validation ช่วยในการประเมินประสิทธิภาพของแบบจำลองอย่างเหมาะสมโดยไม่ต้องแบ่งข้อมูลออกเป็นชุดฝึกและชุดทดสอบ เพราะทุกครั้งจะใช้ข้อมูลทั้งหมดที่มีอยู่ในการทดสอบแบบจำลองอย่างแน่นอน ซึ่งสามารถช่วยลดปัญหาการแบ่งข้อมูลที่อาจเป็นไปได้หรือปัญหาที่เกิดขึ้นจากการแบ่งข้อมูลออกเป็นชุดฝึกและชุดทดสอบ

2.6 การวัดประสิทธิภาพ (Measure Evaluation)

2.6.1 ตัววัดอัตราคำแนะนำที่ถูกต้อง (Hit Rate)

การวัดผลด้วย Hit rate เป็นหนึ่งในวิธีการประเมินประสิทธิภาพของระบบแนะนำ โดยพิจารณาจากความสำเร็จในการแนะนำสิ่งของที่ผู้ใช้งานจริง ๆ ต้องการหรือสนใจ ซึ่งมีการใช้งานจริงในบางรูปแบบเป็นหลัก

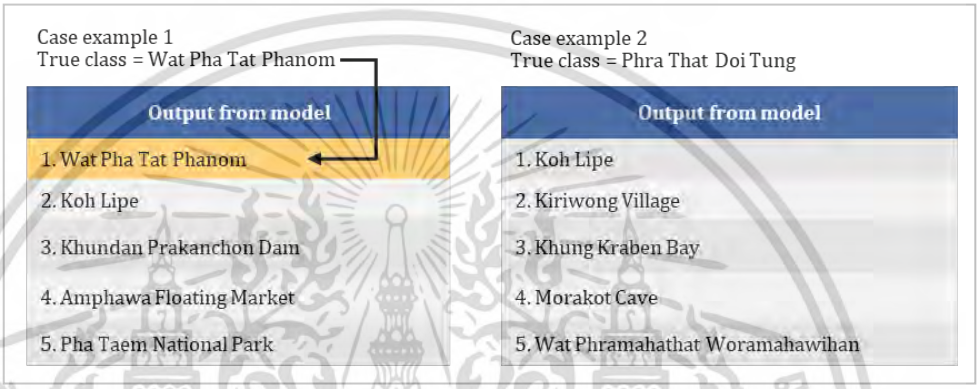
Hit rate คำนวณจากจำนวนของการแนะนำที่ถูกต้อง (hits) หรือสิ่งของที่ผู้ใช้งานจริง ๆ เลือกหรือดำเนินการต่อ มาตรฐานสำหรับการวัดนี้คืออัตราส่วนของการแนะนำที่ถูกต้องต่อจำนวนรวมของการแนะนำทั้งหมด เป็นสัดส่วนหรือเปอร์เซ็นต์ที่บ่งชี้ถึงประสิทธิภาพของระบบแนะนำในการทำนายและแนะนำสิ่งของให้กับผู้ใช้งานได้ถูกต้องและเหมาะสม สำหรับการคำนวณ Hit rate สามารถทำได้โดยใช้สมการที่ 2.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$Hit\ Rate = \frac{\sum_{i=1}^{N} HR@K}{N} \quad (2.1)$$

โดยที่ k คือจำนวนคำแนะนำที่ถูกแนะนำโดยระบบแนะนำให้แก่ผู้ใช้

N คือจำนวนของการแนะนำทั้งหมด



รูปที่ 2.7 ตัวอย่างผลลัพธ์ที่ได้จากแบบจำลอง

จากรูปที่ 2.7 แสดงผลลัพธ์สำหรับผู้ใช้สองคนที่แตกต่างกัน ในตัวอย่างที่ 1 มีการแนะนำจากระบบแนะนำทั้งหมด 5 สถานที่ ในขณะคลาสที่เป็นคำตอบแท้จริงสำหรับผู้ใช้นี้คือวัดพระธาตุพนม ดังนั้นค่า $HR@5$ คือ 1 เนื่องจากวัดพระธาตุพนมปรากฏขึ้นในรายการที่แนะนำ ส่วนในตัวอย่างที่ 2 คลาสที่ถูกต้องคือพระธาตุดอยตุง แต่สถานที่ดังกล่าวไม่อยู่ในรายการที่ระบบแนะนำ ดังนั้น $HR@5$ คือ 0 โดยคำนวณผลรวมทั้งสองตัวอย่างจะได้ $Hit\ Rate$ คือ (1+0)/2 = 0.5

โดยข้อดีของ Hit rate มีดังนี้

1. ง่ายต่อการเข้าใจและใช้งาน: Hit rate เป็นวิธีการวัดผลที่ง่ายต่อการเข้าใจและใช้งาน ซึ่งสามารถอธิบายได้ง่ายๆ ว่ามีกี่เหตุการณ์ที่ระบบแนะนำถูกต้องต่อจำนวนรวมของการแนะนำทั้งหมด
2. จำนวนผลลัพธ์ที่สอดคล้อง: Hit rate ช่วยให้เข้าใจได้ถึงสัดส่วนของผลลัพธ์ที่ถูกต้องหรือสอดคล้องกับความต้องการของผู้ใช้งาน ซึ่งเป็นข้อมูลที่สำคัญในการปรับปรุงคุณภาพของระบบแนะนำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. เหมาะสำหรับการทำนายและการแนะนำที่มีจำกัด: ในบางกรณี เช่น ในระบบการแนะนำสินค้าหรือเพลง ผู้ใช้งานอาจมีความสนใจในเพียงเล็กน้อยจากจำนวนที่แนะนำ ในกรณีนี้ Hit rate จะช่วยให้เข้าใจได้ถึงประสิทธิภาพของระบบแนะนำต่อผู้ใช้งานได้ดี
4. การใช้งานในระบบที่มีข้อมูลจำกัด: สำหรับระบบที่มีข้อมูลจำกัดหรือยากต่อการเข้าถึง การใช้ Hit rate เป็นวิธีการวัดผลที่เหมาะสม เนื่องจากสามารถจำกัดจำนวนข้อมูลที่ต้องใช้ในการวัดผลได้ในขณะที่ยังคงให้ข้อมูลที่มีประสิทธิภาพ

2.6.2 ตัววัดอัตราการจัดลำดับคำแนะนำที่ถูกต้อง (NDCG)

NDCG (Normalized Discounted Cumulative Gain) เป็นวิธีการวัดความสมบูรณ์ของการจัดลำดับผลลัพธ์จากระบบค้นหาหรือระบบแนะนำ โดยคำนวณค่าความสมบูรณ์ของลำดับผลลัพธ์ตามความสนใจของผู้ใช้ ค่า NDCG มีค่าระหว่าง 0 ถึง 1 โดยค่าที่สูงขึ้นแสดงถึงความสมบูรณ์ของลำดับผลลัพธ์ที่ดีขึ้น การคำนวณ NDCG จะพิจารณาการจัดลำดับของผลลัพธ์ โดยคำนวณค่า Discounted Gain (ความสนใจที่ลดลงเมื่อได้รับผลลัพธ์ในตำแหน่งที่สูงขึ้น) และทำการปรับมาตราส่วนให้ได้ค่า NDCG ที่มีความสมบูรณ์ โดยสามารถคำนวณได้ตามสมการที่ 2.2 – 2.4

$$NDCG = \frac{DCG}{IDCG} \tag{2.2}$$

$$DCG = \sum_{i=1}^{N} \frac{\text{relevance}_i}{log_2(i+1)} \tag{2.3}$$

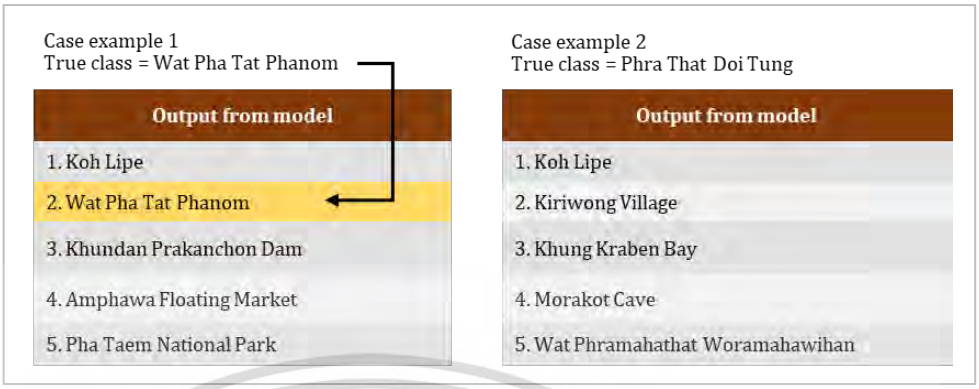
$$IDCG = \sum_{i=1}^{N} \frac{2^{\text{relevance}_i}-1}{log_2(i+1)} \tag{2.4}$$

โดยที่ relevance_i คือจำนวนคำแนะนำที่แนะนำโดยระบบแนะนำให้แก่ผู้ใช้

N คือจำนวนของการแนะนำทั้งหมด

i คือจำนวนดัชนีของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.8 ตัวอย่างผลลัพธ์ที่ได้จากแบบจำลอง

จากรูปที่ 2.8 แสดงผลลัพธ์สำหรับผู้ใช้สองคนที่แตกต่างกัน ในตัวอย่างที่ 1 มีการแนะนำจากระบบแนะนำทั้งหมด 5 สถานที่ ดังนั้น $\text{relevance}_i = 5$ ในขณะคลาสที่เป็นคำตอบแท้จริงสำหรับผู้ใช้นี้คือวัดพระธาตุพนม และถูกแนะนำอยู่ในอันดับที่ 2 ดังนั้นจะได้ค่า DCG = 1.26 และค่า $IDCG$ = 1.89 นำค่ามาหา $NDCG$ = 1.26/1.89 = 0.67 ในส่วนของตัวอย่างที่ 2 ระบบไม่สามารถแนะนำสถานที่ที่ถูกต้องใน 5 รายการได้ ดังนั้นค่า $NDCG$ จึงมีค่าเป็น 0

การใช้งาน NDCG มักจะเหมาะสำหรับระบบค้นหาหรือระบบแนะนำที่ต้องการลำดับผลลัพธ์ที่ดี โดยคำนึงถึงความสนใจของผู้ใช้ ซึ่งมักใช้ในการวัดและปรับปรุงคุณภาพของระบบแนะนำในสถานการณ์ต่าง ๆ อย่างกว้างขวาง อีกทั้งยังเป็นวิธีการที่แก่ผู้พัฒนาระบบและนักวิจัยในการทดสอบและปรับปรุงความสามารถในการแนะนำของระบบได้อย่างมีประสิทธิภาพ โดยที่ NDCG มีข้อดีหลายประการที่ทำให้เป็นเครื่องมือที่มีความสำคัญในการประเมินและปรับปรุงคุณภาพของระบบค้นหาหรือระบบแนะนำได้ดังนี้

1. ความเป็นธรรมและความเหมาะสม NDCG มีการคำนวณที่คำนึงถึงความสนใจของผู้ใช้ โดยการให้น้ำหนักสูงกับผลลัพธ์ที่มีความสนใจสูง ซึ่งทำให้การประเมินนั้นเป็นธรรมและเหมาะสมกับความต้องการของผู้ใช้
2. การใช้งานในระบบที่มีลำดับผลลัพธ์ NDCG เหมาะสำหรับระบบที่มีความสำคัญในการจัดลำดับผลลัพธ์ เช่น ระบบค้นหาหรือระบบแนะนำ โดยช่วยให้ผู้พัฒนาระบบสามารถปรับปรุงการจัดลำดับให้มีความสมบูรณ์และเหมาะสมกับผู้ใช้งานได้อย่างมีประสิทธิภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ความสามารถในการจำแนกผลลัพธ์ที่สนใจ NDCG ช่วยให้ผู้พัฒนาระบบสามารถจำแนกผลลัพธ์ที่มีความสนใจและมีความสำคัญสูงจากผลลัพธ์ที่ไม่สนใจหรือมีความสำคัญต่ำ เนื่องจากน้ำหนักที่ใช้ในการคำนวณนั้นสามารถปรับได้ตามความสนใจของผู้ใช้
4. ความยืดหยุ่นและสามารถปรับปรุงได้ NDCG สามารถใช้ได้กับหลายรูปแบบของการจัดลำดับและการคำนวณ และสามารถปรับแต่งการคำนวณได้ตามความต้องการของระบบและการใช้งาน
5. ความทันสมัยและความน่าเชื่อถือ เนื่องจาก NDCG เป็นวิธีการวัดผลที่ได้รับความนิยมในวงกว้างในวงการวิจัยและการพัฒนาระบบ การใช้ NDCG ในการประเมินและปรับปรุงคุณภาพของระบบค้นหาหรือระบบแนะนำจึงเป็นที่ยอมรับและน่าเชื่อถือ

2.7 งานวิจัยที่เกี่ยวข้อง

ในระบบแนะนำ (Recommendation System) มี 3 ประเภทหลักๆ ดังที่ผู้วิจัยกล่าวข้างต้น

2.7.1 ระบบแนะนำที่ใช้ข้อมูลเชิงเนื้อหา (Content-Based Recommendation Systems)

เทคนิคนี้มีผู้วิจัยและได้ประสิทธิผลที่ดีเป็นจำนวนมาก โดยมีงานวิจัยของ Carla Binucci และคณะ (2017) เรื่อง Designing the Content Analyzer of a Travel Recommender System ได้ศึกษาเกี่ยวกับระบบแนะนำการเดินทางตามเนื้อหา (Content-Based) โดยจะแนะนำสถานที่ท่องเที่ยวโดยพิจารณาจากความเหมาะสมระหว่างความชอบของผู้ใช้และกลุ่มจุดสนใจที่ตรงกันมากที่สุด (Points of Interests) ซึ่งเรียกสั้นๆ ว่า POI และต้องกำหนดความสัมพันธ์ของ POI ให้เกี่ยวข้องกับหัวข้อที่น่าสนใจ (Topics of Interests) หรือเรียกสั้นๆ ว่า TOI เรื่อง Using POI functionality and accessibility levels for delivering personalized tourism recommendations ของ Filipe Santos และคณะ (2019) งานวิจัยนี้มุ่งเน้นศึกษาเพื่อเสนอระบบคำแนะนำที่พิจารณาระดับการทำงานของผู้ใช้เกี่ยวกับข้อจำกัดทางร่างกายหรือจิตใจ งานวิจัยนี้อธิบายชุดของแบบจำลองและอัลกอริทึมที่ใช้ภายใต้ระบบแนะนำการท่องเที่ยวตามผู้ใช้และโปรไฟล์จุดที่น่าสนใจ (POI)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.2 ระบบแนะนำการกรองร่วมกัน (Collaborative Filtering Recommendation Systems)

เทคนิคระบบแนะนำการกรองร่วมกัน ถูกนำไปใช้อย่างแพร่หลาย และหลากหลาย โดยขึ้นอยู่กับข้อมูลที่มี และวัตถุประสงค์ที่นำไปใช้ โดยในส่วนของเทคนิคนี้สามารถแบ่งออกเป็น 2 ประเภท คือ

1) Memory-Based Collaborative Filtering

โดยงานวิจัยเรื่อง Traveler preferences from online reviews: Role of travel goals, class and culture ของ Swagato Chatterjee และ Prasenjit Mandal (2020) ในการศึกษานี้ ได้สำรวจความชอบ พฤติกรรม และการประเมินหลังการซื้อแตกต่างกันของนักท่องเที่ยว โดยขึ้นอยู่กับเป้าหมายการเดินทาง ระดับการเดินทาง และวัฒนธรรมของประเทศบ้านเกิดของนักท่องเที่ยว ซึ่งผลลัพธ์ที่ได้เป็นการวัดประสิทธิภาพของสายการบินและมีการแนะนำของสายการบิน รวมถึงช่วยให้สายบินมีการจัดการได้ดียิ่งขึ้น นอกเหนือจากนี้ยังมีงานวิจัยโดยใช้ข้อมูลจาก Facebook เพื่อวิเคราะห์หาพฤติกรรมความชื่นชอบของการท่องเที่ยว โดย K. Kesorn และคณะ (2017) ได้ศึกษาวิจัยโดยใช้ข้อมูลจาก Facebook ได้แก่ สถานที่เช็คอิน, จุดหมายปลายทาง, ฤดูกาลท่องเที่ยว, และการจัดกลุ่มเพื่อนสนิทบน Facebook โดยวิเคราะห์จากความถี่ในการกดถูกใจ และการแสดงความคิดเห็นต่อกันและกัน เพื่อค้นหาน้ำหนักของจุดหมายปลายทาง และแนะนำจุดหมายปลายทางที่เหมาะสมให้แก่ผู้ใช้ Merlinda Sumardi และคณะ (2017) ได้พัฒนาแอปพลิเคชันชื่อ TripBuddy โดยใช้เทคนิค K-mean เพื่อหาจัดกลุ่มพฤติกรรมการของนักทองเที่ยว โดยแอปพลิเคชันนี้เป็นการแนะนำจุดหมายปลายทางที่เหมาะสม พร้อมเวลาในการเดินรถ และค่าใช้จ่าย เพื่อง่ายต่อการวางแผนการเดินทางของนักท่องเที่ยว Zahra Abbasi-Moud และคณะ (2021) ได้ศึกษาเรื่อง Tourism recommendation system based on semantic clustering and sentiment analysis งานวิจัยนี้ได้รวบรวมความคิดเห็นจากแอปพลิเคชัน Trip Advisor และนำมาจัดกลุ่มเพื่อหากลุ่ม และวิเคราะห์หาความชื่นชอบของนักท่องเที่ยว สุดท้ายระบบจะแนะนำสถานที่ท่องเที่ยวให้เหมาะ โดยเปรียบเทียบจากการจัดกลุ่มให้ใกล้เคียงที่สุด งานวิจัยของ Lin Lia และคณะ (2019) ได้รวบรวมสถานที่ออนไลน์เพื่อการวิจัยที่มีจุดมุ่งหมายเพื่อกำหนดและยืนยันองค์ประกอบการเรียนรู้ที่มีอิทธิพลต่อความนิยมของคำแนะนำ "สถานที่ที่น่าเยี่ยมชมที่สุด (BP2V)" โดยอิงตามแบบจำลองการศึกษาพฤติกรรมเชิงระบบ (HSM) นอกจากนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

งานวิจัยนี้ยังศึกษาผลกระทบของประเภทจุดหมายปลายทาง (เช่น สถานที่ท่องเที่ยว ร้านอาหาร แหล่งช้อปปิ้ง และกิจกรรม) เกี่ยวกับคุณสมบัติ งานนี้ใช้สถานที่ 565 แห่งในสหรัฐอเมริกาจากผู้ใช้ที่แนะนำบน Qyer.com ซึ่งเป็นชุมชนท่องเที่ยวออนไลน์ที่สำคัญของจีน งานวิจัยของ Swapnil Shende และคณะ (2023) ศึกษาความเคลื่อนไหวของผู้คนในเมืองเดลี เปรียบเทียบช่วงก่อนโควิดกับช่วงระบาดของโควิด การใช้ข้อมูลจาก Twitter เทคนิคการเรียนรู้ของเครื่องสามอย่าง ได้แก่ Support Vector Machine (SVM), K-Nearest Neighbors (kNN) และ Decision Tree เพื่อแยกสถานที่ตามที่เช็คอินไว้ก่อนหน้านี้ และจัดกลุ่มโดยใช้ K-Means, DBSCAN และ Mean-Shift

2) Model-Based Collaborative Filtering

งานวิจัยของ Guixiang Zhu และคณะ (2020) มีการเสนอวิธีการแนะนำแพ็คเกจการเดินทางแบบเน้นความสนใจทางระบบประสาท (Neural Attentive Travel Package Recommendation: NATR) ระบบใช้การผสมผสานระหว่างวิธี memory-base และ model-base เพื่อสร้างแพ็คเกจการเดินทางตามพฤติกรรมของนักเดินทาง งานวิจัยเรื่อง An extensive study on the evolution of context-aware personalized travel recommender systems ของ Shini Renjith และคณะ (2020) นักวิจัยหลายคนในขอบเขตของระบบการแนะนำการเดินทางและระบบการรับรู้บริบทตั้งแต่ปี 1997 ถึง 2018 ได้รับการศึกษาและอธิบายอัลกอริทึมการจำแนกประเภทและตัวชี้วัดที่ใช้ในการประเมินประสิทธิภาพของระบบการแนะนำ งานวิจัยของ Subramaniyaswamy V และคณะ (2015) ได้พัฒนาระบบแนะนำการเดินทางอัจฉริยะโดยอาศัยการแชร์รูปภาพบนแพลตฟอร์มโซเชียลมีเดีย ระบบค้นหาจุดหมายปลายทางที่ผู้ใช้ชอบเดินทางจากภาพถ่ายที่ผู้ใช้ส่งมาของจุดหมายปลายทางนั้น นอกจากนี้ การศึกษานี้ยังได้อธิบายวิธีการจำแนกประเภทและรูปแบบการเรียนรู้แบบเบย์ (Bayesian learning) ตามความต้องการของผู้ใช้อีกด้วย Toly Chen และ Yu-Cheng Wang ได้ศึกษาเรื่อง A calibrated piecewise-linear FGM approach for travel destination recommendation during the COVID-19 pandemic (2021) โดยนักวิจัยมีการเสนออัลกอริทึมค่าเฉลี่ยเรขาคณิตคลุมเครือแบบแยกส่วนเชิงเส้น (Fuzzy Geometric Mean: FGM) ที่ปรับเทียบแล้วเพื่อแนะนำสถานที่ท่องเที่ยวที่เหมาะสมในช่วงสถานการณ์โควิด-19 การระบาดใหญ่ของการเดินทางในภูมิภาคในไต้หวัน ในส่วนแรกของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

งานนี้ ได้มีการศึกษาปัจจัยที่มีอิทธิพลต่อการเลือกจุดหมายปลายทางในช่วงสถานการณ์โควิด-19 ส่วนที่สองคือการปรับปรุงประสิทธิภาพของกระบวนการลำดับชั้นการวิเคราะห์แบบคลุมเครือที่มีอยู่ (FAHP) และยังมีงานวิจัย Hsiu-Sen Chiang และ Tien-Chi Huang (2013) ได้พัฒนาระบบการวางแผนการเดินทางส่วนบุคคลถูกผลิตขึ้นเพื่อตอบสนองความต้องการของผู้ใช้ทุกประเภท และให้บริการวางแผนตารางการเดินทางอัตโนมัติแก่ผู้ใช้ ผู้ใช้สามารถให้คะแนนแก่การเดินทางที่ไม่น่าพึงพอใจได้ ซึ่งระบบจะตอบรับ และช่วยให้ผู้ใช้รายใหม่มีอัตราความแม่นยำที่สูงขึ้นสำหรับแผนการเดินทางครั้งต่อไป Balamurugan S. และ Sankaran M. (2023) ศึกษาการใช้อัลกอริทึมการเรียนรู้ของเครื่องแบบต้นไม้ เช่น Decision Tree, Random Forest, Adaboost และ Xgboost เพื่อทำนายความเหนื่อยล้าของคนขับรถบรรทุกทางไกลในอินเดีย เพื่อช่วยให้กรมทางหลวงเตรียมความพร้อมด้านความปลอดภัยที่เพิ่มขึ้น ในงานวิจัยเรื่อง A Machine Learning-Based Tourist Path Prediction (2016) ได้สร้างระบบแนะนำสถานที่ท่องเที่ยวโดยข้อมูลเกี่ยวกับระยะเวลาและระยะเส้นทางของนักท่องเที่ยว โดยใช้อัลกอริทึม ได้แก่ Random Forests, LambdaMART, Ranking SVM, RankBoost และ ListNet ผลการทดลองคืออัลกอริทึม Random Forests ให้ประสิทธิภาพสูงสุด

2.7.3 ระบบแนะนำผสมผสาน (Hybrid Recommendation System)

โดยงานวิจัยเรื่อง A hybrid context-aware approach for e-tourism package recommendation based on asymmetric similarity measurement and sequential pattern mining ของ Maral Kolahkaj และคณะ (2022) ได้พัฒนาระบบการแนะนำเพื่อแก้ไขปัญหาข้อมูลล้นเกินและแนะนำเนื้อหาที่เกี่ยวข้องแก่ผู้ใช้ โดยพิจารณาข้อมูลเวลา จุดหมายปลายทาง การประเมินโดยนัยของผู้ใช้ ลักษณะการท่องเที่ยว และรูปแบบการเคลื่อนที่ของการเดินทางตามลำดับ โดยอิงจากภาพที่ติดแท็กภูมิศาสตร์ งานวิจัยของ Jiangning Hea และคณะ (2016) งานนี้นำเสนอแบบจำลองที่มุ่งเน้นศึกษาด้านความชอบส่วนบุคคล และเสนอคำแนะนำตามประวัติการเดินทางของแต่ละคน ผลกระทบทางสังคมของผู้ร่วมเดินทาง และความสัมพันธ์ทางสังคม เช่น เพื่อนร่วมงาน เพื่อนร่วมโรงเรียน ฯลฯ งานวิจัยของ Joel P. Lucas และคณะ (2013) ได้นำวิธีการแนะนำสำหรับการท่องเที่ยวมาใช้ ในงานวิจัยนี้ มีการใช้การจำแนกประเภทตามเทคนิคการเชื่อมโยง (Association Technique) เพื่อสร้างการทำนายจากแบบจำลอง งานวิจัยเรื่อง Post-pandemic shared mobility and active travel in Alabama: A machine learning analysis of COVID-19 survey data ของ Ningzhe Xu และคณะ (2023) ได้ศึกษาการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คาดการณ์การเปลี่ยนแปลงพฤติกรรมการเดินทางในอนาคต การศึกษาพฤติกรรมการเดินทางในรัฐอลาบามาได้ดำเนินการโดยใช้การสำรวจหน่วยตัวอย่าง 481 หน่วยและอัลกอริทึมที่หลากหลาย รวมถึง Random Forest, Adaptive Boosting, Support Vector Machine, K-Nearest Neighbours และ Artificial Neural Network หลังจากปัญหาโควิดแพร่ระบาดก็จะเกิดขึ้นและแจ้งการขนส่งถึงผลกระทบจากการเดินทาง

2.7.4 การวัดประสิทธิภาพด้วย Hit Rate และ NDCG

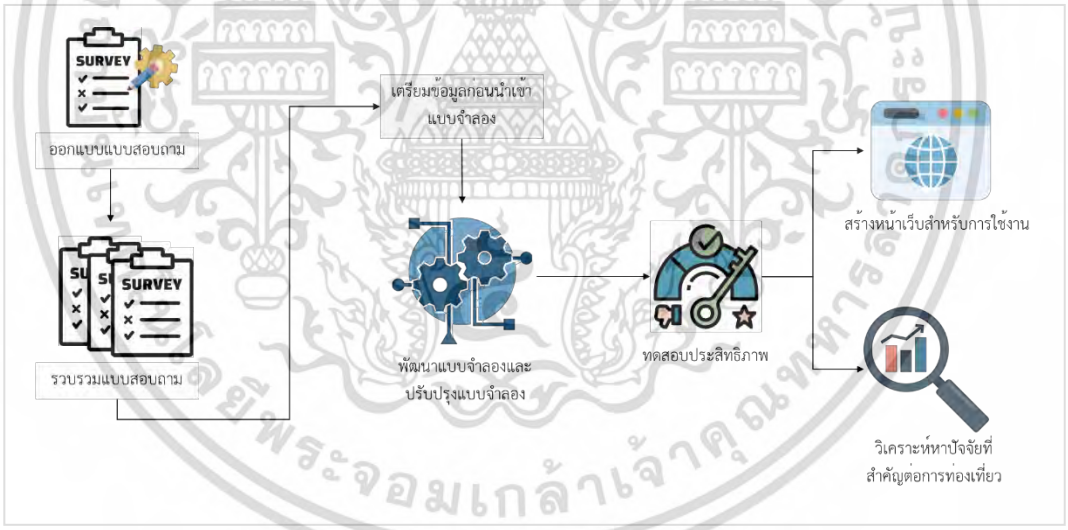
โดยงานวิจัยของ Yue Li และ Endong Wang ได้นำเสนอระบบแนะนำสำหรับวิดีโอเกมส์เพื่อการศึกษา ด้วยใช้วิธี knowledge graph และวัดประสิทธิภาพของระบบแนะนำด้วยค่า Hit Rate และ NDCG โดยที่ค่า k = {10, 20} ได้ผลคือ HR@10 = 0.7256, HR@20 = 0.7990, NDCG@20 = 0.646 และ NDCG@20 = 0.6046 นอกจากนี้ยังมีงานวิจัยเรื่อง Session-based recommendations with sequential context using attention-driven LSTM ของ Chhotelal Kumar และ Mukesh Kumar ได้นำเสนอระบบแนะนำที่ใช้เทคนิค LSTM-Attention ในการสร้างระบบแนะนำเพลงให้แก่ผู้ใช้ โดยวัดด้วยค่า HR@1, HR@5 และ HR@10 ได้ค่า 0.2689, 0.3675 และ 0.4297 ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

วิธีดำเนินงานวิจัย

ในเนื้อหาของบทนี้เป็นการอธิบายถึงขั้นตอนการศึกษาและพัฒนาระบบแนะนำสถานที่ท่องเที่ยวในเมืองรองของประเทศไทย โดยงานวิจัยนี้ได้ใช้วิธีการเรียนรู้ของเครื่อง ทั้งหมด 4 วิธี ประกอบด้วย 1. อัลกอริทึมต้นไม้ตัดสินใจ 2. การสุ่มแบบป่าไม้ 3. การหาเพื่อนบ้านใกล้เคียง (K-Nearest Neighbors: k-NN) และ 4. เพอร์เซ็ปตรอนหลายชั้น โดยขั้นตอนในการทำงานของงานวิจัยนี้มี 6 ขั้นตอน คือ 1. สร้างแบบสอบถามเกี่ยวกับพฤติกรรมและประสบการณ์การท่องเที่ยวในเมืองรอง 2. เก็บและรวบรวมข้อมูลที่ได้จากแบบสอบถาม 3. ออกแบบและพัฒนาแบบจำลองของระบบแนะนำและปรับปรุง 4. วัดประสิทธิภาพ 5. นำผลการทดลองที่ได้มาวิเคราะห์หาปัจจัยที่สำคัญและส่งผลต่อการกระตุ้นการท่องเที่ยว 6. นำไปใช้งานโดยการสร้างหน้าเว็บแอปพลิเคชันสำหรับการใช้ระบบแนะนำ ดังรูปที่ 3.1



รูปที่ 3.1 โครงสร้างการทำดำเนินงานวิจัยของระบบแนะนำสถานที่ท่องเที่ยวในเมืองรอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

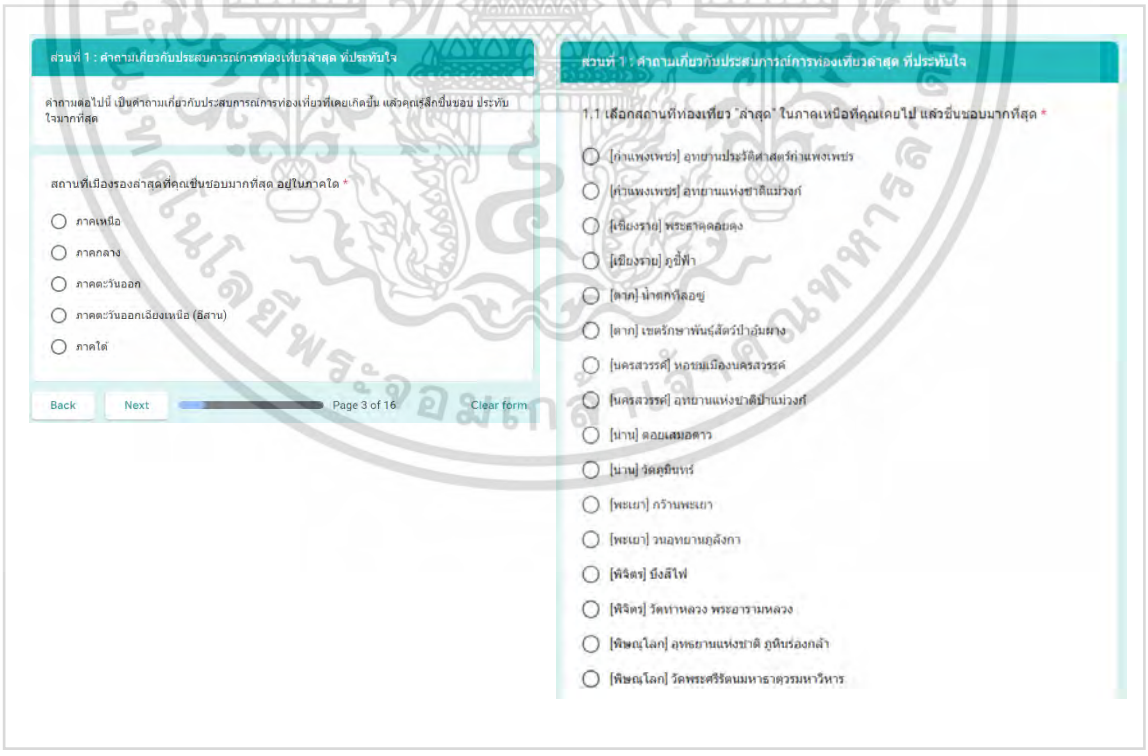
3.1 การเก็บรวบรวมชุดข้อมูล

โดยชุดข้อมูลที่ผู้วิจัยนำมาใช้เป็นการเก็บแบบสอบผ่าน Google Forms โดยเป็นคำถามเกี่ยวกับประสบการ์การท่องเที่ยวเมืองรองในประเทศไทยของนักท่องเที่ยวชาวไทย โดยเก็บรวบรวมข้อมูลตัวอย่างได้ 429 หน่วยตัวอย่าง กระจายอยู่ใน 5 ภาคของประเทศไทย แสดงดังตารางที่ 3.1

ตารางที่ 3.1 จำนวนข้อมูลหน่อยตัวอย่างในแต่ละภูมิภาคของประเทศไทย

ภูมิภาค	จำนวนหน่วยตัวอย่าง
ภาคเหนือ	157
ภาคใต้	85
ภาคตะวันออก	70
ภาคตะวันออกเฉียงเหนือ	70
ภาคกลาง	47

จากตารางที่ 3.1 เป็นการแสดงจำนวนหน่วยตัวอย่างที่นำมาวิจัยในครั้งนี้โดยแบ่งเป็น 5 ภูมิภาคตามภูมิศาสตร์ของประเทศไทย



รูปที่ 3.2 ข้อคำถามใน Google Forms

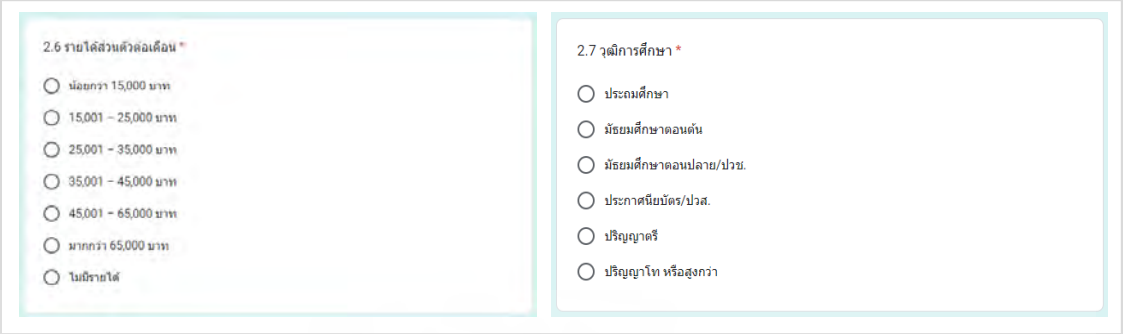
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 3.3 ข้อคำถามใน Google Forms

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 3.4 ข้อคำถามใน Google Forms

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.5 ข้อคำถามใน Google Forms

จากรูปที่ 3.2 – 3.5 แสดงถึงหน้าจอของแบบสอบถามที่ใช้ในการวิจัยครั้งนี้ โดยเก็บผ่าน Google Form โดยข้อคำถามที่นำมาใช้ในแบบสอบถามมีทั้งหมด 19 คำถาม โดยแบ่งเป็น 2 ส่วนหลักคือ ข้อคำถามเกี่ยวกับประสบการณ์ที่ประทับใจที่สุดในการท่องเที่ยวเมืองรองและข้อคำถามเกี่ยวกับข้อมูลส่วนบุคคลของนักท่องเที่ยว โดยมีรายละเอียดดังตารางที่ 3.2

ตารางที่ 3.2 รายละเอียดในแต่ละข้อคำถามที่ใช้ในแบบสอบถาม

ข้อคำถาม	คำอธิบายข้อคำถาม	จำนวนคำตอบ	ตัวเลือกในคำถาม
เดือนที่เดินทางท่องเที่ยว	เดือนที่ผู้ตอบแบบสอบถามเดินทางไปยังสถานที่ท่องเที่ยว	12	{เดือนมกราคม, เดือนกุมภาพันธ์, เดือนมีนาคม, เดือนเมษายน, เดือนพฤษภาคม, เดือนมิถุนายน, เดือนกรกฎาคม, เดือนสิงหาคม, เดือนกันยายน, เดือนตุลาคม, เดือนพฤศจิกายน, เดือนธันวาคม}
ลักษณะการเดินทาง	รูปแบบการเดินทางที่ผู้ตอบแบบสอบถามเลือกใช้ไปยังสถานที่ท่องเที่ยว	5	{รถส่วนตัว, รถประจำทาง, เครื่องบิน, รถไฟ, รถเช่า}
งบประมาณการเดินทาง	จำนวนงบประมาณที่ผู้ตอบแบบสอบถามใช้จ่ายเพื่อเดินทางมายังสถานที่ท่องเที่ยว	8	{≤1,000฿, 1,001-2,000฿, 2,001-3,000฿, 3,001-4,000฿, 4,001-5,000฿, 5,001-6,000฿, 6,001-7,000฿, >7,000฿}

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อคำถาม	คำอธิบายข้อคำถาม	จำนวนคำตอบ	ตัวเลือกในคำถาม
ผู้ร่วมเดินทาง	ผู้ที่ผู้ตอบแบบสอบถามเดินทางด้วยในการมายังสถานที่ท่องเที่ยว	6	{เดินทางคนเดียว, เพื่อน, แฟน, ครอบครัว, สามี/ภรรยา, เพื่อนร่วมงาน}
กิจกรรม	ลักษณะกิจกรรมที่ผู้ตอบแบบสอบถามได้ร่วมในการไปเที่ยวยังสถานที่ท่องเที่ยวนั้นๆ	8	{ดำน้ำ, เดินป่า, เรียนรู้วัฒนธรรมท้องถิ่น, เรียนรู้ประวัติศาสตร์, ไปสถานบันเทิง, ซื้อสินค้า เช่น ของฝาก หรือของดังประจำท้องถิ่น, ชื่นชมธรรมชาติ, ไหว้พระ/ปฏิบัติธรรม}
อาหารริมทาง	ผู้ตอบแบบสอบถามมีความชื่นชอบทานอาหารริมทางเมื่อไปท่องเที่ยวยังสถานที่นั้นๆ หรือไม่	2	{ชอบทาน, ไม่ชอบทาน}
งบประมาณอาหาร	จำนวนงบประมาณที่ผู้ตอบแบบสอบถามใช้จ่ายเกี่ยวกับค่าอาหารเมื่อมายังสถานที่ท่องเที่ยว	6	{≤1,000฿, 1,001-2,000฿, 2,001-3,000฿, 3,001-4,000฿, 4,001-5,000฿, >5,000฿}
แพ็คเกจ	ผู้ตอบแบบสอบถามได้ซื้อแพ็คเกจในการท่องเที่ยวครั้งนี้หรือไม่	2	{ซื้อ, ไม่ได้ซื้อ}
พักค้างแรม	ผู้ตอบแบบสอบถามมีการพักค้างรามที่สถานที่ท่องเที่ยวหรือไม่	2	{ค้างคืน, ไม่ได้ค้างคืน}
จำนวนคืนที่พักค้างแรม	จำนวนคืนที่ผู้ตอบแบบสอบถามพักค้างแรมที่สถานที่ท่องเที่ยว	7	{ประกอบด้วย 1 คืน, 2 คืน, 3 คืน, 4 คืน, 5 คืน, มากกว่า 5 คืน}

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อคำถาม	คำอธิบายข้อคำถาม	จำนวนคำตอบ	ตัวเลือกในคำถาม
ลักษณะที่พักแรม	ลักษณะที่พักสำหรับค้างแรมที่ผู้ตอบแบบสอบถามเลือกในสถานที่ท่องเที่ยว	6	{โรงแรม/รีสอร์ท (Resort) , เกสเฮาส์ (Guesthouse) /บังกะโล (Bangalow) , เต็นท์ (Tent) , บ้านญาติ, และ อพาร์ทเมนท์ (Apartment)}
งบประมาณสำหรับที่พักค้างแรม	จำนวนงบประมาณที่ผู้ตอบแบบสอบถามใช้จ่ายเกี่ยวกับค่าที่พักค้างแรมที่สถานที่ท่องเที่ยว	10	{≤1,000฿, 1,001-2,000฿, 2,001-3,000฿, 3,001-4,000฿, 4,001-5,000฿, 5,001-6,000฿, 6,001-8,000฿, 8,001-10,000฿,
เพศ	เพศของผู้ตอบแบบสอบถาม	3	{ชาย, หญิง, เพศทางเลือก (LGBTQ }
อายุ	อายุของผู้ตอบแบบสอบถาม	6	{≤20 ปี, 21-30 ปี, 31-40 ปี, 41-50 ปี, 51-60 ปี, >60 ปี}
ศาสนา	ศาสนาของผู้ตอบแบบสอบถาม	4	{พุทธ, คริสต์, อิสลาม, ไม่นับถือศาสนา}
สถานภาพ	สถานภาพของผู้ตอบแบบสอบถาม	6	{โสด, สมรส (มีบุตร), สมรส (ไม่มีบุตร), หม้าย, หย่าร้าง (มีบุตร), หย่าร้าง (ไม่มีบุตร)}
อาชีพ	อาชีพของผู้ตอบแบบสอบถาม	9	{นักเรียน/นักศึกษา, พ่อบ้าน/แม่บ้าน, รับจ้าง, เกษตรกร, ข้าราชการ/รัฐวิสาหกิจ, ค้าขาย/ธุรกิจส่วนตัว, พนักงานบริษัทเอกชน, เกษียณ, ว่างงาน}
รายได้	รายได้ต่อเดือนของผู้ตอบแบบสอบถาม	7	{≤15,000฿, 15,001-25,000฿, 25,001-30,000฿, 30,001-35,000฿, 40,001-45,000฿, 45,001-65,000฿, >65,000฿, No income}

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อคำถาม	คำอธิบายข้อคำถาม	จำนวนคำตอบ	ตัวเลือกในคำถาม
ระดับการศึกษา	ระดับการศึกษาสูงสุดของผู้ตอบแบบสอบถาม	6	{ประถมศึกษา, มัธยมศึกษาตอนต้น, มัธยมศึกษาตอนปลาย/ปวช., ประกาศนียบัตร/ปวส., ปริญญาตรี, ปริญญาโท หรือสูงกว่า}
Class label	สถานที่ท่องเที่ยวที่ผู้ตอบแบบสอบถามประทับใจที่สุด	82	{ตลาดน้ำอัมพวา (สมุทรสาคร), เกาะช้าง (ตราด), อุทยานประวัติศาสตร์พนมรุ้ง (บุรีรัมย์)}

3.2 การเตรียมข้อมูล

3.2.1 การเลือกคุณลักษณะ (Feature Selection)

การเลือกคุณลักษณะ (Feature Selection) เป็นขั้นตอนสำคัญในการเตรียมข้อมูลก่อนที่จะทำการประมวลผลต่อไป การดำเนินการนี้สามารถเพิ่มประสิทธิภาพในการทำนายของแบบจำลองได้ โดยในงานวิจัยนี้เราได้ใช้วิธีการเลือกคุณลักษณะจำนวน 3 วิธีเพื่อดำเนินการทดลองดังต่อไปนี้

1. เทคนิคการเลือกคุณลักษณะที่สำคัญ (Feature Importance) ในการศึกษาครั้งนี้ ผู้วิจัยได้ใช้วิธีการทางอัลกอริทึมของเทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคการสุ่มแบบป่า (Random Foerst) เพื่อวิเคราะห์คุณลักษณะที่มีความสำคัญต่อการตัดสินใจของนักท่องเที่ยวในการเลือกสถานที่ท่องเที่ยวปลายทาง โดยการเลือกคุณลักษณะที่สำคัญผู้วิจัยได้พิจารณาจากจากคุณลักษณะที่มีค่ามากกว่า 0.05 มาใช้ในการวิจัย
2. เทคนิคไคสแควร์ (Feature Importance) เพื่อให้เข้าใจความสัมพันธ์ระหว่างคุณลักษณะและการเลือกจุดหมายปลายทางของนักท่องเที่ยว ในการวิจัยนี้เราได้เลือกใช้วิธีการสถิติไคสแควร์ โดยเกณฑ์สำหรับการเลือกคือค่า p-value ที่ต่ำกว่า 0.01
3. การอินเตอร์เซ็ค (Intersect) วิธีการนี้เกี่ยวข้องกับการทับซ้อนของคุณลักษณะที่ถูกเลือกจากเทคนิคการเลือกคุณลักษณะที่สำคัญ (Feature Importance) และ เทคนิคไคสแควร์ (Feature Importance)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2 การแปลงข้อมูลโดยใช้เทคนิคการเข้ารหัสวันฮอด (One-hot encoding)

ในงานวิจัยนี้ ข้อมูลที่ที่นำมาวิจัยจากการสำรวจเป็นข้อมูลประเภทแบบหมวดหมู่ (categorical data) ทั้งหมดยกเว้นข้อมูล อายุ (age) ผู้เข้าร่วมการสำรวจต้องกรอกปีเกิดในแบบสอบถาม และจากนั้นอายุของผู้ใช้จะถูกคำนวณขึ้นมาหลังจากนั้น และเพื่อให้เข้ากับลักษณะแบบหมวดหมู่ของคุณลักษณะอื่น ๆ ผู้วิจัยจึงใช้กระบวนการการเข้ารหัสแบบหมวดหมู่ (discretization process) เพื่อแปลงคุณลักษณะ อายุ เป็นรูปแบบแบบหมวดหมู่ ตามที่แสดงในภาพที่ 3.6 (ก) โดยค่าอายุถูกแบ่งเป็นช่วงอายุ 6 กลุ่มได้แก่ น้อยกว่า 20 ปี, 21-30 ปี, 31-40 ปี, 41-50 ปี, 51-60 ปี, และมากกว่า 60 ปี และในภาพที่ 3.6 (ข) แสดงให้เห็นถึงทุกคุณลักษณะได้รับการเข้ารหัสแบบหมวดหมู่ (one-hot encoding) เพื่อนำเข้าไปยังแบบจำลองต่อไป

ID	Meaning
RES00001	51-60 years old
RES00002	21-30 years old
RES00003	More than 60 years old
RES00004	Less than 20 years old
RES00005	31-40 years old
RES00006	21-30 years old
RES00007	41-50 years old

(ก)

ID	less than 20 years old	21-30 years old	31-40 years old	41-50 years old	51-60 years old	more than 60 years old
RES00001	0	0	0	0	1	0
RES00002	0	1	0	0	0	0
RES00003	0	0	0	0	0	1
RES00004	1	0	0	0	0	0
RES00005	0	0	1	0	0	0
RES00006	0	1	0	0	0	0
RES00007	0	0	0	1	0	0

(ข)

รูปที่ 3.6 ตัวอย่างการแปลงคุณลักษณะอายุเป็นประเภทหมวดหมู่โดยใช้การเข้ารหัสวันฮอด

3.3 การพัฒนาแบบจำลอง

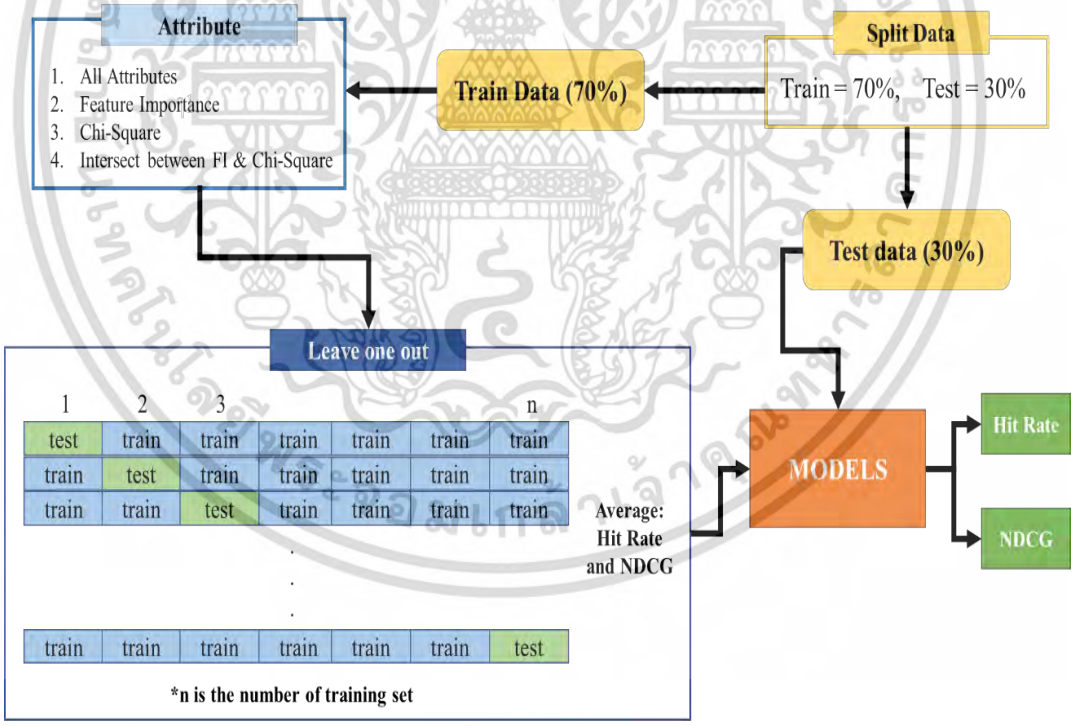
ภาพรวมของการทดลองของงานวิจัยถูกแสดงในรูปที่ 3.7 โดยข้อมูลถูกแบ่งเป็นสองส่วนสำหรับการฝึกแบบจำลองและทดสอบแบบจำลอง โดยใช้อัตราส่วน 70:30 ซึ่งในขั้นตอนการฝึก ใช้เทคนิค leave one out cross-validation และใช้เทคนิคการเลือกคุณลักษณะที่แตกต่างกัน เพื่อเพิ่มประสิทธิภาพในการทำนายของแบบจำลอง โดยมีรายละเอียดการใช้อัลกอริทึมดังตารางที่ 3.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.3 รายละเอียดของอัลกอริทึมที่ใช้ในการวิจัย

อัลกอริทึม	การตั้งค่าพารามิเตอร์
Decision Tree/ Random Forest	Gini/ Entropy
kNN	k = {1, 3, 5, 7, 9}
MLP (ReLU, sigmoid)	อัตราการเรียนรู้ = 0.1 Hidden layer = 100 nodes

โดยแบบจำลองจะแสดงลำดับสถานที่ที่แนะนำด้วยการเลือกจากค่าความน่าจะเป็น ซึ่งจะแนะนำสถานที่ทั้งหมด 5 อันดับ โดยเรียงจากค่าความน่าจะเป็นมากที่สุด 5 อันดับมาแสดง โดยแบบจำลองที่มีผลทำนายที่ดีที่สุดในขั้นตอนของการฝึกด้วยเทคนิค leave one out จะถูกเลือกมาการประยุกต์ใช้ในระบบแนะนำ และในท้ายที่สุดชุดข้อมูลทดสอบจะถูกนำมาวัดประสิทธิภาพของแบบจำลอง โดยใช้ตัววัดประสิทธิภาพ 2 ค่าคือ Hit rate และ NDCG โดยค่า k ในงานวิจัยครั้งนี้มีค่าเป็น 5



รูปที่ 3.7 ภาพรวมของกระบวนการทดลอง

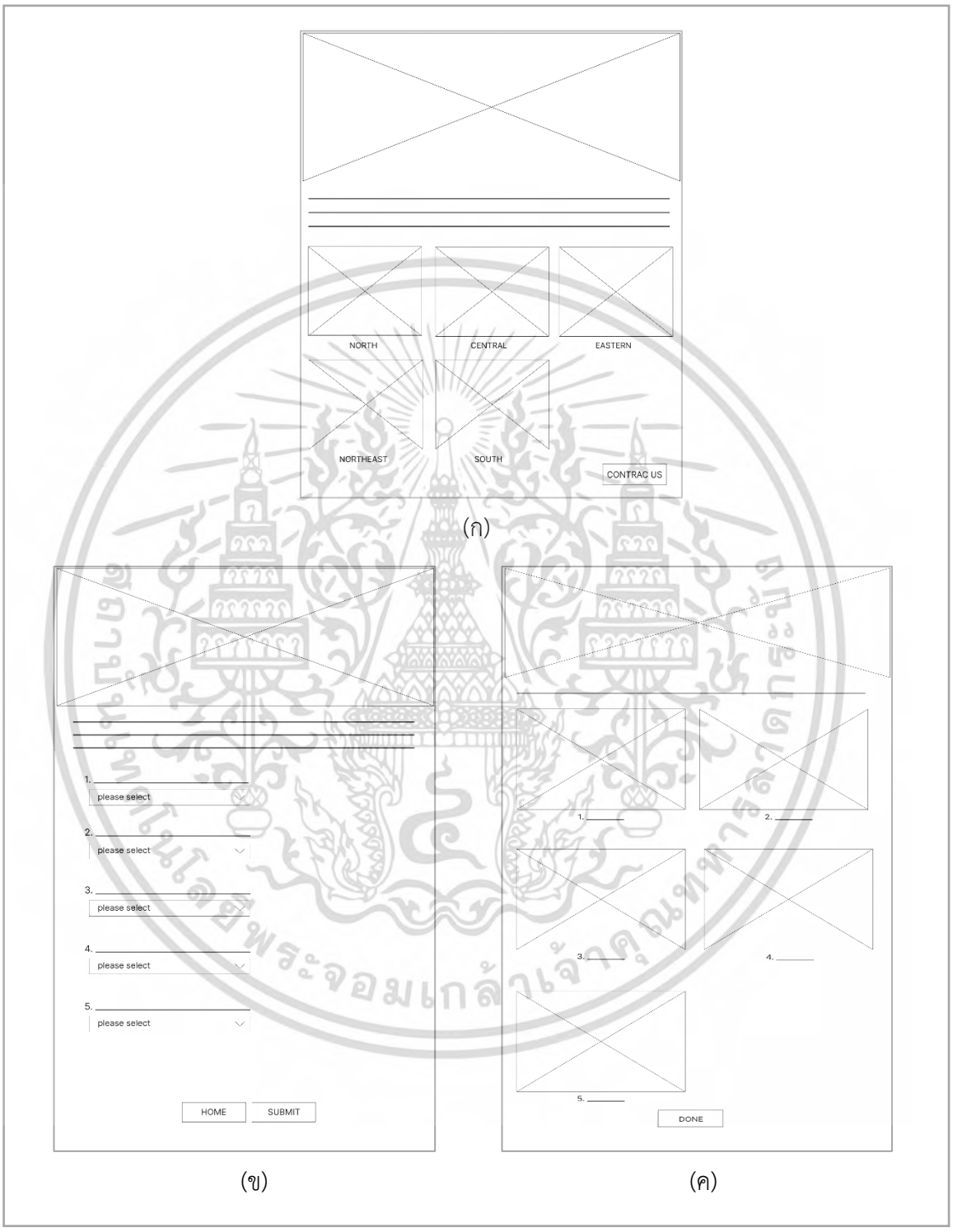
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.7 แสดงถึงภาพรวมถึงโครงสร้างและขั้นตอนการสร้างแบบจำลอง โดยมีขั้นตอนการดำเนินการทั้งหมด 5 ขั้นตอน ดังนี้

1. นำข้อมูลที่มีมาแบ่งออกเป็น 2 ส่วน ด้วยอันตราส่วน 70:30 โดยให้ส่วน 70 เป็นข้อมูลสำหรับการฝึกฝนของแบบจำลอง และส่วน 30 เป็นข้อมูลสำหรับการทดสอบแบบจำลอง
2. นำข้อมูลฝึกฝนมาทำการเลือกคุณลักษณะที่สำคัญทั้ง 4 วิธีคือ 1.) การเลือกคุณลักษณะทั้งหมดที่มี (All Attributes) 2.) การเลือกโดยใช้เทคนิคการเลือกคุณลักษณะที่สำคัญ (Feature Importance) 3.) การเลือกโดยใช้เทคนิคไคสแควร์ (Feature Importance) 4.) การเลือกโดยใช้วิธีการอินเตอร์เซ็ค (Intersect)
3. จากนั้นนำข้อมูลที่ถูกเลือกจากข้อ 2. มาดำเนินการฝึกฝนแบบจำลองด้วยเทคนิค Leave-One-Out Cross Validation โดยจะใช้ข้อมูลที่มีทั้งหมดเป็นทั้งข้อมูลฝึกฝนและเป็นข้อมูลทดสอบ เพื่อนำเข้า Machine Learning ทั้ง 4 วิธีที่กล่าวข้างต้น เพื่อเพิ่มประสิทธิภาพการทำนายของแบบจำลอง
4. จากนั้นจะนำค่า Hit rate และค่า NDCG ที่ได้ทั้งหมดจากการทำ Leave-One-Out Cross Validation มาหาค่าเฉลี่ย เพื่อเลือกแบบจำลองที่มีผลลัพธ์ที่ดีที่สุดมาใช้ในระบบ
5. หลังจากนั้นจะดำเนินการนำแบบจำลองที่เลือกมาทดสอบด้วยข้อมูลส่วน 30 ที่ถูกแบ่งเป็นข้อมูลทดสอบไว้ในขั้นตอนที่ 1. เพื่อวัดประสิทธิภาพของ Hit rate และ NDCG อีกครั้งจากชุดข้อมูลทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 การออกแบบส่วนประสานงานผู้ใช้



รูปที่ 3.8 แผนผังและโครงสร้าง (Wire Frame) ก่อนนำไปพัฒนาเว็บแอปพลิเคชัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.8 แสดงโครงสร้างภาพรวมเบื้องต้นก่อนที่ผู้วิจัยจำนำไปพัฒนาเว็บแอปพลิเคชัน โดยประกอบด้วย 3 หน้า ได้แก่

- **หน้าที่ 1:** ดังรูปที่ 3.8 (ก) ในส่วนของด้านบนสุดจะเป็นการใส่ภาพที่สื่อถึงการท่องเที่ยว และถัดลงมาจะเป็นประโยคข้อความเพื่อสื่อให้ผู้ใช้เลือกภูมิภาคที่ผู้ใช้สนใจไปท่องเที่ยว โดยประกอบด้วย 5 ภูมิภาคดังแสดงในรูป
- **หน้าที่ 2:** เมื่อผู้ใช้เลือกภูมิภาค จากนั้นจะแสดงหน้าที่ 2 ดังรูปที่ 3.8 (ข) โดยในส่วนด้านบนสุดจะเป็นรูปภาพเดียวกับหน้าที่ 1 ถัดลงมาจะเป็นข้อคำถามที่แบบจำลองจำเป็นต้องใช้ในการทำนายโดยในแต่ละภูมิภาคจะมีคำถามที่แตกต่างกันออกไปตามการเลือกคุณลักษณะ (Attributes) ที่เหมาะสมที่สุด เมื่อผู้ใช้ตอบข้อคำถามครบทั้งหมดและกดปุ่ม Submit จะเข้าสู่หน้าที่ 3
- **หน้าที่ 3:** ในส่วนของด้านบนสุดจะเป็นรูปภาพเดียวกันกับหน้าที่ 1 และ 2 ถัดลงจะเป็นข้อความเพื่อสื่อถึงผลลัพธ์การทำนายสถานที่ท่องเที่ยวจากแบบจำลอง โดยแบบจำลองจะใช้ข้อมูลที่ได้ในหน้าที่ 2 เพื่อทำนายสถานที่ท่องเที่ยวที่เหมาะสม และผลลัพธ์ที่ได้จะถูกแสดงออกเป็นชื่อสถานที่ท่องเที่ยวพร้อมรูปภาพทั้งหมด 5 สถานที่ ดังรูปที่ 3.8 (ค)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการวิจัยและการอภิปรายผล

งานวิจัยนี้ได้ทำการศึกษาและพัฒนาแบบจำลองสำหรับระบบแนะนำสถานที่ท่องเที่ยวในเมืองรองของประเทศไทยจากการเรียนรู้ของเครื่องเพื่อแนะนำสถานที่ท่องเที่ยวเมืองรองอย่างเหมาะสมรายบุคคล โดยมีผลการวิจัยและอภิปรายดังนี้

4.1 ผลการวิจัย

ในงานวิจัยนี้ได้ดำเนินการแบ่งแบบจำลองออกเป็น 5 แบบจำลองเพื่อแนะนำสถานที่ท่องเที่ยวในเมืองรองของประเทศไทยโดยแบ่งออกตามภูมิภาคของประเทศไทย ได้แก่ ภาคเหนือ ภาคใต้ ภาคตะวันออก ภาคตะวันออกเฉียงเหนือ และภาคกลาง โดยจำนวนข้อมูลที่ถูกแบ่งเป็นข้อมูลเรียนรู้และข้อมูลทดสอบในแต่ละภูมิภาคแสดงดังตารางที่ 4.1 โดยที่จำนวนสถานที่ท่องเที่ยวจะหมายถึงจำนวนสถานที่ท่องเที่ยวในแต่ละภูมิภาคหรือหมายถึงจำนวนคลาสในแต่ละแบบจำลอง

ตารางที่ 4.1 รายละเอียดของชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบในแต่ละภูมิภาค

อันดับ	แบบจำลอง	จำนวนข้อมูลเรียนรู้	จำนวนข้อมูลทดสอบ	จำนวนสถานที่ท่องเที่ยว
1	ภาคเหนือ	109	48	26
2	ภาคใต้	59	26	14
3	ภาคตะวันออก	49	21	12
4	ภาคตะวันออกเฉียงเหนือ	49	21	18
5	ภาคกลาง	32	15	12

สำหรับการสร้างแบบจำลอง ใช้อัลกอริทึมการเรียนรู้ของเครื่อง 4 แบบดังที่ได้กล่าวไปข้างต้น ได้ผลการทดลองดังตารางที่ 4.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 ผลการทดลองในแต่ละแบบจำลองโดยแบ่งตามภูมิภาค

ภูมิภาค	การเลือกคุณลักษณะ	อัลกอริทึมการเรียนรู้ของเครื่อง	Hit Rate	NDCG
ภาคเหนือ	All Attributes	Decision Tree (Entropy)	0.486	0.509
		Decision Tree (Gini)	0.422	0.508
		Random Forest (Entropy)	0.560	0.484
		Random Forest (Gini)	0.550	0.491
		MLP (Relu)	0.541	0.454
		MLP (Sigmoid)	0.505	0.533
		KNN (1)	0.422	0.644
		KNN (3)	0.523	0.425
		KNN (5)	0.495	0.396
		KNN (7)	0.440	0.575
		KNN (9)	0.459	0.619
	Feature Importance	Decision Tree (Entropy)	0.422	0.418
		Decision Tree (Gini)	0.413	0.481
		Random Forest (Entropy)	0.569	0.451
		Random Forest (Gini)	0.569	0.408
		MLP (Relu)	0.523	0.435
		MLP (Sigmoid)	0.541	0.458
		KNN (1)	0.431	0.583
		KNN (3)	0.495	0.395
		KNN (5)	0.505	0.364
		KNN (7)	0.486	0.524
		KNN (9)	0.486	0.521
	Chi-Square	Decision Tree (Entropy)	0.468	0.498
		Decision Tree (Gini)	0.477	0.531
		Random Forest (Entropy)	0.587	0.564
		Random Forest (Gini)	0.550	0.440
		MLP (Relu)	0.569	0.447

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภูมิภาค	การเลือกคุณลักษณะ	อัลกอริทึมการเรียนรู้ของเครื่อง	Hit Rate	NDCG
ภาคเหนือ	Chi-Square	MLP (Sigmoid)	0.541	0.458
		KNN (1)	0.468	0.570
		KNN (3)	0.514	0.511
		KNN (5)	0.523	0.513
		KNN (7)	0.569	0.452
		KNN (9)	0.550	0.481
	Intersect	Decision Tree (Entropy)	0.523	0.570
		Decision Tree (Gini)	0.440	0.390
		Random Forest (Entropy)	0.560	0.449
		Random Forest (Gini)	0.569	0.484
		MLP (Relu)	0.514	0.472
		MLP (Sigmoid)	0.541	0.424
		KNN (1)	0.486	0.587
		KNN (3)	0.495	0.487
		KNN (5)	0.505	0.373
		KNN (7)	0.505	0.477
		KNN (9)	0.514	0.507
ภาคใต้	All Attributes	Decision Tree (Entropy)	0.710	0.390
		Decision Tree (Gini)	0.660	0.430
		Random Forest (Entropy)	0.780	0.410
		Random Forest (Gini)	0.780	0.360
		MLP (Relu)	0.690	0.430
		MLP (Sigmoid)	0.678	0.410
		KNN (1)	0.590	0.440
		KNN (3)	0.640	0.490
		KNN (5)	0.640	0.470
		KNN (7)	0.680	0.420
		KNN (9)	0.630	0.480

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภูมิภาค	การเลือกคุณลักษณะ	อัลกอริทึมการเรียนรู้ของเครื่อง	Hit Rate	NDCG
ภาคใต้	Feature Importance	Decision Tree (Entropy)	0.660	0.460
		Decision Tree (Gini)	0.630	0.390
		Random Forest (Entropy)	0.760	0.360
		Random Forest (Gini)	0.780	0.360
		MLP (Relu)	0.630	0.470
		MLP (Sigmoid)	0.695	0.415
		KNN (1)	0.630	0.390
		KNN (3)	0.580	0.440
		KNN (5)	0.680	0.410
		KNN (7)	0.660	0.430
		KNN (9)	0.690	0.380
	Chi-Square	Decision Tree (Entropy)	0.680	0.440
		Decision Tree (Gini)	0.750	0.370
		Random Forest (Entropy)	0.760	0.440
		Random Forest (Gini)	0.780	0.400
		MLP (Relu)	0.690	0.450
		MLP (Sigmoid)	0.780	0.390
		KNN (1)	0.630	0.380
		KNN (3)	0.710	0.480
		KNN (5)	0.710	0.510
		KNN (7)	0.730	0.530
		KNN (9)	0.710	0.610
	Intersect	Decision Tree (Entropy)	0.630	0.350
		Decision Tree (Gini)	0.730	0.390
		Random Forest (Entropy)	0.760	0.380
		Random Forest (Gini)	0.760	0.420
		MLP (Relu)	0.680	0.370
		MLP (Sigmoid)	0.746	0.412

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภูมิภาค	การเลือกคุณลักษณะ	อัลกอริทึมการเรียนรู้ของเครื่อง	Hit Rate	NDCG
ภาคใต้	Intersect	KNN (1)	0.630	0.340
		KNN (3)	0.710	0.430
		KNN (5)	0.710	0.390
		KNN (7)	0.690	0.410
		KNN (9)	0.690	0.440
นอกภาคใต้	All Attributes	Decision Tree (Entropy)	0.373	0.642
		Decision Tree (Gini)	0.569	0.609
		Random Forest (Entropy)	0.784	0.510
		Random Forest (Gini)	0.784	0.490
		MLP (Relu)	0.804	0.385
		MLP (Sigmoid)	0.843	0.415
		KNN (1)	0.392	0.759
		KNN (3)	0.529	0.507
		KNN (5)	0.569	0.419
		KNN (7)	0.667	0.446
		KNN (9)	0.725	0.416
	Feature Importance	Decision Tree (Entropy)	0.430	0.480
		Decision Tree (Gini)	0.310	0.640
		Random Forest (Entropy)	0.820	0.410
		Random Forest (Gini)	0.840	0.370
		MLP (Relu)	0.800	0.490
		MLP (Sigmoid)	0.843	0.394
		KNN (1)	0.350	0.670
		KNN (3)	0.490	0.730
		KNN (5)	0.590	0.590
		KNN (7)	0.650	0.490
		KNN (9)	0.710	0.460
	Chi-Square	Decision Tree (Entropy)	0.410	0.680

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภูมิภาค	การเลือกคุณลักษณะ	อัลกอริทึมการเรียนรู้ของเครื่อง	Hit Rate	NDCG
ภาคตะวันออกเฉียงเหนือ	Chi-Square	Decision Tree (Gini)	0.390	0.410
		Random Forest (Entropy)	0.760	0.450
		Random Forest (Gini)	0.800	0.380
		MLP (Relu)	0.800	0.460
		MLP (Sigmoid)	0.824	0.508
		KNN (1)	0.350	0.800
		KNN (3)	0.490	0.580
		KNN (5)	0.610	0.490
		KNN (7)	0.690	0.470
		KNN (9)	0.730	0.390
	Intersect	Decision Tree (Entropy)	0.410	0.570
		Decision Tree (Gini)	0.350	0.510
		Random Forest (Entropy)	0.840	0.400
		Random Forest (Gini)	0.820	0.400
		MLP (Relu)	0.760	0.460
		MLP (Sigmoid)	0.824	0.464
		KNN (1)	0.370	0.810
		KNN (3)	0.510	0.750
		KNN (5)	0.670	0.430
		KNN (7)	0.670	0.480
		KNN (9)	0.780	0.370
ภาคตะวันออก	All Attributes	Decision Tree (Entropy)	0.570	0.550
		Decision Tree (Gini)	0.490	0.470
		Random Forest (Entropy)	0.690	0.560
		Random Forest (Gini)	0.730	0.480
		MLP (Relu)	0.670	0.540
		MLP (Sigmoid)	0.673	0.552
		KNN (1)	0.570	0.570

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภูมิภาค	การเลือกคุณลักษณะ	อัลกอริทึมการเรียนรู้ของเครื่อง	Hit Rate	NDCG
อินเดียตะวันออกเฉียงเหนือ	All Attributes	KNN (3)	0.590	0.590
		KNN (5)	0.590	0.660
		KNN (7)	0.650	0.450
		KNN (9)	0.610	0.600
	Feature Importance	Decision Tree (Entropy)	0.610	0.530
		Decision Tree (Gini)	0.510	0.380
		Random Forest (Entropy)	0.690	0.490
		Random Forest (Gini)	0.650	0.680
		MLP (Relu)	0.710	0.570
		MLP (Sigmoid)	0.633	0.794
		KNN (1)	0.570	0.530
		KNN (3)	0.610	0.580
		KNN (5)	0.690	0.560
		KNN (7)	0.690	0.520
		KNN (9)	0.670	0.560
	Chi-Square	Decision Tree (Entropy)	0.550	0.450
		Decision Tree (Gini)	0.610	0.430
		Random Forest (Entropy)	0.760	0.500
		Random Forest (Gini)	0.730	0.620
		MLP (Relu)	0.760	0.470
		MLP (Sigmoid)	0.673	0.465
		KNN (1)	0.570	0.570
		KNN (3)	0.630	0.450
		KNN (5)	0.690	0.590
		KNN (7)	0.710	0.490
		KNN (9)	0.690	0.520
	Intersect	Decision Tree (Entropy)	0.570	0.490
		Decision Tree (Gini)	0.510	0.410

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภูมิภาค	การเลือกคุณลักษณะ	อัลกอริทึมการเรียนรู้ของเครื่อง	Hit Rate	NDCG
ภาคตะวันออกเฉียงเหนือ	Intersect	Random Forest (Entropy)	0.690	0.490
		Random Forest (Gini)	0.730	0.510
		MLP (Relu)	0.710	0.500
		MLP (Sigmoid)	0.653	0.496
		KNN (1)	0.570	0.530
		KNN (3)	0.590	0.510
		KNN (5)	0.610	0.680
		KNN (7)	0.650	0.540
		KNN (9)	0.650	0.560
ภาคกลาง	All Attributes	Decision Tree (Entropy)	0.56	0.39
		Decision Tree (Gini)	0.56	0.36
		Random Forest (Entropy)	0.56	0.57
		Random Forest (Gini)	0.59	0.67
		MLP (Relu)	0.50	0.57
		MLP (Sigmoid)	0.50	0.60
		KNN (1)	0.44	0.32
		KNN (3)	0.44	0.47
		KNN (5)	0.47	0.37
		KNN (7)	0.47	0.52
		KNN (9)	0.47	0.83
	Feature Importance	Decision Tree (Entropy)	0.53	0.41
		Decision Tree (Gini)	0.56	0.45
		Random Forest (Entropy)	0.63	0.55
		Random Forest (Gini)	0.59	0.84
		MLP (Relu)	0.63	0.31
		MLP (Sigmoid)	0.56	0.45
		KNN (1)	0.47	0.38
		KNN (3)	0.47	0.73

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภูมิภาค	การเลือกคุณลักษณะ	อัลกอริทึมการเรียนรู้ของเครื่อง	Hit Rate	NDCG
ภาคกลาง	Feature Importance	KNN (5)	0.50	0.57
		KNN (7)	0.63	0.38
		KNN (9)	0.63	0.72
	Chi-Square	Decision Tree (Entropy)	0.47	0.38
		Decision Tree (Gini)	0.56	0.32
		Random Forest (Entropy)	0.72	0.46
		Random Forest (Gini)	0.69	0.45
		MLP (Relu)	0.59	0.52
		MLP (Sigmoid)	0.66	0.39
		KNN (1)	0.47	0.38
		KNN (3)	0.47	0.45
		KNN (5)	0.50	0.60
		KNN (7)	0.56	0.41
		KNN (9)	0.66	0.44
	Intersect	Decision Tree (Entropy)	0.47	0.38
		Decision Tree (Gini)	0.53	0.35
		Random Forest (Entropy)	0.72	0.40
		Random Forest (Gini)	0.69	0.40
		MLP (Relu)	0.59	0.52
		MLP (Sigmoid)	0.66	0.36
		KNN (1)	0.44	0.37
		KNN (3)	0.47	0.43
		KNN (5)	0.53	0.55
		KNN (7)	0.63	0.39
		KNN (9)	0.63	0.38

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.2 แสดงผลการทดลองในแต่ละภูมิภาคของประเทศไทยด้วยเทคนิค Leave one out สำหรับการสร้างแบบจำลอง โดยในแต่ละภูมิภาคได้ทำการทดลองด้วยวิธีการเลือกคุณลักษณะที่แตกต่างกัน 4 วิธี และในแต่ละวิธีการเลือกคุณลักษณะจะมีการสร้างแบบจำลองด้วยอัลกอริทึมการเรียนรู้ของเครื่องทั้งหมด 4 แบบดังที่กล่าวไว้ในข้างต้น โดยในวิธีต่างๆ ของอัลกอริทึมจะแสดงอยู่ในวงเล็บ เช่น Decision Tree จะแสดงวิธี Entropy และ Gini อยู่ในวงเล็บด้านหลัง เป็นต้น

ประสิทธิภาพในการทำนายสถานที่ท่องเที่ยวของแบบจำลองบนข้อมูลทดสอบในแต่ละภูมิภาคแสดงผลดังตารางที่ 4.3 โดยแบบจำลองที่ให้ค่า Hit Rate สูงสุดในแต่ละภูมิภาคจะถูกนำมาใช้ในการประเมินประสิทธิภาพบนข้อมูลทดสอบและนำไปปรับใช้กับเว็บแอปพลิเคชันสำหรับระบบแนะนำ

ตารางที่ 4.3 ประสิทธิภาพการทำงานของแบบจำลองในแต่ละภูมิภาค

ภูมิภาค	**การเลือกคุณลักษณะ**	**อัลกอริทึมการเรียนรู้ของเครื่อง**	**Hit Rate**	**NDCG**
ภาคเหนือ	Chi-Square	Random Forest	0.69	0.53
ภาคใต้	All Attributes	Random Forest	0.81	0.45
ภาคตะวันออก	All Attributes	MLP (sigmoid)	0.86	0.55
ภาคตะวันออกเฉียงเหนือ	Chi-Square	Random Forest	0.76	0.69
ภาคกลาง	Chi-Square	Random Forest	0.87	0.74
		ค่าเฉลี่ย	**0.81**	**0.61**

เมื่อพิจารณาผลการทดลองในตารางที่ 4.3 พบว่าอัลกอริทึม Random Forest ให้ประสิทธิภาพในการทำนายสูงสุดในเกือบทุกภูมิภาค เนื่องจากอัลกอริทึมนี้เป็น Ensemble classifier จากผลการทดลองทั้ง 5 ภูมิภาค พบว่าภาคกลางมีค่า Hit Rate สูงสุดที่ 0.87 รองลงมาคือภาคตะวันออก ภาคใต้ ภาคตะวันออกเฉียงเหนือ และภาคเหนือตามลำดับ โดยมีค่า Hite Rate เฉลี่ยอยู่ที่ 0.81 ในส่วนของค่า NDCG มีค่าสูงสุดที่ภาคกลางคือ 0.74 ตามมาด้วยภาคตะวันออกเฉียงเหนือ ภาคตะวันออก ภาคเหนือ และภาคใต้ตามลำดับ โดยมีค่าเฉลี่ยของ NDCG คือ 0.61

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 อภิปรายผล

ในส่วนของการอภิปรายผลได้กล่าวถึง 3 หัวข้อ ได้แก่ การวิเคราะห์ปัจจัยที่มีผลต่อการท่องเที่ยว ความรู้ที่ได้จากแบบจำลอง และการนำแบบจำลองไปพัฒนาระบบ

4.2.1 การวิเคราะห์ปัจจัยที่มีผลต่อการท่องเที่ยว

ในส่วนนี้จะเป็นการวิเคราะห์ปัจจัยที่ส่งผลต่อการเลือกจุดหมายปลายทางการเดินทางของนักท่องเที่ยว โดยรายการคุณลักษณะที่เลือกได้ถูกแยกออกตามภูมิภาคแสดงดังตารางที่ 4.4 โดยคุณลักษณะที่ถูกเลือกมาในแต่ละภูมิภาคทำการเลือกโดยเทคนิคที่กล่าวในตารางที่ 4.3 เมื่อพิจารณาตารางด้านล่างจะพบว่าในแต่ละภูมิภาคจะมีจำนวนคุณลักษณะที่แตกต่างกันออกไป เช่น ภาคกลางมีคุณลักษณะ 8 ประการ โดยการถูกเลือกด้วยวิธีไคสแควร์ ในขณะที่ภาคเหนือถูกเลือก 9 ประการ เป็นต้น โดยในรายการที่แสดงในแต่ละภูมิภาคได้ถูกจัดอันดับตามค่าของ p-value ของแต่ละคุณลักษณะโดยเรียงลำดับจากน้อยไปหามาก

ตารางที่ 4.4 คุณลักษณะที่ถูกเลือกมาใช้ในแบบจำลองโดยแบ่งตามภูมิภาค

ภูมิภาค	ภาคเหนือ	ภาคกลาง	ภาคตะวันออก	ภาคตะวันออกเฉียงเหนือ	ภาคใต้
คุณลักษณะที่ถูกเลือก	งบประมาณสำหรับที่พัก	กิจกรรม	งบประมาณการเดินทาง	ระดับการศึกษา	จำนวนคืนที่พักค้างแรม
	อายุ	งบประมาณอาหาร	เดือนที่เดินทางท่องเที่ยว	ลักษณะที่พักแรม	งบประมาณการเดินทาง
	กิจกรรม	อาชีพ	อายุ	งบประมาณการเดินทาง	เดือนที่เดินทางท่องเที่ยว
	ระดับการศึกษา	เดือนที่เดินทางท่องเที่ยว	ระดับการศึกษา	งบประมาณอาหาร	ลักษณะที่พักแรม
	ลักษณะที่พักแรม	อายุ	รายได้	กิจกรรม	ผู้ร่วมเดินทาง
	เดือนที่เดินทางท่องเที่ยว	ลักษณะที่พักแรม	อาชีพ	สถานภาพ	รายได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภูมิภาค	ภาคเหนือ	ภาคกลาง	ภาคตะวันออก	ภาคตะวันออกเฉียงเหนือ	ภาคใต้
คุณลักษณะที่ถูกเลือก	สถานภาพ	แพ็คเก็จ	ผู้ร่วมเดินทาง	ผู้ร่วมเดินทาง	ระดับการศึกษา
	งบประมาณอาหาร	รายได้	งบประมาณอาหาร	งบประมาณสำหรับที่พัก	งบประมาณสำหรับที่พัก
	งบประมาณการเดินทาง		งบประมาณสำหรับที่พัก		กิจกรรม
			จำนวนคืนที่พักค้างแรม		งบประมาณอาหาร
			ลักษณะที่พักแรม		อาชีพ
			กิจกรรม		อายุ
			เพศ		สถานภาพ
			ศาสนา		เพศ
			สถานภาพ		ลักษณะการเดินทาง
			ลักษณะการเดินทาง		ศาสนา
			พักค้างแรม		พักค้างแรม
			แพ็คเก็จ		อาหารริมทาง
			อาหารริมทาง		แพ็คเก็จ

จากการวิเคราะห์คุณลักษณะในการจัดอันดับ 5 อันดับแรกของแต่ละภูมิภาค พบว่ามีคุณลักษณะที่เหมือนกันในหลายภูมิภาค ซึ่งมีคุณลักษณะที่เหมือนกันอยู่ 3 ภูมิภาคจาก 5 ภูมิภาค โดยมีทั้งหมด 6 คุณลักษณะ ได้แก่ กิจกรรม เดือนที่เดินทางท่องเที่ยว อายุ ระดับการศึกษา งบประมาณการเดินทาง และลักษณะสถานที่พักแรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภูมิภาค	ภาคกลาง	ภาคตะวันออก	ภาคเหนือ	ภาคตะวันออกเฉียงเหนือ	ภาคใต้
กิจกรรม	100%	100%	100%	100%	100%
ไหว้พระ/ปฏิบัติธรรม	13%	4%	8%	9%	13%
ชื่นชมธรรมชาติ	36%	70%	62%	41%	38%
ดำน้ำ	2%	12%	0%	0%	31%
ไปสถานบันเทิง	0%	1%	1%	0%	0%
เดินป่า	4%	3%	9%	9%	0%
เรียนรู้ประวัติศาสตร์	6%	3%	11%	20%	2%
เรียนรู้วัฒนธรรมท้องถิ่น	23%	0%	4%	16%	5%
ซื้อสินค้า เช่น ของฝาก ของดังประจำท้องถิ่น	15%	7%	5%	6%	12%

(ก) กิจกรรม

ภูมิภาค	ภาคกลาง	ภาคตะวันออก	ภาคเหนือ	ภาคตะวันออกเฉียงเหนือ	ภาคใต้
เดือนที่เดินทางท่องเที่ยว	100%	100%	100%	100%	100%
มกราคม	4%	7%	16%	7%	6%
กุมภาพันธ์	9%	7%	8%	4%	9%
มีนาคม	2%	10%	3%	6%	8%
เมษายน	21%	26%	4%	13%	26%
พฤษภาคม	2%	7%	0%	1%	9%
มิถุนายน	0%	3%	2%	0%	0%
กรกฎาคม	2%	5%	1%	0%	0%
สิงหาคม	0%	0%	3%	1%	0%
กันยายน	2%	0%	1%	1%	1%
ตุลาคม	19%	10%	7%	13%	14%
พฤศจิกายน	17%	16%	22%	27%	14%
ธันวาคม	21%	10%	33%	26%	12%

(ข) เดือนที่เดินทางท่องเที่ยว

รูปที่ 4.1 ค่าสัดส่วนของคุณลักษณะในแต่ละภูมิภาค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภูมิภาค	ภาคกลาง	ภาคตะวันออก	ภาคเหนือ	ภาคตะวันออกเฉียงเหนือ	ภาคใต้
อายุ	100%	100%	100%	100%	100%
≤20 ปี	2%	1%	4%	1%	4%
21-30 ปี	55%	51%	45%	23%	52%
31-40 ปี	17%	25%	18%	27%	21%
41-50 ปี	15%	7%	11%	27%	14%
51-60 ปี	11%	14%	18%	16%	9%
>60 ปี	0%	3%	3%	6%	0%

(ค) อายุ

ภูมิภาค	ภาคกลาง	ภาคตะวันออก	ภาคเหนือ	ภาคตะวันออกเฉียงเหนือ	ภาคใต้
งบประมาณการเดินทาง	100%	100%	100%	100%	100%
≤1,000฿	9%	5%	1%	4%	4%
1,001-2,000฿	21%	27%	13%	23%	18%
2,001-3,000฿	30%	25%	16%	29%	11%
3,001-4,000฿	23%	10%	17%	6%	16%
4,001-5,000฿	15%	14%	14%	16%	20%
5,001-6,000฿	2%	7%	8%	9%	7%
6,001-7,000฿	0%	0%	10%	3%	1%
>7,000฿	0%	12%	20%	11%	24%

(ง) งบประมาณการเดินทาง

รูปที่ 4.1 (ต่อ) ค่าสัดส่วนของคุณลักษณะในแต่ละภูมิภาค

รูปที่ 4.1 (ก) – (ง) แสดงค่าสัดส่วนของคุณลักษณะที่กระจายอยู่ในภูมิภาคเดียวกัน โดยแสดงเป็นค่าเปอร์เซ็นต์ เมื่อพิจารณาที่รูป 4.2 (ก) สรุปได้ว่าการเลือกสถานที่ท่องเที่ยวของนักท่องเที่ยวขึ้นอยู่กับกิจกรรมที่นักท่องเที่ยวต้องการ เช่น หากนักท่องเที่ยวชื่นชอบความสวยงามของธรรมชาติมักจะนิยมเดินทางไปภาคตะวันออกและภาคเหนือ ในทางกลับกันนักท่องเที่ยวที่สนใจกิจกรรมดำน้ำจะเลือกเดินทางไปยังภาคใต้ ในขณะเดียวกันนักท่องเที่ยวที่เดินทางไปภาค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตะวันออกเฉียงเหนือมีแนวโน้มที่จะสนใจเกี่ยวกับกิจกรรมการเรียนรู้ประวัติศาสตร์ และสุดท้ายนักท่องเที่ยวที่มีความชื่นชอบและหลงในวัฒนธรรมท้องถิ่นมักนิยมเดินทางไปยังภาคกลาง

เมื่อพิจารณารูปที่ 4.1 (ข) เห็นได้ชัดว่าเดือนที่นักท่องเที่ยวนิยมเดินทางเป็นจำนวนมาก คือเดือนเมษายน พฤศจิกายน ธันวาคม และตามมาด้วยเดือนตุลาคมและมกราคม นักท่องเที่ยวมักนิยมเดินทางไปภาคตะวันออกและภาคใต้ในช่วงฤดูร้อนคือเดือนเมษายน เนื่องจากภูมิภาคเหล่านี้มีสถานที่ท่องเที่ยวเชิงทะเล และหมู่เกาะต่างๆ เป็นจำนวนมาก ในทางตรงกันข้ามในฤดูหนาวคือตั้งแต่เดือนพฤศจิกายนจนถึงเดือนมกราคม นักท่องเที่ยวมีแนวโน้มที่จะออกเดินทางไปยังภาคเหนือและภาคตะวันออกเฉียงเหนือ เนื่องจากมีสภาพอากาศที่หนาวเย็นและมีภูมิประเทศเป็นเทือกเขาโดยส่วนใหญ่ นอกจากนี้ในช่วงฤดูฝนคือเดือนตุลาคมพบว่านักท่องเที่ยวนิยมเดินทางไปภาคกลางมากกว่าภูมิภาคอื่นๆ เนื่องจากมีกิจกรรมทางน้ำมากมาย เช่น ล่องแก่ง และ ชื่นชมน้ำตก เป็นต้น

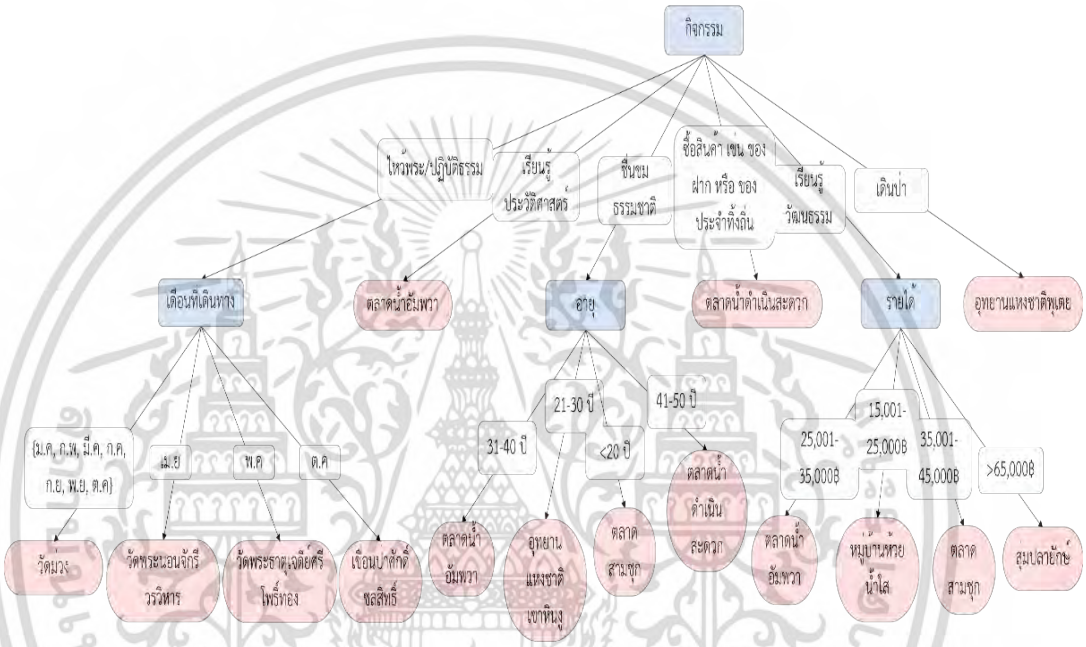
เมื่อพิจารณาอายุของนักท่องเที่ยวจากรูปที่ 4.1 (ค) พบว่านักท่องเที่ยวอายุ 21 – 30 ปี นิยมเดินทางไปยังภาคกลาง ภาคใต้ และภาคตะวันออกตามลำดับ ในทางกลับกันนักท่องเที่ยวอายุ 31 – 50 ปี มีความชื่นชอบในการเดินทางไปยังภาคตะวันออกเฉียงเหนือมากกว่าภาคอื่นๆ

จากรายละเอียดข้อมูลในรูปที่ 4.1 (ง) สามารถสรุปออกมาได้ว่างบประมาณสำหรับการเดินทางของนักท่องเที่ยวที่ใช้ในการเดินทางไปท่องเที่ยวยังภูมิภาคต่างๆ มีความแตกต่างกันอย่างมีนัยสำคัญ นักท่องเที่ยวที่มีงบประมาณระหว่าง 1,001 – 2,000 บาท มักเลือกภาคตะวันออกเป็นจุดหมายปลายทางในการเดินทาง ส่วนผู้ที่มีงบประมาณระหว่าง 2,001 – 3,000 บาท มักนิยมเดินทางไปท่องเที่ยวภาคกลางและภาคตะวันออกเฉียงเหนือ ในส่วนของนักท่องเที่ยวที่มีงบประมาณเกิน 7,000 บาท นิยมเดินทางไปยังภาคใต้และภาคเหนือตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 ความรู้ที่ได้จากแบบจำลอง

จากการนำความความรู้จากระบบการแนะนำสถานที่ท่องเที่ยวถูกนำเสนออยู่ในรูปแบบต้นไม้ตัดสินใจ โดยแสดงตัวอย่างของภาคกลางดังรูปที่ 4.2 เมื่อพิจารณาต้นไม้ตัดสินใจพบ 4 ปัจจัยที่มีอิทธิพลต่อการเลือกสถานที่ท่องเที่ยวในภาคกลาง ได้แก่ กิจกรรม เดือนที่เดินทางท่องเที่ยว อายุ และรายได้



รูปที่ 4.2 ตัวอย่างต้นไม้ตัดสินใจของภาคกลาง

เมื่อพิจารณาต้นไม้ตัดสินใจดังรูปที่ 4.2 จะพบสถานที่ท่องเที่ยวประเภทตลาดประจำท้องถิ่นที่มีชื่อเสียงหลักๆ 3 แห่งในภาคกลาง ได้แก่ ตลาดน้ำอัมพวา ตลาดน้ำดำเนินสะดวก และตลาดสามชุก โดยในแต่ละตลาดมีนักท่องเที่ยวที่มีลักษณะแตกต่างกัน เช่น นักท่องเที่ยวที่มาเยือนยังตลาดน้ำอัมพวามีพฤติกรรมที่เพลิดเพลินกับการเรียนรู้ประวัติศาสตร์ ในขณะเดียวกันนักท่องเที่ยวที่เลือกไปท่องเที่ยวยังตลาดน้ำดำเนินสะดวกมักจะเพลิดเพลินไปกับการซื้อสินค้าท้องถิ่นโดยส่วนใหญ่ ซึ่งความรู้ที่ได้จากต้นไม้ตัดสินใจนี้สามารถเชื่อถือได้เนื่องจากมีศาสนสถานและสถานที่ที่สำคัญทางประวัติศาสตร์หลายแห่งตั้งอยู่ใกล้บริเวณตลาดน้ำอัมพวา นอกจากนี้สังเกตได้ว่านักท่องเที่ยวในกลุ่มอายุต่างๆ มักจะเลือกท่องเที่ยวตลาดที่แตกต่างกันตามความชอบในสำหรับกลุ่มนักท่องเที่ยวที่นิยมไปเพื่อชื่นชมธรรมชาติ นักท่องเที่ยวกลุ่มวัยรุ่น (อายุต่ำว่า 20 ปี)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

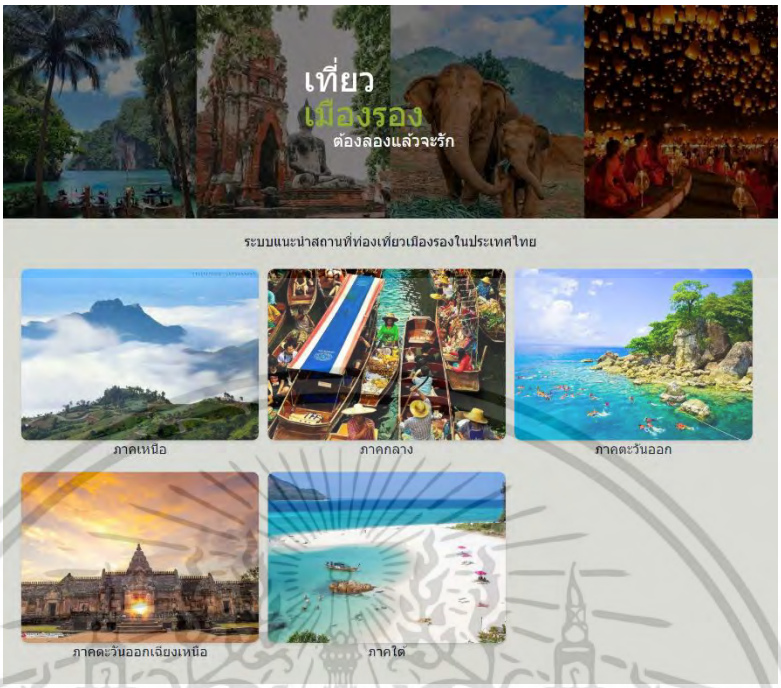
มักชอบเดินทางไปยังตลาดสามชุก ในขณะที่นักท่องเที่ยวกลุ่มวัย 31 – 40 ปี นักนิยมเดินทางไปยังตลาดน้ำอัมพวา และในทางกลับกันนักท่องเที่ยวที่มีอายุ 41 – 50 ปี มีความชื่นชอบไปยังตลาดน้ำดำเนินสะดวกมากกว่า และสำหรับลักษณะของนักท่องเที่ยวที่มาเยี่ยมชมตลาดน้ำอัมพวาพบว่าคือผู้ที่ชอบเรียนรู้วัฒนธรรมท้องถิ่น และมีรายได้อยู่ในช่วง 25,001 – 35,000 บาท

จากความรู้ที่ได้รับจากการวิเคราะห์สามารถนำไปใช้จัดการการท่องเที่ยวในท้องถิ่นของตลาดน้ำอัมพวาได้ เช่น ออกแบบแพ็คเกจโฮมสเตย์ในราคาประมาณ 10% ของรายได้ต่อเดือน อยู่ที่ราคาประมาณ 2,500 – 3,000 บาท ซึ่งแพ็คเกจเหล่านี้อาจรวมถึงกิจกรรมที่เน้นถึงการเรียนรู้วัฒนธรรมท้องถิ่น เช่น การเยี่ยมชมสวนมะพร้าว การเยี่ยมชมวิธีการผลิตน้ำตาลมะพร้าว ทัวร์ล่องเรือเพื่อสัมผัสชีวิตริมแม่น้ำประจำท้องถิ่น และการนั่งเรือไปตลาดน้ำ เป็นต้น โดยมีจุดประสงค์เพื่อดึงดูดนักท่องเที่ยวให้มีเพิ่มมากขึ้น และที่สำคัญยังมีส่วนช่วยในการสร้างรายได้ให้กับชุมชนท้องถิ่นอีกด้วย

4.2.3 การนำแบบจำลองไปพัฒนาระบบ

ในงานวิจัยนี้ได้พัฒนาระบบแนะนำสถานที่ท่องเที่ยวในรูปแบบของเว็บแอปพลิเคชัน โดยมีรูปแบบการใช้งานที่ง่ายต่อการทำความเข้าใจแสดงดังรูปที่ 4.3 โดยในขั้นตอนแรกผู้ใช้จำเป็นต้องเลือกภูมิภาคในประเทศไทยที่ต้องการรับคำแนะนำสำหรับการท่องเที่ยวส่วนบุคคลจากระบบ ดังแสดงในรูป 4.3 (ก) หลังจากนั้นผู้ใช้จะไปยังหน้าถัดไปแสดงดังรูป 4.3 (ข) ซึ่งผู้ใช้ต้องมีส่วนร่วมในการตอบคำถามที่เกี่ยวข้องกับการเดินทางที่ต้องการ จากนั้นระบบจะทำการนำข้อมูลที่ได้เข้าสู่แบบจำลองและแสดงผลการแนะนำสถานที่ท่องเที่ยวที่เหมาะสม 5 อันดับให้แก่ผู้ใช้ โดยจัดลำดับตามความน่าจะเป็นของสถานที่ท่องเที่ยวที่เกิดขึ้นจากการทำนายของแบบจำลองโดยจะเรียงลำดับจากมากไปน้อยซึ่งถูกแสดงในรูปที่ 4.3 (ค)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก) หน้าเลือกภูมิภาค

(ข) หน้ากรอกข้อมูล

รูปที่ 4.3 ตัวอย่างเว็บแอปพลิชันของโปรแกรมแนะนำสถานที่ท่องเที่ยวในภาคเหนือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(ค) หน้าแนะนำสถานที่ท่องท่อง 5 อันดับ

รูปที่ 4.3 (ต่อ) ตัวอย่างเว็บแอปพลิชันของโปรแกรมแนะนำสถานที่ท่องเที่ยวในภาคเหนือ

งานวิจัยนี้ได้สำรวจความพึงพอใจต่อการใช้เว็บแอปพลิเคชันสำหรับโปรแกรมแนะนำสถานที่ท่องเที่ยวในเมืองรองด้วยแบบสอบถามดังรูปที่ 4.4 โดยได้เก็บความพึงพอใจจากผู้ใช้จำนวน 37 คน โดยแบ่งเป็นเพศชายจำนวน 8 คน เพศหญิงจำนวน 26 คน และ LGBTQ จำนวน 3 คน ในส่วนของอายุมีสัดส่วนคือ น้อยกว่า 21 ปีจำนวน 1 คน 21 – 30 ปี จำนวน 12 คน 31 – 40 ปี จำนวน 14 คน 41 - 50 ปี จำนวน 3 คน 51 – 60 ปี จำนวน 6 คน และมากกว่า 60 ปี จำนวน 1 คน โดยให้ผู้ใช้ได้ใช้โปรแกรมจริงโดยเลือกภูมิภาคที่ต้องการคนละหนึ่งภูมิภาค และกรอกข้อมูลที่ต้องการ จากนั้นระบบจะแนะนำสถานที่ท่องเที่ยวรายบุคคลให้แก่ผู้ใช้ทั้ง 37 คน และให้ผู้ใช้ตอบความพึงพอใจต่อโปรแกรมแนะนำทั้งหมด 5 ด้าน ได้แก่ ด้านความพึงพอใจต่อคำแนะนำสถานที่ท่องเที่ยว ด้านความเหมาะสมของจำนวนข้อคำถาม ด้านความง่ายต่อการใช้โปรแกรม ด้านความเหมาะสมของรูปแบบการแสดงหน้าจอ และด้านความสวยงามของโปรแกรม จากผลการสำรวจความพึงพอใจพบว่าคะแนนจากทุกด้านอยู่ระหว่าง 4.6 ถึง 4.8 และคะแนนเฉลี่ยในทุกด้านของโปรแกรมเท่ากับ 4.7 (พอใจมากที่สุด) แสดงดังตารางที่ 4.5 ดังนั้นสามารถสรุปได้ว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รายการสถานที่ท่องเที่ยวที่แนะนำไปยังผู้ใช้งานในแต่ละบุคคลาสามารถสร้างความพึงพอใจได้มากที่สุด

แบบประเมินความพึงพอใจการใช้งานระบบแนะนำ
สถานที่ท่องเที่ยวเมืองรองในประเทศไทยรายบุคคล

คำชี้แจง แบบสอบถามชุดนี้จัดทำขึ้นเพื่อประเมินความพึงพอใจของผู้ใช้งานระบบแนะนำ ซึ่งข้อมูลที่ได้จะเป็นประโยชน์อย่างยิ่งต่อการพัฒนาและปรับปรุงการให้บริการให้มีความเหมาะสมต่อไป

กรุณาทำเครื่องหมาย ✓ ในช่องที่ท่านเลือก โดยคำถามมีทั้งหมด 2 ส่วน

ส่วนที่ 1 ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม

1. เพศ ☐ ชาย ☐ หญิง
2. อายุ ☐ ต่ำกว่า 21 ปี ☐ 21 – 30 ปี ☐ 31 – 40 ปี ☐ 41 – 50 ปี ☐ 51 – 60 ปี ☐ มากกว่า 60 ปี
3. ระดับการศึกษา
☐ ต่ำกว่า ป. ตรี ☐ ป. ตรี ☐ ป. โท ☐ สูงกว่า ป.โท
4. ภาคที่เลือกในโปรแกรม
☐ ภาคเหนือ ☐ ภาคใต้ ☐ ภาคตะวันออก ☐ ภาคตะวันออกเฉียงเหนือ ☐ ภาคกลาง

ส่วนที่ 2 ระดับความพึงพอใจในการใช้งานระบบแนะนำ โดยมีเกณฑ์วัดระดับดังนี้

5 = มากที่สุด 4 = มาก 3 = ปานกลาง 2 = น้อย 1 = น้อยที่สุด 0 = ไม่พึงพอใจ

ประเด็นวัดระดับความพึงพอใจ		ระดับความพึงพอใจ					ไม่พึงพอใจ
		5	4	3	2	1	0
1.	ความพึงพอใจต่อคำแนะนำสถานที่ท่องเที่ยว						
2.	จำนวนข้อคำถามมีความเหมาะสม						
3.	โปรแกรมมีความง่ายต่อการใช้งาน						
4.	การจัดรูปแบบการแสดงหน้าจอมีความเหมาะสม						
5.	ความสวยงามของโปรแกรม						

ข้อเสนอแนะ.....

.....

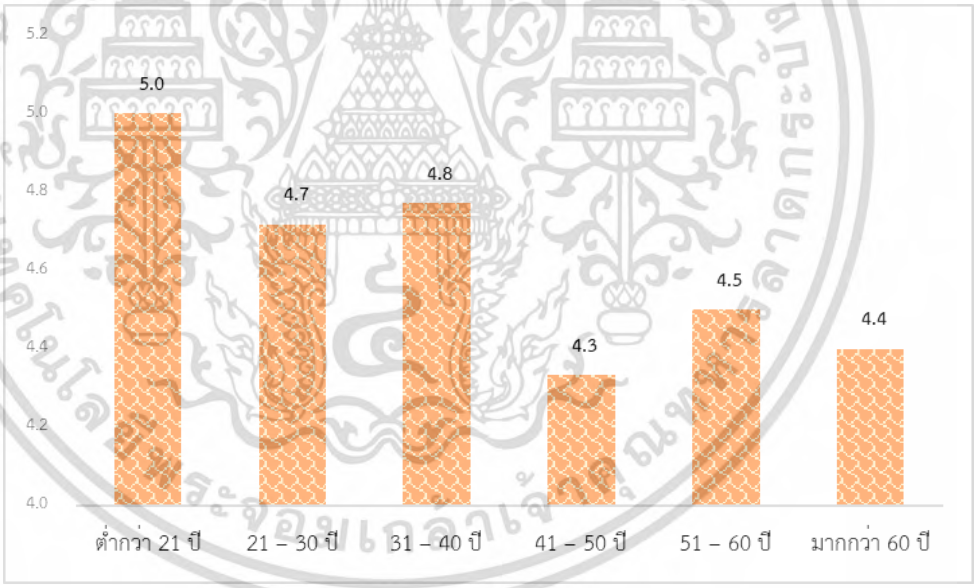
รูปที่ 4.4 แบบสอบถามที่ใช้ในการสำรวจความพึงพอใจต่อโปรแกรมแนะนำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 คะแนนความพึงพอใจต่อโปรแกรมแนะนำสถานที่ท่องเที่ยว

ข้อ	คำถาม	คะแนนเฉลี่ย
1	ความพึงพอใจต่อคำแนะนำสถานที่ท่องเที่ยว	4.6
2	จำนวนข้อคำถามมีความเหมาะสม	4.6
3	โปรแกรมมีความง่ายต่อการใช้งาน	4.8
4	การจัดรูปแบบการแสดงหน้าจอมีความเหมาะสม	4.8
5	ความสวยงามของโปรแกรม	4.6
	เฉลี่ยรวม	4.7

เมื่อพิจารณาความพึงพอใจของผู้ใช้ที่มีต่อโปรแกรมแนะนำตามช่วงอายุ พบว่าผู้ใช้ในช่วงอายุที่น้อยกว่าหรือเท่ากับ 40 ปี อยู่ในคะแนนเฉลี่ยมากกว่า 4.5 ซึ่งอยู่ในระดับที่พึงพอใจมากที่สุด ดังรูปที่ 4.5



รูปที่ 4.5 กราฟเปรียบเทียบความพึงพอใจตามช่วงอายุ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ในงานวิจัยนี้ใช้อัลกอริทึมการเรียนรู้ของเครื่องเพื่อสร้างระบบแนะนำสถานที่ท่องเที่ยว โดยมีขอบเขตของสถานที่ที่แนะนำคือเป็นสถานที่ท่องเที่ยวที่อยู่ในเมืองรองของประเทศไทย ในการรวบรวมข้อมูลทั้งหมดเป็นการรวบรวมผ่านแบบสอบถามออนไลน์ที่ดำเนินการผ่าน Google Form โดยประสิทธิภาพของแบบจำลองทำการประเมินประสิทธิภาพด้วยสองตัวชี้วัด ได้แก่ Hit Rate และ NDCG ซึ่งหากพิจารณาในห้าภูมิภาคของประเทศไทย ค่าเฉลี่ยของ Hit Rate และ NDCG อยู่ที่ 0.81 และ 0.61 ตามลำดับ นอกจากนี้ยังพบว่าการแนะนำสถานที่ท่องเที่ยวในภาคกลางมีประสิทธิภาพสูงสุดทั้งในด้าน Hit Rate และ NDCG เมื่อเทียบกับภาคอื่นๆ

จากการวิเคราะห์ปัจจัยที่มีอิทธิพลต่อการเลือกจุดหมายปลายทาง พบว่าปัจจัยสำคัญในแต่ละภูมิภาคมีความแตกต่างกันเล็กน้อยเมื่อจัดอันดับตามค่า p-value จากต่ำสุดไปสูงสุด อย่างไรก็ตามจากการวิเคราะห์ทั้งห้าภูมิภาค พบว่ามีปัจจัยบางอย่างที่ปรากฏในการจัดอันดับห้าอันดับแรกถึงสามภูมิภาค โดยปัจจัยเหล่านี้ได้แก่ กิจกรรม เดือนที่เดินทาง อายุ การศึกษา งบประมาณการเดินทาง และที่พัก นอกจากนี้ในงานวิจัยนี้ได้มีการวิเคราะห์พฤติกรรมการเดินทางของนักท่องเที่ยวโดยเน้นที่ปัจจัยเหล่านี้เพิ่มเติมเพื่อทำความเข้าใจถึงตัวแปรและปัจจัยที่มีอิทธิพลต่อการเลือกจุดหมายปลายทางของนักท่องเที่ยวในภูมิภาคต่างๆ ของประเทศไทย

ในงานวิจัยนี้ได้ประยุกต์ใช้การเรียนรู้ของเครื่องเพื่อสร้างระบบแนะนำสถานที่ท่องเที่ยวเมืองรองในประเทศไทย และระบบนี้ได้รับการพัฒนาเป็นเว็บแอปพลิเคชันให้ผู้ใช้สามารถป้อนข้อมูลพฤติกรรมการเดินทางที่ต้องการหรือความชอบของผู้ใช้เข้าสู่ระบบ จากนั้นระบบจะแสดงคำแนะนำสถานที่ท่องเที่ยวที่สอดคล้องกับข้อมูลที่ผู้ใช้ โดยเป็นการแสดงการแนะนำ 5 รายการ จากการสำรวจความพึงพอใจของผู้ใช้งานโปรแกรมพบว่ามีคะแนนความพึงพอใจมากที่สุดในทุกด้าน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 ข้อเสนอแนะ

จากการพัฒนาเว็บแอปพลิเคชันที่ได้กล่าวข้างต้น ในการทำงานในอนาคตมีแผนพัฒนาฟังก์ชันการทำงานของแอปพลิเคชันให้ก้าวหน้าต่อไป ตัวอย่างเช่น การสร้างระบบแชทบอทโดยใช้เทคนิคการประมวลผลภาษาธรรมชาติ (Natural language processing : NLP) เพื่อให้คำแนะนำและคำตอบเกี่ยวกับการท่องเที่ยวในเมืองรองของประเทศไทย นอกจากนี้อาจมีการปรับปรุงให้ระบบสามารถออกแบบแผนการเดินทางให้เหมาะสมกับไลฟ์สไตล์และความชอบส่วนบุคคล และเนื่องจากหน่วยตัวอย่างในงานวิจัยนี้อาจยังไม่ครอบคลุมพฤติกรรมการท่องเที่ยวของทุกช่วงวัย เช่น พฤติกรรมการท่องเที่ยวของเด็กที่อายุน้อยกว่า 15 ปี หรือพฤติกรรมการท่องเที่ยวของนักท่องเที่ยวที่มีช่วงอายุที่มากกว่า 70 ปี ส่งผลให้แบบจำลองมีสถานที่ท่องเที่ยวไม่ครอบคลุมสถานที่ที่เหมาะสมกับช่วงวัยเหล่านี้ ในอนาคตอาจต้องมีการเน้นเก็บข้อมูลของกลุ่มตัวอย่างเหล่านี้เพิ่มเติมเพื่อสร้างแบบจำลองให้เหมาะสมและสามารถใช้ได้กับทุกช่วงวัยอายุมากยิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

ไกรศักดิ์ เกษร, วนารัตน์ จุฬพันธ์ทอง และ อนงค์พร ไศลวรากุล. 2560. การวิเคราะห์การเคลื่อนย้ายของนักท่องเที่ยว. สำนักงานการวิจัยแห่งชาติ (วช.)

Carla Binucci, Felice De Luca, Emilio Di Giacomo, Giuseppe Liotta and Fabrizio Montecchiani. 2017. **Designing the Content Analyzer of a Travel Recommender System**. Expert Systems with Applications, vol. 87, pp. 199-208.

Filipe Santos, Ana Almeida, Constantino Martins, Ramiro Gonçalves and José Martins. 2019. **Using POI functionality and accessibility levels for delivering personalized tourism recommendations**. Computers, Environment and Urban Systems, vol. 77.

Swagato Chatterjee and Prasenjit Mandal. 2020. **Traveler preferences from online reviews: Role of travel goals, class and culture.** Tourism Management, vol. 80.

Merlinda Sumardi, Jufery, Frenky, Rini Wongso and Ferdinand Ariandy Luwinda. 2017. **TripBuddy" Travel Planner with Recommendation based on User's Browsing Behaviour**. Procedia Computer Science, vol. 116, pp. 326–333.

Zahra Abbasi-Moud, Hamed Vahdat-Nejad and Javad Sadri. 2021. **Tourism recommendation system based on semantic clustering and sentiment analysis.** Expert Systems with Applications, vol. 167.

Lin Lia, Kyung Young Leeb and Sung-Byung Yanga. 2019. **Exploring the effect of heuristic factors on the popularity of unsecured 'Best places to visit' recommendations in an online travel community.** Information Processing and Management, vol. 56, pp. 1391-1408.

Swapnil Shende, Eeshan Bhaduri, and Arkopal Kishore Goswami. 2023. **Analyzing changes in travel patterns due to Covid-19 using Twitter data in India.** Case Studies on Transport Policy, vol. 12.

Guixiang Zhu, Youquan Wang, Jie Cao, Zhan Bu, Shuxin Yang, Weichao Liang and Jingting Liu. 2020. **Neural Attentive Travel package Recommendation via exploiting long-term and short-term behaviors.** Knowledge-Based Systems, vol. 211.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

Shini Renjith, A. Sreekumar and M. Jathavedan. 2020. **An extensive study on the evolution of context-aware personalized travel recommender systems**. Information Processing and Management, vol. 57.

Subramaniyaswamy V, Vijayakumar V, Logesh Rc and Indragandhi V.2015. **Intelligent travel recommendation system by mining attributes from community contributed photos**. Procedia Computer Science, vol. 50, pp. 447-455.

Toly Chen and Yu-Cheng Wang. 2021. **A calibrated piecewise-linear FGM approach for travel destination recommendation during the COVID- 19 pandemic.** Applied Soft Computing, vol. 109.

Hsiu-Sen Chiang and Tien-Chi Huang. 2013. **User-adapted travel planning system for personalized schedule recommendation.** Information Fusion, vol. 21, pp. 3-17.

Balamurugan S., and Sankaran M. 2023. **Prediction and exploration of contributory factors related to fatigue driving among long-haul truck drivers traveling across India: A tree-based machine learning approach.** Journal of Transport & Health, vol. 32.

Maral Kolahkaj, Ali Harounabadi, Alireza Nikravanshalmani and Rahim Chinipardaz. 2022. **A hybrid context-aware approach for e-tourism package recommendation based on asymmetric similarity measurement and sequential pattern mining.** Electronic Commerce Research and Applications, vol. 42.

Jiangning Hea, Hongyan Liua and Hui Xiong. 2016. **SocoTraveler: Travel-package recommendations leveraging social influence of different relationship types.** Information & Management, vol. 53, pp. 934-950.

Joel P. Lucas, Nuno Luz , María N. Moreno , Ricardo Anacleto, Ana Almeida Figueiredo and Constantino Martins. 2013 **.A hybrid recommendation approach for a tourism system.** Expert Systems with Applications, vol. 40, pp. 3532-3550.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

Ningzhe Xu, Qifan Nie, Jun Liu, and Steven Jones. 2023. **Post-pandemic shared mobility and active travel in Alabama: A machine learning analysis of COVID-19 survey data.** Travel Behaviour and Society, vol. 32.

Yue Li and Endong Wang. 2024. **Construction of personalized recommendation model for educational video game resources based on knowledge graph**. Entertainment Computing, vol 50.

Chhotelal Kumar and Mukesh Kumar. 2024. **Session-based recommendations with sequential context using attention-driven LSTM**. Computers and Electrical Engineering, vol 115.

Siwen Zheng, Yu Liu and Zhenchao Ouyang. 2016. **A Machine Learning-Based Tourist Path Prediction**. 4th IEEE International Conference on Cloud Computing and Intelligence Systems, pp. 38–42.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ	นางสาวสุมิตรา กองเป็ง
วัน เดือน ปีเกิด	27 ตุลาคม 2536
ที่อยู่ปัจจุบัน	14/61 เดอะเบสท์ ลำลูกกาคลอง 6 หมู่ 15 ตำบลบึงคำพร้อย อำเภอลำลูกา จังหวัดปทุมธานี 12150
ประวัติการศึกษา	(2558) วิทยาศาสตรบัณฑิต สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (2566) วิทยาศาสตรมหาบัณฑิต สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ผลงานทางวิชาการ	S. Kongpeng and A. Haskunatai, “Tourist Destination Recommendation System based on Machine Learning”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้