

การประยุกต์ระบบจีพีเอส และการทำเหมืองข้อมูลเพื่อทำนาย และป้องกันการ
เกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด

นายเกรียงไกร เอี่ยมเลิศวงศ์

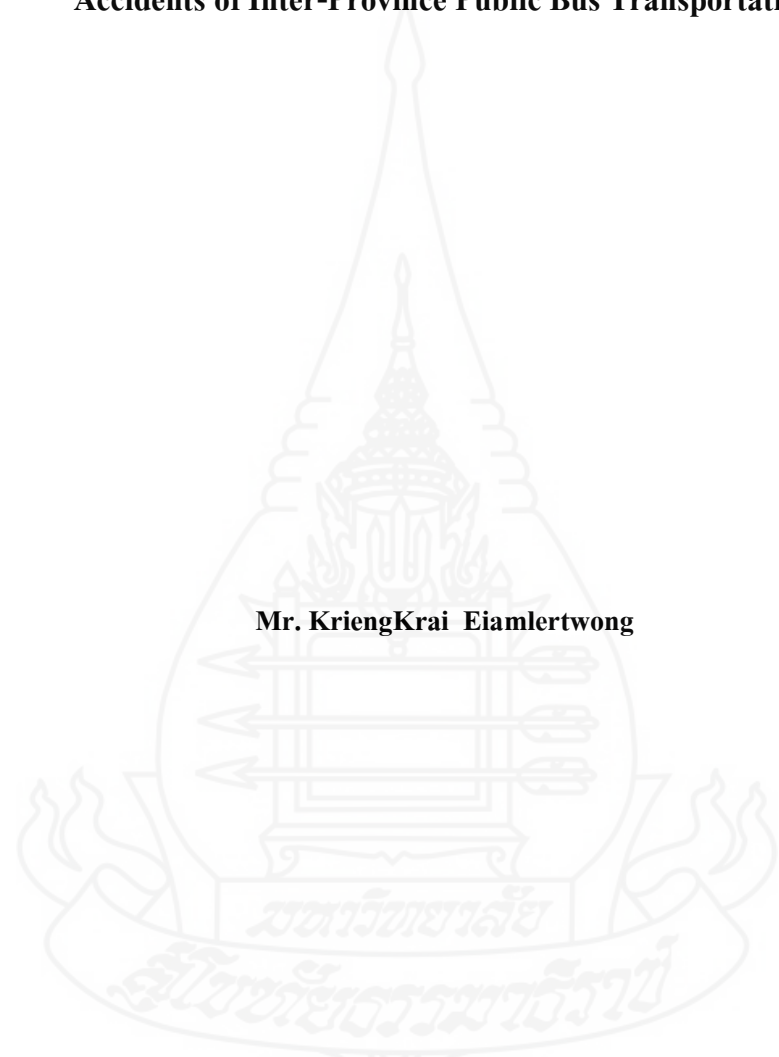


วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
แขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

พ.ศ. 2559

**The Application of GPS System and Data Mining to Predict and Prevent
Accidents of Inter-Province Public Bus Transportation**

Mr. Kriengkrai Eiamlertwong



A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of Master of Science in Information and Communication Technology

School of Science and Technology
Sukhothai Thammathirat Open University

2016

หัวข้อวิทยานิพนธ์	การประยุกต์ระบบจีพีเอส และการทำเหมืองข้อมูลเพื่อทำนาย และ ป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด
ชื่อและนามสกุล	นายเกรียงไกร เอี่ยมเลิศวงศ์
แขนงวิชา	เทคโนโลยีสารสนเทศและการสื่อสาร
สาขาวิชา	วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
อาจารย์ที่ปรึกษา	1. รองศาสตราจารย์ ดร. วิภา เจริญกัณธารักษ์ 2. อาจารย์ ดร. ดวงดาว วิชาดากุล

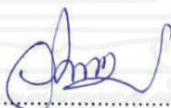
วิทยานิพนธ์นี้ได้รับความเห็นชอบให้เป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรระดับปริญญาโท เมื่อวันที่ 21 ตุลาคม 2559

คณะกรรมการสอบวิทยานิพนธ์



..... ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. ประจวบ วานิชชัชวาล)



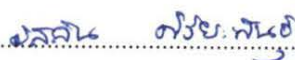
..... กรรมการ

(รองศาสตราจารย์ ดร. วิภา เจริญกัณธารักษ์)



..... กรรมการ

(อาจารย์ ดร. ดวงดาว วิชาดากุล)



..... ประธานกรรมการบัณฑิตศึกษา

(รองศาสตราจารย์ รตนา ศิริยะพันธ์)

ชื่อวิทยานิพนธ์ การประยุกต์ระบบจีพีเอส และการทำเหมืองข้อมูลเพื่อทำนาย และ

ป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด

ผู้วิจัย นายเกรียงไกร เอี่ยมเลิศวงศ์ **รหัสนักศึกษา** 2569600139

ปริญญา วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีสารสนเทศและการสื่อสาร)

อาจารย์ที่ปรึกษา (1) รองศาสตราจารย์ ดร. วิภา เจริญภักดิ์ (2) อาจารย์ ดร. ดวงดาว วิชาดากุล
ปีการศึกษา 2559

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์ 1) สร้างระบบจัดการข้อมูลจีพีเอส โดยใช้เทคโนโลยีของบิกดาตา 2) พัฒนาแบบจำลองเพื่อทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะด้วยการทำเหมืองข้อมูล 3) สร้างโปรแกรมคัดกรองพนักงานขับรถโดยสารสาธารณะ

วิธีดำเนินการวิจัย ประกอบด้วย 1) สร้างระบบจัดการข้อมูลจีพีเอส จากข้อมูลติดตามรถโดยสารสาธารณะระบบจีพีเอส ระหว่าง พ.ศ. 2555-2559 จำนวน 121,226,674 ระเบียบ และข้อมูลการเกิดอุบัติเหตุรถโดยสารสาธารณะ ระหว่าง พ.ศ. 2556-2559 จำนวน 406 ระเบียบ โดยการแบ่งชุดข้อมูล 2 แบบ คือ ข้อมูลสถิติการเกิดอุบัติเหตุรถโดยสารสาธารณะ และข้อมูลบุคคลเครื่องมือที่ใช้ คือ มงโกตีบี รุ่น 2.6.7 2) สร้างแบบจำลองเพื่อทำนายการเกิดอุบัติเหตุ ใช้วิธีแบ่งข้อมูลทดสอบ ด้วยอัลกอริทึมเคมีนส์ เพื่อจัดกลุ่มเสี่ยง และอัลกอริทึมอะไพโอริ เพื่อสร้างกลุ่มฝ้าระวัง และต้นไม้ตัดสินใจ เพื่อจำแนกองค์ประกอบอุบัติเหตุของกลุ่มเสี่ยง เครื่องมือที่ใช้คือ เคนไนม์ รุ่น 2.12 3) สร้างโปรแกรมคัดกรองพนักงานขับรถเพื่อป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะ เครื่องมือที่ใช้ คือ ไมโครซอฟท์ วิซวล ฟอกซ์ โปร รุ่น 9.0

ผลการวิจัยพบว่า 1) คลังข้อมูลระบบจีพีเอส มีประสิทธิภาพสามารถรองรับข้อมูลบิกดาตา การประมวลผลข้อมูลเชิงวิเคราะห์ในรูปแบบรายงานมีความรวดเร็ว 2) สร้างแบบจำลองเพื่อทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ พบว่าการสร้างกลุ่มเสี่ยงด้วยวิธีแบ่งข้อมูลทดสอบได้ผลลัพธ์ของ ข้อมูลสร้างแบบจำลอง และข้อมูลทดสอบ เหมือนกัน และการหาความสัมพันธ์เพื่อสร้างกลุ่มฝ้าระวัง จากการทดสอบประสิทธิภาพได้ค่าสหสัมพันธ์ มากกว่าหนึ่งแสดงว่าภูมิความสัมพันธ์มีความน่าเชื่อถือสำหรับสร้างแบบจำลองทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ เพื่อประกอบการตัดสินใจ และคัดกรองพนักงานขับรถ เพื่อป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะ

คำสำคัญ จีพีเอส เหมืองข้อมูล รถโดยสาร อุบัติเหตุ

Thesis title: The Application of GPS System and Data Mining to Predict and Prevent Accidents of Inter-Province Public Bus Transportation

Researcher: Mr. Kriengkrai Eiamlertwong; **ID:** 2569600139;

Degree: Master of Science (Information and Communication Technology);

Thesis advisor: (1) Dr. Vipaporn Jaroenpuntaruk, Associate Professor;

(2) Dr. Duangdao Wichadakul; **Academic year:** 2016

Abstract

The objectives of this research were 1) To create GPS repository using Big Data Technology; 2) To propose the prediction model of inter-province public bus transportation accidents using data mining; and 3) To develop the application based on the model to screen bus drivers of public bus transportation. The research methodology consisted of 1) Generating the GPS repository: data of the research was obtained from the GPS vehicle tracking system during 2012-2016 with 121,226,674 records and public bus accident data during 2013-2016 with 406 records. The data consisted of bus accident statistics and personal data. Tool for this step was MongoDB v2.6.7; 2) Generating the model prediction used Split Test and KMeans algorithm for Risk Group Clustering, Apriori algorithm for Surveillance Group Clustering; and Decision Tree for Accident Risk Classification. Data mining tool was KNIME v2.12; and 3) Developing program based on the model to screen the bus driver to prevent the public bus transportation accidents: tool for this process was Microsoft Visual FoxPro v9.0.

The research finding showed that 1) The GPS data warehouse could effectively handle large data and the analytical data processing in report format was efficiently; and 2) The prediction model was based on Split Test and surveillance data with training data set and testing data set. The association rules of surveillance data with the lift result was more than one approved that the prediction model was precise and appropriate. Moreover, the prediction model of inter province public bus accidents was applied to screen bus drivers for accident prevention of public bus inter province transportation.

Keywords: GPS, Data Mining, Public Bus, Accident

กิตติกรรมประกาศ

การทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณา และความอนุเคราะห์จาก รองศาสตราจารย์ ดร.วิภา เจริญภัณฑารักษ์ อาจารย์ที่ปรึกษาการทำวิทยานิพนธ์ ที่ได้ให้ความกรุณาให้คำแนะนำ และติดตามการทำวิทยานิพนธ์ครั้งนี้อย่างใกล้ชิดตลอดมา นับตั้งแต่เริ่มต้น จนกระทั่งสำเร็จเรียบร้อยสมบูรณ์ ผู้วิจัยรู้สึกซาบซึ้งในความกรุณาของท่านเป็นอย่างยิ่ง

ขอขอบพระคุณ กรรมการบริษัทฯ เคนรตสายตะวันออกเฉียงเหนือ ขอสงวนนามบริษัทฯ เพื่อให้ไม่มีผลกระทบต่อการใช้ข้อมูลในการวิจัย และอนุญาตนำข้อมูลโปรแกรมคอมพิวเตอร์ และขอขอบคุณเจ้าหน้าที่แผนกบัญชี แผนกจีพีเอส แผนกอุบัติเหตุ แผนกซ่อมบำรุง บริษัทฯ ทุกท่านที่เสียสละเวลาและให้ความร่วมมือในการเก็บรวบรวมข้อมูลภาคสนาม

ขอขอบพระคุณเจ้าของเอกสารบทความ หนังสือทุกท่านที่ผู้วิจัยใช้ในการสืบค้นข้อมูล ที่ไม่ได้กล่าวนามไว้ ณ ที่นี้

นอกจากนี้ผู้วิจัยขอขอบพระคุณคณาจารย์ สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช เพื่อนักศึกษาและผู้ที่มีส่วนเกี่ยวข้องในการทำวิทยานิพนธ์ครั้งนี้ทุกท่านที่ได้กรุณาให้การสนับสนุนช่วยเหลือ และให้กำลังใจตลอดมา

เกรียงไกร เอี่ยมเลิศวงศ์

ตุลาคม 2559



สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญตาราง.....	ฅ
สารบัญภาพ.....	ฉ
บทที่ 1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
วัตถุประสงค์การวิจัย.....	2
ขอบเขตของการวิจัย.....	2
วิธีดำเนินการวิจัย.....	3
ประโยชน์ที่คาดว่าจะได้รับ.....	4
นิยามศัพท์เฉพาะ.....	5
บทที่ 2 วรรณกรรมที่เกี่ยวข้อง.....	8
โครงสร้างสถาปัตยกรรมข้อมูลขนาดใหญ่ (Big data Architecture).....	8
ฐานข้อมูล โนเอสคิวแอล(NoSQL Database).....	11
การทำเหมืองข้อมูล(Data Mining).....	14
ระบบการติดตามรถยนต์ GPS.....	20
การพัฒนาระบบสารสนเทศ.....	24
งานวิจัยที่เกี่ยวข้อง.....	26
บทที่ 3 วิธีดำเนินการวิจัย.....	31
ประชากรและกลุ่มตัวอย่าง.....	31
เครื่องมือที่ใช้ในการดำเนินวิจัย.....	32
ขั้นตอนการดำเนินงาน.....	33

สารบัญ (ต่อ)

	หน้า
บทที่ 4 การสร้างแบบจำลองและพัฒนาระบบ.....	41
สร้างระบบจัดการข้อมูล GPS.....	41
สร้างแบบจำลองการทำนาย.....	58
สร้างโปรแกรมใช้งาน.....	95
ประเมินโปรแกรมใช้งาน.....	109
บทที่ 5 สรุปการวิจัย และข้อเสนอแนะ.....	116
สรุปผลการวิจัย.....	116
การอภิปรายผล.....	120
ปัญหาและอุปสรรค.....	122
ข้อเสนอแนะ.....	123
บรรณานุกรม.....	124
ภาคผนวก.....	127
ก แบบประเมินความพึงพอใจของผู้ใช้ระบบ.....	128
ข แบบประเมินความเชื่อมั่นคำทำนายระบบ.....	132
ค คู่มือการติดตั้งโปรแกรม.....	135
ง คู่มือการใช้โปรแกรมทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ.....	143
ประวัติผู้วิจัย.....	147

สารบัญตาราง

	หน้า
ตารางที่ 2.1	25
ตารางที่ 4.1	44
ตารางที่ 4.2	45
ตารางที่ 4.3	45
ตารางที่ 4.4	45
ตารางที่ 4.5	46
ตารางที่ 4.6	46
ตารางที่ 4.7	46
ตารางที่ 4.8	48
ตารางที่ 4.9	50
ตารางที่ 4.10	51
ตารางที่ 4.11	51
ตารางที่ 4.12	52
ตารางที่ 4.13	53
ตารางที่ 4.14	53
ตารางที่ 4.15	54
ตารางที่ 4.16	54
ตารางที่ 4.17	61
ตารางที่ 4.18	62
ตารางที่ 4.19	63
ตารางที่ 4.20	76
ตารางที่ 4.21	80
ตารางที่ 4.22	82
ตารางที่ 4.23	84
ตารางที่ 4.24	86
ตารางที่ 4.25	90

สารบัญตาราง(ต่อ)

	หน้า
ตารางที่ 4.26 สรุปผลการทดสอบของกลุ่มเสี่ยงและการเกิดอุบัติเหตุ.....	92
ตารางที่ 4.27 การประเมินกลุ่มเฝ้าระวัง.....	93
ตารางที่ 4.28 สรุปผลการทดสอบกลุ่มเฝ้าระวัง.....	94
ตารางที่ 4.29 VPANALYTICS สำหรับเก็บข้อมูลบันทึกผลการทำนายพนักงานขับรถ.....	96
ตารางที่ 4.30 AC_SECSION สำหรับเก็บข้อมูลภูมิภาค.....	97
ตารางที่ 4.31 AC_EDUCATE สำหรับเก็บข้อมูลการศึกษา.....	97
ตารางที่ 4.32 DRIVER สำหรับเก็บข้อมูลประวัติพนักงานขับรถ.....	97
ตารางที่ 4.33 AC_PROVINCE สำหรับเก็บข้อมูลชื่อจังหวัด.....	98
ตารางที่ 4.34 VP_ACD สำหรับเก็บข้อมูลบันทึกช่วงเวลา.....	99
ตารางที่ 4.35 แสดงผลการประเมินความพึงพอใจการใช้งานระบบ.....	110
ตารางที่ 4.36 เกณฑ์การกำหนดระดับความพึงพอใจ.....	111
ตารางที่ 4.37 ผลการวิเคราะห์ข้อมูลจากการตอบแบบสอบถามเพื่อประเมินความพึงพอใจ.....	111
ตารางที่ 4.38 แสดงผลการประเมินความเชื่อมั่นของระบบการทำนาย.....	113
ตารางที่ 4.39 ผลการวิเคราะห์ข้อมูลจากการประเมินความเชื่อมั่นของระบบการทำนาย.....	114
ตารางที่ 5.1 สรุปผลการทดสอบของกลุ่มเสี่ยง.....	118
ตารางที่ 5.2 คุณลักษณะบุคคลกลุ่มเฝ้าระวัง.....	119

สารบัญภาพ

	หน้า
ภาพที่ 2.1 องค์ประกอบข้อมูลขนาดใหญ่ Big Data 5 V.....	9
ภาพที่ 2.2 รูปแบบผลิตภัณฑ์ต่างๆของ NoSQL Database.....	10
ภาพที่ 2.3 ตัวอย่างการจัดเก็บข้อมูลแบบ Key-Value Data Model.....	12
ภาพที่ 2.4 ตัวอย่างการจัดเก็บข้อมูลแบบ Document Data Model.....	12
ภาพที่ 2.5 ตัวอย่างการจัดเก็บข้อมูลแบบ Document Data Model.....	13
ภาพที่ 2.6 แสดงถึงข้อมูลแบบ Graph Data Model.....	14
ภาพที่ 2.7 ขั้นตอนมาตรฐาน คริสป์-ดีเอ็ม CRISP-DM.....	15
ภาพที่ 2.8 การค้นหารูปแบบที่เกิดควบคู่ในรายการ.....	17
ภาพที่ 2.9 ความสัมพันธ์ตะกร้าสินค้า.....	17
ภาพที่ 2.10 การจำแนกข้อมูลด้วยต้นไม้ตัดสินใจ.....	18
ภาพที่ 2.11 การจัดกลุ่มด้วย Clustering.....	19
ภาพที่ 2.12 โครงสร้างวงโคจรของดาวเทียม 24 ดวง.....	21
ภาพที่ 2.13 ตำแหน่งที่ตั้งของสถานีดาวเทียม.....	21
ภาพที่ 2.14 การทำงานของ GPS.....	22
ภาพที่ 2.15 โครงสร้างพื้นฐานของระบบการติดตามรถโดยสารด้วย GPS.....	22
ภาพที่ 2.16 ตัวอย่างข้อมูลจาก GPS ของรถ.....	23
ภาพที่ 2.17 ขั้นตอนวงจรการพัฒนาาระบบ SDLC.....	24
ภาพที่ 3.1 ขั้นตอนแสดงขั้นตอนการพัฒนาลังข้อมูล.....	34
ภาพที่ 3.2 กระบวนการสร้างตัวแบบทำนายการใช้อัตราความเร็วของรถโดยสารสาธารณะ.....	36
ภาพที่ 3.3 กระบวนการสร้างตัวแบบทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ.....	37
ภาพที่ 4.1 แสดงความสัมพันธ์ข้อมูลเพื่อสร้างคลังข้อมูลไม่มีโครงสร้าง.....	47
ภาพที่ 4.2 แสดงความสัมพันธ์ข้อมูลอุบัติเหตุรถโดยสาร.....	55
ภาพที่ 4.3 ขั้นตอนการพัฒนาลังข้อมูล GPS.....	55

สารบัญภาพ(ต่อ)

	หน้า
ภาพที่ 4.29 โปรแกรม Microsoft Visual FoxPro v9.0.....	100
ภาพที่ 4.30 แผนผังโปรแกรม.....	100
ภาพที่ 4.31 หน้าจอเข้าระบบงาน.....	101
ภาพที่ 4.32 หน้าจอหลัก.....	102
ภาพที่ 4.33 หน้าจอหัวข้อการทำนายความเสี่ยงโดยกำหนดคุณสมบัติส่วนบุคคล.....	103
ภาพที่ 4.34 หน้าจอการทำนายความเสี่ยงโดยกำหนดองค์ประกอบการเดินรถ.....	104
ภาพที่ 4.35 ออกแบบหน้าจอการทำนายความเสี่ยงโดยกำหนดคุณสมบัติส่วนบุคคล.....	105
ภาพที่ 4.36 การแสดงผลการทำงานการทำนายความเสี่ยงโดยกำหนด คุณสมบัติส่วนบุคคล.....	105
ภาพที่ 4.37 รายงานผลการทำนายความเสี่ยงโดยกำหนดคุณสมบัติส่วนบุคคล.....	106
ภาพที่ 4.38 ออกแบบหน้าจอการทำนายความเสี่ยงโดยกำหนดองค์ ประกอบการเดินรถ.....	107
ภาพที่ 4.39 แสดงผลการทำงานการทำนายความเสี่ยงโดย กำหนดองค์ประกอบการเดินรถ.....	108

บทที่ 1

บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

การเกิดอุบัติเหตุจากรถโดยสารสาธารณะ ที่มีการนำเสนอข่าวจากสื่อพิมพ์ต่างๆ หรือจากสถิติอุบัติเหตุรถโดยสารของกรมการขนส่งทางบก พบว่าสาเหตุส่วนใหญ่มาจาก พนักงานขับรถ รถโดยสาร และถนน รวมทั้งสิ่งแวดล้อมอื่นๆ ซึ่งปัจจัยหลักมาจากพนักงานขับรถ เช่น การขับรถเร็วเกินกว่าที่กฎหมายกำหนด หรือปล่อยให้พนักงานขับรถเกินเวลาที่กำหนดในมาตรฐานความปลอดภัย ดังนั้นกรมการขนส่งทางบกกำหนดให้ผู้ประกอบการเดินรถโดยสารติดตั้ง GPS รถโดยสารสาธารณะ เพื่อติดตาม และควบคุมพฤติกรรมรถของพนักงานขับรถด้วยเทคโนโลยี GPS แบบเรียลไทม์ เช่น ข้อมูลการใช้ความเร็ว ชั่วโมงการขับขี่ และตำแหน่งพิกัดของรถ โดยเชื่อมโยงข้อมูลเข้าสู่กรมการขนส่งทางบก ปัญหาที่ผู้ประกอบการเดินรถโดยสารพบเจอ คือข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS มีข้อมูลมากขึ้น การทำงานเพื่อประมวลผลข้อมูลทำได้ช้าเมื่อมีปริมาณข้อมูลเพิ่มขึ้น ซึ่งสาเหตุมาจากระบบ GPS ไม่ได้ถูกออกแบบเพื่อรองรับข้อมูลขนาดใหญ่ ทำให้ผู้ให้บริการระบบ GPS ต้องปรับลดขนาดข้อมูล ส่งผลให้ข้อมูลขาดความต่อเนื่องในการนำข้อมูลมาทำรายงาน หรือการประมวลผลข้อมูลเชิงวิเคราะห์ เพื่อนำข้อมูลไปประกอบการวางแผนและการตัดสินใจ

ผู้วิจัยเล็งเห็นถึงความสำคัญของการประยุกต์ระบบ GPS เพื่อพัฒนาลงข้อมูลระบบ GPS ตามหลักการ Big Data และ ปัญหาการเกิดอุบัติเหตุรถโดยสารสาธารณะ จึงนำเทคนิคการทำเหมืองข้อมูลมาประยุกต์ใช้เพื่อการทำนายการเกิดอุบัติเหตุรถโดยสารเพื่อใช้เป็นเครื่องมือสนับสนุนในการพิจารณาและคัดกรองบุคคลมาทำหน้าที่การขับรถโดยสารสาธารณะ

2. วัตถุประสงค์การวิจัย

2.1 สร้างระบบการจัดการข้อมูลระบบ GPS โดยใช้เทคโนโลยีของ Big Data เพื่อใช้ในการวิเคราะห์พฤติกรรมการขับรถของพนักงานขับรถและการเกิดอุบัติเหตุของรถโดยสารสาธารณะ

2.2 สร้างแบบจำลอง เพื่อทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ โดยใช้หลักการทำเหมืองข้อมูล

2.3 สร้างโปรแกรมคัดกรองพนักงานขับรถ สำหรับบุคคลที่จะทำหน้าที่ขับรถโดยสาร เพื่อป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะ

3. ขอบเขตของการวิจัย

การวิจัยเพื่อประยุกต์ระบบ GPS และการทำเหมืองข้อมูลเพื่อทำนายและป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด ซึ่งงานวิจัยนี้สามารถที่จะใช้งานร่วมกันของหลายหน่วยงาน โดยมีขอบเขตดังนี้

3.1 ขอบเขตด้านผู้ใช้งาน

- 3.1.1 ผู้บริหาร
- 3.1.2 เจ้าหน้าที่บริหารการเดินรถ
- 3.1.3 เจ้าหน้าที่อุบัติเหตุ
- 3.1.4 เจ้าหน้าที่บุคคล
- 3.1.5 ผู้ดูแลระบบ

3.2 ขอบเขตด้านระบบงาน

3.2.1 การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลเพื่อใช้ในงานวิจัยนี้ ได้รับความอนุเคราะห์จากบริษัทเดินรถโดยสารระหว่างจังหวัด เพื่อใช้สำหรับงานวิจัย ได้แก่ ข้อมูลการติดตามรถโดยสารสาธารณะระบบ GPS ระหว่างพ.ศ. 2555-2559 จำนวนทั้งสิ้น 121,226,674 ระเบียบ ขนาดข้อมูล 58,009.27 ล้านไบต์ เป็นฐานข้อมูลเชิงโครงสร้างโดยทำการรวบรวมเป็นฐานข้อมูลขนาดใหญ่ และข้อมูลการเกิดอุบัติเหตุรถโดยสารสาธารณะระหว่างพ.ศ. 2556 - 2559 จำนวนทั้งสิ้น 406 ระเบียบ

3.2.2 เครื่องมือสนับสนุนการจัดการข้อมูลโดยใช้เทคโนโลยี Big Data ได้แก่

โปรแกรม MongoDB v2.6.7 มีคุณสมบัติในการจัดการข้อมูลจำนวนมากได้
 อย่างดี มีประสิทธิภาพด้านการประมวลผลการทำงานรวดเร็ว เช่น บันทึก ปรับปรุง ลบข้อมูล
 โครงสร้างข้อมูลมีความยืดหยุ่นไม่ต้องประกาศโครงสร้างในการจัดเก็บข้อมูลก่อน สามารถรองรับ
 ในส่วนขยายพื้นที่จัดเก็บข้อมูล และเป็นซอฟต์แวร์ฐานข้อมูลฟรีแบบโอเพนซอร์ส (Open Source)

โปรแกรม KNIME v2.12 เป็นเครื่องมือที่ถูกนำมาใช้ในด้านวิเคราะห์
 ข้อมูล และการทำเหมืองข้อมูล มีอินเตอร์เฟซที่เป็นปลั๊กอินเพื่อติดต่อข้อมูลหลายรูปแบบ รวมถึง
 ฐานข้อมูลของ MongoDB ด้วย รูปแบบการทำงานซอฟต์แวร์เป็นแบบ Data Flow ที่แสดงขั้นตอน
 การทำงานของข้อมูล เป็นซอฟต์แวร์โอเพนซอร์ส (Open Source) ให้ใช้ฟรีโดยไม่มีค่าใช้จ่าย

โปรแกรม Microsoft Visual FoxPro v9.0 สามารถสร้างแอปพลิเคชัน
 สำหรับการใช้งาน สร้างรูปแบบฟอร์มการรับข้อมูล สร้างรายงาน และมีฐานข้อมูลเพื่อรองรับการ
 บันทึกข้อมูล

4. วิธีดำเนินการวิจัย

4.1 สร้างระบบจัดการข้อมูล GPS จากข้อมูลติดตามรถโดยสารสาธารณะ ซึ่งเป็น
 ข้อมูลมีโครงสร้างเชิงสัมพันธ์ (Relational Database) เปลี่ยนข้อมูลไม่มีโครงสร้าง (Unstructured
 Data) โดยใช้มองโกดีบี (MongoDB) รองรับข้อมูลระบบ GPS และนำโปรแกรม KNIME v2.12
 สร้างตัวแบบจำลองการใช้ความเร็วรถโดยสารและวิเคราะห์การใช้อัตราความเร็วของรถโดยสาร

4.2 สร้างแบบจำลองทำนาย โดยใช้กระบวนการสร้างตัวแบบตามขั้นตอน CRISP-
 DM โดยนำเทคนิคการจัดแบ่งกลุ่ม (Clustering) เพื่อแบ่งกลุ่มพนักงานขับรถ เทคนิคแผนภูมิต้นไม้
 ตัดสินใจ (Decision Tree) เพื่อหาค่าประกอบที่เป็นปัจจัยก่อให้เกิดอุบัติเหตุของแต่ละกลุ่ม และ
 การใช้กฎความสัมพันธ์ (Association Rule) ด้วยวิธี Apriori หาความสัมพันธ์การเกิดอุบัติเหตุกับ
 คุณลักษณะเฉพาะบุคคลที่ไม่ถูกจัดอยู่ในกลุ่ม เพื่อลดความเสี่ยงการเกิดอุบัติเหตุอีกวิธีหนึ่ง

4.3 สร้างโปรแกรมการใช้งาน ด้วยโปรแกรม Microsoft Visual FoxPro v9.0 เพื่อ
 ทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ สำหรับคัดกรองบุคคลในตำแหน่งพนักงานขับรถ

4.4 เครื่องมือที่ใช้ในงานวิจัย มีดังนี้

4.4.1 เครื่องคอมพิวเตอร์

- หน่วยความจำ 32 GB (Gigabyte)
- HardDisk 1 - 4 TB (Terabyte)

4.4.2 ระบบปฏิบัติการ Windows 8.1 64 bit

4.4.3 โปรแกรม *KNIME v2.12.0 (64 bits)*

4.4.4 โปรแกรม *MongoDB v2.6.7*

4.5.5 โปรแกรม *Microsoft Visual FoxPro v9.0*

5. ประโยชน์ที่คาดว่าจะได้รับ

การประยุกต์ระบบ GPS และการทำเหมืองข้อมูลเพื่อทำนายและป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด ประโยชน์ที่คาดว่าจะได้รับ มีดังนี้

5.1 การพัฒนาระบบจัดการข้อมูล GPS เป็นฐานข้อมูลไม่มีโครงสร้าง (Unstructured Data) ด้วยโปรแกรม MongoDB v2.6.7 แทนฐานข้อมูลเชิงสัมพันธ์ (Relational Database) เดิม มีขีดจำกัดของด้านขนาดข้อมูล การประมวลผลที่ช้าเมื่อขนาดใหญ่ขึ้น ซึ่งแตกต่างกันมากกับระบบข้อมูล GPS ที่ใช้ฐานข้อมูลไม่มีโครงสร้างด้วยโปรแกรม MongoDB v2.6.7 สามารถรองรับข้อมูลจำนวนมากได้อย่างดี มีประสิทธิภาพด้านการประมวลผลการทำงานรวดเร็ว เช่น บันทึก ปรับปรุง ลบข้อมูลโครงสร้างข้อมูลมีความยืดหยุ่นสูงของข้อมูล

5.2 สร้างแบบจำลองที่สามารถทำนายผลลัพธ์ การเกิดอุบัติเหตุรถโดยสารสาธารณะ และสามารถคัดกรองพนักงานขับรถ และจำแนกบุคคลในกลุ่มเสี่ยง หรือกลุ่มเฝ้าระวัง

5.3 สร้างโปรแกรม แอปพลิเคชัน เพื่อคัดกรองพนักงานขับรถ ป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะ

6. นิยามศัพท์เฉพาะ

6.1 Big Data คือข้อมูลปริมาณมาก มหาศาล ในระดับ TeraByte (TB) หรือ ระดับ Peta Byte (PB) ซึ่งก็คือ 1000 ขนาด TeraByte และ Petabyte ซึ่งมีขนาดเท่ากับ 10^{15} เท่าของความจุ 1 byte หรือ 1.5 ล้าน Gigabyte ที่มีคุณสมบัติของข้อมูล เช่น ปริมาณ (Volume) ความหลากหลายของชนิด (Variety) ข้อมูลความเร็วของข้อมูล (Velocity)

6.2 ฐานข้อมูล RDBMS ระบบจัดการฐานข้อมูลเชิงสัมพันธ์ (Relational DBMS; RDBMS) ข้อมูลในฐานข้อมูล และถูกเก็บใน โครงสร้างแบบตารางเรียกว่า ตารางความสัมพันธ์ (Relations) ซึ่งมีโครงสร้างคล้ายโครงสร้างข้อมูลในแฟ้มข้อมูลทั่วไป แต่มีความสามารถในการเลือกข้อมูลจากตารางหลายตารางฯ เข้ามาใช้งานร่วมกันได้โดยง่าย ทำให้มีผู้เปรียบเทียบตารางความสัมพันธ์ว่าเหมือนกับแฟ้มข้อมูล

6.3 จีพีเอส (GPS) ระบบกำหนดตำแหน่งบนพื้นโลกผ่านดาวเทียม (ย่อมาจาก Global Positioning System) โดยพิกัดบนพื้นโลกที่ได้ จะมาจากการคำนวณสัญญาณนาฬิกาที่ส่งจากดาวเทียม มาที่เครื่องรับสัญญาณ GPS ส่วนดาวเทียม GPS ที่สามารถให้ระบุตำแหน่งได้นั้น จะถูกออกแบบมาโดยเฉพาะให้โคจรรอบโลก เพื่อส่งข้อมูลที่จะนำไปใช้คำนวณพิกัดออกมาตลอดเวลา

6.4 จีพีเอส แทรคกิ้ง (GPS TRACKING) เป็นระบบติดตาม รถ คน สัตว์ สิ่งของ แสดงข้อมูลการติดตามตลอดเวลา หรือที่เรียกว่า REAL TIME ประกอบด้วย หน่วยรับข้อมูลจากดาวเทียม GPS ประมวลผลตำแหน่งที่อยู่ ส่งข้อมูลไปยัง SERVER ของผู้ให้บริการ และแสดงผลผ่านเว็บเบราว์เซอร์ ตลอดเวลา สามารถนำไปประยุกต์ใช้งานได้หลายวัตถุประสงค์ เช่น ตรวจสอบติดตามการทำงาน ติดตามพฤติกรรมการใช้รถ หรือไว้ตรวจสอบเหตุ และคล้ายกับอุปกรณ์ของเครื่องบิน เปรียบได้กับ กล้องดำ ซึ่งกล้องดำจะเก็บข้อมูลทุกอย่าง ไม่ว่าจะเป็นเวลาที่เครื่องขึ้น การไต่ ระดับความสูง ความเร็วลม ความเร็วเครื่อง เป็นต้น สำหรับรถยนต์ GPS TRACKING มีหน่วยความจำภายในเมื่อนำมาใช้กับรถยนต์ เรียกว่า CAR TRACKING หรือ ระบบติดตามรถ นอกจากนี้ GPS CAR TRACKING จะทำงานร่วมกับ SOFTWARE ที่เรียกว่า GPS TRACKING SYSTEM ซึ่งพัฒนาและออกแบบแตกต่างกันไปตาม บริษัทผู้ให้บริการ โดยหลักการ SOFTWARE TRACKING จะนำค่าแสดงตำแหน่งจากดาวเทียม GPS มาแสดงผล โดยวิธีการซ้อนทับของข้อมูล OVERLAY ด้วยระบบ GIS ผู้ให้บริการบางรายมีแผนกจัดทำแผนที่และจัดทำแผนที่ด้วยตนเอง บางรายใช้แผนที่ร่วมผู้ให้บริการรายอื่นๆ เช่น GOOGLE MAP API บางรายก็ให้บริการแผนที่รายละเอียดสูง และเป็นปัจจุบัน หรือ แผนที่ในอนาคต

6.5 โอเพนซอร์ส เทคโนโลยี (Open Source Technology) คือซอฟต์แวร์ลิขสิทธิ์ที่มีไลเซนส์แบบโอเพนซอร์ส ซึ่งมีลักษณะต่างจากไลเซนส์ของซอฟต์แวร์ทั่วไป คือผู้พัฒนาเจ้าของซอฟต์แวร์จะอนุญาตให้ผู้ใช้ติดตั้งและใช้งานได้อย่างไม่จำกัดทั้งจำนวน และรูปแบบการใช้งาน ไม่ว่าจะเป็นการใช้งานส่วนตัว ในเชิงการค้า หรือในองค์กร นอกจากนี้ยังอนุญาตและสนับสนุนให้เรียนรู้ทำความเข้าใจการทำงานของซอฟต์แวร์ โดยการเผยแพร่ต้นฉบับ (Source Code) ของซอฟต์แวร์ออกมา และอนุญาตให้แก้ไขดัดแปลงให้ตรงความต้องการได้

6.6 Distribute Data Storage การจัดเก็บข้อมูลในระบบฐานข้อมูลแบบกระจายมีอยู่หลายวิธีด้วยกัน เช่น การทำสำเนา (Replication , การแยกกรีเลชัน (Fragmentation), วิธี Replication และ Fragmentation เป็นการแยกกรีเลชันออกเป็นหลาย ๆ ส่วน และแต่ละส่วนก็จะมีจัดการแบบสำเนา

6.7 ดาต้าไมนิ่ง (Data Mining) การทำเหมืองข้อมูล เป็นการค้นหาความสัมพันธ์และรูปแบบทั้งหมด ที่มีอยู่จริงในฐานข้อมูล ซึ่งความสัมพันธ์และรูปแบบเหล่านั้น ได้ถูกซ่อนไว้ภายในข้อมูลจำนวนมาก Data Mining จะทำการสำรวจ และวิเคราะห์ข้อมูลให้อยู่ในรูปแบบที่เต็มไปด้วยความหมายและอยู่ในรูปของกฎ โดยความสัมพันธ์เหล่านั้น แสดงให้เห็นถึงความรู้ต่าง ๆ (Knowledge) ที่ มีประโยชน์ในฐานข้อมูลในปัจจุบันองค์กรธุรกิจส่วนใหญ่เผชิญกับปัญหาของข้อมูลดิบจำนวนมากแต่ ข้อมูลที่ประยุกต์ใช้ได้นั้นมีน้อย การนำ Data Mining มาใช้ประยุกต์ใช้ โดยสามารถดึงความรู้ออกมาจากข้อมูลจำนวนมากที่ถูกเก็บสะสม และซ่อนไว้

6.8 Classification เป็นกระบวนการสร้างโมเดลจัดการข้อมูลออกเป็นประเภทตามลักษณะของแต่ละกลุ่มข้อมูล มักใช้กับลักษณะประจำที่มีค่าไม่ต่อเนื่อง ผลลัพธ์จะได้ตัวแบบจัดจำแนกประเภท(Classifier) เพื่อนำไปใช้แยกประเภทของตัวอย่างที่ไม่ทราบคลาส โดยจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนด อย่างเช่น แผนภูมิต้นไม้ตัดสินใจ(Decision Tree) เป็นลักษณะที่เข้าใจง่าย คล้ายแผนภูมิต้องการ โดยแต่ละ โหนดแสดง Attribute แต่ละกิ่งแสดงถึงผลการทดสอบ และลิฟ โหนดแสดงคลาสที่กำหนดไว้ ผลลัพธ์ที่ได้เขียนเป็นกฎให้เข้าใจง่าย และมีความถูกต้องสูง

6.9 Clustering เป็นเทคนิคการลดขนาดของข้อมูลด้วยการรวมกลุ่มของข้อมูล ที่มีความคล้ายคลึงกัน หรือมีลักษณะเดียวกันไว้ด้วยกัน การทำเทคนิค Clustering จะใช้ตัวแปรทุกตัว มาวิเคราะห์ให้ได้กลุ่มข้อมูล จากนั้นจึงนำตัวแปรข้อมูลใกล้เคียงกันมารวมเป็นกลุ่มข้อมูล

6.10 กฎความสัมพันธ์ (Association Rule) การค้นหาความสัมพันธ์ของข้อมูลที่มีอยู่ เพื่อช่วยในการวิเคราะห์และตัดสินใจ เช่น การหารายการ (Item) ใน Transaction เดียวกัน ซึ่งสามารถบอกได้ว่ารายการใดที่มีแนวโน้มที่จะสามารถพบด้วยกันใน Transaction เดียวกัน ตัวอย่าง เช่น การเกิดอุบัติเหตุรถยนต์ มีความเป็นไปได้จากการขับรถความเร็วสูง

6.11 คัดกรองพนักงานขับรถ การองค์ประกอบของบุคคลในการทำหน้าที่ขับ ที่มีความเสี่ยงในการเกิดอุบัติเหตุทางรถยนต์ เช่น อายุ เพศ การศึกษา น้ำหนักตัว ส่วนสูง ประสบการณ์ขับรถ และรายได้

6.12 ปัจจัยที่ก่อให้เกิดอุบัติเหตุทางรถโดยสาร การหาองค์ประกอบการเกิดอุบัติเหตุทางรถยนต์ นอกจากการคัดกรองพนักงานขับรถแล้ว ยังมีปัจจัยอื่นๆ เช่น เส้นทางการเดินรถ ความเร็วของการขับรถ จำนวนชั่วโมงการขับรถ และประเภทรถ เป็นต้น



บทที่ 2

วรรณกรรมที่เกี่ยวข้อง

การประยุกต์ระบบ GPS และการทำเหมืองข้อมูลเพื่อทำนายและป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด ผู้วิจัยได้ทำการศึกษาค้นคว้าเอกสาร และงานวิจัยที่เกี่ยวข้อง หลักการของทฤษฎี และเทคโนโลยีต่างๆ ที่มีความจำเป็นต่อการพัฒนาระบบ ในด้านของแนวคิด ทฤษฎี และเป็นแนวทางการพัฒนาระบบ โดยมีรายละเอียดดังนี้

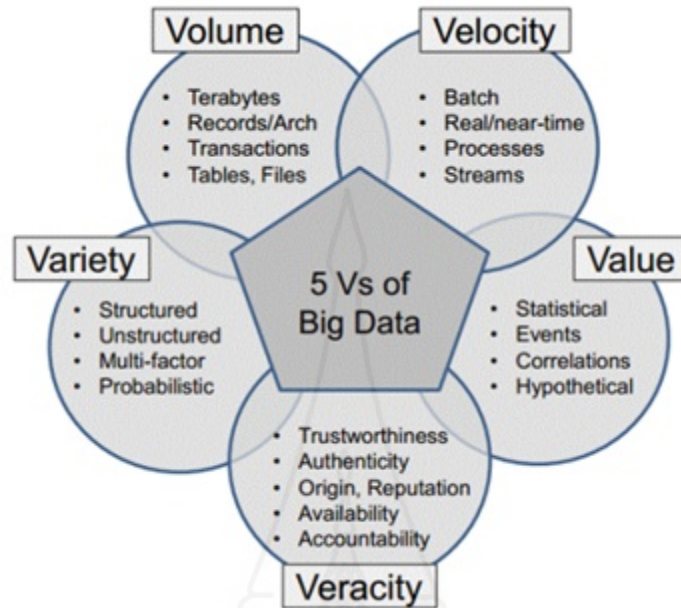
1. โครงสร้างสถาปัตยกรรมข้อมูลขนาดใหญ่ (Big Data Architecture)

Parth Chandarana และคณะ (2014) ศึกษาวิจัยเรื่อง "Big Data Analytics Frameworks" กล่าวว่า บิ๊กดาตา (Big Data) เป็นข้อมูลขนาดใหญ่ที่มีความหลากหลายและครอบคลุมเทคโนโลยีคลาวด์ โดยข้อมูลมีที่มาจากหลายแหล่ง เช่น ข้อมูลสภาพภูมิอากาศ โพสต์จากเว็บไซต์ เครือข่ายสื่อสังคมออนไลน์ เซ็นเซอร์ การทำธุรกรรมธนาคาร ข้อมูลการจราจร เป็นต้น ซึ่ง Big Data มีลักษณะที่แตกต่างจากข้อมูลทั่วไป ทางด้านข้อมูลขนาดใหญ่ (VOLUME) อยู่ในระดับ Terabytes ขึ้นไป ความเร็ว (VELOCITY) หมายถึง ข้อมูลจะมีการเปลี่ยนแปลงตลอดเวลาและรวดเร็ว รูปแบบหลากหลาย (VARIETY) หมายถึง รูปแบบมีความหลากหลายของรูปแบบข้อมูล ซึ่งอาจจะเป็นรูปแบบที่มีโครงสร้าง (Structure Data) ไม่มีโครงสร้าง (Unstructure Data) และกึ่งมีโครงสร้าง (SemiStructure Data)

โสภณ พินิจกิจเจริญกุล (2559) ได้กล่าวความหมาย Big Data คือ ข้อมูลขนาดใหญ่ ซึ่งข้อมูลเหล่านี้เกิดจากการพัฒนาของเทคโนโลยี และการเก็บรวบรวมข้อมูลไว้อย่างมากมายมหาศาล ในยุคแรกๆ นั้น ข้อมูลที่ถูกเรียกว่า Big data จะต้องประกอบด้วยคุณลักษณะ 3 อย่าง หรือที่เรียกว่า 3 V คือ ข้อมูลที่มีปริมาณมาก (VOLUME), ข้อมูลที่มีการเปลี่ยนแปลงตลอดเวลาและรวดเร็ว (VELOCITY) และข้อมูลมีความหลากหลาย (VARIETY) แต่ในปัจจุบัน ได้มีการกำหนดคุณลักษณะเพิ่มเติมขึ้นมาอีก 2 อย่าง คือ คุณภาพและความน่าเชื่อถือของข้อมูล (VARACITY) และข้อมูลที่มีคุณค่า (VALUE)

Joseph O.Chan (2013) กล่าวถึง "โครงสร้าง และหลักการของ Big Data คุณสมบัติของข้อมูล Big Data มี 5 ประเภท ได้แก่ VOLUME, VELOCITY, VARIETY, VERACITY, VALUE และเครื่องมือสนับสนุนร่วมกับหลักการนำการข้อมูล Big Data มาทำการวิเคราะห์ข้อมูลเชิงลึก"

องค์ประกอบของข้อมูลขนาดใหญ่ (Big Data)



ภาพที่ 2.1 องค์ประกอบข้อมูลขนาดใหญ่ Big Data 5 V

ที่มา: www.patrickcheesman.com/wp-content/uploads/2014/08/Big-Data-5V.gif

1) ปริมาณ (VOLUME) หมายถึง ข้อมูลที่มีจำนวนมหาศาลที่เพิ่มขึ้นในทุกขณะ จาก Terabytes (1 TB เท่ากับหนึ่งล้านล้าน ไบต์) กลายเป็น Petabytes และกลายเป็น Zettabytes และมากขึ้นไปเป็นลำดับ ตัวอย่างของ Big Data เช่น ข้อมูลบันทึกการใช้งานเว็บ, TFID, เครื่องขายเซ็นเซอร์, เครื่องขายส่งคอมออนไลน์, เอกสารและข้อความบนอินเทอร์เน็ต, การทำดัชนีค้นหาบนอินเทอร์เน็ต, การบันทึกการใช้งานโทรศัพท์, ดาราศาสตร์, สภาพอากาศ, การวิจัยทางชีวเคมี, ชีววิทยา และการวิจัยทางวิทยาศาสตร์ที่ซับซ้อน โดยเฉพาะการมีโทรศัพท์ระบบ 3G และ 4G ยิ่งทำให้ปริมาณข้อมูลเพิ่มขึ้นอย่างรวดเร็ว ตัวอย่างเช่น มีอีเมลที่ส่งกันทุกวันๆละประมาณ 294 ล้านฉบับ, ผู้ใช้เว็บเครื่องขายส่งคอมออนไลน์ อย่าง Facebook มากกว่า 1,000 ล้านคน คิดเป็นร้อยละ 75 อยู่นอกประเทศสหรัฐอเมริกา มีรูปภาพ 250 ล้านรูปที่มีการอัปโหลดขึ้นบนเว็บไซต์เฟซบุ๊กทุกวัน เมื่อปริมาณข้อมูลได้เพิ่มขึ้นจำนวนมหาศาลดังกล่าว จึงมีการหาแนวทางในการพัฒนาแพลตฟอร์ม หรือเทคโนโลยีใหม่ๆ ขึ้นมา เพื่อแก้ปัญหา และจัดการกับข้อมูลจำนวนมากเหล่านี้ให้ได้

2) ความหลากหลายของประเภทข้อมูล (VARIETY) หมายถึง ข้อมูลที่มาจากแหล่งข้อมูลที่แตกต่างกัน และมีหลากหลายประเภทได้แก่ ประเภทข้อมูลที่มีโครงสร้าง (Structure) ได้แก่ กลุ่มข้อมูลที่มีความสัมพันธ์กัน ถูกจัดเก็บเอาไว้ในฐานข้อมูลที่มีกฎเกณฑ์ เช่น การส่งใบแจ้งเดือนเกี่ยวกับบัญชีชำระหนี้ไปให้ลูกค้า, ข้อมูลประเภทกึ่งโครงสร้าง (Semi-Structure) เป็นข้อมูล

การตัดสินใจเกี่ยวกับปัญหาที่ไม่สามารถ ระบุกระบวนการหรือวิธีการตัดสินใจได้ล่วงหน้าแต่ไม่มากพอที่จะนำไปตัดสินใจได้อย่างแน่นอน ประเภทข้อมูลที่ไม่มีโครงสร้าง (UnStructure) เป็นข้อมูลการตัดสินใจเกี่ยวกับปัญหาที่ไม่สามารถกำหนดกระบวนการตัดสินใจ ได้ล่วงหน้า เช่น การลงทุนซื้อหุ้น รวมไปถึงข้อมูลที่เป็นข้อความ รูปภาพ อีเมล เสียง และวิดีโอ เป็นต้น

ด้วยความหลากหลายของประเภทข้อมูลดังกล่าว ทำให้เกิดความท้าทายกับองค์กร และบุคคลทั่วไปว่า ทำอย่างไรถึงจะทำการสกัดเอาข้อมูลที่มีความจำเป็น มาใช้ในการตัดสินใจกับธุรกิจได้ จากการสำรวจพบว่าร้อยละ 90 เป็นข้อมูลที่ไม่มีโครงสร้าง การส่งสารสนเทศที่ถูกต้องและถูกเวลามีแค่ร้อยละ 2 เท่านั้น และมีถึงร้อยละ 90 ที่ไม่สามารถจัดการ และสกัดข้อมูลออกมาใช้ได้ร้อยละ 43 ที่มีการโพสต์กันในเว็บเครือข่ายสังคมออนไลน์ ส่วนใหญ่เป็นเรื่องที่เกี่ยวกับผลิตภัณฑ์สินค้า และการสร้างแบรนด์ และร้อยละ 90 ลูกคามีความเชื่อถือเกี่ยวกับเรื่องราวที่โพสต์เหล่านั้น กระบวนการแก้ปัญหาเหล่านี้ ได้มีการพัฒนาระบบขึ้นมาแก้ปัญหาที่เรียกว่า NoSQL (Not Only SQL) สำหรับฐานข้อมูลประเภทไม่มีโครงสร้าง ได้แก่ Cassandra, CouchBase, Hbase, MongoDB สำหรับความหมายของ NoSQL หมายถึง ระบบฐานข้อมูลที่ไม่ใช่ภาษา SQL มีความสามารถในการรับข้อมูล ทั้งที่เป็น ข้อมูลกึ่งโครงสร้าง และไม่มีโครงสร้าง เช่น ผลิตภัณฑ์ Hbase, MongoDB เป็นต้น

Type	Example
Key-Value Store	redis, riak
Wide Column Store	HBASE, cassandra
Document Store	mongoDB, CouchDB
Graph Store	Neo4j, InfiniteGraph

ภาพที่ 2.2 รูปแบบผลิตภัณฑ์ต่างๆ ของ NoSQL Database

ที่มา: What is NoSQL and Benefits of Using it for App Development (www.algoworks.com)

3) ความรวดเร็ว (VELOCITY) หมายถึง อัตราความเร็วที่เพิ่มขึ้นของข้อมูล เป็นไปอย่างรวดเร็วใกล้เคียงเวลาเป็นจริง (Real-Time Information) เช่น การส่งข้อความที่สนทนากันบนเว็บเครือข่ายสังคมออนไลน์ การใช้โทรศัพท์มือถือและแท็บเล็ตถ่ายรูป ถ่ายวิดีโอ อัปเดตกิจกรรมที่เกิดขึ้น เช่นการเดินทางไปท่องเที่ยว สามารถที่จะอัปโหลดรูปภาพขึ้นเฟซบุ๊กได้ในทันที รวมไปถึงการอ่านแถบป้ายของ RFID มาตรฐานอัจฉริยะ รวมไปถึงการประมวลผลการสั่งซื้อสินค้าผ่านระบบออนไลน์ การขนส่งสินค้าในระบบห่วงโซ่อุปทาน และข้อมูลการบริการทั้งหมด

4) คุณภาพและความน่าเชื่อถือของข้อมูล (VARACITY) เพราะถ้าข้อมูลไม่มีคุณภาพ ไม่มีความน่าเชื่อถือ การนำข้อมูลมาวิเคราะห์ ผลลัพธ์ที่ได้ไม่ดีอยู่ดี

5) ข้อมูลที่มีคุณค่า (VALUE) คือ ข้อมูลที่มีคุณค่า สามารถนำไปใช้ประโยชน์ได้ หรือ มีมูลค่าและความสำคัญต่อธุรกิจในการนำมาใช้ประโยชน์

2. ฐานข้อมูลโนเอสคิวแอล (NoSQL Database)

Wei and Chi (2012) ได้กล่าวถึง "การไม่สนับสนุนในเรื่องของการเชื่อมความสัมพันธ์ของข้อมูลหลายตารางการจัดการข้อมูลทำได้ยากต่างจากฐานข้อมูลประเภท NoSQL ที่เน้นการจัดการเก็บข้อมูลให้ง่ายต่อการสอบถามข้อมูล" NoSQL มาจากการรวมคำศัพท์ คือ No และ SQL ผู้ที่นำคำศัพท์ NoSQL มาใช้ตอนแรกต้องการที่จะใช้คำว่า No RDBMS หรือ No relational แต่ด้วยการออกเสียงที่ดูน่าฟังกว่า จึงใช้คำว่า NoSQL มีบางคนพยายามนำเสนอว่า NoSQL นั้นย่อมาจาก Not Only SQL แต่ไม่ว่าจะหมายถึงอย่างไร NoSQL ก็ถูกใช้แทนฐานข้อมูลและการจัดเก็บข้อมูลที่ไม่เหมือนกับหลักการ ของฐานข้อมูลเชิงสัมพันธ์ การออกแบบการจัดเก็บข้อมูลในฐานข้อมูล NoSQL นั้น ไม่ต้องมีกระบวนการลดความซ้ำซ้อนของข้อมูล (Normalize) เพราะเป็นฐานข้อมูลที่สามารถเก็บข้อมูลที่มีความสัมพันธ์กันไว้ภายในระเบียบเดียวกันไม่มีการกำหนดโครงสร้างการจัดเก็บข้อมูลที่เฉพาะเจาะจง"เป้าหมายของฐานข้อมูล NoSQL คือลดความยุ่งยากซับซ้อนในการจัดการข้อมูลเพื่อตอบสนองการทำงานได้อย่างรวดเร็วขึ้นและรองรับการขยายขนาดของข้อมูล"Sadalage and Fowler(2013)

ฐานข้อมูล NoSQL สามารถจำแนกเป็น 4 แบบได้ดังต่อไปนี้

- 1 Key-Value Data Model
- 2 Document Data Model
- 3 Column-Family Data Model
- 4 Graph Data Model

Key-Value Data Model ฐานข้อมูล NoSQL แบบ Key-Value เป็นรูปแบบที่มีการนำคีย์ (Key) หรือ ID (Identity) เข้ามาใช้ในการจัดเก็บ การระบุระเบียบข้อมูลและการดึงข้อมูลออกมาใช้งานในแต่ละคอลัมน์จะเก็บคีย์คู่กับค่าของข้อมูลเสมอ ตัวอย่างการจัดเก็บข้อมูลแบบ Key-Value ดังภาพที่ 2.3

Example of Key-Value Data Model of NoSQL		
Key : 1	ID : KM001	Name : Mr. Nirut
Key : 2	Degree : Master degree	Faculty : SIT
Key : 3	Province : Bangkok	Country : Thailand

ภาพที่ 2.3 ตัวอย่างการจัดเก็บข้อมูลแบบ Key-Value Data Model

จากภาพที่ 2.3 แสดงการจัดเก็บข้อมูลแบบ Key-Value สังกัดได้ว่าในหนึ่งระเบียนของแต่ละคอลัมน์จะมีการเก็บค่าคีย์คู่กับค่าข้อมูลเสมอ

Document Data Model เป็นฐานข้อมูล NoSQL แบบ Document Data Model เป็นรูปแบบที่มีการจัดเก็บข้อมูลในลักษณะแบบคอกคิวเมนต์ที่มีโครงสร้างทางมิติเชิงลึกและเชิงกว้างในการจัดเก็บข้อมูล ซึ่งการจัดเก็บข้อมูลนั้น สามารถเก็บในรูปแบบ XML (Extensible Markup Language), JSON (Java Script Object Notation), BSON (Binary JSON) การจัดเก็บข้อมูลในฐานข้อมูลนั้นใช้หลักการใส่คีย์ในคอกคิวเมนต์ เพื่อเป็นการอ้างอิงข้อมูล โครงสร้างการจัดเก็บข้อมูลเป็นแบบลำดับชั้นซึ่งสามารถเก็บข้อมูลเชิงเดี่ยวและ ข้อมูลเชิงลึกได้ สามารถเก็บค่าที่เป็นอาร์เรย์ (Array) ได้ฐานข้อมูล NoSQL ที่มีลักษณะการจัดการข้อมูลแบบ Document Data Model ได้แก่ MongoDB และ CouchDB เป็นต้น ตัวอย่าง การจัดเก็บข้อมูลแบบ Document Data Model ดังภาพที่ 2.4

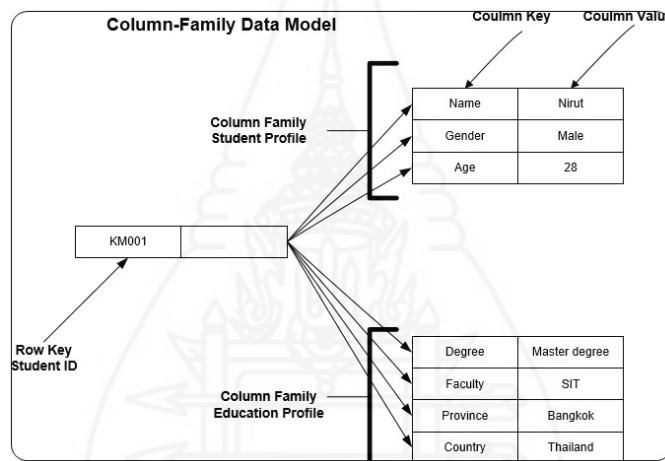
```
{
  _id: "KM001",
  name: "Mr. Nirut",
  education: {
    degree: "Master degree",
    faculty: "SIT",
    province: "Bangkok"
    country: "Thailand"
  }
}
```

ภาพที่ 2.4 ตัวอย่างการจัดเก็บข้อมูลแบบ Document Data Model

จากภาพที่ 2.4 แสดงการจัดเก็บข้อมูลแบบ Document Data Model ของฐานข้อมูล NoSQL จะเห็นได้ว่าการจัดเก็บเป็นลักษณะการจัดเก็บข้อมูลแบบโครงสร้างของข้อมูลหนึ่งระเบียน มีคีย์ กำกับค่าของข้อมูล ข้อมูลในแต่ละฟิลด์สามารถเก็บเป็นข้อมูลเชิงเดี่ยวได้เช่น ฟิลด์ Name เก็บข้อมูล

ตัวอักษร “Mr. Nirut” บางฟิลด์ก็สามารถเก็บข้อมูลเชิงลึกได้เช่น ฟิลด์ Education สามารถเก็บข้อมูล Degree, Faculty, Province และ Country เป็นชั้นข้อมูลซ้อนลงไปได้

Column-Family Data Model ฐานข้อมูล NoSQL แบบ Column-Family Data Model เป็นรูปแบบที่มีการจัดเก็บข้อมูลใน ลักษณะแบบตารางที่มีการจัดกลุ่มของคอลัมน์ โดยแต่ละตารางสามารถประกอบด้วยหลายแถวและ แต่ละแถวสามารถประกอบด้วยจำนวนคอลัมน์ที่แตกต่างกันได้ ไม่จำเป็นว่าทุกแถวจะมีจำนวนคอลัมน์คงที่เท่ากันตลอดทุกแถว ในแต่ละแถวนั้นจะมีคีย์กำกับ การดำเนินการจัดเก็บข้อมูลจะมีการ จัดเก็บข้อมูลแบบเรียงลำดับ โดยจะใช้คีย์ของแถวข้อมูลเป็นตัวจัดเรียงค่าข้อมูล ฐานข้อมูล NoSQL ที่มีลักษณะการจัดการข้อมูลแบบ Column-Family Data Model ได้แก่ Cassandra และ HBase เป็นต้น ดังภาพที่ 2.5

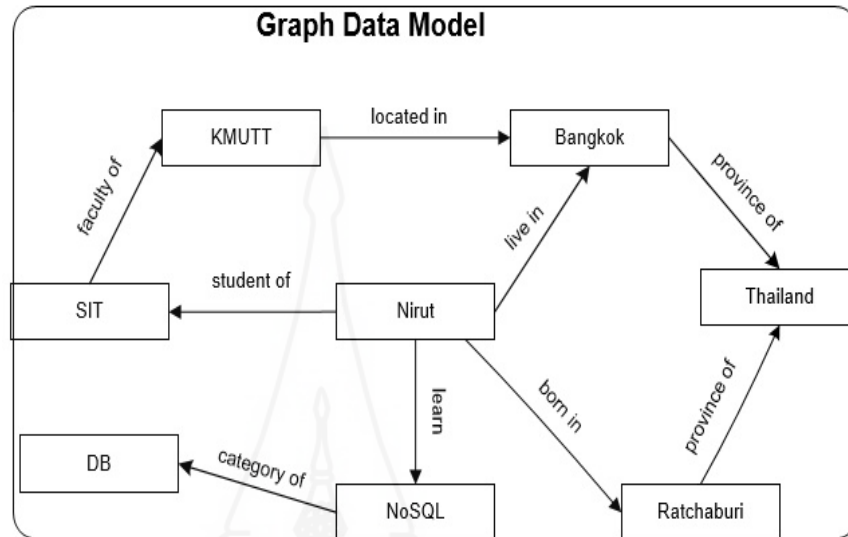


ภาพที่ 2.5 ตัวอย่างการจัดเก็บข้อมูลแบบ Document Data Model

จากรูปที่ 2.5 แสดงถึงการจัดเก็บข้อมูลแบบ Column-Family Data Model ในตารางข้อมูล ประกอบด้วยแถวข้อมูลที่มีคีย์กำกับ ในที่นี้ข้อมูลมีคีย์ของแถวข้อมูลคือ KM001 แต่ละแถวจะมีคีย์ที่เชื่อมไปยังกลุ่มของคอลัมน์ ในที่นี้มีกลุ่มคอลัมน์อยู่สองกลุ่มคือกลุ่มคอลัมน์ประวัติส่วนตัวและกลุ่มคอลัมน์ประวัติ ซึ่งแต่ละกลุ่มของคอลัมน์จะประกอบด้วยฟิลด์ของข้อมูล แต่ละฟิลด์ของข้อมูลประกอบด้วยไปด้วยคีย์ของคอลัมน์และค่าข้อมูลในคอลัมน์นั้น

Graph Data Model ฐานข้อมูล NoSQL แบบ Graph Data Model เป็นรูปแบบที่มีการจัดเก็บข้อมูลที่มีความสัมพันธ์ ในลักษณะแบบกราฟซึ่งมีลักษณะการเชื่อมโยง โหนด (Node) ด้วยจุดเชื่อมต่อ (Edge) เช่นข้อมูล ประเภทเส้นทางแผนที่ ข้อมูลการเดินทางสาธารณะ เป็นต้น การจัดเก็บข้อมูลชนิดนี้จะเป็นการเก็บค่า ข้อมูล โหนดและค่าข้อมูลจุดเชื่อมต่อ การจัดเก็บข้อมูลแบบ Graph Data Model สนับสนุนการจัดเก็บ ค่าที่มีความสัมพันธ์แบบ Single-Depth การค้นหาข้อมูลสามารถท่อง โหนดลง

ไปได้เรื่อยๆตาม ความสัมพันธ์ของข้อมูล ฐานข้อมูล NoSQL ที่มีลักษณะการจัดการข้อมูลแบบ Graph Data Model ได้แก่ Neo4J และ FlockDB เป็นต้น ตัวอย่างดังภาพที่ 2.6



ภาพที่ 2.6 แสดงถึงข้อมูลแบบ Graph Data Model

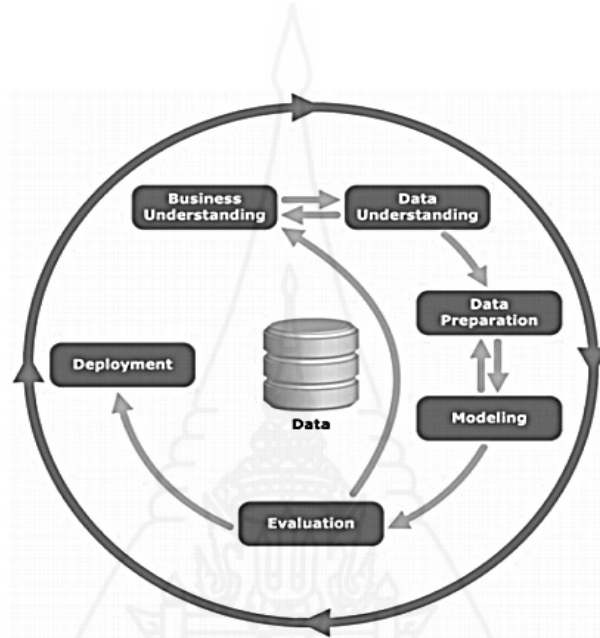
จากภาพที่ 2.6 แสดงถึงข้อมูลแบบ Graph Data Model เป็นตัวอย่างความสัมพันธ์ของข้อมูลในรูปแบบ ที่เป็นการเชื่อมโยงแบบกราฟข้อมูลแต่ละส่วนล้วนเกี่ยวข้องกัน ซึ่งฐานข้อมูล NoSQL สามารถ รองรับการจัดเก็บข้อมูลที่มีความสัมพันธ์แบบนี้ได้ " Islam (2011)

3. การทำเหมืองข้อมูล(Data Mining)

เหมืองข้อมูล (Data Mining) หมายถึง กระบวนการกระทำกับข้อมูลที่มีปริมาณมาก หรือข้อมูลขนาดใหญ่ เพื่อทำการค้นหารูปแบบแนวทางความสัมพันธ์ที่ซ่อนอยู่ในข้อมูล ซึ่ง จำเป็นต้องอาศัยหลักของสถิติ หลักคณิตศาสตร์การเรียนรู้ของเครื่อง การทำเหมืองข้อมูลหรือการ ค้นหาความรู้จากข้อมูลปริมาณมาก มีชื่อเรียกทางศัพท์เทคนิค คือ Knowledge Discovery in Database - KDD

Data Mining ตามศัพท์ที่ราชบัณฑิตยสถานกำหนดไว้หมายถึง การสกัดหรือวิเคราะห์ ค้นหาข้อมูลที่ต้องการจากข้อมูลจำนวนมากได้ หรือกล่าวอีกนัยหนึ่งว่า Data Mining คือ ชุด Software วิเคราะห์ข้อมูลที่ได้ถูกออกแบบมาเพื่อระบบสนับสนุนความต้องการของผู้ใช้ในการ ค้นหาข้อมูลที่ต้องการจากข้อมูลจำนวนมากได้

การทำเหมืองข้อมูลมีขั้นตอนหรือกระบวนการที่ประกอบด้วยรายละเอียดต่างๆ เพื่อให้มีแนวทางการทำงานที่ชัดเจนและแนวทางเดียวกัน มีการกำหนดมาตรฐานในการดำเนินการนี้ว่า คริสป์-ดีเอ็ม (CRISP-DM - Cross Industry Standard Process for Data Mining) โดยมีกรอบงานแยกเป็นแต่ละส่วนที่ช่วยให้ผู้ที่ทำเหมืองข้อมูลใช้เป็นแนวทาง และการทำงานที่มีประสิทธิภาพมากขึ้น



ภาพที่ 2.7 ขั้นตอนมาตรฐาน คริสป์-ดีเอ็ม CRISP-DM

ที่มา: (C. Shearer, 2006)

จากภาพที่ 2.7 เป็นภาพแสดงขั้นตอนมาตรฐาน คริสป์-ดีเอ็ม CRISP-DM หรือ Cross Industry Standard Process for Data Mining โดยพัฒนาขึ้นโดยบริษัทชั้นนำ 3 บริษัทอย่าง Daimler AG, SPSS, NCR (Chapman, 2000) ซึ่งประกอบไปด้วย 6 ขั้นตอนมาตรฐาน ดังต่อไปนี้

1 การเข้าใจในธุรกิจ (Business Understanding) เป็นขั้นตอนที่วิเคราะห์ ความต้องการทำเหมืองข้อมูลในเชิงธุรกิจ เมื่อทราบ จุดประสงค์แล้วก็ระบุผลลัพธ์ หรือ เป้าหมายที่ต้องการ ที่จะได้จากการวิเคราะห์ และขั้นตอนนี้ยังเป็นขั้นตอน การวางแผนงานอีกด้วย

2 การเข้าใจในข้อมูล (Data Understanding) เป็นขั้นตอนที่มีการรวบรวมข้อมูลที่เกี่ยวข้อง นำเชื่อถือ ในจำนวนที่มากเพียงพอที่จะนำมาใช้ในการวิเคราะห์ เนื่องจากข้อมูลเป็นปัจจัยที่สำคัญที่สุดในกระบวนการ

3 การเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนที่ใช้เวลายาวนานที่สุด เนื่องจากความถูกต้อง ที่จะได้จากการทำการวิเคราะห์ข้อมูลนั้นขึ้นอยู่กับ คุณภาพของข้อมูล ซึ่งประกอบไป

ด้วยขั้นตอนย่อยๆ อีก คือ การคัดเลือกข้อมูล (Data Selection) เป็นการกำหนดเป้าหมายว่าสิ่งที่ต้องการจะวิเคราะห์คืออะไร

1) การกลั่นกรองข้อมูล (Data Cleaning) เป็น การทำให้ข้อมูลมีความถูกต้อง โดยจะมีการลบข้อมูลที่ ซ้ำซ้อนกัน ซ่อมแซมข้อมูลที่ขาดหายไป รวมไปถึงแก้ไข ข้อมูลที่มีข้อผิดพลาด การตัดข้อมูลที่ไม่มีอยู่ในช่วงออก เป็นต้น

2) การบูรณาการข้อมูล (Data Integration) เป็นการรวบรวมข้อมูลจากหลายๆ แหล่ง เข้าด้วยกัน เนื่องจากในบางครั้งข้อมูลมาจากหลายแหล่ง

3) การลดขนาดข้อมูล (Data Reduction) เป็นการลดขนาดข้อมูล บางครั้งข้อมูลที่มากเกินไป

จะทำให้การทำงานในการสร้างแบบจำลองมีความช้า และเสียพื้นที่ในการเก็บข้อมูลอย่างมาก อีกทั้งข้อมูลประเภท Text ก็ไม่สามารถใช้ในอัลกอริทึมประเภทการจำแนกประเภท ของข้อมูลได้ ต้องมีการแปลงรูปข้อมูล ซึ่งในบางครั้ง เรียกว่า การสกัด ฟีเจอร์ (Feature Extraction) กระทำ ในขั้นตอนนี้ การแปลงข้อมูลให้อยู่ในช่วงที่กำหนดเอง (Normalization)

4) การทำให้ข้อมูลเป็นช่วง (Data Discretization) เป็นการทำให้ข้อมูลที่เป็นตัวเลข อยู่ในจำนวนของช่วง ที่ต้องการ

4 การสร้างโมเดล (Modeling) เป็นขั้นตอน ในการใช้อัลกอริทึมในการสร้างโมเดล

5 การตรวจสอบและประเมินผล (Evaluation) เป็นขั้นตอนในการวัดประสิทธิภาพของโมเดลวิเคราะห์ข้อมูลเชิงทำนาย ดูจากความแม่นยำ ในการทำนาย การสามารถนำโมเดลไปใช้ในทางธุรกิจ ได้จริง

6 การนำไปใช้งาน (Deployment) เป็น ขั้นตอนที่น่าโมเดลผลที่ได้จากการวิเคราะห์ข้อมูล ไปใช้จริง เพื่อช่วยในการตัดสินใจทางธุรกิจ

ประเภทการทำเหมืองข้อมูล

การทำเหมืองข้อมูลเป็นการค้นหารูปแบบที่เกิดขึ้นกับข้อมูลที่มีความสนใจ โดยแบ่งตามลักษณะงานที่ดำเนินการ เป็นประเภทต่างๆ ได้แก่

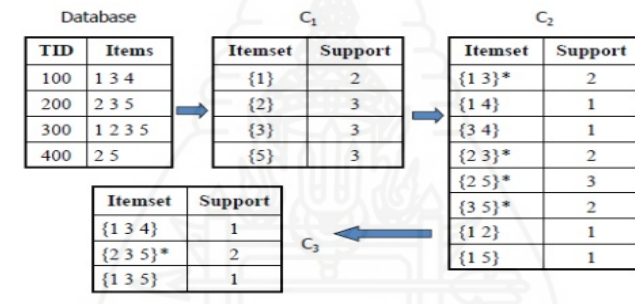
1 กฎความสัมพันธ์ (Association Rule) หมายถึง การค้น การความสัมพันธ์ที่ประกอบด้วย การค้นหารูปแบบที่เกิดควบคู่กันในรายการเดียวกันเสมอ เช่นการวิเคราะห์หรือค้นหาความสัมพันธ์ข้อมูลที่เกิดควบคู่กันในแต่ละรายการ (Transaction) ตัวอย่าง เช่น เมื่อจำนวนชั่วโมงการขับ มักจะเกิดการชนหรือเกิดอุบัติเหตุด้วยเสมอ และประสบการณ์ในการขับขีรถโดยสารสาธารณะ มักจะเกิดการชนหรือเกิดอุบัติเหตุด้วยเสมอด้วยเช่นกัน เป็นต้น หลังจากค้นหารูปแบบ

ที่เกิดขึ้น จะสร้างกฎความสัมพันธ์ โดยมีค่า Support เป็นจำนวนรายการข้อมูลที่น่ามาดำเนินการ และค่า Confidence เป็นค่าความน่าจะเป็นที่ใช้ในการวิเคราะห์ความสัมพันธ์ เช่น

ตัวอย่างความสัมพันธ์การเกิดอุบัติเหตุโดยสารสาธารณะ

"Accident(X,"AC_GWORK_1") => Accident(X,"AC_CHON") (support = 10%,Confidence=60%)"

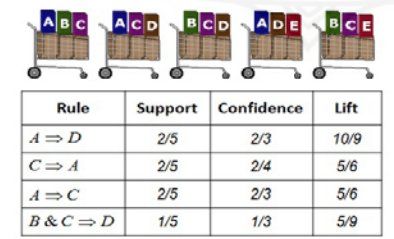
ความหมายว่าร้อยละ 10 ของจำนวนข้อมูลทรานแซกชันที่น่ามาดำเนินการค้นหา (Support=10%) พบว่าโอกาส หรือ ความน่าจะเป็นของการเกิดอุบัติเหตุทางรถโดยสารสาธารณะ คือ AC_GWORK_1 (ประสบการณ์ขยับขึ้นน้อยกว่าหนึ่งปี) จะมีโอกาสเกิดการชน AC_CHON (การชน) ร้อยละ 60 (confidence = 60%) ซึ่งผลที่ได้ทำให้ผู้วิเคราะห์มีความเข้าใจความเชื่อมโยงสัมพันธ์ของข้อมูล



ภาพที่ 2.8 การค้นหาแบบที่ก่อให้เกิดคววคูในรายการ

ที่มา: Dr. Saed Sayad,2010-2016 (www.saedsayad.com/association_rules.htm)

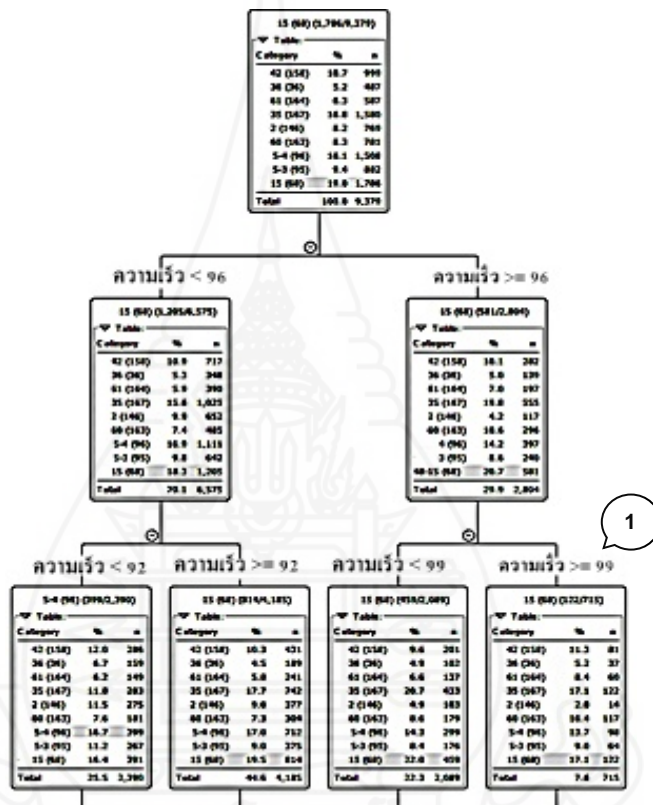
เครื่องมือในการทำเหมืองข้อมูลประเภทนี้เป็นเทคนิคหรืออัลกอริทึม เช่น การหาจำนวนความถี่การเกิดรายการ(Frequent itemset)ด้วยอัลกอริทึมอะไริออริ(Apriori) อัลกอริทึมฟรีเวนแพตเทิร์นโกรธ(Frequent Pattern Growth - FP-growth) กฎความสัมพันธ์ เป็นต้น



ภาพที่ 2.9 ความสัมพันธ์ตะกร้าสินค้า

ที่มา: Dr. Saed Sayad,2010-2016 (www.saedsayad.com/association_rules.htm)

2 กฎการจำแนกประเภท (Classification Rule) หมายถึง การจำแนกข้อมูลที่ต้องการวิเคราะห์ให้อยู่ในประเภทใด (Class) โดยอาศัยแบบจำลองที่สร้างขึ้นมา จะใช้จำนวนข้อมูลจำนวนหนึ่งเพื่อสร้างแบบจำลอง (Training Data Set) หลังจากที่ได้แบบจำลองแล้ว ก็จะต้องทำการทดสอบแบบจำลองด้วยข้อมูลอีกชุดหนึ่งหรืออีกกลุ่ม (Testing Data Set) เพื่อตรวจสอบความแม่นยำว่ามีความน่าเชื่อถือของแบบจำลอง ส่วนข้อมูลที่จะนำมาใช้จะเป็นข้อมูลที่มีค่าไม่ต่อเนื่อง (Discrete Data) เช่น ข้อมูลระบุผลการตัดสินใจ ใช่ หรือ ไม่ใช่ เลี้ยง หรือ ไม่เลี้ยง



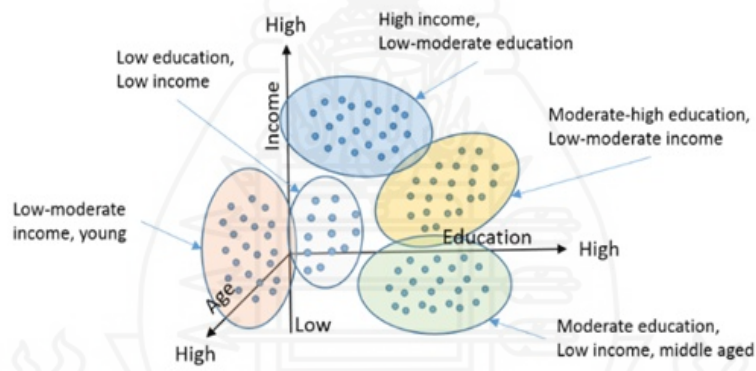
ภาพที่ 2.10 การจำแนกข้อมูลด้วยต้นไม้ตัดสินใจ

จากภาพที่ 2.10 ตัวอย่างการใช้ความเร็วของรถ เป็นการแยกประเภทของกลุ่มรถกับการใช้อัตราความเร็วในการขับรถโดยสาธารณะ ซึ่งเกิดจากการการสร้างตัวแบบจำลองอัลกอริทึมด้วยต้นไม้ตัดสินใจ สามารถแจกแจงกลุ่มหรือประเภทรถ ตามการใช้อัตราความเร็วรถโดยสาธารณะ ตัวอย่าง การจัดประเภท (Class) ของรถโดยสาธารณะ ตามอัตราความเร็วรถ เช่น การใช้ความเร็วรถ โดยสาธารณะที่อัตราความเร็ว มากกว่าหรือเท่ากับ 99 กม.ต่อชั่วโมง มีรถ หรือ กลุ่มรถหมายเลขอะไรบ้าง เป็นต้น

เครื่องมือที่นำมาใช้แบบจำแนกประเภท ได้แก่ การใช้อัลกอริทึมต้นไม้การตัดสินใจ (Decision Tree) ทฤษฎีเบย์ (Bayes Theorem) นาอิวเบย์ (Naive Bayes) โครงข่ายความเชื่อแบบเบย์

(Bayes Belief network) การวิเคราะห์แบบอิงกฎ (Rule Base) ในรูปแบบ IF_THEN Rule โครงข่ายประสาทเทียมหรือนิวรัลเน็ตเวิร์ก (Neural Network) เอสวีเอ็ม (SVM - Support Vector Machine) กฎความสัมพันธ์ (Association Rule) การหาเหตุและผลเพื่อตอบสนองสมมติฐานโดยใช้วิธี Case Base Reasoning การหาคุณสมบัติที่มีความคล้ายคลึงหรือใกล้เคียงกัน โดยวิธี K-Nearest Neighbor เป็นต้น

3 การจัดกลุ่ม (Clustering) หมายถึง ข้อมูลประเภทจัดกลุ่ม ที่ไม่ได้กำหนดหรือแบ่งกลุ่มไว้ล่วงหน้า โดยใช้เทคนิคจัดกลุ่มสมาชิกที่มีความคล้ายคลึงกันสูงสุด (Nearest Neighbor) เพื่อจำแนกข้อมูลแต่ละหน่วยของชุดข้อมูล โดยวัดความห่างจากจุดศูนย์กลาง การวัดค่าเบี่ยงเบน ซึ่งเป็นการรวมหน่วยที่คล้ายคลึงกันมากที่สุดเป็นกลุ่มเดียวกัน ตัวอย่างเช่น การทำเหมืองข้อมูลวิเคราะห์ตำแหน่งที่อยู่ของพนักงานบริษัท เพื่อจัดกลุ่มว่าอาศัยในเขตพื้นที่ใด การจัดกลุ่มประเภทนี้ไม่ได้กำหนดไว้ล่วงหน้า โดยกลุ่มเหล่านี้แต่ละกลุ่มมีคุณลักษณะเป็นไปในทิศทางเดียวกันจึงรวมตัวกันเป็นกลุ่มๆ ได้ ซึ่งต่างจากการจำแนกประเภท ที่มีการกำหนดกลุ่มหรือประเภทไว้ล่วงหน้า



ภาพที่ 2.11 การจัดกลุ่มด้วย Clustering

ที่มา: Jeffrey Strickland, Ph.D., CMSP, 2015(www.linkedin.com)

เครื่องมือที่ใช้ในประเภทการจัดกลุ่ม (Clustering) เช่นการใช้การวิเคราะห์ค่าเฉลี่ยแบบเค หรือการวิเคราะห์เคมีน (K-Mean Analysis)

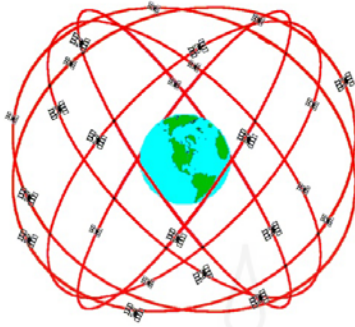
4. ระบบการติดตามรถยนต์ GPS

สมัยแรกการเดินทางมักใช้วิธีสังเกตจากดวงดาว ซึ่งใช้ได้ดีเพราะดาวห่างจากโลกเรามาก ทำให้สามารถมองเห็นกลุ่มดาวจากที่ต่างๆ ในบริเวณกว้าง แต่การวัดดาวทำได้เฉพาะตอนกลางคืนและต้องเป็นคืนที่ท้องฟ้าแจ่มใสเท่านั้น เครื่องมือที่ทันสมัยในยุคอิเล็กทรอนิกส์ ซึ่งมนุษย์สร้างสำหรับการเดินเรือชื่อระบบ LORAN ใช้ความถี่คลื่นวิทยุติดตั้งตามพื้นที่ส่วนต่างๆ และอีกระบบต่อมาใช้ดาวเทียมเหมือนระบบ GPS คือระบบ TRANSIT SYSTEM หรือ SATNAV ทั้งสองระบบที่กล่าวมา ปัจจุบันได้ยกเลิกแล้วเนื่องจากมีข้อบกพร่องในการบอกตำแหน่ง คือระบบ LORAN สามารถบอกตำแหน่งได้เพียงบริเวณหนึ่งๆ เท่านั้น ไม่ครอบคลุมทั้งหมด ส่วนระบบ TRANSIT นั้น สามารถบอกตำแหน่งครอบคลุมพื้นที่ได้มากกว่า แต่มีข้อบกพร่อง คือ วงโคจรดาวเทียมของระบบอยู่ในระดับต่ำและมีจำนวนน้อยเกินไป และเนื่องจากระบบ TRANSIT ใช้วิธีการวัดคลื่นแบบ Doppler ซึ่งถ้ามีการเคลื่อนไหวเครื่องรับสัญญาณเพียงเล็กน้อย ก็จะเกิดความคลาดเคลื่อนในการบอกตำแหน่งไปได้มาก

ระบบ GPS กระทรวงกลาโหม ประเทศสหรัฐอเมริกา ได้ดำเนินโครงการ Global Positioning System หรือ GPS ขึ้น จำนวนดาวเทียม 24 ดวง โคจรอยู่ในระดับสูงที่พ้นจากคลื่นวิทยุรบกวนและวิธีการที่สามารถให้ความถูกต้อง บอกตำแหน่งได้ทุกแห่งบนโลกตลอดเวลา 24 ชั่วโมง จากการพัฒนาทางด้านอุปกรณ์คอมพิวเตอร์ทำให้สามารถผลิตเครื่องรับ GPS ที่มีขนาดเล็ก และมีราคาถูกกว่าเครื่องรับระบบ TRANSIT เดิมเป็นอันมาก ปัจจุบันมีการนำ GPS มาใช้งานในหลายสาขาวิชาที่เกี่ยวข้องกับงานสำรวจ อาทิเช่น ภูมิศาสตร์ วิศวกรรมศาสตร์ สิ่งแวดล้อม ได้แก่ การนำ GPS มาใช้ในการกำหนดขอบเขตและจุดที่แน่นอนของป่าสงวนและอุทยาน ใช้สำรวจภูมิประเทศเพื่อทำแผนที่เส้นชั้นความสูง เป็นต้น

ส่วนประกอบของระบบ GPS ลักษณะทั่วไปของระบบ GPS ประกอบด้วย 3 ส่วน ได้แก่ ส่วนอวกาศ สถานีควบคุม และผู้ใช้งาน

1) ส่วนอวกาศ ระบบดาวเทียม GPS มี 24 ดวง โดยดาวเทียมจำนวน 21 ดวงจะใช้บอกค่าพิกัด ส่วนที่เหลือ 3 ดวง จะสำรองเอาไว้ ดาวเทียมทั้ง 24 ดวงจะมีวงโคจร 6 วงโคจร โดยแบ่งจำนวนดาวเทียมวงโคจรละ 4 ดวง และมีรัศมีสูงจากพื้น โลกประมาณ 20,200 กิโลเมตร (12,600 ไมล์) วงโคจรทั้ง 6 จะเอียงทำมุมกับเส้นศูนย์สูตรเป็นมุม 55 องศา ดาวเทียมแต่ละดวงใช้เวลาในการโคจรครบรอบ 12 ชั่วโมง คือคาบของการโคจรเป็น 12 ชั่วโมงต่อรอบ ความถี่ที่ใช้ในการบอกตำแหน่งค่าพิกัดดาวเทียมแต่ละดวงมี 2 ความถี่ คือ ความถี่ L1:1,575.42 MHz และ ความถี่ L2:1,270.60 MHz



ภาพที่ 2.12 โครงสร้างวงโคจรของดาวเทียม 24 ดวง
ที่มา: มหาวิทยาลัยสงขลานครินทร์ (WWW.RS.PSU.AC.TH)

2) สถานีควบคุม (Control Station Segment) ประกอบด้วย 5 สถานีย่อย ตั้งอยู่ที่เมือง Diego Garcia, Ascension Island, Kwajalein และ Hawaii ส่วน สถานีหลักตั้งที่เมือง Colorado Springs สหรัฐอเมริกา สถานีต่างๆทำหน้าที่คอยติดต่อสื่อสารกับดาวเทียม ทำการคำนวณ เพื่อบอกตำแหน่งของดาวเทียมแต่ละดวง และส่งข้อมูลที่ไปยังดาวเทียมอย่างต่อเนื่อง ทำให้ข้อมูลทันสมัยอยู่เสมอ



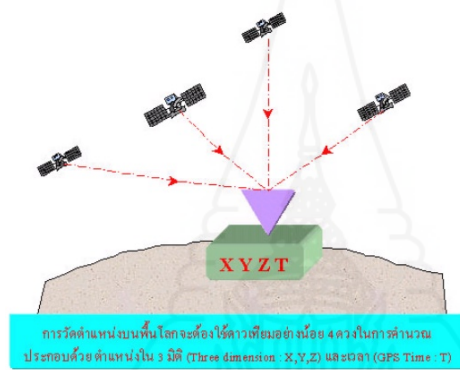
ภาพที่ 2.13 ตำแหน่งที่ตั้งของสถานีดาวเทียม
ที่มา: มหาวิทยาลัยสงขลานครินทร์ (WWW.RS.PSU.AC.TH)

3) ผู้ใช้ (Use Segment) แบ่งเป็น 2 ส่วน คือ พลเรือน และทหาร ในส่วนผู้ใช้งานมีหน้าที่พัฒนาเครื่องรับสัญญาณให้ทันสมัย และสะดวกแก่การใช้งาน สามารถที่จะใช้ได้ทุกแห่งในโลก และให้ค่าที่มีความถูกต้องสูง

การทำงานของ GPS หลักการพื้นฐานของ GPS เป็นเรื่องง่ายๆ แต่อุปกรณ์ของเครื่องมือถูกสร้างขึ้นด้วยวิทยาการขั้นสูง การทำงาน GPS แบ่งออก 5 ขั้นตอน คือ

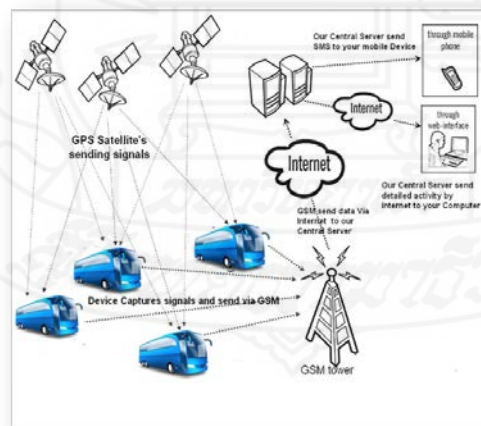
1. การรับสัญญาณจากดาวเทียมหลักการรูปสามเหลี่ยมระหว่างดาวเทียมกับเครื่องรับ
2. GPS วัดระยะโดยใช้เวลาเดินทางของคลื่นวิทยุ
3. ดาวเทียมและเครื่องรับจำเป็นต้องมีนาฬิกาที่ละเอียดสูง
4. นอกจากระยะทางแล้วต้องทราบตำแหน่งของดาวเทียมที่อยู่ในอวกาศด้วย
5. ในชั้นบรรยากาศไอโอโนสเฟียร์ และชั้นบรรยากาศโลก ความเร็วคลื่นวิทยุเดินทาง

ได้ช้าลง จึงต้องทำการแก้ไขจุดนี้ด้วย



ภาพที่ 2.14 การทำงานของ GPS

ที่มา: มหาวิทยาลัยสงขลานครินทร์ (WWW.RS.PSU.AC.TH)



ภาพที่ 2.15 โครงสร้างพื้นฐานของระบบการติดตามรถโดยสารด้วย GPS

ที่มา: EyLine Online GPS Tracing Solution

จากรูปภาพที่ 2.15 เป็นตัวอย่างการใช้งานจริงโครงสร้างพื้นฐานของระบบการติดตามรถโดยสารด้วย GPS ของบริษัทรถรุ่งเรืองจำกัด โดยโครงสร้างระบบการติดตามรถยนต์ด้วย GPS เป็น

อุปกรณ์เครื่องรับสัญญาณ GPS หรือกล่องดำ (Black box) ถูกติดตั้งภายในรถ และรับสัญญาณจากดาวเทียม เพื่อรับข้อมูลพิกัดละติจูดและลองจิจูด ภายในเครื่องรับสัญญาณ GPS จะมีอุปกรณ์ที่เรียกว่า ซิม การ์ด (Sim Card) บรรจุอยู่ใช้สำหรับการส่งข้อมูลตัวรถ ผ่านเครือข่ายผู้ให้บริการ GSM หรือ 3 จี(3G) ซึ่งมีการเชื่อมต่อระบบอินเทอร์เน็ตตลอดเวลา โดยทำการส่งข้อมูลไปยังเครื่องของลูกค้าผู้ให้บริการ จากนั้นข้อมูลที่ได้ผ่านกระบวนการประมวลผลด้วยค่าที่กำหนด แสดงผลผ่านหน้าจอ และออกรายงานให้กับผู้บริหารได้รับทราบข้อมูล

จากปริมาณข้อมูลการติดตามรถยนต์ด้วย GPS ที่ถูกส่งเข้ามาในระบบของรถแต่ละคัน ทุกช่วงเวลาของรถแต่ละคัน ผสมกับปริมาณจำนวนรถมาก จึงเป็นผลให้จำนวนทรานแซกชันหรือข้อมูลการติดตามรถมีปริมาณมากและเพิ่มขึ้นทุกวัน ปริมาณจำนวนทรานแซกชันของฐานข้อมูลมากขึ้นตามไปด้วย มีประสิทธิภาพในการประมวลผลที่ช้า ใช้เวลานาน ซึ่งเป็นจุดอ่อนของระบบฐานข้อมูล RDBMS และมีผลกระทบต่อการทำงานนำเสนอรายงาน จากปริมาณข้อมูลที่มีจำนวนมากว่าล้านเรคคอร์ด หรือจำนวนข้อมูลเป็นพันพันล้านเรคคอร์ด (Megabyte Gigabyte , Terabyte , Petabyte , Exabyte) รวมเป็นฐานข้อมูลขนาดใหญ่ จึงยากที่ระบบฐานข้อมูลทั่วไปจะรองรับ หรือสนับสนุนการประมวลผลของข้อมูลให้มีประสิทธิภาพได้

วันที่พิมพ์ : 26/11/2557 11:01:15

ชื่อรายงาน : Vehicle Now :26/11/2557 11:01:15

ลำดับ	หมายเลขรถ	สถานะ	วัน เวลา	ความเร็ว	ตำบล	อำเภอ	จังหวัด	เวลาจากกล่อง	เบอร์กล่อง
1	48-42 (158)	รถวิ่ง	10:41:02	92	บางพลีน้อย	บางบ่อ	สมุทรปราการ	10:41:02	144180100120
2	48-36 (36)	จอดรถ	09:58:10	เอกมัย	พระโขนง	คลองเตย	กรุงเทพมหานคร	10:36:22	144180100221
3	48-61 (164)	จอดรถ	09:33:12	อุบางนา	บางนา	บางนา	กรุงเทพมหานคร	10:36:20	144180100322
4	48-35 (167)	จอดรถ	21/11/2557	อุบางนา	บางนา	บางนา	กรุงเทพมหานคร	10:32:29	144180100423
5	48-2 (146)	รถวิ่ง	10:41:06	85	บางพลีใหญ่	บางพลี	สมุทรปราการ	10:41:06	144180100524
6	48-60 (163)	จอดไม่ดับเครื่อง	10:41:20	พิทยา	นาเกลือ	บางละมุง	ชลบุรี	10:41:20	144180100625
7	9905-4 (96)	จอดไม่ดับเครื่อง	09:10:39	หมอชิต	บางซื่อ	บางซื่อ	กรุงเทพมหานคร	10:41:06	144180100726
8	9905-3 (95)	รถวิ่ง	10:40:49	0	จตุจักร	จตุจักร	กรุงเทพมหานคร	10:40:49	144180100827
9	48-15 (68)	จอดไม่ดับเครื่อง	10:27:15	0	คลองตันเหนือ	วัฒนา	กรุงเทพมหานคร	10:37:55	144180101020
10	48-56 (56)	จอดรถ	18/12/2556	อุบางนา	บางนา	บางนา	กรุงเทพมหานคร	19/12/2556	144180101222

ภาพที่ 2.16 ตัวอย่างข้อมูลจาก GPS ของรถยนต์

จากภาพที่ 2.16 เป็นตัวอย่างข้อมูลจาก GPS ของรถยนต์ ซึ่งถูกจัดเก็บสะสมอยู่ในรูปแบบของระบบฐานข้อมูลจนถึงปัจจุบัน ข้อมูลจึงมีปริมาณมหาศาลมากขึ้น และทำให้การเข้าถึงข้อมูลซ้ำและประมวลผลข้อมูลก็ช้ามากขึ้นตามไปด้วย ถึงแม้จะมีการปรับแต่ง (Tuning) ฐานข้อมูลเพื่อเพิ่มการประมวลผลและการเข้าถึงแล้วก็ตาม แต่ก็ช่วยเพิ่มประสิทธิภาพได้เพียงระยะเวลาหนึ่งเท่านั้น แต่เมื่อระบบงานดำเนินต่อไป ข้อมูลก็จะเพิ่มขึ้นเรื่อยๆ จนทำให้มีปริมาณมหาศาลมากกว่าเดิม ซึ่งจะลดประสิทธิภาพการเข้าถึงและประมวลผลลงอีก

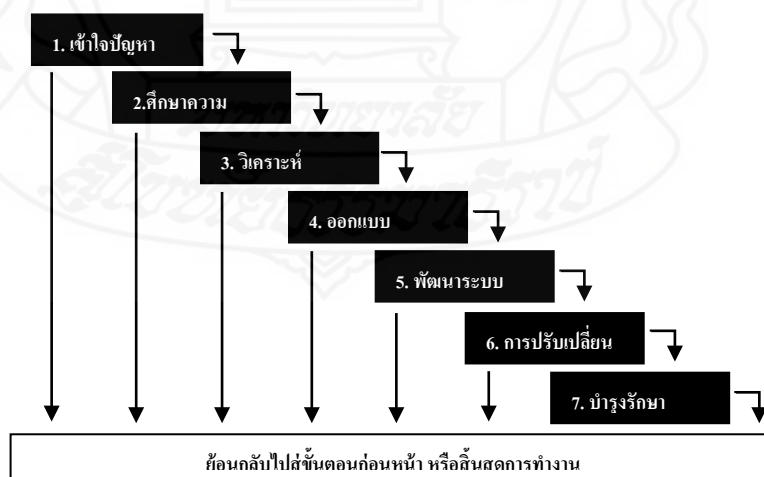
5. การพัฒนาระบบสารสนเทศ

อำเภอ ประเสริฐสกุล (2544) สรุปไว้ว่า กระบวนการพัฒนาระบบสารสนเทศ คือ การพัฒนาระบบสารสนเทศ เป็นการสร้างระบบงานใหม่หรือปรับเปลี่ยนระบบงานเดิมที่มีอยู่แล้ว ให้สามารถแก้ปัญหาการดำเนินงานธุรกิจได้ตามความต้องการของผู้ใช้งาน โดยอาจนำคอมพิวเตอร์ มาช่วยในการนำข้อมูลเข้าสู่ระบบเพื่อประมวลผล จัดเก็บทำให้ได้ผลลัพธ์ตามต้องการ ซึ่งการที่จะ ทำให้ได้ระบบตามที่ต้องการพัฒนามีความเป็นไปได้สูงสุดและสำเร็จจนนำไปสู่การใช้งานในที่สุด จะต้องดำเนินการตามขั้นตอนวงจรการพัฒนาระบบ (System Development Life Cycle : SDLC)

ขั้นตอนการพัฒนาระบบมีอยู่ด้วยกัน 7 ขั้นตอน ได้แก่

- 1) เข้าใจปัญหา (Problem Recognition)
- 2) ศึกษาความเป็นไปได้ (Feasibility Study)
- 3) วิเคราะห์ (Analysis)
- 4) ออกแบบ (Design)
- 5) การสร้างหรือพัฒนาระบบ (Construction)
- 6) การปรับเปลี่ยน (Conversion)
- 7) บำรุงรักษา (Maintenance)

ทั้ง 7 ขั้นตอนของการพัฒนาระบบสารสนเทศ นักวิเคราะห์ระบบ (System Analysis : SA) จะต้อง ปฏิบัติหรือทำอะไรบ้าง สามารถสรุปขั้นตอนได้ดังรูปที่ 2.17



ภาพที่ 2.17 ขั้นตอนวงจรการพัฒนาระบบ SDLC

ตารางที่ 2.1 วงจรการพัฒนาระบบ

หน้าที่	ความหมาย
1. เข้าใจปัญหา	- ตระหนักว่ามีปัญหาในระบบ
2. ศึกษาความเป็นไปได้	- รวบรวมข้อมูล - คาคะเนค่าใช้จ่ายและผลประโยชน์อื่นๆ - ตัดสินใจว่าจะเปลี่ยนระบบหรือไม่
3. วิเคราะห์	- ศึกษาระบบเดิม - กำหนดความต้องการของระบบ - แผนภาพระบบเก่าและระบบใหม่ - สร้างระบบทดลองของระบบใหม่
4. ออกแบบ	- เลือกอุปกรณ์ ฮาร์ดแวร์ และซอฟต์แวร์ - เปลี่ยนแผนภาพการวิเคราะห์เป็นแผนภาพลำดับชั้น - กำเนึงถึงความปลอดภัยของระบบ - ออกแบบ Input และ Output - ออกแบบ ไฟล์ ฐานข้อมูล
5. พัฒนา	- เตรียมสถานที่ - เขียนโปรแกรม - ทดสอบโปรแกรม - เตรียมคู่มือการใช้ และฝึกอบรม
6. นำมาใช้งานจริง	- ป้อนข้อมูล - เริ่มใช้งานระบบใหม่
7. บำรุงรักษา	- เข้าใจปัญหา - ศึกษาสิ่งที่ต้องแก้ไข - ตัดสินใจว่าจะแก้ไขหรือไม่ - แก้ไขเอกสารคู่มือ - แก้ไขโปรแกรม - ใช้งานระบบที่แก้ไขแล้ว

6. งานวิจัยที่เกี่ยวข้อง

พล โกสุภวัฒน์ (2556) ศึกษาเรื่อง "การสร้างแบบจำลองและทำนายผลการเรียนของ นักศึกษานิติศาสตร์มหาวิทยาลัยธรรมศาสตร์ โดยการใช้เหมืองข้อมูล " มีวัตถุประสงค์งานวิจัย ในการใช้ข้อมูลของนักศึกษานิติศาสตร์มหาวิทยาลัยธรรมศาสตร์ ที่เก็บไว้ก่อนเข้าการศึกษา และผล การเรียนชั้นปีที่ 1 และ ชั้นปีที่ 2 มาทำนายผลการเรียนตอนสำเร็จการศึกษาของนักศึกษา และหาว่า กลุ่มวิชาใดที่มีผลต่อการเรียนของนักศึกษา โดยใช้ขั้นตอน CRISP-DM (Cross-Industry Standard Process for Data Mining) มาทำการวิเคราะห์ เพื่อแบ่งกลุ่มนักศึกษา (Clustering) แล้วนำผลที่ได้มา ผลการเรียนเฉลี่ยของนักศึกษาที่สำเร็จการศึกษา (Classification) ด้วยแผนภูมิต้นไม้ตัดสินใจ และ หากกฎความสัมพันธ์ (Association) ของรายวิชาที่มีผลต่อการสำเร็จการศึกษตามหลักสูตร

สุรเดช บุญลือ (2550) ศึกษาเรื่อง “ระบบบริการผู้โดยสารอัจฉริยะผ่านเทคโนโลยี ตรวจสอบพิกัดจากดาวเทียม” เป็นระบบที่สามารถช่วยให้ □ การเดินทางไปในสถานที่ต่างๆได้ □ อย่างปลอดภัย มีประสิทธิภาพ และสะดวกขึ้น ในระหว่างที่ผู้ □ ใช้ □ เดินทาง เครื่องรับพิกัดจาก ดาวเทียมจะมีการรับสัญญาณเป็นระยะๆ เพื่อนำสัญญาณนั้นมาใช้ □ ในการดึงข้อมูลต่าง จากแหล่ง เก็บข้อมูลที่ถูกจัดเก็บไว้ □ เป็นฐานข้อมูลภายในรถยนต์ □ โดยจะนำเสนอข้อมูล ต่างๆ เช่น ขณะที่ ผู้ □ ใช้ □ อยู่ □ ที่ไหน ได้ □ เดินทางผ่านสถานที่ใดมาบ้าง และผู้ใช้ □ สามารถเลือกได้ □ ว่า ต้องการจะดูสถานที่ใด เช่น ปั้มน้ำมัน สถานพยาบาล เป็นต้น รวมทั้งสามารถจัดการทำรายการต่างๆ ของสถานประกอบการ ที่จัดให้ □ บริการไว้ □ เช่น การจองห้องพัก หรือจองตั๋วเครื่องบิน

วรพล ปัญญศรีประการ (2553) ศึกษาเรื่อง “ปัจจัยการยอมรับการนำระบบติดตาม รถยนต์ GPS มาใช้ร่วมบริษัทประกันภัย” ได้กล่าวถึง GPS คือภาคกำเนิดสัญญาณเวลาความแม่นยำ สูง เป็นนาฬิกาอะตอมมิกแบบซีเซียม ออกแบบ และผลิต โดย บริษัท Datum Incorporation USA จากเทคโนโลยีการติดตามรถยนต์ GPS บริษัทประกันภัยนำเสนอให้กับลูกค้าที่สนใจใช้บริการ ระบบการติดตามรถยนต์ โดยเมื่อเกิดเหตุฉุกเฉิน หรืออุบัติเหตุรถชน เพียงกดปุ่มฉุกเฉินขอความช่วยเหลือ เจ้าหน้าที่จะทำการติดต่อกลับไป พร้อมกันนั้นก็ส่งตำแหน่งของเจ้าของรถยนต์ให้กับ พนักงาน ที่อยู่ใกล้จุดเกิดเหตุที่สุด เพื่อเดินทางไปยังสถานที่เกิดเหตุ โดยทางศูนย์จะรายงาน ตำแหน่งและระยะเวลาของเจ้าหน้าที่แก่ผู้ให้บริการตลอด

ชัชชัย แก้วตา และคณะ (2553) ศึกษาเรื่อง "การวินิจฉัยคดีด้วยเทคนิคต้นไม้ตัดสินใจ " ได้กล่าวถึง วิธีการวินิจฉัยคดีโดย ประยุกต์ใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree Techniques) เพื่อจำแนก ความผิดซึ่งถูกอธิบายด้วยชุดของคุณลักษณะ (Attributes) ออกเป็นมาตรา ต่างๆที่ เหมาะสมกับคดีความนั้น และเปรียบเทียบความถูกต้องของการจำแนกข้อมูลด้วยขั้นตอนวิธี ID3

กับขั้นตอนวิธี C4.5 โดยจากผลการ ทดลองพบว่าต้นไม้ตัดสินใจที่สร้างจากขั้นตอนวิธี C4.5 สามารถจำแนก ความผิดได้ถูกต้องมากกว่าขั้นตอนวิธี ID3 และสามารถนำไปประกอบการ พิจารณาวินิจฉัยคดี

บุญมา เฟ่งชวน (2548) ศึกษาวิจัยเรื่อง "การใช้เทคนิคเหมืองข้อมูลเพื่อพัฒนาระบบ สนับสนุนการตัดสินใจ ด้านการผลิตบัณฑิตระดับปริญญาตรี " ได้กล่าวถึง การนำข้อมูลภาวะการณ มีงานทำของบัณฑิตมาวิเคราะห์โดยใช้เทคนิคเหมืองข้อมูล คือการค้นหาหาความสัมพันธ์ เพื่อ ศึกษาและพัฒนาตัวแบบในการทำนายแนวโน้มเลือกอาชีพแรกหลังสำเร็จการศึกษาของนักศึกษา ระดับปริญญาตรี ปีการศึกษา 2548 เพื่อวิเคราะห์ตัวแบบ และการนำเสนอในรูปแบบรายงาน ตารางและกราฟ เพื่อสนับสนุนการตัดสินใจในด้านการ ผลิตบัณฑิต และผลการวิจัยพบว่า สามารถนำตัวแบบที่พัฒนาขึ้นมาหาแนวทางการเลือกอาชีพแรกหลังสำเร็จการศึกษาของนักศึกษา ระดับปริญญาตรีได้

พิจิตร จอมศรี (2549) ศึกษาเรื่อง "การทำนายเนื้อหาของเว็บ โดยใช้เทคนิคเหมือง ข้อมูล กรณีศึกษามหาวิทยาลัยศิลปากร" ได้กล่าวถึง การใช้เทคนิคของกฎความสัมพันธ์ ที่ให้ค่า ความเชื่อมั่น และค่าสนับสนุนเป็นตัวพิจารณา โดยอาศัยข้อมูลวัน เวลา หมวดเว็บ และเว็บ ซึ่งจาก ผลการสร้างตัวแบบสามารถทำนายเนื้อหาเว็บถูกต้องร้อยละ 66.67

พวงทิพย์ แทนแสง (2550) ศึกษาเรื่อง "การทดสอบประสิทธิภาพการทำงานของ อัลกอริทึม การไมนิ่งกฎสำหรับจำแนก" ได้กล่าวถึง กฎการจำแนก เป็นกระบวนการในการจัดแบ่ง ข้อมูล ขึ้นอยู่กับลักษณะของวัตถุประสงค์นั้นๆ ส่วนการจำแนกในด้านการทำเหมืองข้อมูล เป็นการ วิเคราะห์เซตของกลุ่มข้อมูล (Data object) ที่ยังไม่มีการจัดแบ่งประเภท เพื่อสร้างแบบจำลอง (Model) ออกเป็นชุดข้อมูล (Class) ซึ่งลักษณะของคลาสถูกอธิบายโดยกลุ่มของคุณสมบัติ (Attribute) และกลุ่มของข้อมูล (Training Data Set) ที่ใช้ในการสร้างโมเดลกฎการจำแนก โดยใช้ อัลกอริทึม C4.5 เป็นหลัก เพื่อทำการทดสอบด้านความแม่นยำ (Accuracy) และการวัดคุณภาพของ กฎที่ได้รับ (Quality of Rule) ซึ่งประกอบด้วยการวัดความสมบูรณ์ของกฎ (Completeness) ความ ทนทานของกฎ (Consistency) และความครอบคลุมของกฎ (Coverage)

ศิริรัตน์ ชำนาญรบ และคณะ (2555) ศึกษาเรื่อง "การพัฒนาระบบสารสนเทศเพื่อ ส่งเสริมกิจกรรมชุมชน กรณีศึกษา ชุมชนนางเลิ้ง กรุงเทพมหานคร" ได้กล่าวถึงเรื่องขั้นตอนการ พัฒนาระบบสารสนเทศใช้วิธีการ วงจรการพัฒนา (System Development Life Cycle:SDLC) ซึ่งมีขั้นตอนการพัฒนาที่สำคัญ คือ การค้นหาและเลือกสรรโครงการ (Project Identification and Selection) การเริ่มต้นและวางแผน โครงการ (Project Initiating and planning) การวิเคราะห์ระบบ (System Analysis) การออกแบบเชิงตรรกะ (Logical Design) ขั้นตอนการออกแบบเชิงกายภาพ

(Physical Design) การพัฒนาและติดตั้งระบบ (System Implementation) การประเมินผล (Evaluation) และการซ่อมบำรุงระบบ (System Maintenance) จากการเก็บรวบรวมและวิเคราะห์ข้อมูลโดยใช้สถิติค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานด้วยโปรแกรมประมวลผลทางสถิติ ผลการศึกษาพบว่า การประเมินความพึงพอใจของกลุ่มผู้ปฏิบัติการที่มีต่อระบบสารสนเทศด้านการนำเข้าข้อมูลมีระดับความพึงพอใจอยู่ในระดับมาก ($x=4.00$) ด้านการกระบวนการทำงาน และรายงานของระบบมีระดับความพึงพอใจในระดับมาก ($x=4.15$) และมีความพึงพอใจต่อระบบโดยรวมอยู่ในระดับมาก ($x=4.08$) สำหรับการประเมินความพึงพอใจของกลุ่มผู้ใช้ทั่วไป ผลปรากฏว่าอยู่ในระดับมาก ($x=4.21$) เช่นกัน ผลการวิจัยจึงสรุปได้ว่าทุกกลุ่มตัวอย่างมีความพึงพอใจโดยรวมอยู่ในระดับมาก ระบบสารสนเทศที่พัฒนาขึ้นสามารถนำไปติดตั้งใช้งานจริงได้ และให้ระบบสารสนเทศตรงตามความต้องการของผู้ใช้

ไพฑูรย์ จันทร์เรือง (2550) ศึกษาเรื่อง "ระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียนของนักศึกษาในระดับปริญญาตรี โดยใช้เทคนิคต้นไม้ตัดสินใจ" ได้กล่าวถึงการพัฒนาตัวแบบในการทำนายแนวโน้มของผลการเรียนของนักศึกษาเมื่อเข้ามาต่อในระดับปริญญาตรี โดยใช้ข้อมูลของนักศึกษาระดับปริญญาตรี คณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ที่รวบรวมไว้เมื่อปีการศึกษา 2550 และพัฒนาระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียนของนักศึกษา โดยใช้ตัวแบบที่สร้างขึ้นภายใต้เทคนิคต้นไม้ตัดสินใจ (Decision Tree Techniques) พบว่า การสร้าง □ ตัวแบบสำหรับพัฒนาระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียนของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคต้นไม้ตัดสินใจนั้น ผู้จัดทำได้แยกสร้างตัวแบบสำหรับแต่ละสาขาการเรียนเนื่องจากคุณสมบัติของผู้เรียนแต่ละสาขา มีความแตกต่างกัน ทำให้ได้ตัวแบบที่สามารถทำนายแนวโน้มของผลการเรียนที่เหมาะสมสำหรับแต่ละสาขา แต่เนื่องจากผลการเรียน ของนักศึกษาที่นำมาพัฒนาตัวแบบนั้น ส □ วนใหญ่จะมีเกณฑ์ คะแนนเกาะกลุ่มกันอยู่ในช่วงกลางของข้อมูล ทำให้ □ ผลการตัดสินใจส่วนใหญ่ □ จะโน้มเอียงไปในเกณฑ์พอใช้ (ช่วงคะแนน 2.00 – 2.49) และปานกลาง (ช่วงคะแนน 2.50 – 2.99)

ประกายมาศ ศรีสุขทักษิณ และคณะ (2557) ศึกษาเรื่อง "การเปรียบเทียบความเร็วในการประมวลผลระหว่างฐานข้อมูลเชิงสัมพันธ์และฐานข้อมูลไม่สัมพันธ์แบบเอกสาร" ได้กล่าวถึงการทดสอบความเร็วในการประมวลผลระหว่างฐานข้อมูลไม่สัมพันธ์ และฐานข้อมูลเชิงสัมพันธ์ ในด้านการเขียน อ่าน แก้ไข และลบข้อมูล ผลการทดลองพบว่าหากปริมาณข้อมูลหรือจำนวนเชรด (Thred) มีน้อย ความเร็วที่ใช้ในการประมวลผลข้อมูลของฐานข้อมูลแต่ละแบบจะใช้เวลาใกล้เคียงกัน แต่เมื่อปริมาณข้อมูลหรือจำนวนเชรดเพิ่มขึ้นจะยังเห็นความแตกต่างได้อย่างชัดเจน ซึ่งฐานข้อมูลไม่สัมพันธ์ใช้เวลาน้อยกว่าฐานข้อมูลเชิงสัมพันธ์หลายเท่า โดยเฉพาะการลบข้อมูล

สามารถลบข้อมูลได้เร็วกว่าฐานข้อมูลเชิงสัมพันธ์ถึงเกือบ 70 เท่า จึงมีประโยชน์มากสำหรับระบบสารสนเทศขนาดใหญ่ที่มีจำนวนการเข้าใช้งานปริมาณมาก และต้องการเวลาในการประมวลผลที่รวดเร็ว

เสกสรรค์ วิสัยลักษณ์ (2558) ศึกษาเรื่อง "การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน" ได้กล่าวถึงการพัฒนาคลังข้อมูลเพื่อการพยากรณ์ผลการเรียน โดยนำข้อมูลนักเรียนระดับมัธยมศึกษา ระหว่างปีการศึกษา 2548 – 2556 จำนวน 525 ระเบียบ ประกอบด้วย 16 คุณลักษณะ ใช้ชุดข้อมูล 2 แบบ คือ ข้อมูลแบบไม่จัดกลุ่ม (Original Data) และข้อมูลแบบจัดกลุ่ม (Data Cluster) จากนั้นนำไปผ่านกระบวนการคัดเลือกคุณลักษณะ (Feature Selection) ซึ่งใช้วิธี Correlation-based Feature Selection (CFS) และวิธี Information Gain (IG) แล้วใช้เทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ซัพพอร์ตเวกเตอร์แมชชีน (SVM) และต้นไม้ตัดสินใจ (Decision Tree) มาสร้างตัวแบบพยากรณ์และเปรียบเทียบตัวแบบด้วยการทดสอบประสิทธิภาพแบบ 10-fold Cross Validation ผลที่ได้จากการพยากรณ์ชุดข้อมูลแบบไม่จัดกลุ่ม นำมาคัดเลือกคุณลักษณะด้วยวิธี Correlation-based Feature Selection (CFS) ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ให้ค่าความถูกต้องสูงที่สุดที่ร้อยละ 94.48 และมีค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดที่ 0.1880 เหมาะสมสำหรับการสร้างระบบพยากรณ์ผลการเรียนเฉลี่ยของนักเรียน ซึ่งนักเรียนสามารถนำผลการพยากรณ์ไปประกอบการตัดสินใจและวางแผนการเรียน

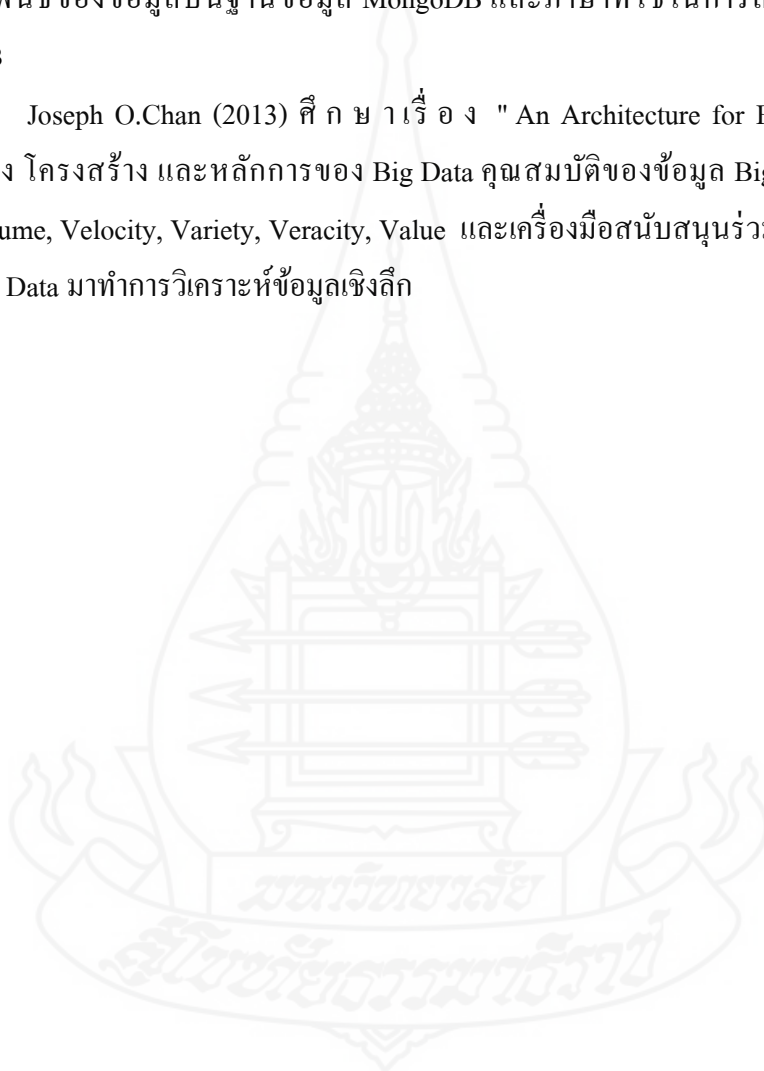
นิรุทธิ์ รวยริน(2557) ศึกษาเรื่อง "การใช้แมพรีดิวซ์เชื่อมคอลเลกชันของฐานข้อมูลมอดโกดบี" ได้กล่าวถึงรูปแบบการสร้างความสัมพันธ์ของฐานข้อมูล NoSQL ขึ้นมาเพื่อช่วยให้การสอบถามข้อมูลที่อยู่คนละส่วนทำได้สะดวกขึ้น การเขียนโปรแกรมทำได้ง่ายขึ้น ลดการสูญเสียเวลาในการโอนถ่ายข้อมูลมาที่ระบบ เพราะกระบวนการทำงานอยู่ที่ฝั่งฐานข้อมูลเป็นหลัก โดยศึกษาและเปรียบเทียบเพื่อวิจัยว่าฐานข้อมูลมอดโกดบีสามารถเข้าถึงข้อมูลได้เร็วกว่าฐานข้อมูล MySQL

Wei and Chi (2012) ศึกษาเรื่อง “Scalable Join Queries in Cloud Data Stores” ได้กล่าวถึงระบบคลาวด์ เป็นระบบที่ไม่สนับสนุนในเรื่องของการเชื่อมความสัมพันธ์ของข้อมูล การสอบถามข้อมูลกับตารางเดียวโดยใช้คีย์หลัก (Primary Key) ในการเข้าถึงข้อมูล เพราะระบบคลาวด์นั้นหากมีการจัดเก็บข้อมูลแยกย่อยเป็นหลายตารางจะทำให้การขยายขนาดพื้นที่จัดเก็บข้อมูลและการจัดการข้อมูลทำได้ยากประกอบกับฐานข้อมูลส่วนใหญ่บนระบบคลาวด์นั้นเป็นฐานข้อมูลประเภท NoSQL ที่เน้นการจัดเก็บข้อมูลให้ง่ายต่อการสอบถามข้อมูล งานวิจัยที่เกี่ยวข้องข้อนี้ก็ได้

แสดงให้เห็นถึงการออกแบบฐานข้อมูลให้แยกเก็บคนละส่วนเพื่อเชื่อมความสัมพันธ์กันในและการขยายขนาดพื้นที่จัดเก็บข้อมูลบนระบบ คลาวด์

Calil and Mello (2012) ศึกษาเรื่อง " SimpleSQL: A Relational Layer for SimpleDB " ได้กล่าวถึงการนำหลักการของแมพรีดิวซ์มาประยุกต์ใช้เพื่อประมวลผลฝั่งฐานข้อมูลเพื่อลดการสูญเสียเวลาในการติดต่อระหว่างฐานข้อมูลกับระบบผู้ใช้งาน และการนำเสนอการเชื่อมความสัมพันธ์ของข้อมูลบนฐานข้อมูล MongoDB และภาษาที่ใช้ในการสอบถามข้อมูลของ MongoDB

Joseph O.Chan (2013) ศึกษาเรื่อง " An Architecture for Big Data Analytics" ได้กล่าวถึง โครงสร้าง และหลักการของ Big Data คุณสมบัติของข้อมูล Big Data มี 5 ประเภท ได้แก่ Volume, Velocity, Variety, Veracity, Value และเครื่องมือสนับสนุนร่วมกับหลักการนำการข้อมูล Big Data มาทำการวิเคราะห์ข้อมูลเชิงลึก



บทที่ 3

วิธีดำเนินการวิจัย

การวิจัย การประยุกต์ระบบ GPS และการทำเหมืองข้อมูลเพื่อทำนายและป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด ศึกษาข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS เพื่อปรับเปลี่ยนจากฐานข้อมูลเชิงสัมพันธ์ (Relational Database) เป็นฐานข้อมูลไม่มีโครงสร้าง (Unstructured Data) ให้รองรับปริมาณข้อมูลขนาดใหญ่ ที่มีประสิทธิภาพด้าน การสืบค้น และการประมวลผลรวดเร็ว และการสร้างตัวแบบจำลอง เพื่อการวิเคราะห์ข้อมูลการติดตามรถโดยสารสาธารณะระบบ GPS และการทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะจากพนักงานขับรถ โดยใช้หลักการทำเหมืองข้อมูล นำผลทำนายไปสร้างโปรแกรมการใช้งาน ดังมีรายละเอียดของวิธีการดำเนินการวิจัย ดังนี้

1. ประชากร และกลุ่มตัวอย่าง
2. เครื่องมือที่ใช้ในการดำเนินงาน
3. ขั้นตอนการดำเนินงาน

1. ประชากร และกลุ่มตัวอย่าง

1.1 ประชากร

ข้อมูลการติดตามรถโดยสารสาธารณะระบบ GPS, ข้อมูลการเกิดอุบัติเหตุของรถโดยสาร และข้อมูลพนักงานขับรถ

1.2 กลุ่มตัวอย่าง

- ฐานข้อมูลการติดตามรถโดยสารสาธารณะระบบ GPS ระหว่างพ.ศ. 2555 - 2559 จำนวนรายการ 121,226,674 ระเบียบ ขนาดข้อมูล 58,009.27 ล้านไบต์
- ฐานข้อมูลการเกิดอุบัติเหตุของรถโดยสาร ระหว่างพ.ศ.2556 - 2559 จำนวนรายการ 406 ระเบียบ
- ฐานข้อมูลพนักงานขับรถ จำนวนรายการ 134 ระเบียบ

2. เครื่องมือที่ใช้ในการดำเนินวิจัย

2.1 ฮาร์ดแวร์ เครื่องคอมพิวเตอร์ ที่มีคุณสมบัติดังนี้

2.1.1 หน่วยประมวลผลกลาง (CPU) Intel(R) Core(TM) i5-4570 3.20 GHz

2.1.2 หน่วยความจำหลัก (MEMORY) 32 GB

2.1.3 อุปกรณ์บันทึกข้อมูล (Hard disk) ความจุ 1 TB ขึ้นไป(4 TB)

2.2 ซอฟต์แวร์

2.2.1 ระบบปฏิบัติการ ไมโครซอฟต์วินโดวส์ 8.1 (Microsoft Windows 8.1 Pro) เป็นระบบปฏิบัติการสำหรับเครื่องคอมพิวเตอร์

2.2.2 โปรแกรมไคนิม์ (KNIME v2.12) มีความสามารถที่หลากหลาย เช่น ด้านเหมืองข้อมูล(Data Mining) ธุรกิจอัจฉริยะ (Business Intelligence) รองรับข้อมูลในรูปแบบต่างๆ เช่น เท็กซ์ไฟล์ (Text, CSV, XLS) รูปแบบการทำงานเป็นเวิร์คโฟลว์ (WORKFLOW)

2.2.3 โปรแกรมมองโกลดีบี (MongoDB v2.6.7) เป็นฐานข้อมูลเพื่อรองรับข้อมูลขนาดใหญ่ หรือ Big Data ซึ่งเป็นฐานข้อมูลแบบไม่มีโครงสร้าง (Unstructured Data) ภายที่ใช้ในการจัดการ ได้แก่ NoSQL ซึ่งต่างจาก SQL หรือ MySQL

2.2.4 โปรแกรม Microsoft Visual FoxPro v9.0 สนับสนุนการเขียนโปรแกรมเพื่อสร้างโปรแกรม แอปพลิเคชัน ที่นำแบบจำลองจากการทำเหมืองข้อมูล เพื่อทำนายผลการคัดกรองพนักงานขับรถ และการป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะ

2.3 แบบสอบถามเพื่อประเมินผลระบบ

2.3.1 แบบสอบถามความพึงพอใจของผู้ใช้ระบบการใช้โปรแกรมที่พัฒนาที่นำแบบจำลองเพื่อทำนายมาใช้งาน

2.3.2 แบบสอบถามความเชื่อมั่นคำทำนายระบบการทำเหมืองข้อมูลเพื่อทำนายและป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด

3. ขั้นตอนการดำเนินงาน

- 1 สร้างระบบจัดการข้อมูล GPS
- 2 สร้างแบบจำลองการทำนาย
- 3 สร้างโปรแกรมใช้งาน
- 4 การประเมินผล
- 5 สรุปผลการวิจัยและข้อเสนอแนะ

3.1 สร้างระบบจัดการข้อมูล GPS

3.1.1 วิเคราะห์ความต้องการของผู้ใช้งาน

วิเคราะห์ความต้องการของผู้ใช้งาน ได้แก่ ผู้บริหาร หัวหน้าฝ่ายบุคคล หัวหน้าฝ่ายการเดินรถ และหัวหน้าฝ่ายอุบัติเหตุ โดยสัมภาษณ์เกี่ยวกับปัญหาและความต้องการ เพื่อนำมาเป็นแนวทางในการออกแบบคลังข้อมูล

3.1.2 การออกแบบคลังข้อมูล

ผู้วิจัยออกแบบคลังข้อมูล โดยออกแบบโครงสร้างคลังข้อมูลที่เกี่ยวข้องสองส่วน ได้แก่ 1) คลังข้อมูลระบบ GPS ด้วยข้อมูลที่ได้จากฐานข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS ซึ่งเป็นขนาดของข้อมูลขนาดใหญ่ เพื่อใช้ในการออกแบบตารางในรูปแบบข้อมูลไร้โครงสร้าง หรือไม่มีโครงสร้าง (UnStructured Data) โดยใช้เทคโนโลยีของ Big Data ที่สนับสนุนข้อมูลขนาดใหญ่ 2) คลังข้อมูลอุบัติเหตุ ด้วยข้อมูลที่ได้จากฐานข้อมูลฝ่ายอุบัติเหตุ ฐานข้อมูลฝ่ายบุคคลแผนกพนักงานขับรถ และผลลัพธ์การประมวลผลจากการวิเคราะห์คลังข้อมูลระบบ GPS เพื่อกำหนดมุมมอง ค่าที่ต้องการวัด ขอบเขต ประเภท และขนาดของตาราง เพื่อใช้ในการออกแบบตาราง และกำหนดความสัมพันธ์ระหว่างตารางจากข้อมูล เพื่อให้ได้ผลลัพธ์ที่ต้องการใช้โครงสร้างคลังข้อมูลอุบัติเหตุ

3.1.3 การพัฒนาคลังข้อมูล

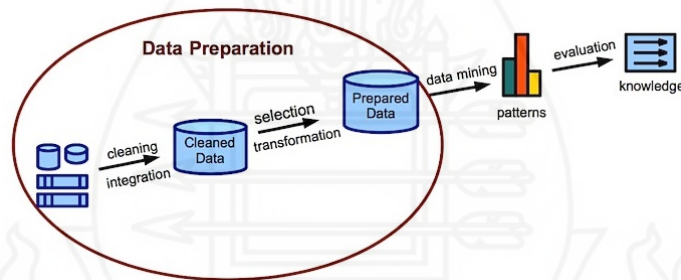
กระบวนการที่สำคัญในการนำข้อมูลเข้าคลังข้อมูลระบบ GPS และคลังข้อมูลอุบัติเหตุ คือ กระบวนการอีทีแอล (Extract Transform and Load : ETL) ซึ่งประกอบด้วย 3 ขั้นตอน ได้แก่

- 1) การคัดแยกข้อมูล (Extract) เป็นการคัดแยกและเลือกข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS และข้อมูลการเกิดอุบัติเหตุ จากฐานข้อมูลฝ่ายอุบัติเหตุ ฝ่ายบุคคลแผนกพนักงานขับรถ และผลการประมวลผลการวิเคราะห์คลังข้อมูลระบบ GPS

2) การแปลงข้อมูล (Transform) เป็นการปรับรูปแบบและโครงสร้างข้อมูล ติดตามรถโดยสารสาธารณะระบบ GPS และข้อมูลการเกิดอุบัติเหตุ เพื่อจัดข้อมูลที่ไม่ถูกต้อง หรือไม่สอดคล้องกันของข้อมูล

3) การนำเข้าข้อมูล (Load) คัดแยกและจัดรูปแบบข้อมูลเรียบร้อยแล้ว เข้าสู่คลังข้อมูลเพื่อวิเคราะห์และแสดงผลตามที่ได้ออกแบบไว้

งานวิจัยนี้ผู้วิจัยนำข้อมูลมาจากรฐานข้อมูลของโปรแกรม ติดตามรถโดยสารสาธารณะระบบ GPS และข้อมูลโปรแกรมการเกิดอุบัติเหตุรถโดยสารสาธารณะระหว่างจังหวัด นำข้อมูลมาผ่านกระบวนการ ETL เพื่อเลือกหรือคัดกรองข้อมูลที่ต้องการ จัดรูปแบบข้อมูล และทำความสะอาดข้อมูลให้มีความถูกต้องสอดคล้องกัน จากนั้นจึงนำข้อมูลที่ได้เข้าสู่คลังข้อมูลเพื่อจัดเก็บ แล้วจึงดำเนินการนำข้อมูลในคลังข้อมูลไปวิเคราะห์ข้อมูล สำหรับเครื่องมือที่ใช้ในการวิจัยนี้คือ โปรแกรม KNIME v2.12 ในการทำเหมืองข้อมูล (Data Mining) ออกแบบรูปแบบจำลอง อัลกอริทึมการทำนาย และโปรแกรม Microsoft Visual FoxPro v9.0 เพื่อสร้างโปรแกรม แอปพลิเคชัน นำเสนอข้อมูลทำนายผลการคัดกรองพนักงานขับรถ และการป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะ



ภาพที่ 3.1 ขั้นตอนแสดงขั้นตอนการพัฒนาคลังข้อมูล

ที่มา: Data Preparator Software, website:www.datapreparator.com/what_is_data_preparation.html

3.2 สร้างแบบจำลองการทำนาย

ผู้วิจัยยึดแนวทางการพัฒนาการวิเคราะห์ตามกระบวนการมาตรฐานอุตสาหกรรม ที่มีขั้นตอนการแก้ปัญหาที่ใช้ในงานธุรกิจ ชื่อว่า CRISP-DM (Cross-Industry Standard Process for Data Mining) โดยนำมาปรับใช้ให้เหมาะสมกับกระบวนการสร้างแบบจำลองประกอบด้วย 6 ขั้นตอน ดังนี้

3.2.1 ทำความเข้าใจปัญหา (Business Understanding)

วัตถุประสงค์ของงานวิจัยนี้เพื่อใช้เทคนิคการทำเหมืองข้อมูล ในการสร้างตัวแบบจำลอง ให้ได้อัลกอริทึมจากข้อมูลอุบัติเหตุ และข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS

เพื่อประโยชน์ในด้านการทำนาย หรือการคาดการณ์การเกิดอุบัติเหตุจากการขับรถโดยสารสาธารณะ ซึ่งเป็นเครื่องมือช่วยสนับสนุนการคัดกรองบุคคลที่จะมาทำการขับรถโดยสารสาธารณะ ให้กับผู้ประกอบการเดินรถโดยสารสาธารณะ

3.2.2 ทำความเข้าใจข้อมูล (Data Understanding)

การเก็บรวบรวมข้อมูลที่นำมาใช้ทำเหมืองข้อมูลประกอบด้วย 2 ส่วน ได้แก่ 1) ข้อมูลการติดตามรถโดยสารสาธารณะระบบ GPS อยู่ภายใต้การจัดการของโปรแกรม Microsoft SQL Server 2008 R2 ผู้วิจัยใช้โปรแกรม KNIME v2.12 เป็นเครื่องมือเพื่อติดต่อกับฐานข้อมูล MS SQL และแปลงข้อมูลเป็นเท็กซ์ไฟล์ (CSV) แล้วใช้คำสั่ง Command Line ของโปรแกรม MongoDB นำข้อมูลเข้าสู่คลังข้อมูล ประกอบด้วยข้อมูลระหว่างพ.ศ. 2555 - 2559 จำนวนทั้งสิ้น 121,226,674 ระเบียบ ขนาดข้อมูล 58,009.27 ล้านไบต์ และ 2) ข้อมูลการเกิดอุบัติเหตุจากพนักงานขับรถระหว่างพ.ศ. 2556 - 2559 จำนวนทั้งสิ้น 406 ระเบียบ เป็นฐานข้อมูลเชิงสัมพันธ์ (Relational Database) ใช้โปรแกรม Microsoft Visual FoxPro v9.0 ทำการแปลงข้อมูลรูปแบบไฟล์ EXCEL

3.2.3 เตรียมข้อมูล (Data Preparation)

ข้อมูลการเดินรถระบบการติดตามรถโดยสารสาธารณะ และข้อมูลการเกิดอุบัติเหตุจากพนักงานขับรถ มีข้อมูลหลายส่วนที่ไม่มีความจำเป็นในการนำมาวิเคราะห์ หรือนำมาทำเป็นเหมืองข้อมูล จึงจำเป็นต้องมีการคัดกรองข้อมูล แยกข้อมูล และตัดข้อมูลที่ไม่ต้องการออกไป เพื่อให้เกิดความเหมาะสม และสะดวกในการนำข้อมูลมาทำการวิเคราะห์ ดังนี้

1) การคัดเลือกข้อมูล (Select Data)

โดยข้อมูลที่ได้จะทำการเลือกเฉพาะข้อมูลที่ครบถ้วน และสามารถนำไปใช้ในการวิเคราะห์และทำนายผลได้จริง โดยที่ข้อมูลที่ไม่ครบถ้วนหรือไม่สมบูรณ์จำเป็นต้องทำการตัดทิ้งหรือนำออกไปจากฐานข้อมูล และคงเหลือไว้ซึ่งข้อมูลที่มีความสมบูรณ์เหมาะสม และต้องสอดคล้องกับงานวิจัย

2) การทำความสะอาดข้อมูล (Clean Data)

จากการตรวจสอบข้อมูล ข้อมูลส่วนใหญ่สมบูรณ์ แต่ยังคงพบว่ามีข้อมูลบางส่วนมีค่าผิดปกติ หรือมีค่าที่ควรจัดการแก้ไข ข้อมูลบางส่วนที่จะนำมาใช้มีการเก็บค่าที่เกินความจำเป็น หรือจัดอยู่ในรูปแบบที่ไม่สามารถนำข้อมูลมาทำการวิเคราะห์หรือการทำนาย และข้อมูลขาดความสมบูรณ์ในการรวบรวมข้อมูลเพื่อการจัดกลุ่ม หรือการแบ่งประเภทของข้อมูล

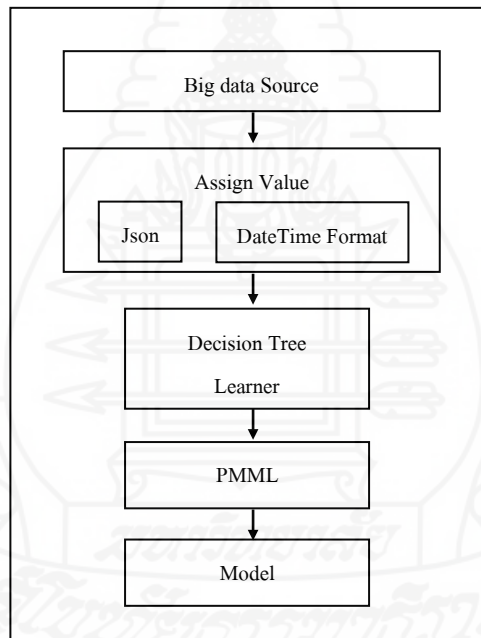
3) การแปลงข้อมูล (Data Transformation) เป็นการแปลงข้อมูลให้เหมาะสมสำหรับการทำเหมืองข้อมูลด้วยโปรแกรม KNIME v2.12 ซึ่งมีโมดูลมองโกดีบี (MongoDB) และ

โมดูล Text File รูปแบบไฟล์ CSV เข้าโปรแกรม KNIME v2.12 เพื่อประโยชน์ในการออกแบบจำลอง อัลกอริทึม การวิเคราะห์ และการทำนาย

3.2.4 สร้างแบบจำลองหรือโมเดล (Modeling)

ผู้วิจัยทำการสร้างแบบจำลองวิเคราะห์การประมวลผลคลังข้อมูล GPS และแบบจำลองทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ ด้วยโปรแกรมไคโนม (KNIME v2.12) เป็นเครื่องมือสำหรับการทำเหมืองข้อมูล สำหรับการออกแบบอัลกอริทึม แบบจำลอง และการทำนาย ประกอบด้วย 2 ส่วน ได้แก่ แบบจำลองการติดตามรถโดยสารสาธารณะระบบ GPS ข้อมูลระหว่าง พ.ศ. 2555 - 2559 จำนวนทั้งสิ้น 121,226,674 ระเบียบ ขนาดข้อมูล 58,009.27 ล้านไบต์ และแบบจำลองการเกิดอุบัติเหตุรถโดยสารสาธารณะ ข้อมูลระหว่างพ.ศ. 2556 - 2559 จำนวนทั้งสิ้น 406 ระเบียบ

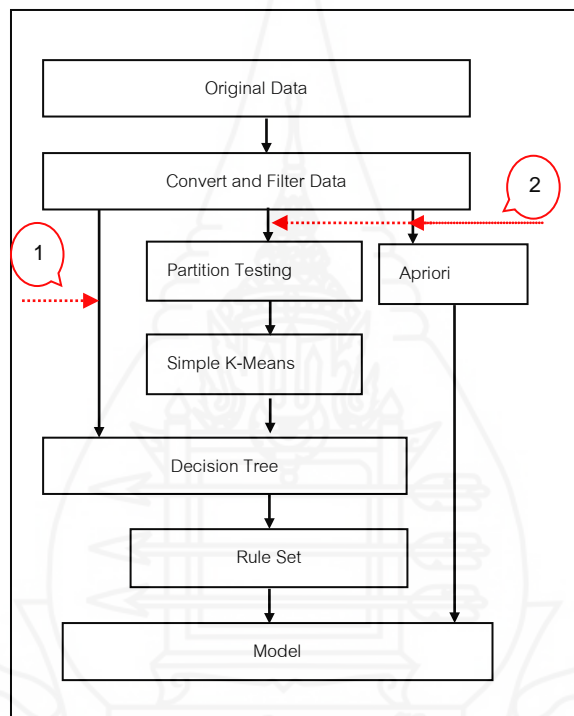
1. แบบจำลองการติดตามรถโดยสารสาธารณะระบบ GPS



ภาพที่ 3.2 กระบวนการสร้างแบบจำลองทำนายการใช้อัตราความเร็วรถโดยสารสาธารณะ
 ภาพที่ 3.2 กระบวนการสร้างแบบจำลองทำนายการใช้อัตราความเร็วรถโดยสารสาธารณะระบบ GPS เริ่มจาก Big Data Source คลังข้อมูลของ MongoDB ที่มีโครงสร้างของข้อมูลเป็นแบบไม่มีโครงสร้าง (None Structured) ส่งข้อมูลเพื่อกำหนดค่า Assign Value สำหรับแปลงชุดข้อมูล JSON (Java Script Object Notation) ซึ่งมีลักษณะคล้ายกับรูปแบบของ XML (Extensible Markup Language) เป็นตารางข้อมูล จากนั้นกำหนดรูปแบบของวันที่ และกำหนดเงื่อนไขของอัตราความเร็วรถเพื่อให้ได้ค่าของข้อมูลที่ต้องการ หลังจากนั้นนำไปผ่านขั้นตอน Decision Tree เพื่อ

จำแนกข้อมูลด้วยเทคนิคการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree Learner) โดยใช้อัลกอริทึมแบบ C4.5 เลือกกำหนดการใช้ค่า Target Column เพื่อการทำนาย และเลือก Include Fields เพื่อการแบ่งกลุ่มข้อมูล และกำหนดคุณลักษณะการวัดค่า Quality Measure Value เป็นแบบ Gain เพื่อวัดค่าคุณภาพในการพิจารณาการจำแนกที่ดีที่สุด ซึ่งทำให้ได้ผลลัพธ์การทำนาย PMML (The Predictive Model Markup Language) ที่ประกอบด้วยค่าของผลลัพธ์ และได้แบบจำลองอัลกอริทึมสามารถนำไปใช้งานได้ 2 รูปแบบ ได้แก่ แบบจำลองกลุ่มรถโดยสารสาธารณะ และ แบบจำลองกลุ่มสถานที่หรือตำแหน่งที่ถูกกำหนดในระบบ GPS

2. แบบจำลองการเกิดอุบัติเหตุรถโดยสารสาธารณะ



ภาพที่ 3.3 กระบวนการออกแบบจำลองการเกิดอุบัติเหตุรถโดยสารสาธารณะ

ภาพที่ 3.3 กระบวนการออกแบบจำลองการเกิดอุบัติเหตุรถโดยสารสาธารณะ Original Data เป็นรูปแบบข้อมูล Excel File ผ่านขั้นตอน Convert and Filter Data เพื่อทำการกำหนดค่า และคัดเลือกข้อมูลที่ระบบต้องการ ข้อมูลแบ่งเป็น 2 ชุด ข้อมูลชุดที่หนึ่ง หมายเลข 1 เป็นข้อมูลองค์ประกอบของปัจจัยการเกิดอุบัติเหตุ ได้แก่ เส้นทางเดินรถ จำนวนชั่วโมงขับรถ ประเภทรถที่ใช้ และสถานที่เกิดเหตุ (บริษัท, สถานี, ภายนอก) ข้อมูลชุดที่สอง หมายเลข 2 เป็นข้อมูลคุณสมบัติบุคคล [11] ได้แก่ อายุ การศึกษา ภูมิลำเนา น้ำหนักตัว ส่วนสูง ประสบการณ์ทำงาน และรายได้ จากภาพที่ 3.3 กระบวนการออกแบบจำลองนั้น ต้องการจำแนกกลุ่มออกเป็นสองกลุ่ม คือ กลุ่มเสี่ยง (เป็นบุคคลที่ถูกจัดให้มีอัตราการก่อให้เกิดอุบัติเหตุการขับรถโดยสารสาธารณะมากที่สุด) และกลุ่มเฝ้าระวัง

(เป็นบุคคลที่ถูกจัดว่าอาจจะก่อให้เกิดอุบัติเหตุการขับรถโดยสาธารณะ) การออกแบบจำลองและทดสอบแบบจำลองของกลุ่มเสี่ยง และกลุ่มเฝ้าระวัง สามารถอธิบายกระบวนการออกแบบได้ ดังนี้ กลุ่มเฝ้าระวังใช้ข้อมูลชุดที่สอง หมายเลข 2 ซึ่งเป็นชุดข้อมูลคุณสมบัติบุคคลนำไปใช้ทดสอบกฎความสัมพันธ์ (Association Rule) ด้วยเทคนิคอโพรไอริ (Apriori) เพื่อหาคุณสมบัติของกลุ่มเฝ้าระวัง ส่วนกลุ่มเสี่ยงเป็นการนำข้อมูลชุดที่หนึ่งหมายเลข 1 เป็นข้อมูลองค์ประกอบของปัจจัยการเกิดอุบัติเหตุ และข้อมูลชุดที่สอง หมายเลข 2 คุณสมบัติบุคคลมาทำการทดสอบแบบจำลอง โดยชุดข้อมูลที่สอง หมายเลข 2 คุณสมบัติบุคคลนำไปทดสอบแบบจำลองด้วย Partition Testing โดยใช้เทคนิค SplitTest ทำการแบ่งข้อมูลเป็นสองชุด ได้แก่ Training Data สำหรับสร้างแบบจำลอง และ Testing Data สำหรับทดสอบแบบจำลองเพื่อเข้าสู่กระบวนการในการจัดกลุ่มข้อมูลด้วยอัลกอริทึม Simple K-Means เพื่อทดสอบข้อมูล Training Data และข้อมูล Testing Data ที่แสดงผลลัพธ์ของข้อมูลคล้ายคลึงกันหรือเหมือนกัน แสดงว่าการทดสอบแบบจำลองการจัดกลุ่มและเป็นจำนวนกลุ่มเหมาะสม ขั้นตอนสุดท้ายนำแต่ละกลุ่มมาใช้ข้อมูลร่วมกับข้อมูลชุดที่หนึ่ง หมายเลข 1 คือ องค์ประกอบของปัจจัยการเกิดอุบัติเหตุ (เส้นทางเดินรถ, จำนวนชั่วโมงขับรถ, ประเภทรถที่ใช้ และสถานที่เกิดเหตุ) ด้วยอัลกอริทึม Decision Tree และ Rule Set เพื่อหาความสัมพันธ์ของข้อมูลชุดที่หนึ่ง หมายเลข 1 และข้อมูลชุดที่สอง หมายเลข 2 เพื่อให้ได้ผลการทำนายที่สามารถนำไปพัฒนาเป็นระบบการทำนายผลการเกิดอุบัติเหตุรถโดยสาธารณะต่อไป

3.2.5 ประเมินประสิทธิภาพ (Evaluation)

ผู้วิจัยใช้โปรแกรม KNIME v2.12 ประเมินประสิทธิภาพแบบจำลองการทำนาย ได้แก่

1) การใช้อัตราความเร็วของรถโดยสาธารณะ ใช้เทคนิคการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree Learner) ในการประเมินประสิทธิภาพโดยกำหนดค่าเป้าหมาย (Target Value) และกำหนดเงื่อนไขการสืบค้น เพื่อสร้างตัวแบบการทำนาย และแสดงค่าความเชื่อมั่น (Confidence Value) ของข้อมูล

2) การเกิดอุบัติเหตุรถโดยสาธารณะ การทดสอบและการประเมินประสิทธิภาพแบบจำลองการทำนายแบ่งเป็น 2 กลุ่ม ได้แก่ กลุ่มเสี่ยง และกลุ่มเฝ้าระวัง สามารถอธิบายกระบวนการได้ดังนี้ กลุ่มเฝ้าระวัง ใช้ชุดข้อมูลคุณสมบัติบุคคลนำไปใช้ทดสอบกฎความสัมพันธ์ (Association Rule) ด้วยเทคนิคอโพรไอริ (Apriori) เพื่อหาคุณสมบัติของกลุ่มเฝ้าระวัง ผู้วิจัยใช้การประเมินประสิทธิภาพของตัวแบบด้วยกฎความสัมพันธ์ (Association Rule) โดยกำหนดค่าความเชื่อมั่น (Confidence) ไว้ที่ร้อยละ 60 และค่าลิฟต์ (Lift) คือค่าสหสัมพันธ์ที่บ่งบอกถึงกฎความสัมพันธ์ที่มีค่ามากกว่า 1 เพื่อคัดเลือกค่าที่เหมาะสมสำหรับการใช้งาน ส่วนกลุ่มเสี่ยง

เป็นการข้อมูลชุดองค์ประกอบของปัจจัยการเกิดอุบัติเหตุ และข้อมูลชุดคุณสมบัติบุคคลมาทำการทดสอบแบบจำลอง โดยชุดข้อมูลคุณสมบัติบุคคลนำไปทดสอบแบบจำลองด้วย Partition Testing โดยใช้เทคนิค SplitTest ทำการแบ่งข้อมูลเป็นสองชุด ได้แก่ Training Data สำหรับสร้างแบบจำลอง และ Testing Data สำหรับทดสอบแบบจำลองเพื่อเข้าสู่กระบวนการในการจัดกลุ่มข้อมูลด้วยอัลกอริทึม Simple K-Means เพื่อทดสอบข้อมูล Training Data และข้อมูล Testing Data ที่แสดงผลลัพธ์ของข้อมูลคล้ายคลึงกันหรือเหมือนกัน แสดงว่าการทดสอบแบบจำลองการจัดกลุ่มและเป็นจำนวนกลุ่มเหมาะสม ขั้นตอนสุดท้ายนำแต่ละกลุ่มมาใช้ข้อมูลร่วมกับข้อมูลชุดองค์ประกอบของปัจจัยการเกิดอุบัติเหตุ(เส้นทางเดินรถ, จำนวนชั่วโมงขับรถ, ประเภทรถที่ใช้ และสถานที่เกิดเหตุ) ด้วยอัลกอริทึม Decision Tree และ Rule Set เพื่อหาความสัมพันธ์ของข้อมูลทั้ง 2 ชุด เพื่อให้ได้ผลการทำนายที่สามารถนำไปพัฒนาเป็นระบบการทำนายผลการเกิดอุบัติเหตุจริง โดยสามารถทำตามขั้นตอนต่อไป

3.2.6 การนำไปใช้งาน (Deployment)

เป็นการนำตัวแบบทำนายที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคเหมือนข้อมูลมาพัฒนาสร้างระบบทำนายการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัดเพื่อนำไปใช้กับบริษัท หรือองค์กร ขนส่งผู้โดยสารรถโดยสารสาธารณะ เพื่อใช้เป็นเครื่องมือประกอบการตัดสินใจในการคัดกรองพนักงานขับรถให้มีประสิทธิภาพดีขึ้น

3.3 สร้างโปรแกรมใช้งาน

การสร้างโปรแกรมใช้งานระบบการทำนายการเกิดอุบัติเหตุจริงโดยสารสาธารณะ ผู้วิจัยใช้โปรแกรม Microsoft Visual FoxPro v9.0 เป็นเครื่องมือสร้างโปรแกรม ให้มีรูปแบบการใช้งานดังนี้ ขั้นตอนหนึ่ง ผู้ใช้สามารถเลือกทำนายบุคคล เช่น ผู้สมัครใหม่ และ พนักงานขับรถ บริษัทฯ ระบุรายละเอียดส่วนบุคคล เช่น ชื่อ นามสกุล วันเกิด อายุ การศึกษา เป็นต้น หากเป็นพนักงานขับรถของบริษัทฯ ให้ระบุรหัสพนักงาน จากนั้นทำการประมวลผลเพื่อแสดงผลการทำนายบุคคลว่าจัดอยู่กลุ่มเสี่ยงหรือกลุ่มเฝ้าระวัง และขั้นที่สองนำผลการทำนายบุคคลมาทำนายอีกครั้งกับปัจจัยอื่น เช่น เส้นทางเดินรถ ประเภทรถ และสถานที่ เป็นต้น เพื่อทำนายผลที่อาจก่อให้เกิดอุบัติเหตุของแต่ละปัจจัยที่นำมาใช้

3.4 การประเมินผล

ในงานวิจัยจะใช้แบบสอบถามเพื่อประเมินความพึงพอใจการประยุกต์ระบบจีพีเอส และแบบประเมินความเชื่อมั่นการทำนายการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด โดยมีขั้นตอนดังนี้

3.4.1 การออกแบบเครื่องมือการประเมิน ในออกแบบเครื่องมือประเมินความพึงพอใจของผู้ใช้ เพื่อประเมินความเชื่อมั่นการทำนายเพื่อป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด ซึ่งใช้แบบสอบถามเป็นเครื่องมือในประเมิน แบบสอบถามที่ใช้เป็นแบบมาตราส่วนแบบ Rating Scale 5 ระดับ คือ มากที่สุด, มาก, ปานกลาง, น้อย และน้อยที่สุด

3.4.2 การประเมินผล เป็นการดำเนินการหลังจากนำแบบสอบถามที่ได้จากการประเมินของกลุ่มผู้ใช้งาน มาวิเคราะห์และแปลผล โดยใช้ค่าสถิติการวัดค่ากลางของข้อมูล ได้แก่ ค่าเฉลี่ยเลขคณิต (Arithmetic Mean)

3.4.3 เกณฑ์การพิจารณา ในการพิจารณาแบบประเมิน ได้กำหนดเกณฑ์การพิจารณาคุณภาพ โดยใช้ค่าเฉลี่ยเลขคณิต (\bar{x}) แล้วนำค่าเฉลี่ยที่ได้มาเปรียบเทียบกับเกณฑ์ของเบสท์ (Best.,1963: 204-208)

3.5 สรุปผลการวิจัยและข้อเสนอแนะ

สรุปผลการวิจัยที่ได้ จากการพัฒนาประยุกต์ระบบ GPS และการใช้เทคนิคทำเหมืองข้อมูล มาสร้างระบบการทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะจากพนักงานขับรถ และอภิปรายผลงานวิจัย รวมทั้งข้อเสนอแนะ

บทที่ 4

การสร้างแบบจำลองและพัฒนาระบบ

การวิจัยการประยุกต์ระบบ GPS และการทำเหมืองข้อมูลเพื่อทำนายและป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด ผู้วิจัยได้พัฒนาระบบคลังข้อมูลในการนำเสนอหลักการ Big Data เพื่อรองรับข้อมูลขนาดใหญ่ระบบ GPS และการทำเหมืองข้อมูลสำหรับทำนายการเกิดอุบัติเหตุของรถโดยสารสาธารณะ โดยนำผลที่ได้ไปสร้างโปรแกรมการใช้งาน มีรายละเอียดดังนี้

- 1 สร้างระบบจัดการข้อมูล GPS
- 2 สร้างแบบจำลองการทำนาย
- 3 สร้างโปรแกรมใช้งาน
- 4 ประเมินผลโปรแกรมใช้งาน

1. สร้างระบบจัดการข้อมูล GPS

1.1 การวิเคราะห์ปัญหาความต้องการของผู้ใช้งาน

จากการสัมภาษณ์และสอบถามข้อมูลจาก ผู้บริหาร ผู้จัดการฝ่ายบุคคล ผู้จัดการฝ่ายการเงิน ผู้จัดการฝ่ายอุบัติเหตุ และเจ้าหน้าที่ที่เกี่ยวข้อง ที่เป็นผู้ใช้งานข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS และการเกิดอุบัติเหตุของรถโดยสารสาธารณะ พบปัญหาและความต้องการของผู้ใช้งานสามารถสรุปได้ดังนี้

1.1.1 ผู้ใช้งานดูรายงานข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS ใช้เวลาประมวลผลข้อมูลนาน และต้องรอเจ้าหน้าที่บริษัทระบบ GPS มาทำการปรับปรุงคลังข้อมูลระบบ GPS เพื่อให้การประมวลผลข้อมูลทำงานเร็วขึ้น แต่ปัญหาการประมวลผลข้อมูลก็ยังคงเกิดขึ้นเสมอ เพราะข้อมูลติดตามรถโดยสารสาธารณะมีปริมาณมาก การรายงานข้อมูลจึงไม่สามารถตอบสนองความต้องการผู้ใช้งานได้อย่างพอเพียง ซึ่งความต้องการของผู้ใช้งานต้องการรายงานที่สามารถดูได้รวดเร็ว

1.1.2 โปรแกรมที่ใช้อยู่ในปัจจุบันรูปแบบคลังข้อมูลหรือโครงสร้างข้อมูลแตกต่าง ซึ่งผู้ใช้งานของแต่ละฝ่ายจะมีโปรแกรมเฉพาะของฝ่ายต่างๆที่ใช้งานอยู่ และมีขั้นตอนที่ยุ่งยากใน

การใช้งาน รวมถึงการเข้าถึงข้อมูลของผู้ใช้งานอย่างผู้บริหาร และผู้จัดการฝ่ายต่างๆ ที่มีหน้าที่เกี่ยวข้อง ซึ่งต้องให้เจ้าหน้าที่นำข้อมูลมารวบรวมสิ่งที่ต้องการก่อนจึงสามารถสรุปข้อมูลการนำเสนอได้ ผู้ใช้งานอยากได้ระบบที่ทำงานในรูปแบบโครงสร้างของข้อมูลเดียวกันเพื่อง่ายต่อการใช้งาน

1.1.3 ผู้ใช้งานดูข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS และการเกิดอุบัติเหตุของรถโดยสารสาธารณะ ที่เป็นรายงานในรูปแบบของกระดาษ และต้องรอให้เจ้าหน้าที่ส่งพิมพ์รายงานข้อมูลออกมาให้เท่านั้น ซึ่งความต้องการของผู้ใช้งานต้องการรายงานที่สามารถดูได้จากคอมพิวเตอร์

1.1.4 ข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS ไม่ได้นำมาใช้ประโยชน์อย่างอื่น ซึ่งผู้ใช้งานต้องการให้ข้อมูลระบบ GPS นำมาใช้ประโยชน์ร่วมกับข้อมูลพนักงานขับรถ เพื่อการคัดเลือกพนักงานขับรถที่มีคุณภาพ และมอบหมายรถโดยสารสาธารณะให้กับพนักงานขับรถที่ไม่เสี่ยงต่อการเกิดอุบัติเหตุในการขับรถโดยสารสาธารณะ เป็นต้น

จากผลที่ผู้วิจัยได้ทำการศึกษา วิเคราะห์ความต้องการ และศึกษาจากงานวิจัยต่างๆ ที่เกี่ยวข้อง ได้พัฒนาค้างข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS เพื่อแก้ปัญหาระบบงานเดิมที่ไม่ได้ออกแบบมารองรับปริมาณข้อมูลมากส่งผลให้การประมวลผลช้า รูปแบบของข้อมูลที่แตกต่างกันจากโปรแกรมที่หน่วยงานนำมาใช้งาน ไม่สามารถนำมาใช้ประโยชน์ในด้านข้อมูลการเกิดอุบัติเหตุของรถโดยสารสาธารณะ จึงพัฒนาค้างข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS มาเพื่อแก้ปัญหา ซึ่งผลจากพัฒนาค้างข้อมูลสามารถเป็นใช้ประโยชน์ของข้อมูลในการตัดสินใจของผู้บริหาร ผู้ใช้งานได้ และนำข้อมูลไปใช้กับข้อมูลการเกิดอุบัติเหตุรถโดยสาร เพื่อประโยชน์ในการทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ ดังนั้นผู้วิจัยจึงพัฒนาค้างข้อมูล GPS และทำเหมืองข้อมูลการทำนายป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด เพื่อแก้ปัญหาที่เกิดขึ้น

1.2 การออกแบบคลังข้อมูล

การออกแบบคลังข้อมูล ผู้วิจัยได้ออกแบบโครงสร้างคลังข้อมูล ประกอบด้วย 2 ส่วน ได้แก่ คลังข้อมูลระบบ GPS และคลังข้อมูลการเกิดอุบัติเหตุรถโดยสารสาธารณะ

1.2.1 คลังข้อมูลระบบ GPS โดยข้อมูลที่นำมาเพื่อออกแบบคลังข้อมูล นำมาจากฐานข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS ซึ่งเป็นฐานข้อมูลเชิงสัมพันธ์ (Relational Database) มีการกำหนดความสัมพันธ์กันของข้อมูล ผู้วิจัยได้ทำการจัดการข้อมูล และวิเคราะห์ข้อมูล ที่จำเป็นต้องทราบ ก่อนทำการปรับเปลี่ยน และนำเข้าเป็นคลังข้อมูลไม่มีโครงสร้าง

(Unstructured Data) หรือ NoSQL จึงจำเป็นต้องทราบถึงลักษณะของข้อมูลที่จะถูกจัดเก็บเพื่อการวิเคราะห์ข้อมูล และลักษณะของข้อมูลที่ถูกจัดเก็บในคลังข้อมูลระบบ GPS

- 1) ข้อมูลรถโดยสารสาธารณะ เช่น รหัสรถ ชื่อหมายเลขรถ ทะเบียนรถ ประเภทรถ
- 2) ข้อมูลกลุ่มรถโดยสารสาธารณะ เช่น รหัสกลุ่ม ชื่อกลุ่มรถ
- 3) ข้อมูลประเภทรถโดยสาร เช่น รหัสประเภทรถ ชื่อประเภทรถ ชื่อยี่ห้อ ชื่อรุ่น
- 4) ข้อมูลสถานี เช่น รหัสสถานี ชื่อสถานี พิกัดตำแหน่งสถานี รัศมีสถานี ชื่อที่ตั้ง
- 5) ข้อมูลสถานีที่ เช่น รหัสสถานีที่ ชื่อสถานีที่ อำเภอ/เขต ตำบล/แขวง จังหวัด ภูมิภาค รหัสไปรษณีย์ ประเทศ พิกัดตำแหน่งสถานี
- 6) ข้อมูลติดตามรถโดยสารสาธารณะ เช่น รหัสซิม GPS สถานะรถยนต์ พิกัดตำแหน่ง ความเร็วรถ วันที่ เวลา รหัสพนักงานขับรถ ระยะทาง

จากข้อมูลที่ได้จากฐานข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS เดิม ทำการกำหนดความสัมพันธ์กันของข้อมูล เพื่อใช้ในการออกแบบตาราง และนำเข้าเป็นคลังข้อมูลไม่มีโครงสร้าง (UnStructured Data) หรือ NoSQL ดังนี้



ตารางที่ 4.1 แสดงข้อมูลติดตามรถโดยสารสาธารณะ แบบไม่มีโครงสร้าง

ชื่อคอลเลกชัน(ตาราง)	GPS_DETAIL			
คำอธิบาย	ข้อมูลติดตามรถโดยสารสาธารณะ แบบไม่มีโครงสร้าง หรือ NoSQL			
คีย์หลัก				
ชื่อฟิลด์	ชนิด	ขนาด	ความหมาย	ตัวอย่างข้อมูล
_ID	Object Id	-		558906e32b571275dca55fbb
R_STATUS	String	-	สถานะรถ	30
R_TIME	String	-	เวลา	2012-07-31T:14:40:11.0
TRUCK_BOX_ID	NumberLong	-	รหัสกล่อง GPS	144180100120
LOCATE_ID	NumberLong	-	รหัสตำแหน่งสถานที่	1011030065
R_LON	Int	-	เส้นลองจิจูด	100.6208170000000000
R_LAT	Int	-	เส้นละติจูด	13.6719000000000010
R_SPEED	Int	-	ความเร็วรถ	50
R_DISTANCE	Int	-	ระยะทาง	5692.50
TRUCK_RADIO_ID	numberLong	-	รหัสซิม	875167730
TRUCK_NAME	String	-	หมายเลขรถ	48-42
TRUCK_LICENSE	String	-	ทะเบียนรถ	10-3485
TRUCK_ID	String	-	รหัสรถ	TREVO-00087
LOCATE_NAME	String	-	ชื่อสถานที่ไทย	บางนา
TAM_ENAME	String	-	ชื่อสถานที่อังกฤษ	BANG NA
TRUCKTYPE_ID	String	-	รหัสกลุ่มรถ	TY-00007
TRUCKTYPE_NAME	String	-	ชื่อกลุ่มรถ	บัสโดยสารประจำทาง

จากตารางที่ 4.1 แสดงข้อมูลติดตามรถโดยสารสาธารณะ แบบไม่มีโครงสร้าง ซึ่งบอกถึงขอบเขตประเภทของคอลเลกชัน (ตาราง) และถูกสร้างจากการกำหนดตารางความสัมพันธ์ของระบบเดิมมีตาราง 6 ตาราง คือ GPS_DATA TABLE, TRUCK TABLE, TRUCK_GROUP TABLE, TRUCKTYPE TABLE, EVENT TABLE และ Locatetion table

ตารางที่ 4.2 GPS_DATA TABLE สำหรับเก็บข้อมูลติดตามรถโดยสารสาธารณะ

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
R_STATUS	Integer	สถานะรถ	"30" = ปกติ
R_TIME	Date	วันที่	2012-07-31T:14:40:11.0
TRUCKBOX_ID	Varchar	รหัสกล่องGPS	144180100120
LOCATE_ID	Integer	รหัสสถานที่	6604012201
R_LON	Float	เส้นลองจิจูด	100.6208170000000000
R_LAT	Float	เส้นละติจูด	13.6719000000000010
R_SPEED	Integer	ความเร็วรถ	85 ก.ม.
DRIVER_ID	Varchar	รหัสคนขับ	1023
R_TIMENOW	Varchar	เวลา	00:21.0
SERIAL_SIM	Varchar	รหัสซิม	144180100120
R_DISTANCE	Float	ระยะทาง	244.1

ตารางที่ 4.3 EVENT TABLE สำหรับเก็บข้อมูลสถานะของรถโดยสาร

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
EVENT_ID	Integer	รหัสสถานะ	"45"
EVENT_NAME	Varchar	ชื่อสถานะ	รถวิ่งความเร็วเกิน 10 นาที

ตารางที่ 4.4 TRUCK TABLE สำหรับเก็บข้อมูลรถโดยสาร

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
TRUCK_ID	Varchar	รหัสรถยนต์	TREVO-00001
TRUCK_NAME	Varchar	ชื่อรถยนต์	V9999-1
TRUCK_REGIS	Varchar	ชื่อทะเบียนรถยนต์	10-9999
TRUCKTYPE_ID	Varchar	รหัสประเภทรถยนต์	TY-00001
TRUCKGROUP_ID	Varchar	รหัสประเภทรถยนต์	TY-00001

ตารางที่ 4.5 TRUCK_GROUP TABLE สำหรับเก็บข้อมูลการจัดกลุ่มรถโดยสาร

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
TRUCKGROUP_ID	Varchar	รหัสกลุ่มรถยนต์	TG-00001
TRUCKGROUP_NAME	Varchar	ชื่อกลุ่มรถยนต์	รถสาย9999

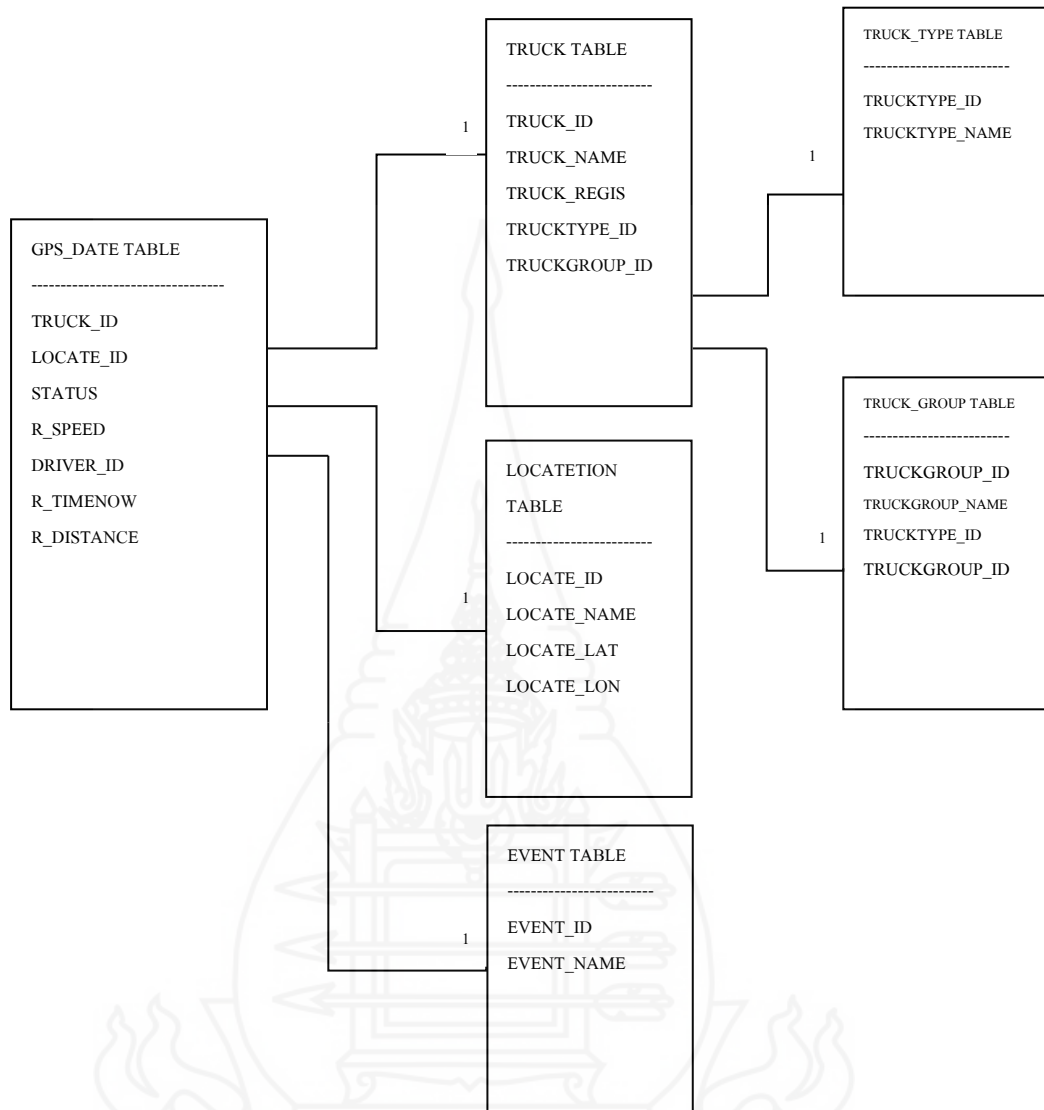
ตารางที่ 4.6 TRUCK_TYPE TABLE สำหรับเก็บข้อมูลการประเภทรถโดยสาร

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
TRUCKTYPE_ID	Varchar	รหัสประเภทรถยนต์	TY-00001
TRUCKTYPE_NAME	Varchar	ชื่อประเภทรถยนต์	รถโดยสารสายตะวันออก
TRUCK_BRAND	Varchar	ชื่อยี่ห้อรถยนต์	SUNLONG,BENZ

ตารางที่ 4.7 LOCATETION TABLE สำหรับเก็บข้อมูลตำแหน่งที่ตั้งต่างๆ

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
LOCATE_ID	Integer	รหัสตำแหน่งที่ตั้ง	6602061001
LOCATE_NAME	Varchar	ชื่อตำแหน่งที่ตั้ง	หนองไทร
LOCATE_LAT	Float	เส้นลองจิจูด	14.513819
LOCATE_LON	Float	เส้นลองจิจูด	102.781023
LOCATE_PROV	Varchar	จังหวัด	บุรีรัมย์
LOCATE_COUNTRY	Varchar	ประเทศ	ไทย

ผู้วิจัยได้ กำหนดการเชื่อมโยงกันข้อมูล 6 ตาราง คือ ตารางติดตามรถโดยสาร
 สาธารณะ (GPS_DATA TABLE) ตารางรถโดยสารสาธารณะ (TRUCK TABLE) ตารางสถานะของ
 รถโดยสารสาธารณะ (EVENT TABLE) และตารางตำแหน่งที่ตั้ง (LOCATETION TABLE)
 เชื่อมโยงกัน และตารางรถโดยสาร (TRUCK TABLE) เชื่อมโยงกับตารางประเภทรถโดยสาร
 สาธารณะ (TRUCK_TYPE TABLE) และตารางจัดกลุ่มรถโดยสารสาธารณะ (TRUCK_GROUP
 TABLE) เพื่อสร้างคลังข้อมูลใหม่ แบบไม่มีโครงสร้าง (Unstructured Data) ดังภาพที่ 4.1



ภาพที่ 4.1 แสดงความสัมพันธ์ข้อมูลเพื่อสร้างคลังข้อมูลไม่มีโครงสร้าง

1.2.2 คลังข้อมูลการเกิดอุบัติเหตุรถโดยสารสาธารณะ เป็นข้อมูลที่ได้จากข้อมูลฝ่ายอุบัติเหตุ ฝ่ายทะเบียนบุคคล และคลังข้อมูลระบบ GPS ได้ทำการปรับปรุงข้อมูล และกำหนดความสัมพันธ์ของตารางข้อมูล ก่อนนำเข้าสู่คลังข้อมูล การวิเคราะห์ข้อมูลสำหรับการออกแบบคลังข้อมูลนั้น จำเป็นต้องแสดงลักษณะของข้อมูล ในฐานข้อมูลการเกิดอุบัติเหตุรถโดยสารสาธารณะ ดังนี้

1) ข้อมูลการเกิดอุบัติเหตุรถโดยสาร เช่น หมายเลขรถ ยี่ห้อรถ รุ่นรถ วันที่ เวลาเกิดเหตุ สถานที่ ชื่อคนขับ คู่กรณี รถคู่กรณี ลักษณะการชน ความเสียหาย และขั้นตอนดำเนินงาน

- 2) ข้อมูลความเสียหายรถคู่กรณี เช่น คู่กรณี รายการความเสียหาย ค่าใช้จ่าย
 - 3) ข้อมูลรถคู่กรณี เช่น รถยนต์ ทะเบียน ยี่ห้อ รุ่น ประกันภัยรถยนต์
 - 4) ข้อมูลประวัติพนักงาน เช่น ชื่อ สกุล อายุ อายุงาน การศึกษา ภูมิลำเนา
 - 5) ข้อมูลความเสียหายของรถโดยสาร เช่น อุบัติเหตุ(รถ/คน/อื่นๆ) มูลค่าความเสียหาย จำนวนคน เป็นต้น
 - 6) ข้อมูลประวัติรถโดยสาร เช่น หมายเลขรถ ทะเบียน ยี่ห้อ รุ่น ประเภทรถ
 - 7) ข้อมูลประวัติประเภทรถโดยสาร เช่น รหัสประเภท เส้นทางเดินรถ
- จากข้อมูลที่ได้ สามารถกำหนดมุมมอง เพื่อใช้การออกแบบและสร้างตาราง ดังนี้
- ตารางที่ 4.8 ข้อมูลโครงสร้าง VPANALYTICS.DBF

ชื่อตาราง	VPANALYTICS			
คำอธิบาย	ข้อมูลบันทึกผลการดำเนินงานพนักงานขับรถ			
คีย์หลัก	VPMYY			
ชื่อฟิลด์	ชนิด	ขนาด	ความหมาย	ตัวอย่างข้อมูล
VP_MYY	Character	7	ช่วงเวลา	08/2016
VP_DATE	Date	8	วันที่	12/08/2016
VP_NEW	Character	1	ผู้สมัครใหม่	NEW11, 222614
VP_DRV	Character	7	พนักงานเก่า	222614
VP_SIGN	Character	3	คำนำหน้าชื่อ	นาย, นาง, นส.
VP_FNAME	Character	25	ชื่อพนักงาน	เดชรินทร์
VP_LNAME	Character	25	นามสกุล	บุญเสมา
VP_BIRTH	Date	8	วันเกิด	01-01-1968
VP_AGE	Character	15	อายุ	48 ปี 7 เดือน
VP_AGE2	Character	5	อายุ*	AGE_4
VP_EDU	Character	4	การศึกษา*	ED_1
VP_PHUMIPAK	Character	4	ภูมิภาค*	PM_1
VP_WEIGHT	Numeric	3	น้ำหนัก	55
VP_WEIGHT2	Character	3	น้ำหนัก*	W_1
VP_HIGH	Numeric	3	ส่วนสูง	160
VP_HIGH2	Character	3	ส่วนสูง*	H_1

ตารางที่ 4.8 (ต่อ)

ชื่อฟิลด์	ชนิด	ขนาด	ความหมาย	ตัวอย่างข้อมูล
VP_TIMEWORK	Numeric	2	ชั่วโมงทำงาน/วัน	8
VP_TIMEWORK2	Character	4	ชั่วโมงทำงาน/วัน*	TK_2
VP_GWORK	Numeric	5.2	ประสบการณ์	3
VP_PROVINCE	Character	2	รหัสจังหวัด	01 = "กรุงเทพมหานคร"
VP_GWORK2	Character	4	ประสบการณ์*	WK_3
VP_ROUTE	Character	5	เส้นทางเดินรถ	b048
VP_TYPEBUS	Character	15	ยี่ห้อรถ	BENZ
VP_CHON	Character	1	ลักษณะอุบัติเหตุ	1 = "รถบรรทุกชนรถอื่น"
VP_PLACE	Character	1	สถานที่อุบัติเหตุ	3 = "สถานที่ภายนอก"
VP_OUTPUT1	Character	100	ทำนาย1	ผ่าน
VP_OUTPUT2	Character	100	ทำนาย2	ผ่าน
VP_OUTPUT3	Character	100	ทำนาย3	ผ่าน

จากตารางที่ 4.8 การออกแบบสร้างตารางข้อมูลสำหรับการทำเหมืองข้อมูลเพื่อการทำนาย และสร้างโปรแกรมทำนายการเกิดอุบัติเหตุรถโดยสาร ซึ่งถูกสร้างจากการกำหนดตารางความสัมพันธ์ของระบบเดิมมี 8 ตาราง คือ ACCIDENT TABLE, UDAMAGEOTH TABLE, UOTHER TABLE, EMPLOYEE1 TABLE, UDAMAGECOM TABLE, UPROCESS TABLE, BUS_NO TABLE และ BUSTYPE TABLE ดังนี้

ตารางที่ 4.9 ACCIDENT TABLE สำหรับเก็บข้อมูลการเกิดอุบัติเหตุรถโดยสาร

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
AC_NO	Character	เลขอุบัติเหตุ	ACD-0001-59
AC_DATE	Date	วันที่บันทึก	10/10/2014
AC_GDATE	Date	วันที่เกิดเหตุ	08/10/2014
AC_GTIME	Character	เวลาเกิดเหตุ	05:00 น
AC_PLACE	Character	รหัสสถานที่	1= บริษัท, 2=สถานี, 3 =ภายนอก
AC_BOOT	Character	ชื่อสถานี	หมอชิต/เอกมัย
AC_ADD	Character	ที่ตั้ง	สี่แยกบางนา
AC_DETAIL	Character	รายละเอียดการชน	ชนท้ายแท็กซี่
AC_CHON	Character	รหัสการชน	1= "รถบริษัทชนรถอื่น"
AC_TF	Character	ความถูกผิด	1=บริษัทถูก,2=บริษัทผิด,3=ผิดร่วม
AC_TELL	Character	การแจ้งเหตุ	1= คนขับแจ้งเหตุ 2=ผู้อื่นแจ้งเหตุ
AC_WORK1	Character	แจ้งบริษัทประกันภัย	1=แจ้ง 0=ไม่แจ้ง
AC_WORK2	Character	ดูสถานที่เกิดเหตุ	1=ไป 0=ไม่ไป
AC_BCODE	Character	รหัสรถโดยสาร	101
AC_SIDE	Character	เบอร์ข้างรถ	99-99
AC_CIRCLE	Character	เบอร์วงกลม	101
AC_REGIS	Character	ทะเบียนรถ	10-9999
AC_TYPE	Character	รหัสประเภทรถ	10
AC_BRAND	Character	ยี่ห้อรถ	BENZ
AC_BUSDET	Character	รายละเอียดรถ	เลขรถ 99-99 ยี่ห้อ SCANIA รุ่น...
AC_DRCODE	Character	รหัสคนขับ	220022
AC_DRNAME	Character	ชื่อคนขับ	นาย สามารถ นาคี
AC_GAGE	Character	อายุ	40 ปี 10 เดือน
AC_GWORK	Character	อายุงาน	5 ปี 3 เดือน
AC_RECCODE	Character	รหัสผู้บันทึก	222011
AC_RECNAME	Character	ชื่อผู้บันทึก	นส. วาสนา นที
AC_JOB	Character	การปฏิบัติงาน	15/10/2015
AC_JDATE	Date	วันที่ปฏิบัติงาน	3=อื่นๆ
AC_PRICE	Numeric	มูลค่าความเสียหาย	60,000

ตารางที่ 4.10 UDAMAGEOTH TABLE สำหรับเก็บข้อมูลความเสียหายของรถยนต์คู่กรณี

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
UD_NO	Character	เลขอุบัติเหตุ	ACD-0001-59
UD_1TYPE	Character	รหัสความเสียหาย1	1=รถ
UD_1DET	Character	ความเสียหาย1	กระจกหักฝั่งขวา
UD_1TOTAL	Numeric	ประเมินความเสียหาย	1,500
UD_2TYPE	Character	รหัสความเสียหาย2	2=คน
UD_2DET	Character	ความเสียหาย2	บาดเจ็บกระดูกขาหัก
UD_2TOTAL	Numeric	ประเมินความเสียหาย	250,000
UD_3TYPE	Character	รหัสความเสียหาย3	3=อื่นๆ
UD_3DET	Character	ความเสียหาย3	รื้อกำแพงพัง
UD_1TOTAL	Numeric	ประเมินความเสียหาย	30ม000
UD_LOGI	Logical	สถานะข้อมูล	.f.= ต้องการ .t.=ยกเลิก

ตารางที่ 4.11 UOTHER TABLE สำหรับเก็บข้อมูลรายละเอียดครรถคู่กรณี

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
UT_NO	Character	เลขอุบัติเหตุ	ACD-0001-59
UT_TYPE	Character	รหัสคู่กรณี	1=รถ, 2=คน, 3=อื่นๆ
UT_1SIDE	Character	หมายเลขรถ	9999-99
UT_1BRAND	Character	ยี่ห้อรถ	BENZ
UT_1COLOR	Character	สีรถ	ขาว/ฟ้า
UT_1PRAKAN	Character	ประกันรถ	ประเภท3
UT_1NAME	Character	ชื่อคนขับ	นาย สม ดีมาก
UT_1TEL	Character	เบอร์ติดต่อ	084-2222-222
UT_2NAME	Character	ชื่อคู่กรณี	นาย สม ดีมาก
UT_2AGE	Numeric	อายุ	40
UT_3OTHER	Character	ความเสียหายอื่นๆ	ชนขอบทางเท้า

ตารางที่ 4.12 EMPLOYEE1 TABLE สำหรับเก็บข้อมูลประวัติพนักงาน

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
RRE_CODE	Character	รหัสพนักงาน	222002
RRE_STATUS	Logical	สถานพนักงาน	.T.=ทำงาน, .F.=ลาออก
RRE_SIGN	Character	ค่าน้ำหน้า	นาง, นาย, นส.
RRE_FNAME	Character	ชื่อ	สามารถ
RRE_LNAME	Character	นามสกุล	กิจการดี
RRE_START	Date	วันที่เริ่มทำงาน	15/01/2013
RRE_END	Date	วันที่ลาออก	30/01/2015
RRE_PASE	Character	เพศ	ชาย, หญิง
RRE_SALARY	Numeric	เงินเดือน	20,000
RRE_MARY	Character	แต่งงาน	แต่งงาน, โสด
RRE_SON	Numeric	ประกันรถ	1
RRE_BIRTH	Date	วันเกิด	30/01/1980
RRE_PASSPORT	Character	เลขบัตรประชาชน	111-2222-222-111
RRE_STDUY	Character	วุฒิการศึกษา	มัธยมต้น
RRE_CADD1	Character	ที่อยู่ปัจจุบัน1	10/20 บางป่อ
RRE_CADD2	Character	ที่อยู่ปัจจุบัน2	บางจาก
RRE_CADD3	Character	ที่อยู่ปัจจุบันจังหวัด	กรุงเทพฯ
RRE_RADD1	Character	ที่อยู่ภูมิลำเนาเดิม1	10 หมู่1 หนองนา
RRE_RADD2	Character	ที่อยู่ภูมิลำเนาเดิม2	เมือง
RRE_RADD3	Character	ภูมิลำเนาเดิมจังหวัด	บุรีรัมย์

ตารางที่ 4.13 UDAMAGECOM TABLE สำหรับเก็บข้อมูลความเสียหายของรถบริษัทฯ

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
UC_NO	Character	เลขอุบัติเหตุ	ACD-0001-59
UC_CAR	Character	สถานะรถ	รถ
UC_CDET	Character	รายละเอียดความเสียหายรถยนต์	กระจกหน้าแตก
UC_CDAMAGE	Numeric	จำนวนความเสียหาย	25,000
UC_KON	Character	สถานะคน	คน
UC_PERSON	NUMERIC	จำนวนคน	3
UC_PDET	Character	รายละเอียดความเสียหายบุคคล	บาดเจ็บนอน รพ. ชื่อ.....

ตารางที่ 4.14 UPROCESS TABLE สำหรับเก็บข้อมูลขั้นตอนทำงานและรายจ่าย

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
UP_NO	Character	เลขอุบัติเหตุ	ACD-0001-59
UP_1COST	Numeric	ค่าใช้จ่ายครั้งที่1	20,000
UP_1DET	Character	รายละเอียดครั้งที่1	ชำระค่ากรณี...ครั้งที่..
UP_2COST	Numeric	ค่าใช้จ่ายครั้งที่2	20,000
UP_2DET	Character	รายละเอียดครั้งที่2	ชำระค่ากรณี...ครั้งที่..
UP_3COST	Numeric	ค่าใช้จ่ายครั้งที่3	20,000
UP_3DET	Character	รายละเอียดครั้งที่3	ชำระค่ากรณี...ครั้งที่..
UP_DATE	Character	วันที่	10/10/2015
UP_CLOSE	Character	ปิดประเด็น	ปิด
UP_TOTAL	Numeric	รวมเป็นเงิน	60,000

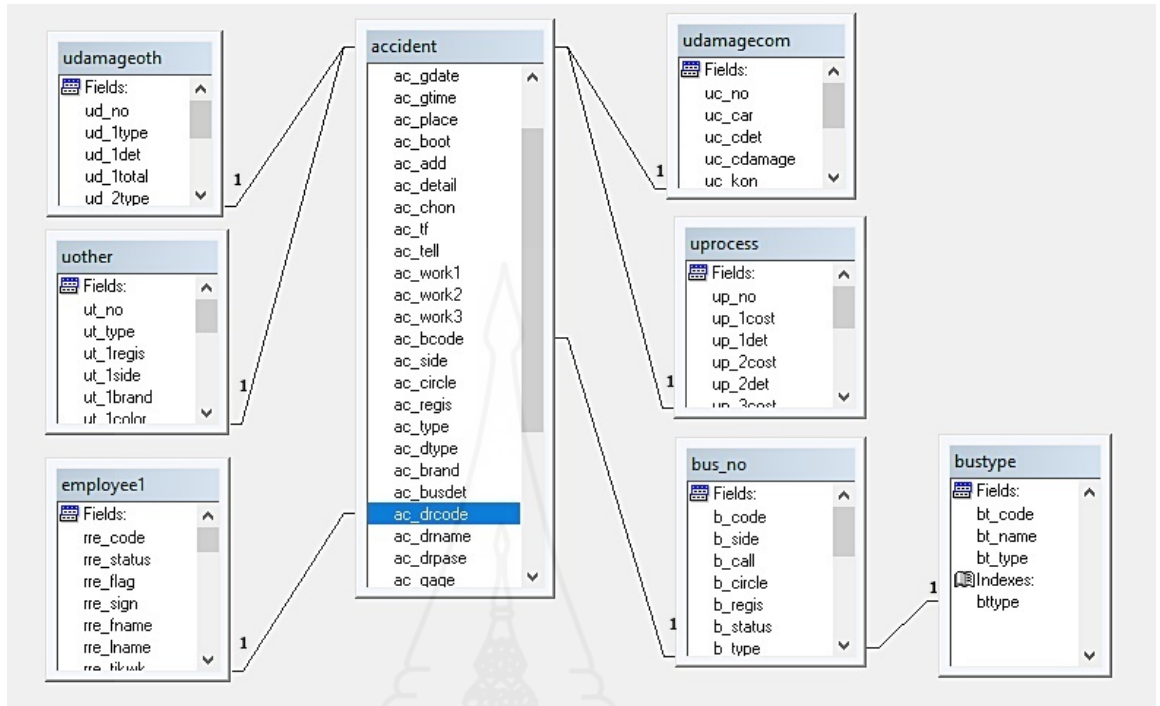
ตารางที่ 4.15 BUS_NO TABLE สำหรับเก็บข้อมูลประวัติรถโดยสาร

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
B_CODE	Character	รหัสรถโดยสาร	101
B_SIDE	Character	หมายเลขข้างรถ	99-99
B_CALL	Character	ชื่อเรียกรถ	99
B_CIRCLE	Character	หมายเลขรถวงกลม	101
B_REGIS	Character	ทะเบียนรถ	10-999
B_STATUS	Logical	สถานะรถ	.F.=ใช้งาน
B_TYPE	Character	ประเภทรถ	รถสาย 99
B_KMNOW	Numeric	จำนวนระยะทาง	90,099

ตารางที่ 4.16 BUSTYPE TABLE สำหรับเก็บข้อมูลประเภทรถโดยสาร

ชื่อฟิลด์	ชนิด	ความหมาย	ตัวอย่างข้อมูล
BT_CODE	Character	รหัสประเภทรถ	10
BT_NAME	Character	ชื่อประเภทรถ	รถสาย99

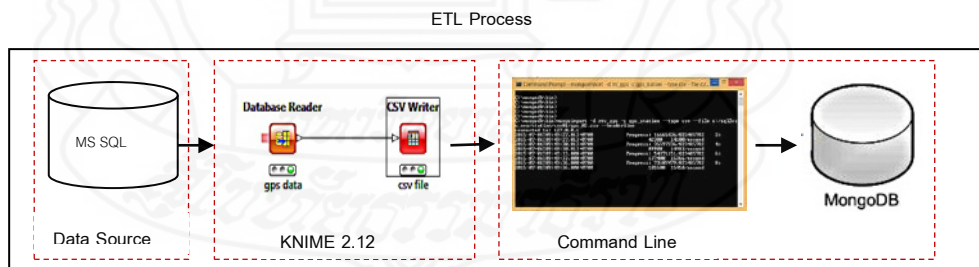
ผู้วิจัยได้ กำหนดการเชื่อมโยงความสัมพันธ์ระหว่างตาราง 8 ตาราง คือ ตารางการเกิดอุบัติเหตุรถโดยสาร (ACCIDENT TABLE) ตารางความเสียหายของรถยนต์ตู้กรณี (UDAMAGEOTH TABLE), ตารางรายละเอียดรถตู้กรณี (UOTHER TABLE), ตารางประวัติพนักงาน(Employee1 Table), ตารางความเสียหายของรถบริษัทฯ (UDAMAGECOM TABLE), ตารางขั้นตอนทำงานและรายจ่าย (UPROCESS TABLE), ตารางประวัติรถโดยสาร (BUS_NO TABLE) เชื่อมโยงกัน และ ตารางประวัติรถโดยสาร (BUS_NO TABLE) เชื่อมโยงกับตารางประเภทรถโดยสาร (BUSTYPE TABLE) ดังภาพที่ 4.2



ภาพที่ 4.2 แสดงความสัมพันธ์ข้อมูลอุบัติเหตุโดยสาร

1.3 การพัฒนาคลังข้อมูล

ผู้วิจัยนำข้อมูลเข้าสู่คลังข้อมูลด้วยกระบวนการ ETL (Extract Transform and Load : ETL) คลังข้อมูลระบบ GPS โดยใช้โปรแกรม KNIME v2.12 และคำสั่ง Command Line ดังภาพที่ 4.3



ภาพที่ 4.3 ขั้นตอนการพัฒนาคลังข้อมูล GPS

ภาพที่ 4.3 แสดงขั้นตอนการพัฒนาคลังข้อมูล GPS เพื่อนำข้อมูลเข้าสู่คลังข้อมูล MongoDB เริ่มจาก Data Source ด้วยการนำข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS จากฐานข้อมูลของ MS SQL จากนั้นผู้วิจัยใช้เครื่องมือ Database Reader Node ของโปรแกรม KNIME v2.12 ทำการติดต่อข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS โดยการกำหนดค่าพารามิเตอร์ของ Database Reader Node ได้แก่ Database Driver, Database URL, UserName for MS SQL, Fetch Metadata, SQL Statement

จากนั้นแปลงรูปแบบข้อมูลเป็นเท็กซ์ไฟล์ (CSV) ด้วย CSV Writer Node ของโปรแกรม KNIME v2.12 สุดท้ายนำเข้าคัดงข้อมูล MongoDB โดยใช้คำสั่ง Command Line ของ โปรแกรม MongoDB v2.6.7

ขั้นตอนการพัฒนาคลังข้อมูล GPS

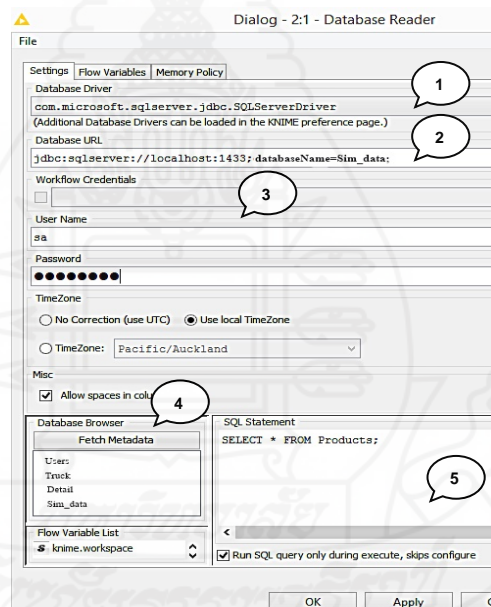
1.3.1 การติดต่อฐานข้อมูลระบบ GPS

1.3.2 การแปลงรูปแบบข้อมูล

1.3.3 การนำข้อมูลเข้าคัดงข้อมูล

การติดต่อฐานข้อมูลระบบ GPS

การนำข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS เดิมเข้าสู่คลังข้อมูลระบบ GPS ผู้วิจัยใช้เครื่องมือ Database Reader Node โปรแกรม KNIME v2.12 ติดต่อฐานข้อมูลของ MS SQL ดังแสดงภาพที่ 4.4



ภาพที่ 4.4 การใช้เครื่องมือ Database Reader Node ติดต่อ MS SQL

ภาพที่ 4.4 อธิบายการกำหนดค่าได้ดังนี้

หมายเลข 1 ระบุ Database Driver เป็น

com.microsoft.sqlserver.jdbc.SQLServerDriver

หมายเลข 2 ระบุ Database URL เป็น

Jdbc:sqlserver://localhost:1433;databaseName=GPS_DATA

หมายเลข 3 ระบุ User Name และ Password ของ MS SQL

หมายเลข 4 คลิก Fetch Metadata เพื่อแสดงตารางของฐานข้อมูล

หมายเลข 5 SQL Statement ระบุคำสั่งรูปแบบประโยคภาษา SQL เพื่อแสดงข้อมูล

ตัวอย่าง

"SELECT

D_GPS.R_STATUS,D_GPS.R_TIME,D_GPS.TRUCK_BOX_ID,D_GPS.ADMIN_ID,D_GPS.R_LON,D_GPS.R_LAT,D_GPS.R_SPEED,D_GPS.R_DISTANCE,
FROM

[GPS_DATA].[DBO].[Z144180100120DATA] AS D_GPS "

การแปลงรูปแบบข้อมูล

นำผลตารางข้อมูล MS SQL ที่ได้จาก Database Reader Node ไปแปลงเป็นเท็กซ์ไฟล์

ด้วย CSV ด้วย เครื่องมือ CSV Writer Node ของโปรแกรม KNIME v2.12

การนำข้อมูลเข้าคลังข้อมูล

ขั้นตอนสุดท้ายคือการนำเข้าคลังข้อมูล ข้อมูลที่ถูกแปลงเป็นรูปแบบของเท็กซ์

ไฟล์(CSV) ใช้คำสั่ง Command Line เพื่อส่งข้อมูลเข้าคลังข้อมูลระบบ GPS ของ MongoDB ดังภาพที่ 4.5

```
C:\mongodb\bin>mongoimport -d rrc_gps -c gps_station --type csv --file c:/sql2csv_new/station/csv01/gps_02.csv --headerline
```

ภาพที่ 4.5 คำสั่งนำเข้าข้อมูลของ MongoDB ด้วย Command Line

พารามิเตอร์ของคำสั่ง "mongoimport"

-d	หมายถึง	ชื่อฐานข้อมูล
-c	หมายถึง	ชื่อ Collection หรือตารางข้อมูล
--type	หมายถึง	นามสกุลของเท็กซ์ไฟล์ (CSV)
--file	หมายถึง	ชื่อของเท็กซ์ไฟล์ เช่น GPS_02.CSV
--headerline	หมายถึง	กำหนดบรรทัดแรกเป็นชื่อฟิลด์

```

" id" : ObjectId("558906e32b571275dca55fbb"),
"R_STATUS" : 30,
"R_TIME" : "2012-07-31T14:40:11.0",
"TRUCK_BOX_ID" : NumberLong(144180100120),
"ADMIN_ID" : NumberLong(1011030065),
"R_LON" : 100.6208170000000000,
"R_LAT" : 13.6719000000000010,
"R_SPEED" : 0,
"R_DISTANCE" : 5692.5000000000000000,
"TRUCK_RADIO_ID" : NumberLong(875167730),
"TRUCK_NAME" : " 42",
"TRUCK_LICENSE" : "10-3485",
"TRUCK_ID" : "TREVO-00087",
"TAM_TNAME" : "บางนา",
"TAM_ENAME" : "BANG NA",
"TRUCKTYPE_ID" : "TY-00007",
"TRUCKTYPE_NAME" : "บัสโดยสารประจำทาง"

```

ภาพที่ 4.6 รูปแบบข้อมูลคอลเลกชัน MongoDB
ข้อมูลคอลเลกชันในคลังข้อมูล MongoDB
["Key ":"Value"]

2. สร้างแบบจำลองการทำนาย

งานวิจัยนี้ยึดแนวทางการทำเหมืองข้อมูลตามกระบวนการมาตรฐานอุตสาหกรรม ที่มีขั้นตอนการสร้างแบบจำลองการทำนาย ชื่อว่า CRISP-DM (Cross-Industry Standard Process for Data Mining) ประกอบด้วย 6 ขั้นตอน ดังนี้

2.1 การทำความเข้าใจปัญหา (Business Understanding)

ผู้วิจัยได้ศึกษาข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS และการเกิดอุบัติเหตุของรถโดยสารสาธารณะ และวิเคราะห์ปัญหาที่เกิดขึ้นกับผู้ประกอบการเดินรถพบว่า

2.1.1 ข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS ถึงแม้ว่าข้อมูลสามารถตอบสนองความต้องการผู้ใช้งาน เช่น การหาตำแหน่งของรถ ความเร็วที่ใช้ในการขับรถ และระบบการตรวจจับต่างๆของระบบ GPS เป็นต้น ที่แสดงผลให้ทราบได้ทันทีจากโปรแกรมระบบ GPS แต่เมื่อข้อมูลถูกบันทึกมากขึ้น มีขนาดใหญ่ หรือที่เรียกว่า บิ๊กดาตา (Big Data) ส่งผลให้ระบบ GPS ทำการประมวล และสืบค้นข้อมูลช้ามาก ต้องทำการปรับลดขนาดข้อมูลลงเพื่อให้ระบบใช้งานได้ปกติจากการปรับลดข้อมูลนั้น ส่งผลให้การวิเคราะห์ข้อมูลที่ต้องการขาดความต่อเนื่อง

2.1.2 การเกิดอุบัติเหตุของรถโดยสารสาธารณะ พบว่าพนักงานขับรถเป็นอันดับต้นๆที่ก่อให้เกิดอุบัติเหตุรถ สาเหตุมาจากหลายปัจจัย เช่น อายุ การศึกษา ประสบการณ์ขับรถโดยสาร การดื่มสุรา จำนวนชั่วโมงขับรถโดยสาร อายุงาน และเส้นทางเดินรถ และระยะทาง เป็น

ต้น ถึงแม้ว่าผู้ประกอบการเดินรถจะมีมาตรการ แผนการป้องกัน หรือนำเทคโนโลยีที่ทันสมัย ก็ไม่สามารถป้องกันการเกิดอุบัติเหตุที่เกิดขึ้นทั้งหมดได้

ดังนั้นการนำหลักการ Big Data มาปรับใช้กับคลังข้อมูลระบบ GPS และเครื่องมือทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ เพื่อช่วยเหลือการตัดสินใจ และคัดเลือกรถบุคคลได้อย่างเหมาะสมที่สุด

2.2 ทำความเข้าใจข้อมูล (Data Understanding)

ผู้วิจัยรวบรวมข้อมูลที่เกี่ยวข้องเพื่อนำไปวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล โดยได้สร้างคลังข้อมูลขึ้นมาดังกล่าวมาแล้วนั้น ได้แก่

2.2.1 คลังข้อมูลระบบ GPS เป็นข้อมูลที่ได้มาจากการใช้เครื่องมือโปรแกรม KNIME v.12 เพื่อนำข้อมูลออกมาให้อยู่ในรูปแบบของเท็กซ์ไฟล์ (CSV) และใช้คำสั่งนำข้อมูลเข้าคลังข้อมูลในรูปแบบ คำสั่ง Command line ของ MongoDB ประกอบด้วยข้อมูลติดตามรถโดยสารสาธารณะระหว่างพ.ศ. 2555-2559 จำนวนทั้งสิ้น 121,226,674 ระเบียบขนาดข้อมูล 58,009.27 ล้านไบต์ ดังภาพที่ 4.7

I_R_STATUS	I_R_TIME	S_TRUCK...	S_ADMIN...	I_R_SPEED	S_TRUCK...	S_TRUCK...	S_TAM_NAME	S_TAM_ENAME	S_TRUCKTYPE_ID	S_TRUCKTYPE_N...
30	30. Jan. 2013	144180100120	6605020505	98	48-42 (158)	10-3485	บางประกง	BANG PAKONG	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6605020505	96	48-42 (158)	10-3485	บางประกง	BANG PAKONG	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6605020505	96	48-42 (158)	10-3485	บางประกง	BANG PAKONG	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6605030801	94	48-42 (158)	10-3485	คลองสำพร	KHLONG TAM RU	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6605030813	98	48-42 (158)	10-3485	หนองไผ่แดง	NONG MAI DAE...	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6605030908	95	48-42 (158)	10-3485	แพมมวง	RAEM CHA BANG	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6605030814	96	48-42 (158)	10-3485	หนองรี	NONG RI	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6605030813	98	48-42 (158)	10-3485	หนองไผ่แดง	NONG MAI DAE...	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6605030801	90	48-42 (158)	10-3485	คลองสำพร	KHLONG TAM RU	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6605020505	98	48-42 (158)	10-3485	บางประกง	BANG PAKONG	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6605020505	97	48-42 (158)	10-3485	บางประกง	BANG PAKONG	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6605020505	96	48-42 (158)	10-3485	บางประกง	BANG PAKONG	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6605020505	98	48-42 (158)	10-3485	บางประกง	BANG PAKONG	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6605020507	93	48-42 (158)	10-3485	บางศรี	BANG WUA	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6605020507	95	48-42 (158)	10-3485	บางศรี	BANG WUA	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6605020507	92	48-42 (158)	10-3485	บางศรี	BANG WUA	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6605020507	92	48-42 (158)	10-3485	บางศรี	BANG WUA	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6605020509	95	48-42 (158)	10-3485	พิมพา	PHIMPHA	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6605020502	90	48-42 (158)	10-3485	ท่าข้าม	THA KHAM	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6605030808	93	48-42 (158)	10-3485	บ้านสวน	BAN SUAN	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6605030812	92	48-42 (158)	10-3485	หนองข้างคอก	NONG KHANG K...	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6604040205	98	48-42 (158)	10-3485	ราชาเทวะ	RACHA THEWA	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6604040205	94	48-42 (158)	10-3485	ราชาเทวะ	RACHA THEWA	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6604040201	98	48-42 (158)	10-3485	บางแก้ว	BANG KAE0	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6604040201	99	48-42 (158)	10-3485	บางแก้ว	BANG KAE0	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6604040201	90	48-42 (158)	10-3485	บางแก้ว	BANG KAE0	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6604040201	91	48-42 (158)	10-3485	บางแก้ว	BANG KAE0	TY-00007	บัสโดยสารประจำทาง
30	29. Jan. 2013	144180100120	6604040201	100	48-42 (158)	10-3485	บางนา	BANG NA	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6604040301	93	48-42 (158)	10-3485	บางเสาธง	BANG SAO THO...	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6604040301	95	48-42 (158)	10-3485	บางเสาธง	BANG SAO THO...	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6604040301	95	48-42 (158)	10-3485	บางเสาธง	BANG SAO THO...	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6604040301	93	48-42 (158)	10-3485	บางเสาธง	BANG SAO THO...	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6604040104	95	48-42 (158)	10-3485	บางบ่อ	BANG BO	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6604040104	94	48-42 (158)	10-3485	บางบ่อ	BANG BO	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6604040104	97	48-42 (158)	10-3485	บางบ่อ	BANG BO	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6604040104	96	48-42 (158)	10-3485	บางบ่อ	BANG BO	TY-00007	บัสโดยสารประจำทาง
30	30. Jan. 2013	144180100120	6605030803	91	48-42 (158)	10-3485	บางป่า	NA PA	TY-00007	บัสโดยสารประจำทาง

ภาพที่ 4.7 แสดงไฟล์ข้อมูลที่ได้จากคลังข้อมูลระบบ GPS

2.2.2 คลังข้อมูลการเกิดอุบัติเหตุรถโดยสารสาธารณะ เป็นข้อมูลที่ได้จากการใช้โปรแกรม Microsoft Visual FoxPro v9.0 เพื่อนำข้อมูลออกมาให้อยู่ในรูปแบบไฟล์เอ็กเซล (EXCEL) ประกอบด้วยข้อมูลการเกิดอุบัติเหตุจากพนักงานขับรถระหว่างพ.ศ. 2556-2559 จำนวนทั้งสิ้น 406 ระเบียบ คัดภาพที่ 4.8

S_ac_no	S_ac_date	S_ac_gdate	I_ac_place	S_ac_boot	S_ac_add	S_ac_detail	I_ac_chon	I_ac_tf	S_ac_bcode	S_ac_side	S_ac_cirde	S_ac_rn
ACD-0092-58	1/10/15	1/10/15	1	non	ลานจอดรถ...	คอกหลังไปชน...	3	2	209	9905-36	A79	10-4098
ACD-0075-57	1/7/14	1/7/14	2	air	ท่าอากาศยาน...	คอกไปโดน...	1	2	179	389-7	A52	33-2305
ACD-0074-57	1/5/14	1/5/14	2	air	บึงกิ้ง...	ล้มโดนรถมือ...	1	2	179	389-7	A52	33-2305
ACD-0073-57	1/3/14	1/3/14	1	non	อุโมงค์...	เอาสิบล้อไป...	4	2	230	389-4	269	10-5967
ACD-0088-57	11/20/14	11/20/14	3	non	ถนนสุขุม...	ชนท้ายเบียดคร...	1	2	230	389-4	269	10-5967
ACD-0042-59	5/14/16	5/14/16	3	non	แยกสน...	เบียดกับรถ...	2	1	245	389-10	280	10-6564
ACD-0038-59	4/7/16	4/7/16	3	non	ไปแดงห...	ชนท้ายรถเก...	1	2	182	389-10	A55	10-3820
ACD-0078-57	2/28/14	2/28/14	3	non	ลาดกระบัง	สิบล้อเบียด...	1	2	230	389-4	269	10-5967
ACD-0079-57	4/21/14	4/21/14	3	non	กม. 46 ...	รถยกเบียด...	1	1	231	389-3	270	10-5968
ACD-0081-57	7/14/14	7/14/14	3	non	ถนนมาช...	รถคันหน้าชน...	1	2	231	389-3	270	10-5968
ACD-0085-58	6/4/15	6/4/15	2	tep	สถานีเท...	เบียดกับรถบิ...	1	2	231	389-3	270	10-5968
ACD-0080-57	7/8/14	7/8/14	2	gas	หน้าปิ่น...	วิ่งออกจากร...	1	2	231	389-3	270	10-5968
ACD-0082-57	10/11/14	10/17/14	2	gas	บึงสีตรา...	ล้มกับรถ...	4	2	229	389-1	268	10-5934
ACD-0087-57	11/21/14	11/21/14	3	non	ซอยเทพ...	ชนท้ายรถบิ...	1	2	229	389-1	268	10-5934
ACD-0011-56	3/24/13	3/24/13	2	pat	ทางออก...	เบียดเสาทาง...	4	2	115	48-26	144	10-3155
ACD-0042-57	7/26/14	7/26/14	3	non	หน้าสำน...	เบียดรถรวม...	1	2	134	48-06	153	10-5997
ACD-0055-56	10/4/13	10/4/13	3	non	แยกห้วย...	รถบรรทุก...	1	2	063	9905-1	93	10-2751
ACD-0088-58	8/27/15	8/27/15	2	air	สนามบิน	โศกครุฑเี...	2	1	233	389-8	272	10-6046
ACD-0086-58	7/26/15	7/26/15	2	tep	สถานีเท...	คอกหลังชน...	1	2	233	389-8	272	10-6046
ACD-0092-57	12/27/14	12/27/14	3	non	ถนนเทพ...	กำลังเบียด...	1	2	233	389-8	272	10-6046
ACD-0007-57	2/9/14	2/9/14	2	air	ในสนาม...	รถที่วิ่งชน...	1	1	063	9905-1	93	10-2751
ACD-0085-57	10/27/14	10/27/15	3	non	ถนนเทพ...	เขี่ยท้าย...	4	2	230	389-4	269	10-5967
ACD-0090-58	10/25/15	10/25/15	3	non	ถนนเทพ...	รถคนจอด...	1	2	248	389-6	283	10-5997
ACD-0076-57	1/17/14	1/17/14	3	non	แยกไฟ...	ไฟแดงเท...	1	2	233	389-8	272	10-6046
ACD-0002-56	1/11/13	1/11/13	3	non	หน้าโรง...	รถบิ...	2	1	139	48-42	158	10-3485
ACD-0011-59	3/6/16	3/6/16	2	pat	ด้านหลัง...	คอกหลังไป...	3	2	138	48-19	157	10-3472
ACD-0010-58	2/4/15	2/4/15	2	pat	หน้าสตา...	เลี้ยวเข้า...	1	2	138	48-19	157	10-3472
ACD-0033-58	4/11/15	4/11/15	3	non	ถนนสา...	เบียดรถ...	1	2	138	48-19	157	10-3472
ACD-0040-58	4/27/15	4/27/15	3	non	บนทางค...	เลี้ยวชน...	1	3	135	48-17	154	10-3471
ACD-0112-58	11/30/15	11/30/15	2	pat	สถานีท่า...	เลี้ยวเข้า...	1	2	237	389-3	270	10-5968
ACD-0014-58	2/21/15	2/21/15	2	oob	อ่าวอุดม	เขี่ยรถ...	1	2	135	48-17	154	10-3471
ACD-0031-57	6/9/14	6/9/14	3	non	บนทางค...	รถเก๋ง...	1	2	135	48-17	154	10-3471
ACD-0035-56	7/9/13	7/9/13	1	non	อุโมงค์...	คอกชน...	3	2	115	48-26	144	10-3155
ACD-0038-56	7/10/13	7/10/13	3	non	สามแยก...	ชนท้ายรถ...	1	2	115	48-26	144	10-3155

ภาพที่ 4.8 แสดงไฟล์ข้อมูลที่ได้จากคลังข้อมูลการเกิดอุบัติเหตุรถโดยสารสาธารณะ

2.3 การเตรียมข้อมูล(Data Preparation)

2.3.1 การคัดเลือกข้อมูล (Select Data) โดยข้อมูลที่ได้จะทำการเลือกเฉพาะข้อมูลที่ครบถ้วน และสามารถนำไปใช้ในการวิเคราะห์และทำนายผลได้จริง โดยที่ข้อมูลที่ไม่ครบถ้วนหรือไม่สมบูรณ์ จำเป็นต้องทำการตัดทิ้งหรือนำออกไปจากฐานข้อมูล และคงเหลือไว้ซึ่งข้อมูลที่มีความสมบูรณ์เหมาะสม และต้องสอดคล้องกับงานวิจัย

2.3.2 การทำความสะอาดข้อมูล (Clean Data) จากการตรวจสอบข้อมูล ข้อมูลส่วนใหญ่สมบูรณ์ แต่ยังคงพบว่ามีข้อมูลบางส่วนมีค่าผิดปกติหรือมีค่าที่ควรจัดการแก้ไข โดยได้ดำเนินการแก้ไขดังนี้

1) ข้อมูลบางส่วนที่จะนำมาใช้มีการเก็บค่าที่เกินความจำเป็น หรือจัดอยู่ในรูปแบบที่ไม่สามารถนำข้อมูลมาทำการวิเคราะห์หรือการทำนาย ได้แก่

- GPS_DETAIL ให้รายละเอียดการวิเคราะห์ทางด้านความเร็วของรถ มีข้อมูลบางส่วนที่ต้องปรับรูปแบบเพื่อประโยชน์ในการใช้งาน คือ วันที่ (R_TIME)

รูปแบบเดิม " R_TIME = 2012-07-31T14:40:11.0 "

รูปแบบใหม่ " R_TIME = 2012-07-31 " หรือ yyyy-MM-dd

2) ข้อมูลขาดความสมบูรณ์ เพื่อที่จะนำไปใช้ในการวิเคราะห์หรือการทำนาย โดยเกิดจากการออกแบบฐานข้อมูลไม่ได้ทำการออกแบบไว้ ผู้วิจัยจึงทำการเพิ่มเติม รวบรวมข้อมูล เพื่อให้ง่ายต่อการจัดกลุ่ม หรือการแบ่งประเภทของข้อมูล ตัวอย่าง เช่น

กรุงเทพมหานคร	ภาค	กรุงเทพฯปริมณฑล
นครสวรรค์	ภาค	ภาคกลาง
ขอนแก่น	ภาค	ภาคตะวันออกเฉียงเหนือ เป็นต้น

3) การลดขนาดข้อมูล การลดขนาดของข้อมูลโดยนำข้อมูลที่ไม่มีผลต่อการวิเคราะห์แยกออกไป ซึ่งในงานวิจัยนี้ใช้ความสามารถของโปรแกรม KNIME v2.12 เป็นตัวคัดกรองข้อมูลออกโดยไม่ผลกระทบต่อข้อมูลเดิม ตารางที่ 4.17 แสดงรายละเอียดการลดขนาดข้อมูล GPS_DETAIL

ชื่อฟิลด์	ชนิด	ขนาด	ความหมาย	ตัวอย่างข้อมูล
_ID	Object Id	-		558906e32b571275dca555fbb
R_LON	Int	-	เส้นลองติจูด	100.6208170000000000
R_LAT	Int	-	เส้นละติจูด	13.6719000000000010
R_DISTANCE	Int	-	ระยะทาง	5692.50
TRUCK_RADIO_ID	numberLong	-	รหัสซิม	875167730
TRUCK_ID	String	-	รหัสรถ	TREVO-00087
R_TIME_2	ISODate	-	เวลา	ISODate("2012-07-31T14:40:11.000Z")

ตารางที่ 4.18 แสดงรายละเอียดการลดข้อมูล GPS_ACD_406CASE.XLSX

ชื่อฟิลด์	ชนิด	ขนาด	ความหมาย	ตัวอย่างข้อมูล
AC_NO	Character		หมายเลขอุบัติเหตุ	ACD-0092-58
AC_DATE	Date		วันที่แจ้ง	2015-01-10
AC_GDATE	Date		วันที่เกิดอุบัติเหตุ	2015-01-10
AC_ADD	Character		ที่ตั้งเกิดเหตุ	ลานจอดรถตู้หน้าอู่
AC_DETAIL	Character		รายละเอียดที่ตั้ง	ถอยหลังไปชนรถเคนในอู่
AC_BCODE	Character		รหัสรถ	209
AC_SIDE	Character		เบอร์รถ	9905-36
AC_CIRCLE	Character		เบอร์วงกลมรถ	A79
AC_REGIS	Character		ทะเบียนรถ	10-4098
AC_DRNAME	Character		ชื่อพนักงานขับรถ	นายไพศาล มัดทอง

4) การเปลี่ยนแปลงรูปแบบข้อมูล จากการตรวจสอบข้อมูลสำหรับการทำเหมืองข้อมูลนั้น ข้อมูล GPS_ACD_406CASE.XLSX ซึ่งเป็นข้อมูลเกี่ยวกับการเกิดอุบัติเหตุโดยพนักงานขับรถ ยังพบว่ามีข้อมูลบางส่วนยังขาดความสมบูรณ์ เนื่องจากรูปแบบของข้อมูลยังไม่เหมาะสมกับการสร้างแบบจำลองด้วยการแบ่งกลุ่ม และการหาความสัมพันธ์ในการทำเหมืองข้อมูล เพราะว่าการทำเหมืองข้อมูลนั้น จำเป็นต้องใช้เทคนิคต่างๆ ซึ่งอยู่ในรูปแบบที่มีความแตกต่างกัน จึงมีความจำเป็นต้องเปลี่ยนแปลง หรือเพิ่มเติมค่าบางอย่าง เช่น การจัดข้อมูลบางส่วนให้เป็นกลุ่ม หรือการจกข้อมูลให้เป็นช่วง (Range)

ตารางที่ 4.19 ลักษณะข้อมูลที่ใช้ในการวิเคราะห์ด้วยโปรแกรม KNIME v2.12

ข้อมูล	ลักษณะข้อมูล	รายละเอียดกำหนดใหม่
PHUMIPAK ภูมิภาค	กรุงเทพฯปริมณฑล	PM_1
	ภาคกลาง	PM_2
	ภาคเหนือ	PM_3
	ภาคใต้	PM_4
	ภาคตะวันออกเฉียงเหนือ	PM_5
	ภาคตะวันออก	PM_6
	ภาคตะวันตก	PM_7
AC_GAGE2 อายุ	30-35	AGE_1
	36-40	AGE_2
	41-45	AGE_3
	46-50	AGE_4
	51-55	AGE_5
	56-60	AGE_6
	มากกว่า 60	AGE_7
AC_GWORK2 ประสบการณ์ขับรถ	น้อยกว่า 1 ปี	WK_1
	1-2	WK_2
	3-4	WK_3
	5-6	WK_4
	ตั้งแต่ 7 ปีขึ้นไป	WK_5
HIGH_2 ส่วนสูง	150 - 160	H_1
	161- 170	H_2
	มากกว่า 170	H_3
WEIGHT_2 น้ำหนักตัว	50-65	W_1
	66-75	W_2
	76-85	W_3
	มากกว่า 85	W_4

ตารางที่ 4.19 (ต่อ)

ข้อมูล	ลักษณะข้อมูล	รายละเอียดกำหนดใหม่
EDUCATION การศึกษา	ระดับชั้นประถม	ED_1
	ระดับมัธยมต้น	ED_2
	ระดับมัธยมปลาย - ปวช.	ED_3
	อื่นๆ - ปริญญาตรี	ED_4
AC_BRAND ประเภทรถ	เบนซ์	BENZ
	สแกนเนีย	SCANIA
	ซันลอง	SUNLONG
	โกลด์เด้น ดากอน	GOLDEN DAGON
	โตโยต้า(รถตู้)	TOYOTA
AC_DTYPE เส้นทาง	รถสาย 48	B048
	รถสาย 9905	M9905
	รถสาย 389	B389
	รถสาย 789	H789
	รถตู้9905	V9905
AC_TF ความถูกต้องอุบัติเหตุ	บริษัทถูก	1
	บริษัทผิด	2
	รับผิดชอบร่วม	3
AC_PLACE สถานที่เกิดเหตุ	บริษัท	1
	สถานี	2
	ภายนอก	3
AC_CHON ลักษณะการชน	รถบริษัท ชน รถอื่น	1
	รถอื่น ชน รถบริษัท	2
	รถบริษัทชนกันเอง	3
	รถบริษัท ชน อื่นๆ	4
TIMEWORK_2 จำนวนชั่วโมงต่อวัน (ขับรถ)	ขับรถน้อยกว่าหรือเท่ากับ 6 ชั่วโมง	TK_1
	ขับรถวันละ 7 -8 ชั่วโมง	TK_2
	ขับรถวันละมากกว่า 8 ชั่วโมง	TK_3

ตารางที่ 4.19 (ต่อ)

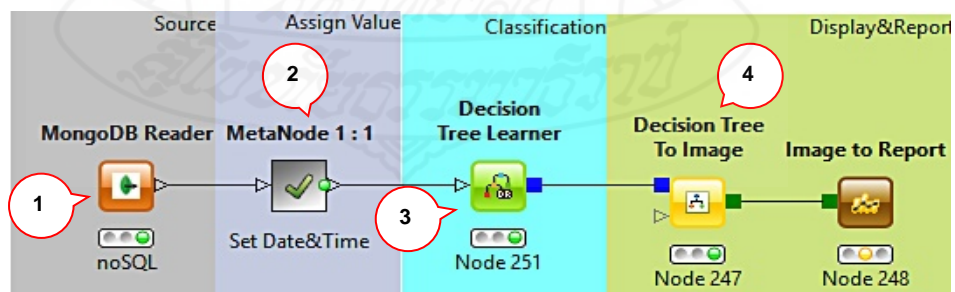
ข้อมูล	ลักษณะข้อมูล	รายละเอียดกำหนดใหม่
	0 - 10,000	IN_1
INCOME_2	10,001 - 15,000	IN_2
รายได้	15,001 - 20,000	IN_3
	20,001 - 25,000	IN_4
	มากกว่า 25,000	IN_5

2.4 การสร้างแบบจำลองทำนาย (Modeling)

ข้อมูลที่ใช้ในการสร้างแบบจำลองเป็นผลมาจากการออกแบบคลังข้อมูล ได้แก่ คลังข้อมูลระบบ GPS เพื่อสร้างแบบจำลองการใช้อัตราความเร็วของรถโดยสารสาธารณะ และคลังข้อมูลการเกิดอุบัติเหตุรถโดยสารสาธารณะเพื่อสร้างแบบจำลองทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ ผลที่ได้จากการวิจัยเป็นดังนี้

2.4.1 การใช้อัตราความเร็วของรถโดยสารสาธารณะ

ผู้วิจัยได้เลือกเทคนิคการทำเหมืองข้อมูล แบบการแบ่งประเภท Classification ในการสืบค้นข้อมูล หรือการสกัดความรู้ที่ซ่อนอยู่ในรูปแบบข้อมูล Big Data โดยได้เลือกวิธีต้นไม้ตัดสินใจ (Decision Tree) จากข้อมูลที่ออกแบบไว้ คือ GPS_DETAIL และสร้างแบบจำลองการใช้อัตราความเร็วของรถโดยสารสาธารณะแบ่งเป็น 2 แบบ ได้แก่ กลุ่มรถโดยสารสาธารณะ และ กลุ่มตำแหน่งที่ตั้ง



ภาพที่ 4.9 ขั้นตอนสร้างแบบจำลองการใช้อัตราความเร็วของรถโดยสารสาธารณะ
ภาพที่ 4.9 แสดงรายละเอียดของขั้นตอนการสร้างแบบจำลองการใช้อัตราความเร็วของรถโดยสารสาธารณะ สามารถอธิบายดังนี้

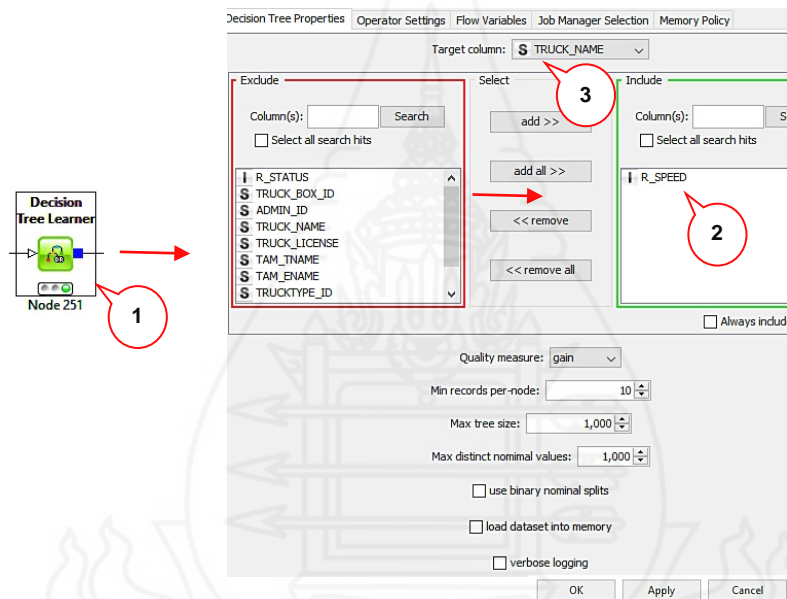
หมายเลข 1 MongoDB Reader Node นำเข้าข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS

หมายเลข 2 Meta Node กำหนดค่าในการทดสอบข้อมูล เช่น จำนวนวัน และความเร็วของรถ เป็นต้น

หมายเลข 3 Decision Tree Learner Node จำแนกประเภทรถโดยสารสาธารณะด้วยเทคนิค Classification

หมายเลข 4 Decision Tree to Image Node และ Image To Report Node แสดงผลลัพธ์ในรูปแบบภาพต้นไม้อัตโนมัติ

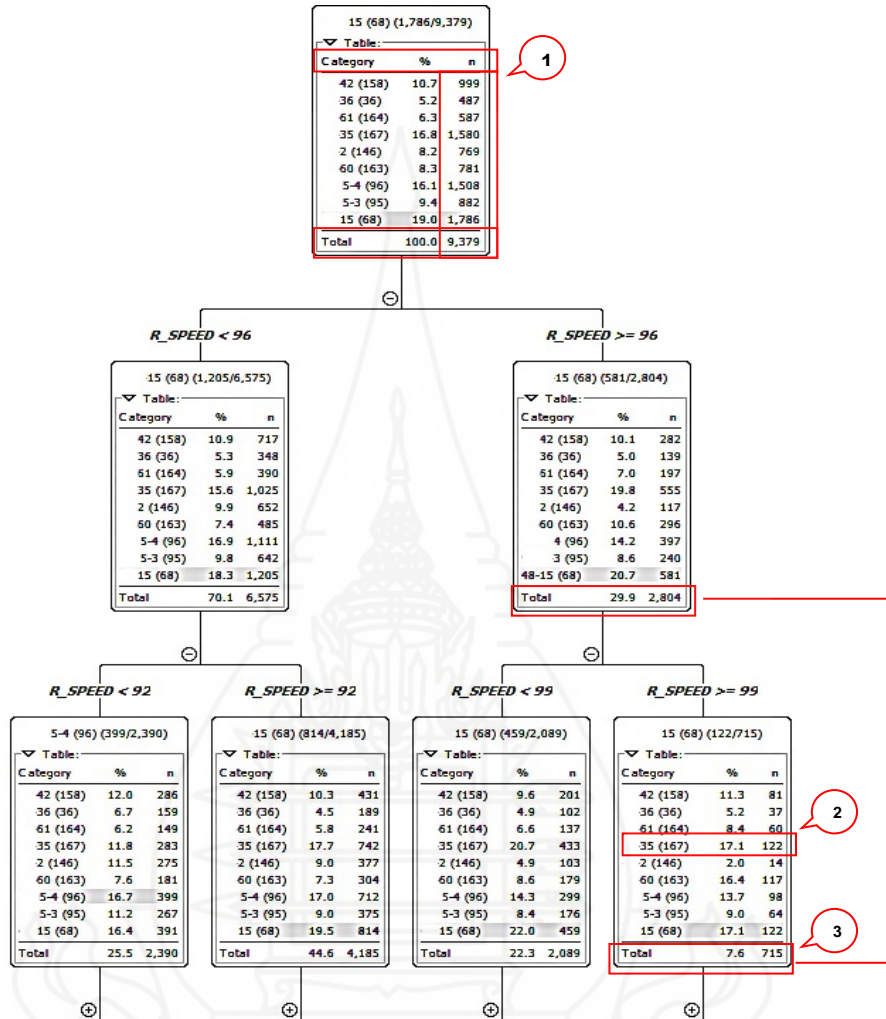
1) การสร้างแบบจำลองกลุ่มรถโดยสารสาธารณะใช้อัตราความเร็ว



ภาพที่ 4.10 กำหนด TARGET ของกลุ่มรถโดยสารสาธารณะใช้อัตราความเร็ว

ภาพที่ 4.10 ดับเบิ้ลคลิก Decision Tree Learner Node หมายเลข 1 กำหนดค่าการจำแนกประเภทรถโดยสารสาธารณะ คลิกเลือกฟิลด์ความเร็ว (R_SPEED) ในช่อง Exclude ไปเพิ่มที่ช่อง Include ตามที่แสดงหมายเลข 2 เป็นอินพุต (INPUT) ของข้อมูล คลิก Target Column หมายเลข 3 เป็นการกำหนดค่าผลลัพธ์เป้าหมาย (TARGET) ที่สนใจคือหมายเลขรถฟิลด์หมายเลขรถ (TRUCK_NAME) คลิกปุ่ม OK เพื่อยืนยันการกำหนดค่าให้กับ Decision Tree Learner Node ลำดับต่อไปต้องการดูผลลัพธ์การจำแนกรถโดยสารสาธารณะ ทำการคลิกขวา Decision Tree to Image Node หมายเลข 4 จากภาพที่ 4.9 เลือกหัวข้อ Execute ประมวลผล จากนั้นคลิกขวาอีกครั้งที่ Decision Tree to Image

Node หมายเลข 4 เลือก View Decision Tree View แสดงรายงานข้อมูลภาพต้นไม้ตัดสินใจ ตามตัวอย่างภาพที่ 4.11



ภาพที่ 4.11 แสดงรายงานของกลุ่มรถโดยสารสาธารณะตามเงื่อนไขความเร็วรถ

ภาพที่ 4.11 แสดงผลลัพธ์การจำแนกรถโดยสารสาธารณะ โดยใช้ความเร็วรถเป็นอินพุต (INPUT) เพื่อการจำแนกประเภทรถโดยสารสาธารณะ อธิบายผลลัพธ์ที่ได้ดังนี้

หมายเลข 1 ความหมายแต่ละส่วน

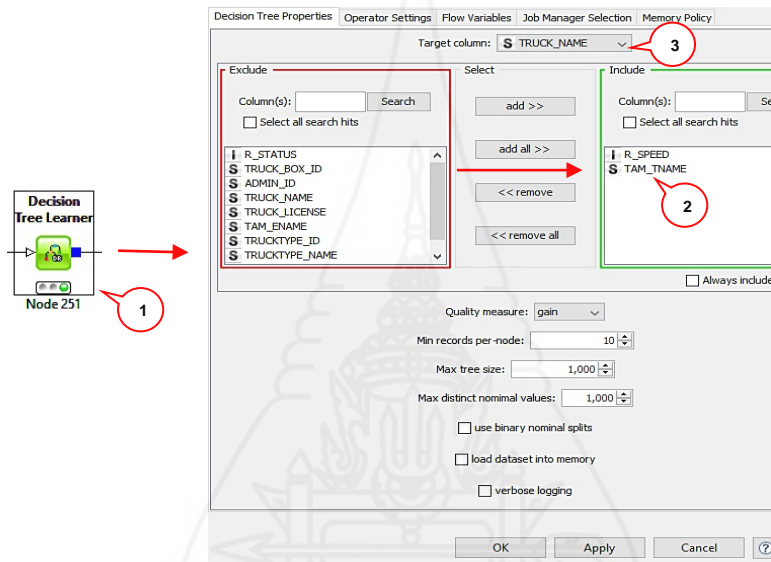
- " Category " หมายถึง กลุ่มรถโดยสาร
- " n " หมายถึง จำนวนครั้งการใช้อัตราความเร็วของรถโดยสาร
- " % " หมายถึง เปรียบเทียบร้อยละกับจำนวนครั้งทั้งหมด เช่น

หมายเลขรถ 42 (158) การใช้ความเร็ว มากกว่าหรือเท่ากับ 90 กม.ต่อชั่วโมง มีจำนวน 999 ครั้ง คิดเป็น ร้อยละ 10.7 ของจำนวนครั้งรวม (TOTAL) 9379 ครั้ง

หมายเลข 2 ความหมายอีกหนึ่งตัวอย่าง หมายเลข 35 (168) ใช้อัตราความเร็วรถมากกว่าหรือเท่ากับ 99 กม.ต่อชั่วโมง คิดเป็น 122 ครั้ง คิดเป็นร้อยละ 17.1 ของจำนวนทั้งหมด 715 ครั้ง

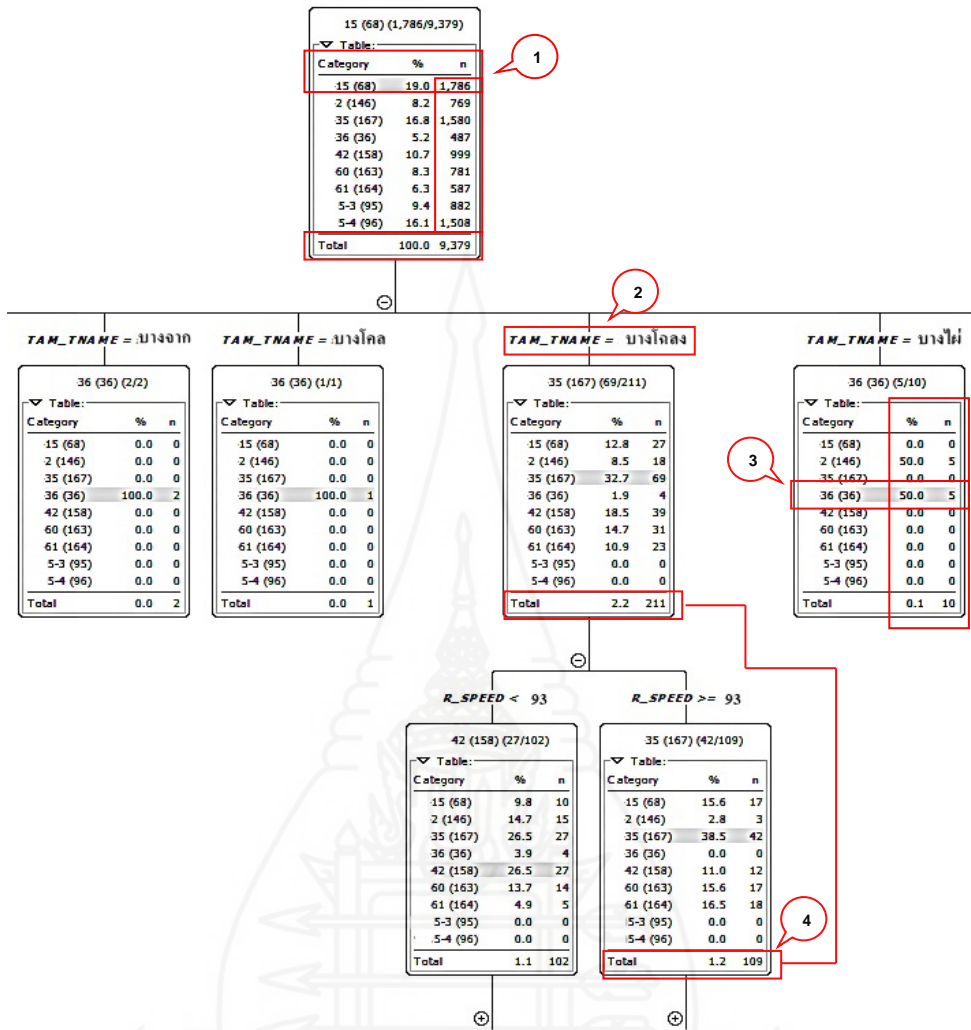
หมายเลข 3 "n" จำนวนครั้งรวม ของความเร็วมากกว่าหรือเท่ากับ 99 กม.ต่อชั่วโมง รวมเป็น 715 ครั้ง คิดเป็นร้อยละ 7.6 ของ 2804 ครั้ง

2) การสร้างแบบจำลองแบ่งกลุ่มอัตราความเร็วของตำแหน่งที่ตั้ง



ภาพที่ 4.12 กำหนด TARGET เป้าหมายของกลุ่ม

ภาพที่ 4.12 ดับเบิลคลิก Decision Tree Learner Node หมายเลข 1 กำหนดค่าการจำแนกประเภทรถโดยสารสาธารณะ คลิกเลือกฟิลด์ความเร็ว (R_SPEED) และสถานที่ (TAM_TNAME) ในช่อง Exclude ไปเพิ่มที่ช่อง Include ตามที่แสดงหมายเลข 2 เป็นอินพุต (Input) เพื่อกำหนดอินพุต(Input) จากนั้นคลิก Target Column หมายเลข 3 เป็นการกำหนดค่าผลลัพธ์เป้าหมาย (TARGET) ที่สนใจคือหมายเลขรถฟิลด์หมายเลขรถ (TRUCK_NAME) คลิกปุ่ม OK เพื่อยืนยันการกำหนดค่าให้กับ Decision Tree Learner Node ลำดับต่อไปต้องการดูผลลัพธ์การจำแนกรถโดยสารสาธารณะ ทำการคลิกขวา Decision Tree to Image Node หมายเลข 4 จากภาพที่ 4.7 ทำการเลือกหัวข้อ Execute ประมวลผล จากนั้นคลิกขวาอีกครั้งที่ Decision tree to Image Node หมายเลข 4 เลือก View Decision Tree View จากภาพที่ 4.9 แสดงรายงานข้อมูลภาพต้นไม้ตัดสินใจตามตัวอย่างภาพที่ 4.13



ภาพที่ 4.13 แสดงรายงานจำแนกประเภทรถโดยสารตามเงื่อนไข ความเร็วรถและสถานที่
ภาพที่ 4.13 การจำแนกประเภทรถโดยสารสาธารณะตามเงื่อนไขความเร็วรถและสถานที่ อธิบาย
ผลลัพธ์ที่ได้ดังนี้

หมายเลข 1 ความหมายแต่ละส่วน

" Category " หมายถึง กลุ่มรถโดยสาร

" n " หมายถึง จำนวนครั้งการใช้อัตราความเร็วของรถโดยสาร

" % " หมายถึง เปรียบเทียบร้อยละ กับจำนวนครั้งทั้งหมด เช่น

หมายเลขรถ 15 (68) การใช้ความเร็ว มากกว่าหรือเท่ากับ 90 กม.ต่อชั่วโมง มีจำนวน 1786 ครั้ง คิด
เป็น ร้อยละ 19 ของจำนวนครั้งรวม (TOTAL) 9379 ครั้ง

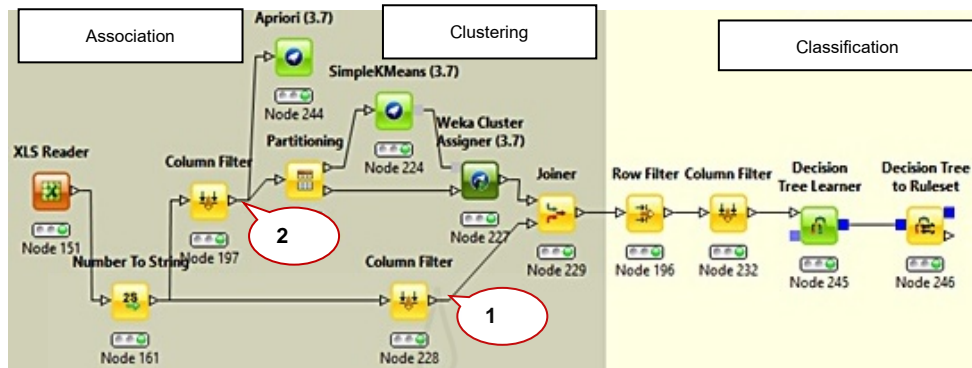
หมายเลข 2 คือ ตำแหน่งหรือสถานที่ (TAM_TNAME)

หมายเลข 3 ความหมายตัวอย่าง หมายเลข 36 (36) ใช้อัตราความเร็วรถผ่านสถานที่บาง
 ไร่จำนวน 5 ครั้งคิดเป็นร้อยละ 50 ของจำนวนทั้งหมด 10 ครั้ง

หมายเลข 4 "n" จำนวนครั้งรวม ของความเร็วมากกว่าหรือเท่ากับ 93 กม.ต่อชั่วโมง รวม
 เป็น 109 ครั้ง คิดเป็นร้อยละ 1.2 ของ 211 ครั้ง

2.4.2 การเกิดอุบัติเหตุรถโดยสารสาธารณะ

การออกแบบจำลองทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ เพื่อ
 ทดสอบแบบจำลองของกลุ่มเสี่ยง และกลุ่มเฝ้าระวัง สามารถอธิบายกระบวนการออกแบบได้ดังนี้
 กลุ่มเฝ้าระวังใช้ชุดข้อมูลคุณสมบัติบุคคลนำไปใช้ทดสอบกฎความสัมพันธ์ (Association Rule)
 ด้วยเทคนิคอปริออริ (Apriori) เพื่อหาคุณสมบัติของกลุ่มเฝ้าระวัง ส่วนกลุ่มเสี่ยงเป็นการนำชุด
 ข้อมูลองค์ประกอบของปัจจัยการเกิดอุบัติเหตุ และชุดข้อมูลคุณสมบัติบุคคลมาทำการทดสอบ
 แบบจำลอง โดยชุดข้อมูล คุณสมบัติบุคคลนำไปทดสอบแบบจำลองด้วย Partition Testing โดยใช้
 เทคนิค SplitTest ทำการแบ่งข้อมูลเป็นสองชุด ได้แก่ Training Data สำหรับสร้างแบบจำลอง และ
 Testing Data สำหรับทดสอบแบบจำลองเพื่อเข้าสู่กระบวนการในการจัดกลุ่มข้อมูลด้วยอัลกอริทึม
 "Simple K-Means" เพื่อทดสอบข้อมูล Training Data และข้อมูล Testing Data ที่แสดงผลลัพธ์ของ
 ข้อมูลคล้ายคลึงกันหรือเหมือนกัน แสดงว่าการทดสอบแบบจำลองการจัดกลุ่มและเป็นจำนวนกลุ่ม
 เหมาะสม ขั้นตอนสุดท้ายนำแต่ละกลุ่มมาใช้ข้อมูลร่วมกับชุดข้อมูลองค์ประกอบของปัจจัยการเกิด
 อุบัติเหตุ (เส้นทางเดินรถ, จำนวนชั่วโมงขับรถ, ประเภทที่ใช้ และสถานที่เกิดเหตุ) ด้วย
 อัลกอริทึม Decision Tree และ Rule Set เพื่อหาความสัมพันธ์ของชุดข้อมูลแต่ละกลุ่มเสี่ยง ให้ได้ผล
 การทำนายที่สามารถนำไปพัฒนาเป็นระบบการทำนายผลการเกิดอุบัติเหตุรถโดยสารต่อไป



ภาพที่ 4.14 แบบจำลองทำนายการเกิดอุบัติเหตุรถโดยสาร

ภาพที่ 7.14 แบบจำลองการทำนายการเกิดอุบัติเหตุรถโดยสารใช้โปรแกรม KNIME v2.12 ข้อมูลถูกแบ่งเป็น 2 ชุด ข้อมูลชุดที่หนึ่ง หมายเลข 1 เป็นข้อมูลองค์ประกอบของปัจจัยการเกิดอุบัติเหตุได้แก่ เส้นทางเดินรถ จำนวนชั่วโมงขับรถ ประเภทที่ใช้ และสถานที่เกิดเหตุ (บริษัท, สถานี, ภายนอก) ส่วนข้อมูลชุดที่สอง หมายเลข 2 เป็นข้อมูลคุณสมบัติบุคคลได้แก่ อายุ การศึกษา ภูมิลำเนา น้ำหนักตัว ส่วนสูง ประสบการณ์ทำงาน และรายได้

กลุ่มเฝ้าระวัง เป็นการนำชุดที่สอง หมายเลข 2 คุณสมบัติบุคคลมาทดสอบ โดยใช้กฎความสัมพันธ์ (Association Rule) ด้วยวิธี Apriori เพื่อหาความสัมพันธ์ของคุณสมบัติบุคคล โดยกำหนดค่าความเชื่อมั่น Confidence ไว้ที่ร้อยละ 60 และค่า 50 จำนวนกฎ (NumRules) เพื่อให้ได้ผลลัพธ์กฎความสัมพันธ์ของคุณลักษณะบุคคลที่แสดงค่าความเชื่อมั่น (Confidence) ของกฎความสัมพันธ์ด้วยจำนวนเป็นร้อยละและค่าลิฟต์ (Lift) คือค่าสหสัมพันธ์ที่บ่งบอกถึงกฎความสัมพันธ์ที่มีค่ามากกว่า 1 เท่านั้น

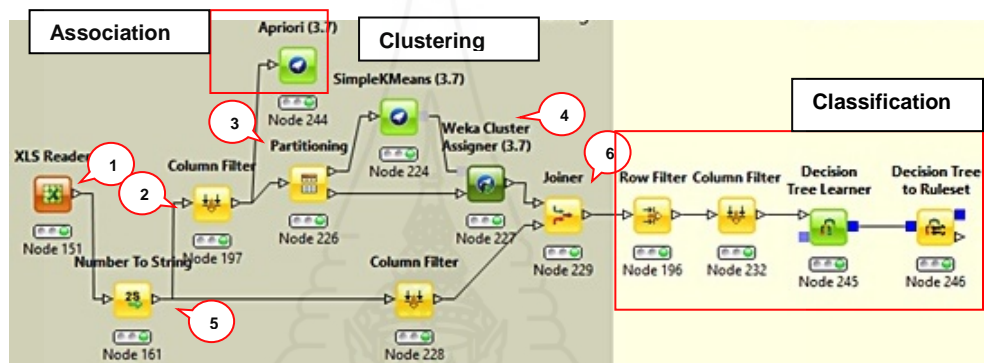
กลุ่มเสี่ยง เป็นการนำข้อมูลชุดที่สอง หมายเลข 2 คุณสมบัติบุคคล ได้แก่ อายุ การศึกษา ภูมิลำเนา น้ำหนักตัว ส่วนสูง ประสบการณ์ทำงาน รายได้ และชุดข้อมูลชุดที่หนึ่ง หมายเลข 1 องค์ประกอบของปัจจัยการเกิดอุบัติเหตุ ได้แก่ เส้นทางเดินรถ จำนวนชั่วโมงขับรถ ประเภทที่ใช้ และสถานที่เกิดเหตุ (บริษัท, สถานี, ภายนอก) มี 2 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 การแบ่งกลุ่มเสี่ยง โดยใช้ข้อมูลชุดที่สอง หมายเลข 2 ทดสอบแบบจำลองด้วยอัลกอริทึม Simple K-Means เพื่อทำการแบ่งกลุ่มเสี่ยงโดยใช้การแบ่งข้อมูลด้วยเทคนิค Split Test ได้แก่ Training Data และ Testing Data

ขั้นตอนที่ 2 นำกลุ่มเสี่ยงแต่ละกลุ่มมาทำการทดสอบร่วมกับชุดข้อมูลชุดที่หนึ่ง หมายเลข 1 องค์ประกอบของปัจจัยการเกิดอุบัติเหตุได้แก่ เส้นทางเดินรถ จำนวนชั่วโมงขับรถ ประเภทที่ใช้ และสถานที่เกิดเหตุ (บริษัท, สถานี, ภายนอก) ด้วยอัลกอริทึม Decision Tree Learner เพื่อหาว่าจะมีองค์ประกอบปัจจัยการเกิดอุบัติเหตุอะไรบ้างในแต่ละกลุ่มเสี่ยงของแต่ละกลุ่มโดยการใช้

Decision Tree to Ruleset Node หาความสัมพันธ์ของกลุ่มเสี่ยงให้ได้ผลลัพธ์การทำนายสามารถนำไปพัฒนาเป็นระบบการทำนายผลการเกิดอุบัติเหตุโดยसारสาธารณะ

รายละเอียดการออกแบบจำลองทำนาย และทดสอบ ผู้วิจัยทำการเปลี่ยนรูปแบบข้อมูลการเกิดอุบัติเหตุโดยसारสาธารณะ เป็นไฟล์เอ็กเซล GPS_ACD_406CASE.XLSX จากข้อมูลภาพที่ 4.8 เพื่อสำหรับทำเหมืองข้อมูลด้วยโปรแกรม KNIME v2.12 มีขั้นตอนและรายละเอียดการสร้างแบบจำลองทำนายการเกิดอุบัติเหตุ ได้แก่ กลุ่มเสี่ยง กลุ่มเฝ้าระวัง และการหาปัจจัยก่อให้เกิดอุบัติเหตุที่มีผลต่อกลุ่มเสี่ยง



ภาพที่ 4.15 รายละเอียดแบบจำลองทำนายการเกิดอุบัติเหตุโดยसार

ภาพที่ 4.15 อธิบายส่วนต่างๆ ดังนี้

หมายเลข 1 คือ XLS Reader Node นำเข้าข้อมูลการเกิดอุบัติเหตุโดยसारสาธารณะ

หมายเลข 2 คือ Column Filter เลือกฟิลด์ข้อมูลพนักงานขับรถ เช่น อายุ (AC_GAGE2), อายุงาน(AC_GWORK2), ความสูง (HIGH2), น้ำหนัก (WEIGHT_2), การศึกษา (EDUCATION), รายได้ (INCOME_2) และ ถิ่นกำเนิด (PHUMIPAK)

หมายเลข 3 คือ Partitioning Node แบ่งข้อมูล 2 ชุด Training data และ Testing data

หมายเลข 4 คือ Simple K-Means Node (3.7) การแบ่งกลุ่มพนักงานขับรถ ใช้เทคนิค Clustering และ Weka Cluster Assigner Node (3.7) กำหนดไว้เพื่อการทดสอบการทำนายข้อมูลทดสอบ

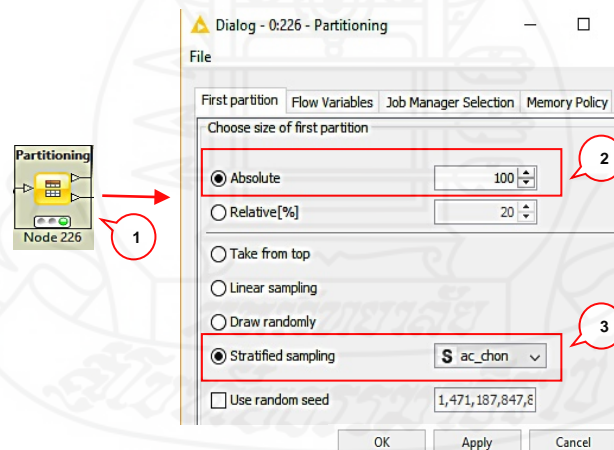
หมายเลข 5 คือ Number to String Node แปลงค่าตัวเลขเป็นค่าข้อความ เพื่อให้มีความเหมาะสมกับการสร้างแบบจำลองของขั้นตอนต่อไป โดยแปลงตัวเลขเป็นข้อความ ได้แก่ สถานที่เกิดเหตุ (AC_PLACE), ลักษณะการชน (AC_CHON), การเกิดอุบัติเหตุใครถูกหรือผิด (AC_TF)

และเส้นทางการเดินรถ (AC_TYPE) เป็นต้น Column Filter Node คัดเลือกฟิลด์ข้อมูลการเกิดอุบัติเหตุ เพื่อไปร่วมกับผลลัพธ์การจัดกลุ่มพนักงานขับรถ (หมายเลข 4) ฟิลด์ที่ทำการคัดเลือกมีดังนี้ สถานที่เกิดเหตุ(AC_PLACE), ตำแหน่ง (AC_BOOT), ลักษณะการชน (AC_CHON), เส้นทางการเดินรถ (AC_DTYPE), ประเภทรถ(AC_BRAND) และชั่วโมงการทำงาน (TIMEWORK_2)

หมายเลข 6 คือ Joiner Node เป็นการรวมข้อมูลที่ได้จากหมายเลข 4 และหมายเลข 5 ให้เป็นข้อมูลเดียวกัน โดยใช้ ฟิลด์ RowId เป็นฟิลด์เชื่อมข้อมูล ส่วน Row Filter Node คัดเลือกกลุ่มข้อมูลที่ผ่านการ Clustering มาแล้วได้ 4 กลุ่ม โดยเลือกแต่ละกลุ่มเพื่อทำขั้นตอนนี้ต่อไป

1) กลุ่มเสี่ยง (Risk Groups)

การสร้างแบบจำลองทำนาย เพื่อสร้างจัดแบ่งกลุ่มเสี่ยง ใช้วิธี **Split Test** แบ่งข้อมูล 2 ชุด ได้แก่ Training Data สำหรับสร้างแบบจำลอง และ Testing Data สำหรับทดสอบแบบจำลอง เพื่อให้ผลลัพธ์การเปรียบเทียบของ Training Data และ Testing Data มีลักษณะข้อมูลคล้ายกันหรือเหมือนกัน โดยใช้ชุดข้อมูลคุณลักษณะบุคคล ได้แก่ อายุ, อายุงาน, ความสูง, น้ำหนัก, การศึกษา, รายได้ และ ถิ่นกำเนิด เป็นข้อมูลสร้างแบบจำลองการจัดกลุ่มเสี่ยง



ภาพที่ 4.16 การแบ่งข้อมูลด้วย Partitioning Node

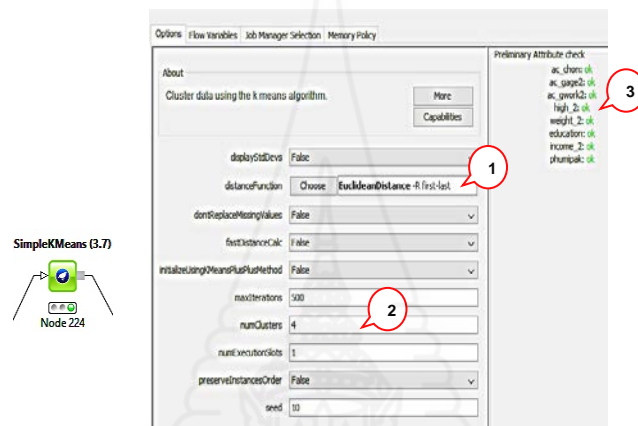
ภาพที่ 4.16 การแบ่งข้อมูลเป็น 2 ส่วน ได้แก่ Training Data และ Testing Data จากข้อมูลอุบัติเหตุจำนวน 406 ทำการกำหนดค่า ดังนี้

หมายเลข 1 คือ Partitioning Node แบ่งข้อมูลอุบัติเหตุเป็น Training Data และ Testing Data

หมายเลข 2 เลือกกำหนด "Absolute" จำนวน 100 รายการ เพื่อทดสอบข้อมูล(Testing Data)

หมายเลข 3 เลือก "Stratified Sampling" ฟังก์ชันลักษณะการชน (AC_CHON) เพื่อทดสอบข้อมูล

ลำดับต่อมากลุ่ม "Apply" ยืนยันค่า และคลิกปุ่ม OK



ภาพที่ 4.17 การจัดกลุ่มด้วย Simple K-Means Node

ภาพที่ 4.17 การกำหนดคุณสมบัติเพื่อทำการแบ่งกลุ่มเสียง โดยใช้ Simple K-Means Node (3.7) ดังนี้

หมายเลข 1 เลือก "EuclideanDistance -R first -last" ซึ่งเป็นมาตรวัดระยะห่างสำหรับการจัดกลุ่ม เพื่อนำข้อมูลที่มีความคล้ายคลึงกันหรือใกล้เคียงกันมากที่สุดจะถูกจัดไว้ในกลุ่มเดียวกัน และยังเป็นมาตรวัดระยะห่างของกลุ่มเป็นที่นิยมนำมาใช้งาน

หมายเลข 2 ระบุจำนวนการแบ่งกลุ่มเหมาะสม เริ่มจาก 3 กลุ่ม ถึง 10 กลุ่ม และนำไปปรึกษากับฝ่ายบุคคลและฝ่ายอุบัติเหตุ โดยหากกลุ่มที่มีลักษณะชัดเจนมีความแตกต่างในแต่ละกลุ่ม

หมายเลข 3 ฟังก์ชันพนักงานขับรถที่นำมาใช้ในการจัดแบ่งกลุ่ม เช่น อายุ (AC_GAGE2), อายุงาน (AC_GWORK2), ความสูง (HIGH2), น้ำหนัก (WEIGHT_2), การศึกษา (EDUCATION), รายได้ (INCOME_2) และ ถิ่นกำเนิด (PHUMIPAK)

kMeans
=====

Number of iterations: 4
Within cluster sum of squared errors: 329.0
Missing values globally replaced with mean/mode

Cluster centroids:

Attribute	Full Data (100)	Cluster#			
		0 (25)	1 (47)	2 (15)	3 (13)
ac_chon	1	1	1	1	1
ac_gage2	AGE1	AGE5	AGE1	AGE4	AGE3
ac_gwork2	WK_1	WK_5	WK_2	WK_5	WK_3
high_2	H_2	H_2	H_3	H_2	H_3
weight_2	W_1	W_1	W_1	W_1	W_3
education	ED_1	ED_1	ED_1	ED_1	ED_1
income_2	IN_3	IN_2	IN_3	IN_5	IN_4
phumipak	PM_5	PM_1	PM_5	PM_2	PM_6

Clustered Instances

0	94 (31%)
1	113 (37%)
2	53 (17%)
3	46 (15%)

ผลลัพธ์เหมือนกันด้วยเทคนิค Split Test

ภาพที่ 4.18 รูปแบบการแจกแจงแต่ละกลุ่มเสี่ยง

ภาพที่ 4.18 ผลจากการกำหนดค่าให้กับ Simple K-Means Node สามารถที่จะแจกแจงคุณลักษณะเฉพาะของแต่ละกลุ่ม ทั้ง 4 กลุ่ม ซึ่งเป็นการนำคุณสมบัติพนักงานขับรถทำการแจกแจงในการจัดแบ่งกลุ่ม โดยมีรายละเอียดของฟิลด์ข้อมูลดังนี้

"AC_CHON"	หมายถึง	ลักษณะการชน(1 รถบริษัทชนรถคันอื่น)
"AC_GAGE2"	หมายถึง	ช่วงอายุพนักงานขับรถที่เกิดอุบัติเหตุ
"AC_GWORK2"	หมายถึง	ช่วงอายุงานของพนักงานขับรถที่เกิดอุบัติเหตุ
"HIGH_2"	หมายถึง	ส่วนสูงของพนักงานขับรถ
"WEIGHT_2"	หมายถึง	น้ำหนักของพนักงานขับรถ
"EDUCATION"	หมายถึง	การศึกษาของพนักงานขับรถ
"INCOME_2"	หมายถึง	รายได้ของพนักงานขับรถ
"PHUMIPAK"	หมายถึง	ถิ่นกำเนิดของพนักงานขับรถ

การจัดแบ่งกลุ่มเสี่ยง ทดสอบด้วยอัลกอริทึม Simple K-Means จากข้อมูลอุบัติเหตุโดยสารจำนวน 406 ระเบียบ แบ่งข้อมูลเป็นสองชุดด้วยเทคนิค Split Test ได้แก่ Training Data จำนวน 306 ระเบียบสำหรับสร้างแบบจำลอง และ Testing Data จำนวน 100 ระเบียบสำหรับทดสอบแบบจำลอง จากการทดสอบแบบจำลองของข้อมูล Training Data และข้อมูล Testing Data ได้ผลลัพธ์ของข้อมูลเหมือนกัน และการแบ่งกลุ่มเสี่ยงเป็นจำนวน 4 กลุ่ม เหมาะสมที่สุด เพราะมี

ลักษณะเฉพาะกลุ่มเสี่ยงชัดเจน และสามารถให้รายละเอียดความหมายของแต่ละกลุ่มเสี่ยงได้ตาม ตารางที่ 4.20 ดังนี้

ตารางที่ 4.20 การแจกแจงคุณลักษณะของแต่ละกลุ่มเสี่ยง

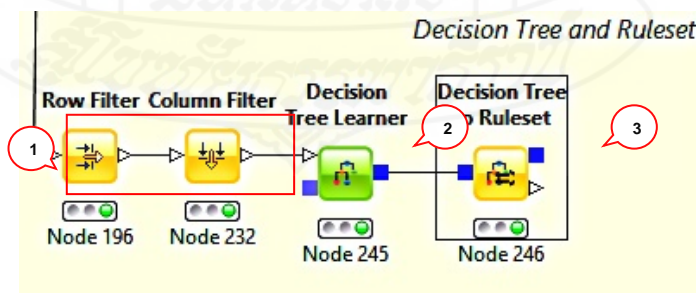
กลุ่มเสี่ยงที่	ลักษณะข้อมูลกลุ่ม	ลักษณะเฉพาะ
1) AC_CHON = 1 AC_GAGE2 = AGE5 AC_GWORK2= WK_5 HIGH_2 = H_2 WEIGHT_2 = W_1 EDUCATION = ED_1 INCOME_2 = IN_2 PHUMIPAK = PM_1	มีภูมิลำเนากรุงเทพฯปริมณฑล การศึกษาระดับชั้นประถม ช่วงอายุ 51-55ปี ประสบการณ์การทำงานในอาชีพขับรถ มากกว่า 7 ปี ส่วนสูง 161 - 170 ซม. น้ำหนักตัว 50 - 65 กก.มีรายได้มากกว่า 10,000บาท แต่ไม่เกิน 15,000บาท	ภูมิลำเนา กรุงเทพฯปริมณฑล อายุพนักงานมาก(51-55ปี) ประสบการณ์การทำงานสูง(>7ปี) การศึกษาน้อย รายได้พอใช้(300- 500บาท)
2) AC_CHON = 1 AC_GAGE2 = AGE1 AC_GWORK2= WK_2 HIGH_2 = H_3 WEIGHT_2 = W_1 EDUCATION = ED_1 INCOME_2 = IN_3 PHUMIPAK = PM_5	มีภูมิลำเนาอยู่ภาคอีสาน การศึกษา ระดับชั้นประถม ช่วงอายุ 30-35ปี ประสบการณ์การทำงานในอาชีพขับรถ 1-2 ปี ส่วนสูง มากกว่า 170 ซม. น้ำหนักตัว 50 - 65 กก.มีรายได้ มากกว่า 15,000บาท แต่ไม่เกิน 20,000บาท	ภูมิลำเนาภาคอีสาน ช่วงอายุรุ่นหนุ่ม(30-35ปี) ประสบการณ์ทำงานน้อย(1-2ปี) การศึกษาน้อย รายได้ปานกลาง(500-700บาท)
3) AC_CHON = 1 AC_GAGE2 = AGE4 AC_GWORK2 = WK_5 HIGH_2 = H_2 WEIGHT_2 = W_1 EDUCATION = ED_1 INCOME_2 = IN_5 PHUMIPAK = PM_2	มีภูมิลำเนา ภาคกลาง การศึกษา ระดับชั้นประถม ช่วงอายุ 46-50 ปี ประสบการณ์การทำงานในอาชีพขับรถมากกว่า 7 ปี ส่วนสูงระหว่าง 161-170 ซม. น้ำหนักตัว 50 - 65 กก.มีรายได้มากกว่า 25,000 บาท	ภูมิลำเนา ภาคกลาง ช่วงอายุเป็นคนหนุ่มใหญ่(46-50ปี) ประสบการณ์ทำงานมาก(>7ปี) การศึกษาน้อย รายได้สูงมาก(มากกว่า800บาท)

ตารางที่ 4.20 (ต่อ)

กลุ่มที่	ลักษณะข้อมูลกลุ่ม	ลักษณะเฉพาะ
4) AC_CHON = 1	มีภูมิลำเนาอยู่ ตะวันออก การศึกษา	ภูมิลำเนาอยู่ ตะวันออก
AC_GAGE2 = AGE3	ระดับชั้นประถม ช่วงอายุ 41-45 ปี	ช่วงอายุหนุ่มใหญ่(41-45)
AC_GWORK2 = WK_3	ประสบการณ์การทำงานในอาชีพขับรถ	ประสบการณ์ทำงานพอใช้
HIGH_2 = H_3	ไม่เกิน 4 ปี ส่วนสูงมากกว่า 170 ซม.	(ไม่เกิน4ปี)
WEIGHT_2 = W_3	น้ำหนักตัว 76 - 85 กก.มีรายได้มากกว่า	การศึกษาน้อย
EDUCATION = ED_1	20,000บาท แต่ไม่เกิน 25,000บาท	รายได้สูง(800บาท)
INCOME_2 = IN_4		
PHUMIPAK = PM_6		

2) การหาปัจจัยก่อให้เกิดอุบัติเหตุที่มีผลต่อกลุ่มเสี่ยง

นำกลุ่มเสี่ยงแต่ละกลุ่มมาทำการทดสอบร่วมกับชุดข้อมูลองค์ประกอบของปัจจัยการเกิดอุบัติเหตุซึ่งได้แก่ เส้นทางเดินรถ จำนวนชั่วโมงขับรถ ประเภทที่ใช้ และสถานที่เกิดเหตุ (บริษัท, สถานี, ภายนอก) ด้วยอัลกอริทึม Decision Tree Learner เพื่อหาว่าจะมีองค์ประกอบปัจจัยการเกิดอุบัติเหตุอะไรบ้างในแต่ละกลุ่มเสี่ยงของแต่ละกลุ่มโดยใช้ Decision Tree to Ruleset Node หากความสัมพันธ์ของกลุ่มเสี่ยงให้ได้ผลลัพธ์การทำนายสามารถนำไปพัฒนาเป็นระบบการทำนายผลการเกิดอุบัติเหตุจริงโดยสาธารณะ ด้วยเทคนิคอัลกอริทึม Classification Role ต้นไม้ตัดสินใจ



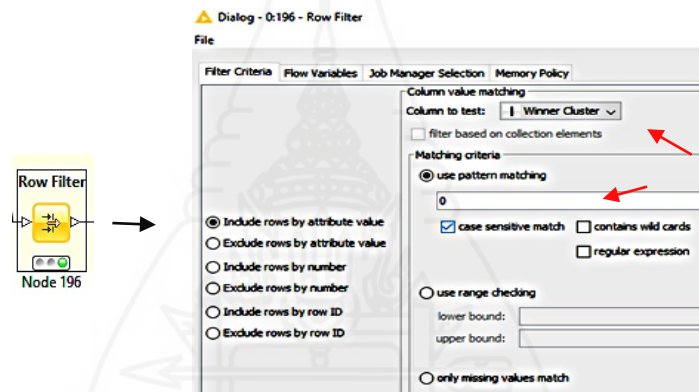
ภาพที่ 4.19 การจำแนกองค์ประกอบการเกิดอุบัติเหตุด้วย Decision Tree Learner Node

ภาพที่ 4.19 แสดงขั้นตอนการใช้เทคนิค Classification ได้ดังนี้

หมายเลข 1 Row Filter Node คัดเลือกกลุ่มที่ได้จัดแบ่งกลุ่มพนักงานขับรถ 4 กลุ่ม นำแต่ละกลุ่มมาทำการทดสอบ ส่วน Column Filter Node คัดเลือกฟิลด์ที่ต้องการให้เป็นอินพุต (INPUT) ให้กับ Decision Tree Learner Node

หมายเลข 2 Decision Tree Learner Node กำหนดค่าของ Class Column เป็นฟิลด์ AC_CHON หมายถึง ลักษณะการชน(1 รถบริษัทชนกับรถผู้อื่น, 2 รถผู้อื่นชนรถบริษัทฯ, 3 รถบริษัทชนกันเอง, 4 รถบริษัทชนอื่นๆ)

หมายเลข 3 Decision Tree to Ruleset แสดงผลลัพธ์ที่ได้รับในรูปแบบของกฎความสัมพันธ์

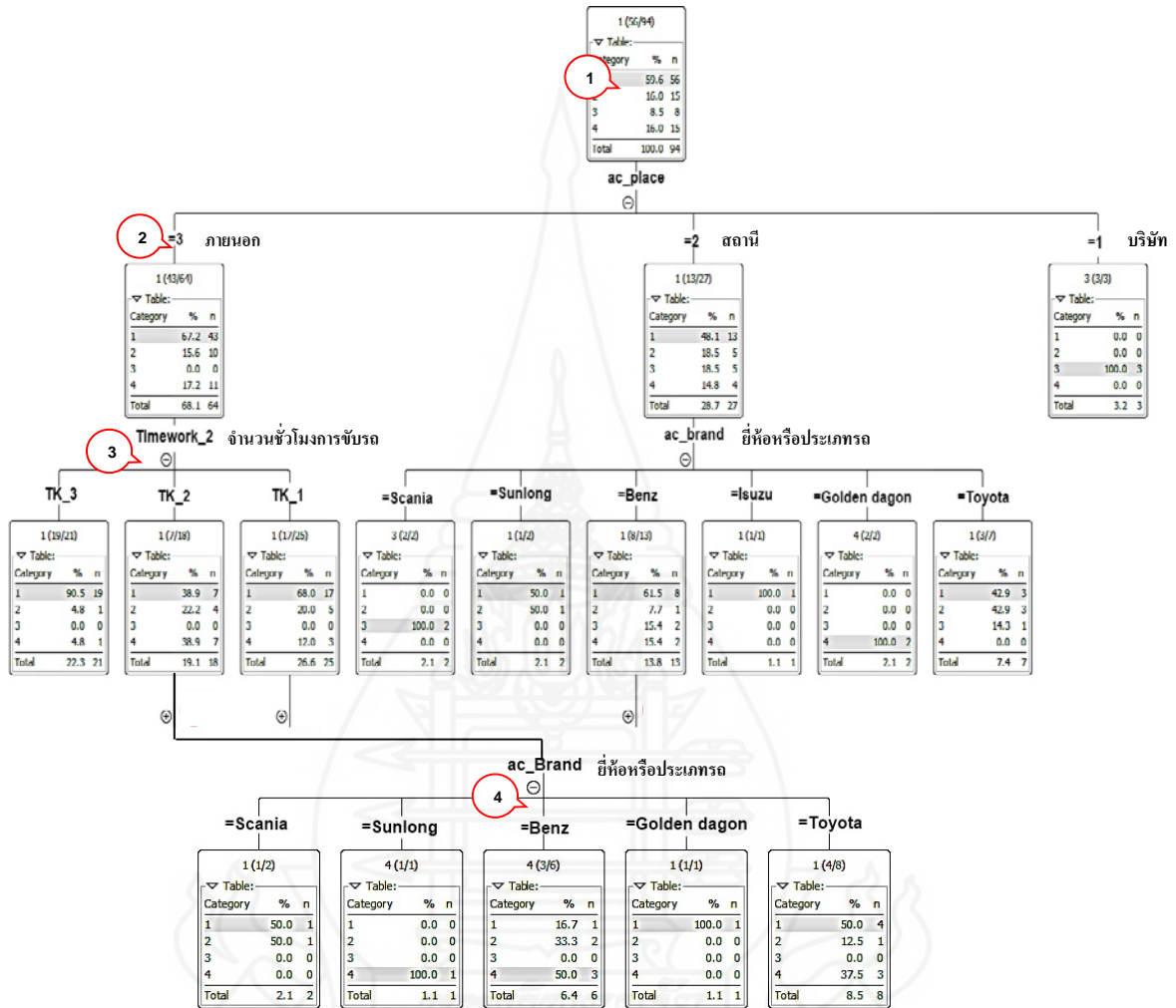


ภาพที่ 4.20 คัดเลือกกลุ่มด้วย Row Filter Node

ภาพที่ 4.20 แสดงวิธีการนำกลุ่มที่ต้องการมาทำการวิเคราะห์ก่อนเข้าสู่ขบวนการต้นไม้ตัดสินใจ โดยใช้ Row Filter Node ที่มีความสามารถในการคัดเลือกแถวตามกลุ่มที่ต้องการ ให้ทำการเลือก Column to Test ระบุ Winner Cluster และเลือก Use Pattern Matching ระบุ 0 (0 - 3)

ในการวิเคราะห์ข้อมูลที่ต้องการหาองค์ประกอบส่วนที่มีผลต่อการเกิดอุบัติเหตุจากการขับรถนั้น ได้ทำการวิเคราะห์แต่ละกลุ่ม มีทั้งสิ้น 4 กลุ่ม ซึ่งสามารถแสดงผลในรูปแบบของต้นไม้ตัดสินใจ ตามตัวอย่างได้ดังนี้

พนักงานขับรถกลุ่มเสี่ยงที่ 1 ชุดข้อมูล Training Data จากข้อมูลจำนวน 94 ข้อมูลของกลุ่มที่ 1 ใช้ Partition Node และการกำหนดข้อมูลการเรียนรู้ กำหนด Class Column เป็นการชน(AC_CHON) และ Quality Measure เป็น Gain Ratio ใช้ Decision Tree Learner Node



ภาพที่ 4.21 การแบ่งประเภทกลุ่มเสี่ยงที่ 1 ด้วยเทคนิค ต้นไม้ตัดสินใจ

ภาพที่ 4.21 อธิบายความหมายได้ดังนี้

หมายเลข 1 "Category" ระบุเป็นคลาส (Class) ความหมาย Category ได้แก่ 1 คือ รถบริษัทชนรถอื่น, 2 คือ รถอื่นชนรถบริษัท, 3 คือ รถบริษัทชนกันเอง, 4 คือ รถบริษัทชนสิ่งอื่นๆ "n" คือ จำนวนครั้งที่ชน และ "%" คือ จำนวนครั้งที่ชนเปรียบเทียบกับจำนวน TOTAL เป็นร้อยละ

หมายเลข 2 คือ สถานที่ที่มีผลหรือเป็นปัจจัยเกิดอุบัติเหตุรถยนต์ ซึ่งสามารถดูจำนวนครั้งของการชนหรือรูปแบบการเกิดอุบัติเหตุที่ Category ส่วนความหมายของสถานที่ที่มีดังนี้

1 หมายถึง ภายในบริษัท, 2 หมายถึง ภายในสถานีต่างๆ, 3 หมายถึง ตามท้องถนนต่างๆ

หมายเลข 3 คือ จำนวนชั่วโมงการขับรถ (TIMEWORK_2) รูปแบบการเกิดอุบัติเหตุ Category ส่วนความหมายจำนวนชั่วโมงการขับรถ ได้แก่ TK_1 คือ ขับรบน้อยกว่าหรือเท่ากับ 6 ชั่วโมงต่อวัน, TK_2 คือ ขับรถ 7-8 ชั่วโมงต่อวัน, TK_3 คือ ขับรถมากกว่า 8 ชั่วโมงต่อวัน

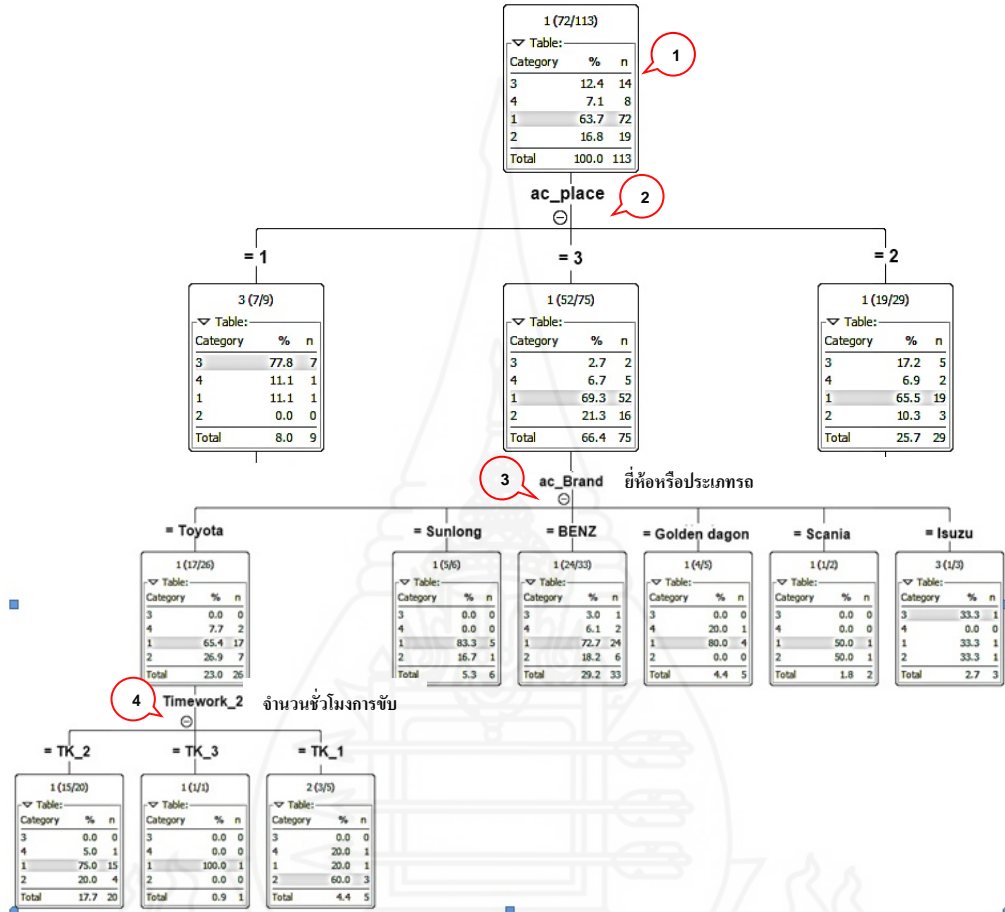
หมายเลข 4 คือ ยี่ห้อรถ (AC_BRAND) เป็นปัจจัยเกิดอุบัติเหตุรถยนต์และสามารถดูจำนวนครั้งของการชนหรือรูปแบบการเกิดอุบัติเหตุที่ Category ตารางที่ 4.21 อธิบายแผนภูมิต้นไม้ตัดสินใจกลุ่มเสี่ยงที่ 1

กลุ่มที่	ลักษณะแผนภูมิ	ผล*	ข้อมูล	ถูก
กลุ่มที่ 1	TIMEWORK_2 = "TK_3" AND AC_PLACE = "3" การขับรถมากกว่า 8 ชม./วันบนเส้นทางถนนมีโอกาสรถบริษัทชนกับรถคันอื่น	1	21	19
			ร้อยละ 90	
	AC_BRAND = "TOYOTA" AND TIMEWORK_2 = "TK_2" AND AC_PLACE = "3" ขับรถตู้โตโยต้าด้วยจำนวนชั่วโมงการขับรถ 7-8 ชม./วันบนเส้นทางถนนมีโอกาสรถบริษัทชนกับรถคันอื่น	1	8	4
			ร้อยละ 50	
	AC_BRAND="BENZ" AND AC_DTYPE = "B048" AND TIMEWORK_2 = "TK_1" AND AC_PLACE = "3" ขับรถโดยสาร(BENZ)บนเส้นทางเดินรถ B048 ด้วยจำนวนชั่วโมงการขับรบน้อยกว่าหรือเท่ากับ 6 ชม./วัน มีโอกาสรถบริษัทชนกับรถคันอื่น	1	11	9
			ร้อยละ 81.8	
	AC_DTYPE = "V9905" AND TIMEWORK_2 = "TK_1" AND AC_PLACE = "3" จำนวนชั่วโมงการขับรบน้อยกว่าหรือเท่ากับ 6 ชม./วัน ด้วยเส้นทางเดินรถ v9905 มีโอกาสรถบริษัทชนกับรถคันอื่น	1	6	3
		ร้อยละ 50		
TIMEWORK_2 = "TK_1" AND AC_BRAND = "BENZ" AND AC_PLACE = "2" จำนวนชั่วโมงการขับรบน้อยกว่าหรือเท่ากับ 6 ชม./วัน ด้วยรถโดยสาร(BENZ)โอกาสรถบริษัทชนกับรถคันอื่นภายในสถานี	1	8	6	
		ร้อยละ 75		
AC_BRAND = "TOYOTA" AND AC_PLACE = "2" ขับรถตู้โตโยต้ามีโอกาสรถบริษัทชนกับรถคันอื่นภายในสถานี	1	7	3	
		ร้อยละ 42.8		

หมายเหตุ ผล* คือ การเกิดอุบัติเหตุ

(1=รถบริษัทชนรถอื่น, 2=รถอื่นชนรถบริษัท, 3=รถบริษัทชนกันเอง, 4=รถบริษัทชนอื่นฯ)

พนักงานขับรถกลุ่มเสี่ยงที่ 2 ชุดข้อมูล Training Data จากข้อมูลจำนวน 113 ข้อมูลของกลุ่มที่ 2 ใช้ Partition Node และการกำหนดข้อมูลการเรียนรู้ กำหนด Class Column เป็นการชน (AC_CHON) และ Quality Measure เป็น Gain Ratio ใช้ Decision Tree Learner Node



ภาพที่ 4.22 การแบ่งประเภทกลุ่มเสี่ยงที่ 2 ด้วยเทคนิค ต้นไม้ตัดสินใจ

ภาพที่ 4.22 อธิบายความหมายได้ดังนี้

หมายเลข 1 "Category" ระบุเป็นคลาส (Class) ความหมาย Category ได้แก่ 1 คือ รถบริษัทชนรถอื่น, 2 คือ รถอื่นชนรถบริษัท, 3 คือ รถบริษัทชนกันเอง, 4 คือ รถบริษัทชนสิ่งอื่นๆ "n" คือ จำนวนครั้งที่ชนหรืออุบัติเหตุ และ "%" คือ จำนวนครั้งที่ชน TOTAL คิดเป็นร้อยละ

หมายเลข 2 สถานที่ที่มีผลหรือเป็นปัจจัยเกิดอุบัติเหตุรถยนต์ ซึ่งสามารถดูจำนวนครั้งของการชนหรือรูปแบบการเกิดอุบัติเหตุที่ Category ส่วนความหมายของสถานที่มี ดังนี้ 1 หมายถึง ภายในบริษัท, 2 หมายถึง ภายในสถานที่ต่างๆ, 3 หมายถึง ภายนอกหรือตามท้องถนนต่างๆ

หมายเลข 3 ยี่ห้อรถหรือประเภทรถ (AC_BRAND) ซึ่งเป็นปัจจัยเกิดอุบัติเหตุรถยนต์ได้เช่นกัน และสามารถดูจำนวนครั้งของการชนหรือรูปแบบการเกิดอุบัติเหตุที่ Category

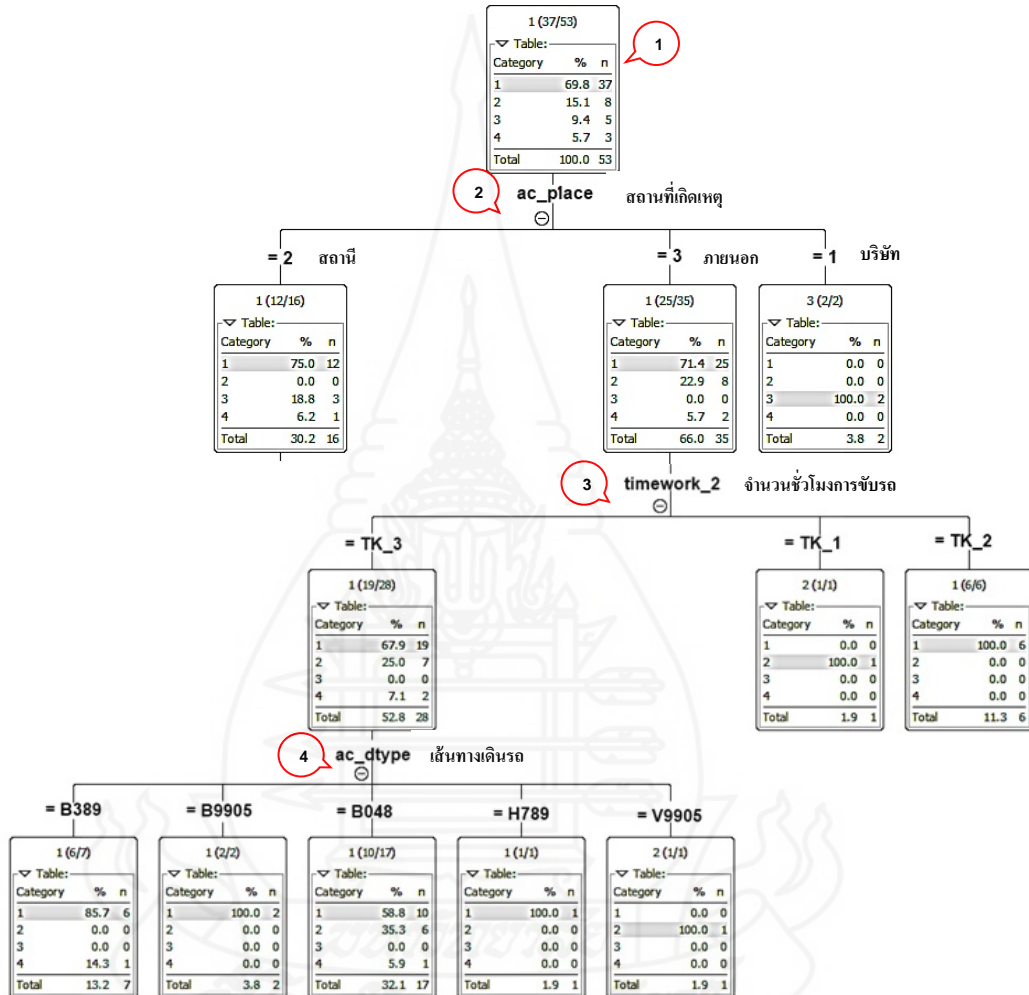
หมายเลข 4 จำนวนชั่วโมงการขับรถ (TIMEWORK_2) ของพนักงานขับรถโดยเฉลี่ยต่อวันซึ่งเป็นปัจจัยเกิดอุบัติเหตุรถยนต์ ซึ่งสามารถดูจำนวนครั้งของการชนหรือรูปแบบการเกิดอุบัติเหตุที่ Category ส่วนความหมายจำนวนชั่วโมงการขับรถ ได้แก่ TK_1 คือขับรถน้อยกว่าหรือเท่ากับ 6 ชั่วโมงต่อวัน, TK_2 คือ ขับรถ 7-8 ชั่วโมงต่อวัน, TK_3 คือขับรถมากกว่า 8 ชั่วโมงต่อวัน ตารางที่ 4.22 อธิบายแผนภูมิต้นไม้ตัดสินใจกลุ่มเสี่ยงที่ 2

กลุ่มที่	ลักษณะแผนภูมิ	ผล*	ข้อมูล	ถูก
กลุ่มที่ 2	(\$AC_BRAND\$ = "BENZ" AND \$AC_PLACES\$ = "1") ขับรถโดยสาร(BENZ) และเกิดเหตุในบริษัทเป็นซึ่งเป็นลักษณะชน กันเอง	3	6	5
			ร้อยละ 83.3	
	\$TIMEWORK_2\$ = "TK_2" AND \$AC_BRAND\$ = "TOYOTA" AND \$AC_PLACES\$ = "3"	1	20	15
	ชั่วโมงการขับรถ 7-8 ชม./วัน ด้วยการขับรถผู้โดยสาร(โตโยต้า) เกิดเหตุ ภายนอกตามท้องถนนซึ่งเป็นลักษณะการชน รถบริษัทชนกับรถอื่นๆ		ร้อยละ 75	
	\$AC_BRAND\$ = "SUNLONG" AND \$AC_PLACES\$ = "3"	1	6	5
	ขับรถโดยสาร(SUNLONG) เกิดเหตุภายนอกตามท้องถนนซึ่งเป็น ลักษณะการชน รถบริษัทชนกับรถอื่นๆ		ร้อยละ 83.3	
\$AC_BRAND\$ = "BENZ" AND \$AC_PLACES\$ = "3"	1	33	24	
ขับรถโดยสาร(BENZ) เกิดเหตุภายนอกตามท้องถนนซึ่งเป็นลักษณะการ ชน รถบริษัทชนกับรถอื่นๆ		ร้อยละ 72.7		
\$AC_BRAND\$ = "GOLDEN DAGON" AND \$AC_PLACES\$ = "3"	1	5	4	
ขับรถโดยสาร(GOLDEN DAGON) เกิดเหตุภายนอกตามท้องถนนซึ่ง เป็นลักษณะการชน รถบริษัทชนกับรถอื่นๆ		ร้อยละ 80		
\$TIMEWORK_2\$ = "TK_2" AND \$AC_PLACES\$ = "2"	1	23	17	
จำนวนชั่วโมงการขับรถ 7-8 ชม./วัน เกิดการชนภายในสถานี เป็นการ ชนลักษณะรถบริษัทชนกับรถอื่น		ร้อยละ 74		

หมายเหตุ ผล* คือ การเกิดอุบัติเหตุ

(1=รถบริษัทชนรถอื่น, 2=รถอื่นชนรถบริษัท, 3=รถบริษัทชนกันเอง, 4=รถบริษัทชนอื่น)

พนักงานขับรถกลุ่มเสี่ยงที่ 3 ชุดข้อมูล Training Data ข้อมูลจำนวน 53 ข้อมูลของกลุ่มที่ 3 ใช้ Partition Node และการกำหนดข้อมูลการเรียนรู้ กำหนด Class Column เป็นการชน (AC_CHON) และ Quality Measure เป็น Gain Ratio ใช้ Decision Tree Learner Node



ภาพที่ 4.23 การแบ่งประเภทกลุ่มเสี่ยงที่ 3 ด้วยเทคนิค ต้นไม้ตัดสินใจ

ภาพที่ 4.23 อธิบายความหมายได้ดังนี้

หมายเลข 1 "Category" ระบุเป็นคลาส (Class) ความหมายของ Category ได้แก่ 1 คือ รถบริษัทชนรถอื่น, 2 คือ รถอื่นชนรถบริษัท, 3 คือ รถบริษัทชนกันเอง, 4 คือ รถบริษัทชนสิ่งอื่นๆ "n" คือ จำนวนครั้งที่ชน และ "%" คือ จำนวนครั้งที่ชนเปรียบเทียบกับจำนวน TOTAL เป็นร้อยละ

หมายเลข 2 สถานที่ที่เป็นปัจจัยเกิดอุบัติเหตุรถยนต์ ซึ่งสามารถดูรูปแบบการเกิดอุบัติเหตุที่ Category ส่วนความหมายของสถานที่มี ดังนี้ 1 คือ ภายในบริษัท, 2 คือ ภายในสถานีต่างๆ, 3 คือ ภายนอกหรือตามท้องถนนต่างๆ

หมายเลข 3 จำนวนชั่วโมงการขับรถ (TIMEWORK_2) เฉลี่ยต่อวัน ซึ่งสามารถดูจำนวนครั้งของการชนหรือรูปแบบการเกิดอุบัติเหตุที่ Category ส่วนความหมายจำนวนชั่วโมงการขับรถ ได้แก่ TK_1 คือ จำนวนชั่วโมงการขับร่น้อยกว่าหรือเท่ากับ 6 ชั่วโมงต่อวัน, TK_2 คือ จำนวนชั่วโมงขับรถ 7-8 ชั่วโมงต่อวัน, TK_3 คือ จำนวนชั่วโมงการขับรถมากกว่า 8 ชั่วโมงต่อวัน

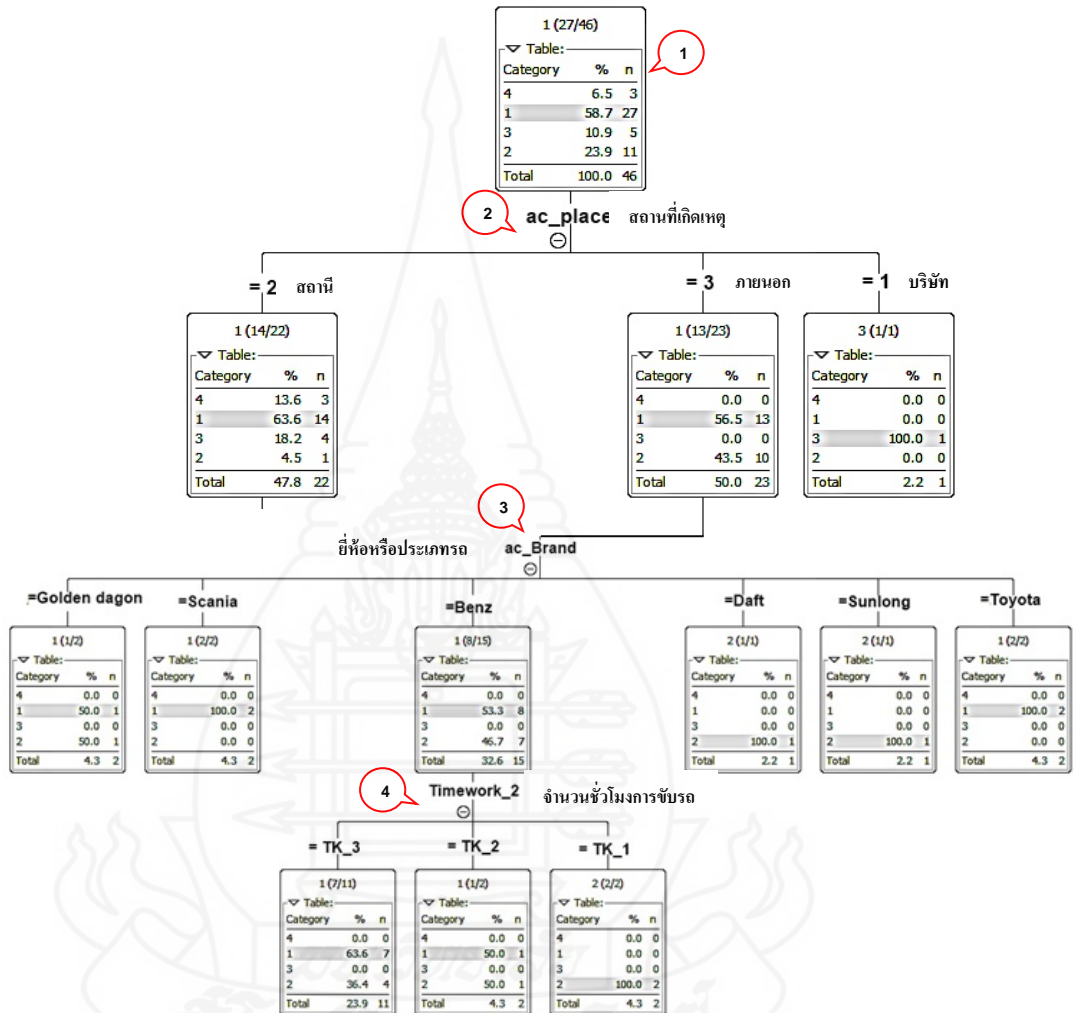
หมายเลข 4 เส้นทางรถของแต่ละเส้นทาง เช่น B389, B9905, B048, V9905 ระบุด้วยฟิลด์ AC_ โดยสามารถดูจำนวนครั้งของการชนหรือรูปแบบการเกิดอุบัติเหตุที่ Category ตารางที่ 4.23 อธิบายแผนภูมิต้นไม้ตัดสินใจกลุ่มเสี่ยงที่ 3

กลุ่มที่	ลักษณะแผนภูมิ	ผล*	ข้อมูล	ถูก
กลุ่มที่ 3	\$AC_DTYPE\$ = "B389" AND \$AC_PLACES\$ = "2"	1	5	5
	ขับรถโดยสารด้วยเส้นทาง B389 ชนภายในสถานี โดยรถบริษัทชนรถอื่น		ร้อยละ 100	
	\$AC_DTYPE\$ = "B048" AND \$AC_PLACES\$ = "2"	1	9	7
	ขับรถโดยสารด้วยเส้นทาง B048 ชนภายในสถานี โดยรถบริษัทชนรถอื่น		ร้อยละ 78	
	\$AC_DTYPE\$ = "B389" AND \$TIMEWORK_2\$ = "TK_3" AND \$AC_PLACES\$ = "3"	1	7	6
	การขับรถโดยสารด้วยเส้นทาง B389 และขับรถมากกว่า 8 ชม./วัน เกิดเหตุบนเส้นทางถนนมีโอกาสรถบริษัทชนกับรถคันอื่น		ร้อยละ 85.7	
	\$AC_DTYPE\$ = "B048" AND \$TIMEWORK_2\$ = "TK_3" AND \$AC_PLACES\$ = "3"	1	17	10
	การขับรถโดยสารด้วยเส้นทาง B048 และขับรถมากกว่า 8 ชม./วัน เกิดเหตุบนเส้นทางถนนมีโอกาสรถบริษัทชนกับรถคันอื่น		ร้อยละ 59	
\$TIMEWORK_2\$ = "TK_2" AND \$AC_PLACES\$ = "3"	1	6	6	
จำนวนขับรถ 7- 8 ชม./วัน เกิดเหตุบนเส้นทางถนนมีโอกาสรถบริษัทชนกับรถคันอื่น		ร้อยละ 100		

หมายเหตุ ผล* คือ การเกิดอุบัติเหตุ

(1=รถบริษัทชนรถอื่น, 2=รถอื่นชนรถบริษัท, 3=รถบริษัทชนกันเอง, 4=รถบริษัทชนอื่นๆ)

พนักงานขับรถกลุ่มเสี่ยงที่ 4 ชุดข้อมูล Training Data จากข้อมูลจำนวน 46 ข้อมูลของกลุ่มที่ 4 ใช้ Partition Node และการกำหนดข้อมูลการเรียนรู้ กำหนด Class Column เป็นการชน (AC_CHON) และ Quality Measure เป็น Gain Ratio ใช้ Decision Tree Learner Node



ภาพที่ 4.24 การแบ่งประเภทกลุ่มเสี่ยงที่ 4 ด้วยเทคนิค ต้นไม้ตัดสินใจ

ภาพที่ 4.24 อธิบายความหมายได้ดังนี้

หมายเลข 1 "Category" ระบุเป็นคลาส (Class) ความหมายของ Category ได้แก่ 1 คือ รถบริษัทชนรถอื่น, 2 คือ รถอื่นชนรถบริษัท, 3 คือ รถบริษัทชนกันเอง, 4 คือ รถบริษัทชนสิ่งอื่นๆ "n" คือ จำนวนครั้งที่ชน และ "%" คือ จำนวนครั้งที่ชนเปรียบเทียบกับจำนวน TOTAL เป็นร้อยละ

หมายเลข 2 สถานที่ที่มีผลหรือเป็นปัจจัยเกิดอุบัติเหตุรถยนต์ ซึ่งสามารถดูจำนวนครั้งของการชนหรือรูปแบบการเกิดอุบัติเหตุที่ Category ส่วนความหมายของสถานที่มี ดังนี้ 1 หมายถึง ภายในบริษัท, 2 หมายถึง ภายในสถานี่ต่างๆ, 3 หมายถึง ภายนอกหรือตามท้องถนนต่างๆ

หมายเลข 3 ยี่ห้อรถหรือประเภทรถ (AC_BRAND) ซึ่งเป็นปัจจัยเกิดอุบัติเหตุรถยนต์ได้เช่นกัน และสามารถดูจำนวนครั้งของการชนหรือรูปแบบการเกิดอุบัติเหตุที่ Category

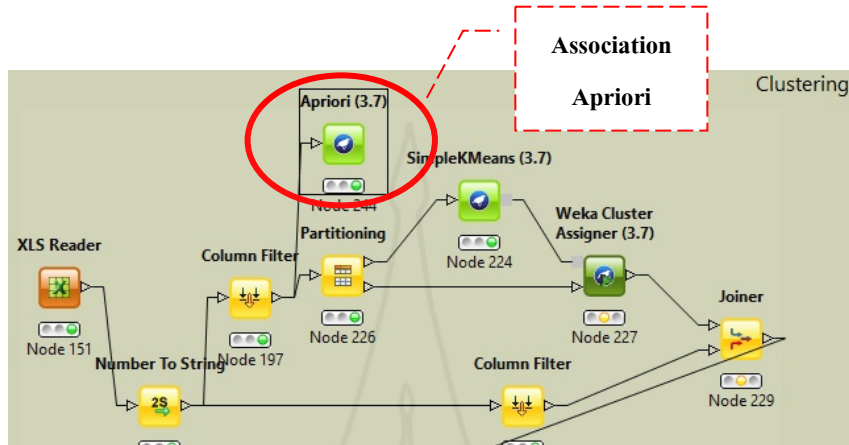
หมายเลข 4 จำนวนชั่วโมงการขับรถ (TIMEWORK_2) ของพนักงานขับรถโดยเฉลี่ยต่อวันซึ่งเป็นปัจจัยเกิดอุบัติเหตุรถยนต์ ซึ่งสามารถดูจำนวนครั้งของการชนหรือรูปแบบการเกิดอุบัติเหตุที่ Category ส่วนความหมายจำนวนชั่วโมงการขับรถ ได้แก่ TK_1 คือ จำนวนชั่วโมงการขับร่น้อยกว่าหรือเท่ากับ 6 ชั่วโมงต่อวัน, TK_2 คือ จำนวนชั่วโมงการขับรถ 7 - 8 ชั่วโมงต่อวัน, TK_3 คือ จำนวนชั่วโมงการขับร่นมากกว่า 8 ชั่วโมงต่อวัน ตารางที่ 4.24 อธิบายแผนภูมิต้นไม้ตัดสินใจกลุ่มเสี่ยงที่ 4

กลุ่มที่	ลักษณะแผนภูมิ	ผล*	ข้อมูล	ถูก
กลุ่มที่ 4	\$AC_BRANDS\$ = "BENZ" AND \$AC_PLACES\$ = "2" ขับรถโดยสาร (BENZ) เกิดเหตุการณ์ชนภายในสถานี่ ซึ่งมีโอกาส รถบริษัทชนรถอื่น	1	16	11
	\$TIMEWORK_2\$ = "TK_3" AND \$AC_BRANDS\$ = "BENZ" AND \$AC_PLACES\$ = "3" จำนวนชั่วโมงขับรถโดยสาร มากกว่า 8 ชั่วโมงต่อวัน และขับรถโดยสาร (BENZ) เกิดอุบัติเหตุบนเส้นทางภายนอก ซึ่งมีโอกาสรถบริษัทชนรถอื่น	1	11	7

หมายเหตุ ผล* คือ การเกิดอุบัติเหตุ

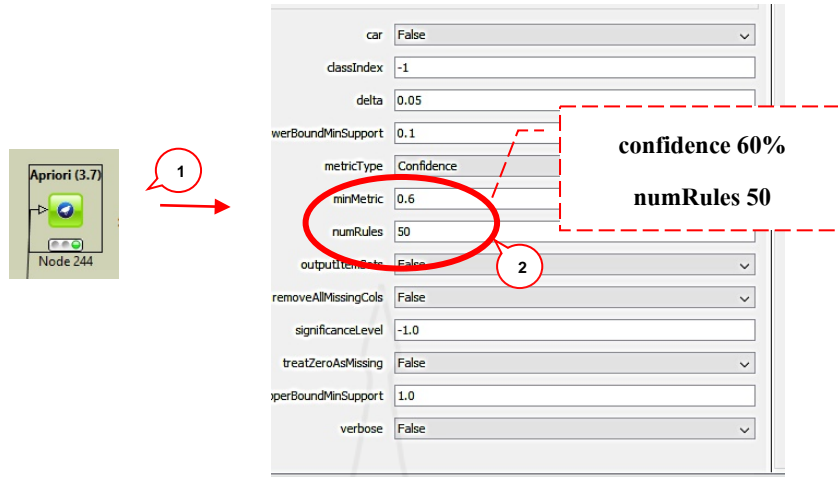
(1=รถบริษัทชนรถอื่น, 2=รถอื่นชนรถบริษัท, 3=รถบริษัทชนกันเอง, 4=รถบริษัทชนอื่นๆ)

3) กลุ่มเฝ้าระวัง (Surveillance Group)



ภาพที่ 4.25 ออกแบบจำลองกลุ่มเฝ้าระวัง

ภาพที่ 4.25 การออกแบบจำลองกลุ่มเฝ้าระวัง ผู้วิจัยทำการสอบถามผู้เชี่ยวชาญในสายงานที่เกี่ยวข้อง ได้แก่ งานด้านบุคคล ด้านการเดินรถโดยสาร และด้านอุบัติเหตุ ได้มีการแสดงความประสงค์เพิ่มเติมว่า ให้นำบุคคลที่มีคุณสมบัติส่วนบุคคลที่ไม่ครบ หรืออาจมีเพียงบางส่วนที่ถูกจัดอยู่ในกลุ่มเสี่ยง นำมาพิจารณาเพิ่มเติม เช่น อายุ การศึกษา ภูมิลำเนา น้ำหนักตัว ส่วนสูง ประสบการณ์ทำงาน และรายได้ เพราะอาจเป็นสาเหตุหนึ่งที่จะก่อให้เกิดอุบัติเหตุได้ และให้นำมาพิจารณาใช้งานของระบบ ด้วยเหตุผลที่ได้กล่าวมาเพิ่มเติมนั้น ผู้วิจัยได้นำเอาแบบจำลองอัลกอริทึมกฎความสัมพันธ์ (Association Rule) ด้วยเทคนิคอโพรไอริ (Apriori) เพื่อหาความสัมพันธ์ของคุณสมบัติส่วนบุคคล โดยกำหนดค่าความเชื่อมั่น (Confidence) ไว้ที่ร้อยละ 60 ค่าจำนวนกฎ (NumRules) 50 ค่า และค่าลิฟต์ (Lift) คือค่าสหสัมพันธ์ที่บ่งบอกถึงกฎความสัมพันธ์ที่มีค่ามากกว่า 1 ให้ได้ผลลัพธ์ความสัมพันธ์ของคุณสมบัติส่วนบุคคลมาใช้งานในระบบ โดยทำการกำหนดค่าให้ Apriori Node (3.7) ดังภาพที่ 4.26



ภาพที่ 4.26 กำหนดค่า Confidence

ภาพที่ 4.26 การกำหนดค่าให้ Apriori Node (3.7) ดังนี้

หมายเลข 1 Apriori Node (3.7) ดับเบิลคลิก Apriori Node

หมายเลข2 ทำการกำหนดขอบเขตและกฎความสัมพันธ์ กำหนดค่า "minMetric" เท่ากับ 0.6 เพื่อเป็นการกำหนดค่า "Confidence" เท่ากับ 60 เปอร์เซ็นต์ และกำหนดค่า "numRules" เท่ากับ 50 เพื่อแสดงผลลัพธ์จากการกำหนดค่าที่ตั้งไว้ 50 รายการ

no	Best rules found:	confidence	lift
1	6. ac_chon=1 phumpak=PM_2_65 ==> education=ED_1_58	0.89	1.17
2	15. ac_chon=1 ac_gwork2=WK_2_68 ==> education=ED_1_54	0.79	1.04
3	18. ac_chon=1 income_2=IN_4_78 ==> education=ED_1_61	0.78	1.03
4	20. ac_chon=1 weight_2=W_2_83 ==> education=ED_1_64	0.77	0.91
5	22. ac_chon=1 weight_2=W_1_125 ==> education=ED_1_94	0.75	0.99
6	24. ac_chon=1_255 ==> education=ED_1_189	0.74	0.97
7	25. ac_chon=2_69 ==> education=ED_1_51	0.74	0.97
8	28. ac_chon=1 ac_gage2=AGE3_56 ==> education=ED_1_41	0.73	0.96
9	30. ac_gage2=AGE4_69 ==> ac_chon=1_50	0.72	1.15
10	31. ac_chon=1 high_2=H_2_126 ==> education=ED_1_91	0.72	0.95
11	32. education=ED_2_57 ==> ac_chon=1_41	0.72	1.15
12	34. ac_chon=1 ac_gage2=AGE1_58 ==> weight_2=W_1_41	0.71	1.46
13	38. ac_chon=1 high_2=H_3_87 ==> education=ED_1_60	0.69	0.91
14	39. ac_chon=1 high_2=H_2_weight_2=W_1_78 ==> education=ED_1_53	0.68	0.89
15	41. ac_gage2=AGE1_weight_2=W_1_61 ==> ac_chon=1_41	0.67	1.07
16	43. ac_chon=1 income_2=IN_3_100 ==> education=ED_1_67	0.67	0.88
17	45. income_2=IN_4_118 ==> ac_chon=1_78	0.66	1.05
18	46. ac_gwork2=WK_1_157 ==> ac_chon=1_103	0.66	1.04
19	47. high_2=H_2_weight_2=W_1_119 ==> ac_chon=1_78	0.66	1.04
20	48. ac_gwork2=WK_1_high_2=H_2_87 ==> ac_chon=1_57	0.66	1.04

ภาพที่ 4.27 กฎความสัมพันธ์ข้อมูลค่า Lift มากกว่า 1

ภาพที่ 4.27 แสดงความสัมพันธ์ โดยมีรายละเอียดของฟิลด์ข้อมูลได้แก่

- "AC_CHON" หมายถึง ลักษณะการชน(1 รถบริษัทชนรถคันอื่น)
- "AC_GAGE2" หมายถึง ช่วงอายุพนักงานขับรถที่เกิดอุบัติเหตุ

"AC_GWORK2"	หมายถึง	ช่วงอายุงานของพนักงานขับรถที่เกิดอุบัติเหตุ
"High_2"	หมายถึง	ส่วนสูงของพนักงานขับรถ
"Weight_2"	หมายถึง	น้ำหนักของพนักงานขับรถ
"EDUCATION"	หมายถึง	การศึกษาของพนักงานขับรถ
"INCOME_2"	หมายถึง	รายได้ของพนักงานขับรถ
"PHUMIPAK"	หมายถึง	ถิ่นกำเนิดของพนักงานขับรถ

ส่วนความหมายของคุณสมบัติพนักงานขับรถแสดงไว้ตารางที่ 4.5

ผลลัพธ์จากโมเดลความสัมพันธ์มาทำการเพิ่มเติมเพื่อหาความสัมพันธ์ โดยใช้ค่าความเชื่อมั่น Confidence เท่ากับหรือมากกว่าร้อยละ 60 และค่า Lift มากกว่า 1 ซึ่งสามารถอธิบายความหมาย ดังตัวอย่าง เช่น

AC_CHON=1 PHUMIPAK=PM_2 65=> EDUCATION=ED_1 58

"CONFIDENCE=0.89" "LIFT=1.17"

ลำดับที่ 1 " การเกิดอุบัติเหตุ (ac_chon=1) รถบริษัทฯ ชนกับรถอื่น มีปัจจัยทางด้านคุณสมบัติพนักงานขับรถที่ทำให้เกิดอุบัติเหตุ ได้แก่ ภูมิลำเนาจากภาคกลาง, และการศึกษาระดับชั้นประถมศึกษา " แสดงค่าความเชื่อมั่นร้อยละ 89 และค่าลิฟต์ 1.17 มีความสัมพันธ์ต่อกัน

2.5 การประเมินประสิทธิภาพแบบจำลอง (Evaluation)

การทดสอบประสิทธิภาพแบบจำลองข้อมูลการเกิดอุบัติเหตุรถโดยสาร โดยข้อมูลอุบัติเหตุรถโดยสารสาธารณะมีจำนวน 406 ระเบียบ ระหว่างพ.ศ. 2556 -2559 ข้อมูลแบ่งเป็น 2 ชุด ข้อมูลชุดที่หนึ่งเป็นข้อมูลองค์ประกอบของปัจจัยการเกิดอุบัติเหตุ ได้แก่ เส้นทางเดินรถ จำนวนชั่วโมงขับรถ ประเภทรถที่ใช้ และสถานที่เกิดเหตุ (บริษัท, สถานี, ภายนอก) และข้อมูลชุดที่สองเป็นข้อมูลคุณสมบัตินักคน ได้แก่ อายุ การศึกษา ภูมิลำเนา น้ำหนักตัว ส่วนสูง ประสบการณ์ทำงาน และรายได้ ผลการทดสอบได้กลุ่มข้อมูลสองลักษณะ คือ กลุ่มเสี่ยง และกลุ่มเฝ้าระวัง ซึ่งสามารถอธิบายผลของการประเมินแบบจำลอง ดังนี้

กลุ่มเสี่ยง การทดสอบประสิทธิภาพมี 2 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 โดยใช้ข้อมูลคุณสมบัตินักคน ทดสอบแบบจำลองด้วยอัลกอริทึม Simple K-Means เพื่อทำการแบ่งกลุ่มเสี่ยงโดยใช้การแบ่งข้อมูลด้วยเทคนิค Split Test ได้แก่ Training Data จำนวน 306 ระเบียบสำหรับสร้างแบบจำลอง และ Testing Data จำนวน 100 ระเบียบ

สำหรับทดสอบแบบจำลอง จากการทดสอบแบบจำลองของข้อมูล Training Data และข้อมูล Testing Data กำหนดกลุ่มทดสอบจำนวน 3-10 กลุ่มด้วยอัลกอริทึม Simple K-Means เพื่อหา ลักษณะเฉพาะของจำนวนกลุ่มที่เหมาะสม และนำเสนอผู้บริหารบริษัทฯในพิจารณาจำนวนกลุ่ม เพื่อคัดเลือก ได้ข้อสรุปการแบ่งกลุ่มเสี่ยงทั้งสิ้น 4 กลุ่ม เหมาะสมที่สุด เพราะมีลักษณะเฉพาะกลุ่มเสี่ยงชัดเจน

ขั้นตอนที่ 2 นำกลุ่มเสี่ยงแต่ละกลุ่มมาทำการทดสอบร่วมกับชุดข้อมูลองค์ประกอบของปัจจัยการเกิดอุบัติเหตุ ได้แก่ เส้นทางเดินรถ จำนวนชั่วโมงขับรถ ประเภทรถที่ใช้ และสถานที่เกิดเหตุ (บริษัท, สถานี, ภายนอก) ด้วยอัลกอริทึม Decision Tree Learner เพื่อหาว่าจะมีองค์ประกอบปัจจัยการเกิดอุบัติเหตุอะไรบ้างในแต่ละกลุ่มเสี่ยงของแต่ละกลุ่ม โดยการใช้ Decision Tree to Ruleset Node หากความสัมพันธ์ของกลุ่มเสี่ยงให้ได้ผลลัพธ์การทำนายสามารถนำไปพัฒนาเป็นระบบการทำนายผลการเกิดอุบัติเหตุโดยสารสนเทศ ตามตารางที่ 4.25

ตารางที่ 4.25 การประเมินค่าความสัมพันธ์ของกลุ่มเสี่ยงและการเกิดอุบัติเหตุ

กลุ่ม	กฎความสัมพันธ์	ความน่าเชื่อถือร้อยละ
	การเกิดอุบัติเหตุรถบริษัทฯชนรถอื่นเกิดจากจำนวนชั่วโมงการขับรถต่อวันโดยเฉลี่ยมากกว่า 8 ชั่วโมงต่อวัน	90.5
กลุ่ม 1	การเกิดอุบัติเหตุรถบริษัทฯชนรถอื่นเกิด โดยใช้รถตู้โตโยต้า และจำนวนชั่วโมงการขับรถต่อวันโดยเฉลี่ย 7 - 8 ชั่วโมงต่อวันในเส้นทางเดินรถ V9905	50
	การเกิดอุบัติเหตุรถบริษัทฯชนรถอื่นเกิด โดยใช้รถปรับอากาศ BENZจำนวนชั่วโมงการขับรถต่อวันเฉลี่ยน้อยกว่า 6 ชั่วโมงต่อวัน โดยเกิดเหตุที่สถานี	75
กลุ่ม 2	การเกิดอุบัติเหตุรถบริษัทฯชนกันเอง ที่เกิดเหตุบริเวณภายในบริษัท โดยรถที่ใช้ในการขับขี่เป็นรถโดยสารยี่ห้อ BENZ	83
	การเกิดอุบัติเหตุรถบริษัทฯชนรถอื่นเกิด โดยใช้รถตู้โตโยต้า และจำนวนชั่วโมงการขับรถต่อวันโดยเฉลี่ย 7 - 8 ชั่วโมงต่อวันเกิดเหตุภายนอกบริษัท	75
	การเกิดอุบัติเหตุรถอื่นชนรถบริษัทฯชนเกิด โดยรถที่ถูกชนเป็นรถตู้โตโยต้า จำนวนชั่วโมงการขับรถต่อวันโดยเฉลี่ยน้อยกว่า 6 ชั่วโมงต่อวันเกิดเหตุภายนอกบริษัท	60
	การเกิดอุบัติเหตุรถบริษัทฯชนรถอื่นเกิด โดยรถยี่ห้อ SUNLONG เกิดเหตุภายนอกบริษัท	83
	การเกิดอุบัติเหตุรถบริษัทฯชนรถอื่นเกิด โดยรถยี่ห้อ BENZ เกิดเหตุภายนอกบริษัท	73
	การเกิดอุบัติเหตุรถบริษัทฯชนรถอื่นเกิด โดยใช้รถปรับอากาศ GOLDEN DAGON เกิดเหตุภายนอกบริษัท	80
	การเกิดอุบัติเหตุรถบริษัทฯชนรถอื่นเกิด โดยจำนวนชั่วโมงการขับรถต่อวันเฉลี่ย 7 - 8 ชั่วโมงต่อวัน โดยเกิดเหตุที่สถานี	74

ตารางที่ 4.25 (ต่อ)

กลุ่ม	กฎความสัมพันธ์	ความน่าเชื่อถือร้อยละ
	การเกิดอุบัติเหตุรถบริษัทชนรถอื่น สำหรับรถโดยสารในเส้นทาง B389 โดยเกิดเหตุภายในสถานี	100
กลุ่ม 3	การเกิดอุบัติเหตุรถบริษัทชนรถอื่น สำหรับรถโดยสารในเส้นทาง B048 โดยเกิดเหตุภายในสถานี	77.8
	การเกิดอุบัติเหตุรถบริษัทชนรถอื่นเกิดในเส้นทาง B389 และจำนวนชั่วโมงการขับรถต่อวันโดยเฉลี่ยมากกว่า 8 ชั่วโมงต่อวันเกิดเหตุภายนอกบริษัท	85
	การเกิดอุบัติเหตุรถบริษัทชนรถอื่นเกิดในเส้นทาง B048 และจำนวนชั่วโมงการขับรถต่อวันโดยเฉลี่ยมากกว่า 8 ชั่วโมงต่อวันเกิดเหตุภายนอกบริษัท	58.8
	การเกิดอุบัติเหตุรถบริษัทชนรถอื่นเกิดจากจำนวนชั่วโมงการขับรถต่อวันโดยเฉลี่ยมากกว่า 7- 8 ชั่วโมงต่อวันเกิดเหตุภายนอกบริษัทฯ	100
กลุ่ม 4	การเกิดอุบัติเหตุรถบริษัทชนรถอื่นเกิดโดยใช้รถปรับอากาศ BENZ เกิดเหตุภายในสถานี	68.7
	การเกิดอุบัติเหตุรถบริษัทชนรถอื่นเกิดโดยใช้รถปรับอากาศ BENZ จำนวนชั่วโมงการขับรถต่อวันเฉลี่ยมากกว่า 8 ชั่วโมงต่อวัน เกิดเหตุภายนอกบริษัทฯ	63.6

ตารางที่ 4.26 สรุปผลการทดสอบของกลุ่มเสี่ยงและการเกิดอุบัติเหตุ

กลุ่มเสี่ยง	ลักษณะเฉพาะกลุ่มบุคคล	ลักษณะการเกิดเหตุ	เกิดเหตุร้อยละ
	ชาย -> กรุงเทพฯปริมณฑล -> อายุ 51-55 ->	- ขับรถมากกว่า 8 ชั่วโมงต่อวัน	90.5
กลุ่ม 1	ประสบการณ์มากกว่า 7 ปี -> การศึกษา (ประถม) -> รายได้เฉลี่ยต่อวัน 300-500 บาท	- ขับรถต่อวัน 7-8 ชั่วโมง รถตู้ V9905 - ขับรถต่อวันน้อยกว่า 6 ชั่วโมง รถ BENZ เกิดเหตุที่สถานี	50 75
	ชาย -> ภาคอีสาน -> อายุ 30-35 ปี ->	- ขับรถ 7-8 ชั่วโมงต่อวัน	74-75
กลุ่ม 2	ประสบการณ์ 1-2 ปี -> การศึกษา(ประถม) -> รายได้เฉลี่ยต่อวัน 500-700 บาท	- BENZ, GOLDEN DAGON, SUNLONG - สถานี, ภายนอก	73 - 83 74 - 75
	ชาย -> ภาคกลาง -> อายุ 46--50 ปี ->	-เส้นทาง B389 เกิดเหตุภายในสถานี	100
กลุ่ม 3	ประสบการณ์มากกว่า 7 ปี -> การศึกษา (ประถม) -> รายได้เฉลี่ยต่อวัน มากกว่า 800 บาท	- เส้นทาง B048 เกิดเหตุภายในสถานี - ขับรถมากกว่า 8 ชั่วโมงต่อวันเส้นทาง B389 - ขับรถมากกว่า 8 ชั่วโมงต่อวันเส้นทาง B048	77.8 85 58.5
	ชาย -> ภาคตะวันออก -> อายุ 41-45 ->	- ขับรถน้อยกว่า 6 ชั่วโมงต่อวัน รถ BENZ	63.6
	กลุ่ม 4	ประสบการณ์น้อยกว่า 4 ปี -> การศึกษา (ประถม) -> รายได้เฉลี่ยต่อวันมากกว่า 800 บาท	- รถ BENZ เกิดเหตุภายในสถานี

ตารางที่ 4.26 สามารถอธิบายผลการทดสอบกลุ่มเสี่ยง โดยสังเขป ดังนี้ กลุ่มเสี่ยงที่ 1 ลักษณะเฉพาะกลุ่มบุคคล (เพศชาย, ภูมิลำเนาที่กรุงเทพฯปริมณฑล, อายุ 51-55 ปี, ประสบการณ์มากกว่า 7 ปี, การศึกษาระดับประถมศึกษา, รายได้ต่อวัน 300-500 บาท) ถ้าขับรถโดยสารสาธารณะมีความเสี่ยงก่อให้เกิดอุบัติเหตุโดยสารสาธารณะ เมื่อขับรถมากกว่า 8 ชั่วโมงต่อวันคิดเป็นร้อยละ 90.5 ส่วนการขับรถโดยสารมีความเสี่ยงก่อให้เกิดอุบัติเหตุ เมื่อขับรถมากกว่า 8 ชั่วโมงต่อวัน คิดเป็นร้อยละ 50 และรถโดยสารยี่ห้อ BENZ มีความเสี่ยงก่อให้เกิดอุบัติเหตุโดยสารภายในสถานี เมื่อขับรถน้อยกว่า 6 ชั่วโมงต่อวัน คิดเป็นร้อยละ 75

กลุ่มเฟ้าะวัง ใช้ข้อมูลคุณสมบัติบุคคล ทดสอบอัลกอริทึมกลุ่มเฟ้าะวัง โดยใช้กฎความสัมพันธ์ (Association Rule) ด้วยวิธี Apriori กำหนดค่าความเชื่อมั่น (Confidence) ไม่ต่ำกว่าร้อยละ 60 และค่าลิฟต์ (Lift) คือค่าสหสัมพันธ์ที่บ่งบอกถึงกฎความสัมพันธ์ที่มีค่ามากกว่า 1 ตารางที่ 4.27

ตารางที่ 4.27 การประเมินกลุ่มเฟ้าะวัง

คำสั่ง	กฎความสัมพันธ์	Confidence ร้อยละ	Lift
1	AC_CHON=1 PHUMIPAK=PM_2 65 ==> EDUCATION=ED_1 58 การเกิดอุบัติเหตุ(AC_CHON=1) รถบริษัทฯชนกับรถอื่น มีคุณสมบัติบุคคล เช่น ภูมิลำเนาจากภาคกลาง, การศึกษาระดับชั้นประถมศึกษา	89	1.17
2	AC_CHON=1 AC_GWORK2=WK_2 68 ==> EDUCATION=ED_1 54 การเกิดอุบัติเหตุ(AC_CHON=1) รถบริษัทฯชนกับรถอื่น มีคุณสมบัติบุคคล เช่น อายุงาน 1-2 ปี, การศึกษาระดับชั้นประถมศึกษา	89	1.04
3	AC_CHON=1 INCOME_2=IN_4 78 ==> EDUCATION=ED_1 61 AC_GWORK2=WK_2 การเกิดอุบัติเหตุ(AC_CHON=1) รถบริษัทฯชนกับรถอื่น มีคุณสมบัติบุคคล เช่น รายได้ 20,001 - 25,000/เดือน, มีการศึกษาระดับชั้นประถมศึกษาและอายุงาน 1-2 ปี	78	1.03
4	AC_GAGE2=AGE4 69 ==> AC_CHON=1 50 การเกิดอุบัติเหตุ(AC_CHON=1) รถบริษัทฯชนกับรถอื่น มีคุณสมบัติบุคคล เช่น อายุอยู่ระหว่าง 46-50 ปี	72	1.15
5	EDUCATION=ED_2 57 ==> AC_CHON=1 41 การเกิดอุบัติเหตุ(AC_CHON=1) รถบริษัทฯชนกับรถอื่น มีคุณสมบัติบุคคล เช่น การศึกษาระดับมัธยมต้น	72	1.15
6	AC_CHON=1 AC_GAGE2=AGE1 58 ==> WEIGHT_2=W_1 41 เกิดอุบัติเหตุ(AC_CHON=1) รถบริษัทฯชนรถอื่น มีคุณสมบัติบุคคล เช่น อายุ 30-35 ปี และมีน้ำหนัก 50-65	71	1.46
7	income_2=IN_4 118 ==> ac_chon=1 78 การเกิดอุบัติเหตุ(AC_CHON=1) รถบริษัทฯชนกับรถอื่น มีคุณสมบัติบุคคล เช่น รายได้ 20,001 - 25,000/เดือน	66	1.05

ตารางที่ 4.27 (ต่อ)

คำสั่ง	กฎความสัมพันธ์	Confidence ร้อยละ	Lift
8	AC_GWORK2=WK_1 157 ==> AC_CHON=1 103 การเกิดอุบัติเหตุ(AC_CHON=1) รถบริษัทฯชนกับรถอื่น มีคุณสมบัติบุคคล เช่น อายุงานขับรถน้อยกว่า 1 ปี	66	1.04
9	HIGH_2=H_2 WEIGHT_2=W_1 119 ==> AC_CHON=1 78 การเกิดอุบัติเหตุ(AC_CHON=1) รถบริษัทฯชนกับรถอื่น มีคุณสมบัติบุคคล เช่น มีส่วนสูง 161- 170 ซม.และน้ำหนัก 50-65 กก	66	1.04
10	AC_GWORK2=WK_1 HIGH_2=H_2 87 ==> AC_CHON=1 57 การเกิดอุบัติเหตุ(AC_CHON=1) รถบริษัทฯชนกับรถมีคุณสมบัติบุคคล เช่น อายุงานในการขับรถน้อยกว่า 1 ปี และมีส่วนสูง 161- 170 ซม.	66	1.04

ตารางที่ 4.28 สรุปผลการทดสอบกลุ่มเป้าหมาย

ลำดับ	ความสัมพันธ์นอกกลุ่มกับอุบัติเหตุ	Confidence ร้อยละ	Lift
1	ภาคกลาง -> การศึกษาประถม	89	1.17
2	ประสบการณ์น้อย(1-2ปี) -> การศึกษาประถม	89	1.04
3	รายได้ต่อวัน(700-800บาท) ->การศึกษาประถม	78	1.03
4	อายุ 30-35 ปี -> น้ำหนักตัว 50 - 65 กก.	72	1.15
5	ส่วนสูง 161-170 ซม. -> น้ำหนักตัว 50- 65 กก.	72	1.15
6	ประสบการณ์ไม่ถึง 1 ปี -> ส่วนสูง 161-170 ซม.	71	1.46
7	อายุ 45 - 50 ปี	66	1.05
8	การศึกษามัธยมต้น	66	1.04
9	รายได้เฉลี่ยต่อวัน 700 - 800บาท	66	1.04
10	ประสบการณ์น้อยมากไม่ถึง 1 ปี	66	1.04

ตารางที่ 4.28 ผลการทดสอบกลุ่มเป้าหมาย จากข้อมูลคุณสมบัติบุคคล สามารถอธิบายได้ดังนี้ เช่น ลำดับที่ 1 บุคคลที่จัดอยู่ในกลุ่มเป้าหมายที่อาจก่อให้เกิดอุบัติเหตุรถโดยสารสาธารณะ เป็นบุคคลที่มีภูมิลำเนาภาคกลาง และมีการศึกษาระดับชั้นประถมศึกษา ค่าความเชื่อมั่นร้อยละ 89 และค่าลิฟต์

1.17 ที่แสดงถึงภูมิลำนาคกลาง และการศึกษามีความสัมพันธ์ต่อกัน ส่วนลำดับที่ 2 บุคคลที่จัดอยู่ในกลุ่มเฝ้าระวังที่อาจก่อให้เกิดอุบัติเหตุโดยสาธารณชน เป็นบุคคลที่มีประสบการณ์ขับรถโดยสาธารณชน 1-2 ปี และมีการศึกษาระดับชั้นประถมศึกษา มีค่าความเชื่อมั่นร้อยละ 89 และค่าลิปต์ 1.04 ที่แสดงถึงประสบการณ์ขับรถโดยสาธารณชน และการศึกษามีความสัมพันธ์ต่อกัน

2.6 การนำไปใช้งาน (Deployment)

ระบบการทำนายการเกิดอุบัติเหตุโดยสาธารณชน ผู้วิจัยใช้โปรแกรม Microsoft Visual FoxPro v9.0 ในการพัฒนาเพื่อความสามารถสะดวกในการใช้งาน ซึ่งเป็นการนำแบบจำลองทำนายผลของกลุ่มเสี่ยง และกลุ่มเฝ้าระวังที่ได้การประเมิน ใช้เป็นเครื่องมือช่วย คัดกรองบุคคลพนักงานขับรถโดยสาร หรือผู้สมัครใหม่ เพื่อหลีกเลี่ยงความเสียหายของรถโดยสาร และชีวิตของผู้โดยสาร

3. สร้างโปรแกรมใช้งาน

ในการพัฒนาเพื่อความสามารถสะดวกในการใช้งาน ผู้วิจัยใช้โปรแกรม Microsoft Visual FoxPro v9.0 สร้างระบบการทำนายการเกิดอุบัติเหตุโดยสาธารณชน มีขั้นตอนดังนี้

- 1 การออกแบบฐานข้อมูล
- 2 เครื่องมือใช้พัฒนา
- 3 แผนผังโปรแกรม
- 4 การออกแบบโครงร่างหน้าจอ
- 5 การออกแบบหน้าจอการทำงาน

3.1 การออกแบบฐานข้อมูล

การออกแบบสร้างตารางสำหรับทำนายการเกิดอุบัติเหตุโดยสาธารณชน กำหนดตาราง 6 ตาราง คือ VPACD TABLE, VPANALYTICS TABLE, DRIVER TABLE, AC_EDUCATE TABLE, AC_PROVINCE TABLE, และ AC_SECSION TABLE ดังนี้

ตารางที่ 4.29 VPANALYTICS TABLE สำหรับเก็บข้อมูลบันทึกผลการทำนายพนักงานขับรถ

ชื่อตาราง	VPANALYTICS.DBF			
คำอธิบาย	ข้อมูลบันทึกผลการทำนายพนักงานขับรถ			
คีย์หลัก	VPMYY			
ชื่อฟิลด์	ชนิด	ขนาด	ความหมาย	ตัวอย่างข้อมูล
VP_MMY	Character	7	ช่วงเวลา	08/2016
VP_DATE	Date	8	วันที่	12/08/2016
VP_NEW	Character	1	ผู้สมัครใหม่	NEW11, 222614
VP_DRV	Character	7	พนักงานเก่า	222614
VP_SIGN	Character	3	คำนำหน้าชื่อ	นาย, นาง, นส.
VP_FNAME	Character	25	ชื่อพนักงาน	เดชรรัตน์
VP_LNAME	Character	25	นามสกุล	บุญเสมอ
VP_BIRTH	Date	8	วันเกิด	01-01-1968
VP_AGE	Character	15	อายุ	48 ปี 7 เดือน
VP_AGE2	Character	5	อายุ*	AGE_4
VP_EDU	Character	4	การศึกษา*	ED_1
VP_PHUMIPAK	Character	4	ภูมิภาค*	PM_1
VP_WEIGHT	Numeric	3	น้ำหนัก	55
VP_WEIGHT2	Character	3	น้ำหนัก*	W_1
VP_HIGH	Numeric	3	ส่วนสูง	160
VP_HIGH2	Character	3	ส่วนสูง*	H_1
VP_TIMEWORK	Numeric	2	ชั่วโมงทำงาน/วัน	8
VP_TIMEWORK2	Character	4	ชั่วโมงทำงาน/วัน*	TK_2
VP_GWORK	Numeric	5.2	ประสบการณ์	3
VP_PROVINCE	Character	2	รหัสจังหวัด	01 = "กรุงเทพมหานคร"
VP_GWORK2	Character	4	ประสบการณ์*	WK_3
VP_ROUTE	Character	5	เส้นทางเดินรถ	b048
VP_TYPEBUS	Character	15	ยี่ห้อรถ	BENZ
VP_CHON	Character	1	ลักษณะอุบัติเหตุ	1 = "รถบริษัทฯชนรถอื่น"
VP_PLACE	Character	1	สถานที่อุบัติเหตุ	3 = "สถานที่ภายนอก"
VP_OUTPUT1	Character	100	ทำนาย1	ผ่าน
VP_OUTPUT2	Character	100	ทำนาย2	ผ่าน

ตารางที่ 4.30 AC_SECSION TABLE สำหรับเก็บข้อมูลภูมิภาค

ชื่อตาราง	Ac_secsion.dbf			
คำอธิบาย	ภูมิภาค			
คีย์หลัก	SCCODE			
ชื่อฟิลด์	ชนิด	ขนาด	ความหมาย	ตัวอย่างข้อมูล
SC_CODE	Character	4	รหัสภูมิภาค	PM_2
SC_NAME	Character	25	ชื่อภูมิภาค	ภาคกลาง

ตารางที่ 4.31 AC_EDUCATE TABLE สำหรับเก็บข้อมูลการศึกษา

ชื่อตาราง	Ac_educate.dbf			
คำอธิบาย	การศึกษา			
คีย์หลัก	EDCODE			
ชื่อฟิลด์	ชนิด	ขนาด	ความหมาย	ตัวอย่างข้อมูล
ED_CODE	Character	4	รหัสการศึกษา	ED_1
ED_NAME	Character	25	ชื่อการศึกษา	ประถมศึกษา

ตารางที่ 4.32 DRIVER TABLE สำหรับเก็บข้อมูลประวัติพนักงานขับรถ

ชื่อตาราง	DRIVER.DBF			
คำอธิบาย	ประวัติพนักงานขับรถ			
คีย์หลัก	RRECODE			
ชื่อฟิลด์	ชนิด	ขนาด	ความหมาย	ตัวอย่างข้อมูล
RRE_CODE	Character	7	รหัสพนักงาน	2221621
RRE_STATUS	Logical	1	สถานะ	.F.
RRE_SIGN	Character	3	คำนำหน้า	นาย
RRE_FNAME	Character	25	ชื่อ	ชุมพบ
RRE_LNAME	Character	25	นามสกุล	จันทร์ล้อม
RRE_START	Date	8	วันที่เริ่มงาน	12/04/2006
BUNJU_LOGI	Logical	1	สถานะบรรจุงาน	.T.
RRE_MARY	Character	5	สถานะส่วนตัว	แต่งงาน

ตารางที่ 4.32 (ต่อ)

ชื่อฟิลด์	ชนิด	ขนาด	ความหมาย	ตัวอย่างข้อมูล
RRE_CODE	Character	7	รหัสพนักงาน	2221621
RRE_SON	Numeric	2	จำนวนบุตร	2
BIRTH	Date	8	วันเกิด	01/03/1954
RRE_PICT	Character	50	ที่อยู่รูปภาพ	G:\picture\221621.jpg
HIGH	Character	3	ส่วนสูง	160
WEIGHT	Character	3	น้ำหนัก	52
RJANGWAT	Character	40	พื้นเพจังหวัด	ลพบุรี
RPHUMIPAK	Character	4	ภูมิภาค	PM_2
STUDY1	Character	60	การศึกษา	ป6
EDUCATE	Character	4	การศึกษา*	ED_1
RRP_CODE	Character	3	รหัสตำแหน่งงาน	114
RRP_NAME	Character	25	ชื่อตำแหน่งงาน	พนักงานขับรถสาย
RRD_CODE	Character	3	รหัสแผนงาน	D15
RRD_NAME	Character	25	ชื่อแผนงาน	คนขับรถ

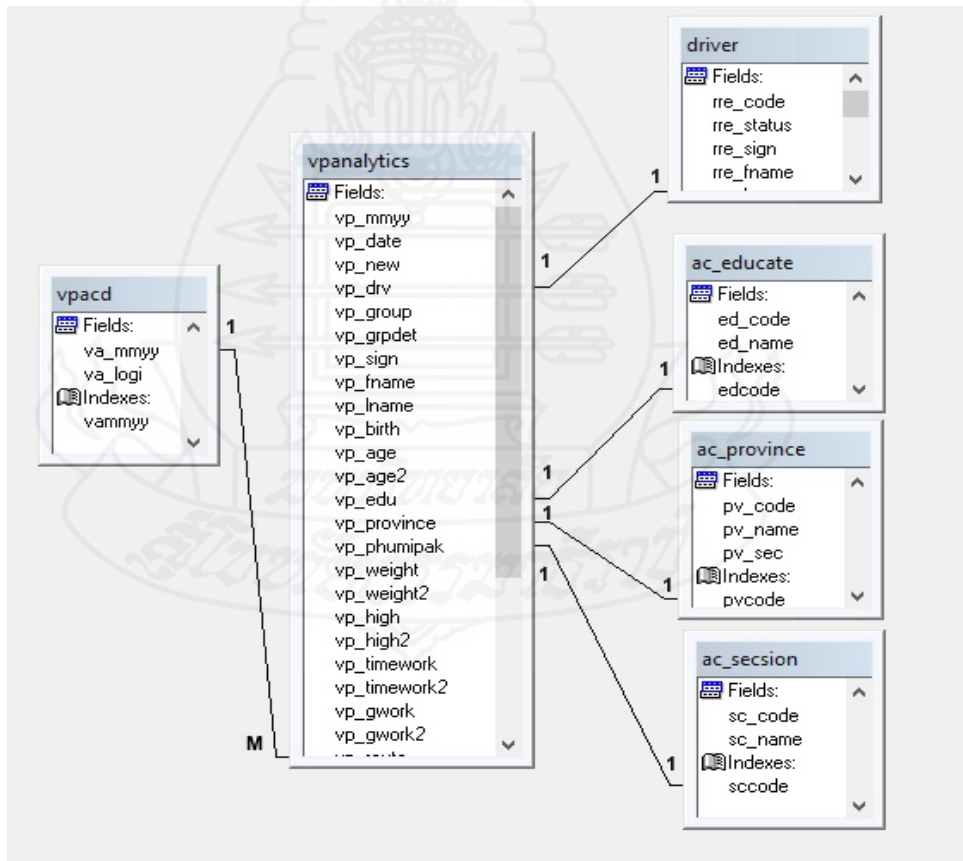
ตารางที่ 4.33 AC_PROVINCE TABLE สำหรับเก็บข้อมูลชื่อจังหวัด

ชื่อตาราง	AC_PROVINCE.DBF			
คำอธิบาย	ชื่อจังหวัด			
คีย์หลัก	PVCODE			
ชื่อฟิลด์	ชนิด	ขนาด	ความหมาย	ตัวอย่างข้อมูล
PV_CODE	Character	2	รหัสจังหวัด	14
PV_NAME	Character	25	ชื่อจังหวัด	ลพบุรี
PV_SEC	Character	10	รหัสภูมิภาค	PM_2

ตารางที่ 4.34 VP_ACD TABLE สำหรับเก็บข้อมูลบันทึกช่วงเวลา

ชื่อตาราง	VP_ACD.DBF			
คำอธิบาย	ข้อมูลบันทึกช่วงเวลา(เดือน/ปี)ผลการทำนายพนักงานขับรถ(บุคคล)			
คีย์หลัก	VAMMY			
ชื่อฟิลด์	ชนิด	ขนาด	ความหมาย	ตัวอย่างข้อมูล
VA_MMY	Character	7	เดือน/ปี	10/2016
VA_LOGI	LOGICAL	1	สถานะ	.f.

กำหนดการเชื่อมโยงความสัมพันธ์ระหว่างตาราง 6 ตาราง คือ VPACD TABLE เชื่อมโยง VPANALYTICS TABLE และ VPANALYTICS TABLE เชื่อมโยงกับ DRIVER TABLE, AC_EDUCATE TABLE, AC_PROVINCE TABLE, และ AC_SECSION TABLE



ภาพที่ 4.28 แสดงความสัมพันธ์ข้อมูลทำนายการเกิดอุบัติเหตุโดยสาร

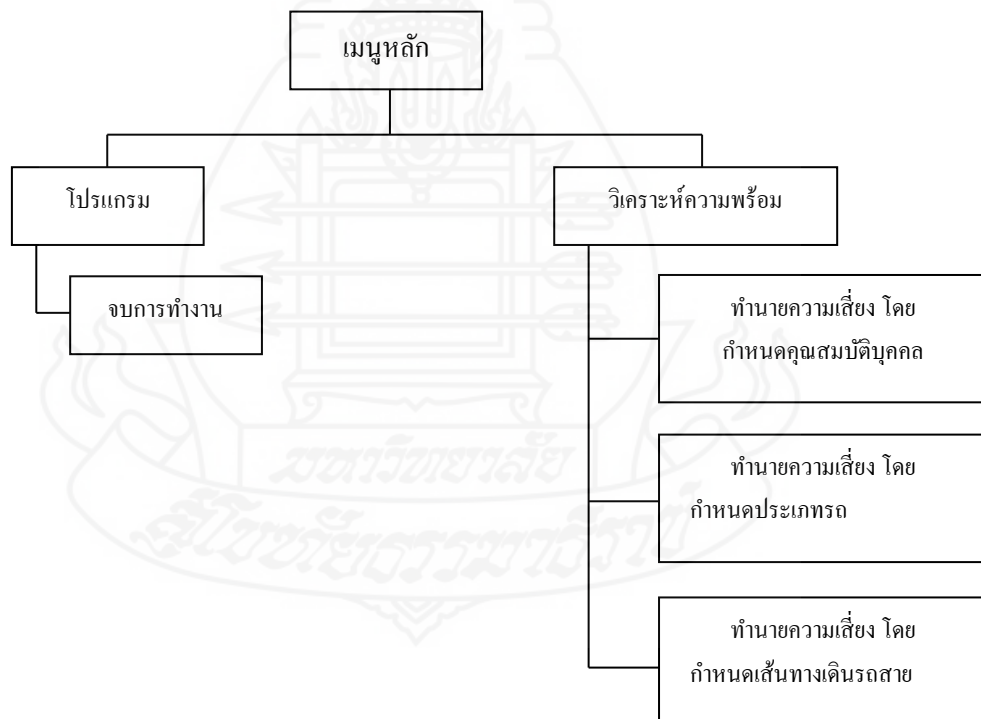
3.2 เครื่องมือใช้พัฒนา

โปรแกรม Microsoft Visual FoxPro v9.0 การใช้งานด้านการเขียน โปรแกรม รูปแบบต่างๆ ที่มีความสามารถ การจัดการฐานข้อมูลด้วยตัวเอง ออกแบบฟอร์มการรับข้อมูล การรายงาน สำหรับงานวิจัยนี้



ภาพที่ 4.29 โปรแกรม Microsoft Visual FoxPro v9.0

3.3 แผนผังโปรแกรม (Site Map Design)



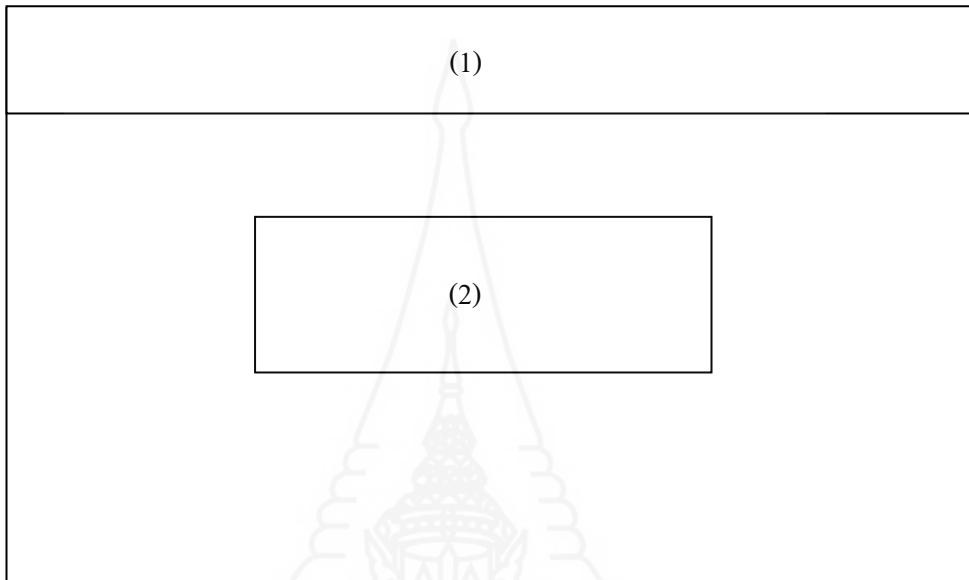
ภาพที่ 4.30 แผนผังโปรแกรม

ภาพที่ 4.30 แสดงแผนผังของโปรแกรมระบบความสัมพันธ์การเกิดอุบัติเหตุของพนักงานขับรถ แสดงผลการวิเคราะห์ ทำนายที่อาจก่อให้เกิดอุบัติเหตุการขับรถโดยสาร สามารถตรวจสอบ

วิเคราะห์ความพร้อม 3 ลักษณะ ได้แก่ ทำนายความเสี่ยงโดยกำหนดคุณสมบัติบุคคล ทำนายความเสี่ยงโดยกำหนดประเภทรถ ทำนายความเสี่ยงโดยกำหนดเส้นทางการเดินรถสาย

3.4 การออกแบบโครงร่างหน้าจอ

3.4.1 หน้าจอเข้าระบบงาน



ภาพที่ 4.31 หน้าจอเข้าระบบงาน

ภาพที่ 4.30 อธิบายส่วนประกอบต่างๆของหน้าจอเข้าระบบงาน ได้แก่

หมายเลข 1 หัวเรื่องแสดงชื่อของระบบงาน

หมายเลข 2 การเข้าสู่ระบบ มีส่วนประกอบ ดังนี้

ช่อง " ชื่อผู้ใช้งาน " ระบุชื่อผู้ใช้งานระบบ

ช่อง " รหัสผ่าน " ระบุรหัสผ่านผู้ใช้งานระบบ

ปุ่ม " เข้าสู่ระบบ " สำหรับการเข้าสู่ระบบ

ปุ่ม " ออกจากระบบ " สำหรับออกจากระบบ

3.4.2 หน้าจอหลัก

(1)
(2)

ภาพที่ 4.32 หน้าจอหลัก

ภาพที่ 4.32 อธิบายส่วนประกอบต่างๆของหน้าจอหลัก ได้แก่

หมายเลข 1 หัวเรื่องแสดงชื่อของระบบงาน

หมายเลข 2 หัวข้อการใช้งาน มีส่วนประกอบดังนี้

กำหนดผู้ใช้งาน ได้แก่ หัวข้อการใช้งาน การทำนายความเสี่ยงโดยระบุคุณสมบัติบุคคล
และการทำนายความเสี่ยงโดยระบุองค์ประกอบการเดินรถ

หัวข้อรายงาน

ออกจากระบบ

3.4.3 หน้าจอหัวข้อการทำนายความเสี่ยงโดยกำหนดคุณสมบัติส่วนบุคคล

การทำนายความเสี่ยงโดยกำหนดคุณสมบัติบุคคล

(1)			
(2)		(3)	
(4)	(5)		(6)
(7)		(11)	
(8)		(12)	
(9)		(13)	
(10)			
		(14)	(15)
		(16)	(17)
(18)			

ภาพที่ 4.33 หน้าจอหัวข้อการทำนายความเสี่ยงโดยกำหนดคุณสมบัติส่วนบุคคล

ภาพที่ 4.33 อธิบายส่วนประกอบต่างๆของหน้าจอหัวข้อการทำนายความเสี่ยงโดยระบุคุณสมบัติส่วนบุคคล ได้แก่

หมายเลข 1 ช่อง " เดือน - พ.ศ. " ระบุช่วงเดือน - พ.ศ.

หมายเลข 2 ช่อง " ผู้สมัครใหม่ " เลือกสำหรับผู้สมัครงานใหม่

หมายเลข 3 ช่อง " พนักงานขับรถบริษัท " เลือกสำหรับพนักงานบริษัท

หมายเลข 4,5,6 ช่อง " คำนำหน้า - ชื่อ - นามสกุล " สำหรับผู้สมัคร

หมายเลข 7 ช่อง " วัน-เดือน-พ.ศ. " สำหรับระบุวันเกิด

หมายเลข 8 ช่อง " อายุ " แสดงอายุจากการคำนวณของช่องหมายเลข 7

หมายเลข 9 ช่อง " การศึกษา " สำหรับระบุประวัติการศึกษา

หมายเลข 10 ช่อง " จังหวัด " สำหรับระบุจังหวัดที่เกิด

หมายเลข 11 ช่อง " น้ำหนักตัว " สำหรับระบุน้ำหนักตัว

หมายเลข 12 ช่อง " ส่วนสูง " สำหรับระบุส่วนสูง

หมายเลข 13 ช่อง " ประสบการณ์ขับรถ " สำหรับระบุประสบการณ์ขับรถ

หมายเลข 14 ปุ่ม " ประมวลผล " สำหรับประมวลผลการทำนาย

หมายเลข 15 ปุ่ม " บันทึก " สำหรับบันทึกผลการประมวลผล

หมายเลข 16 ปุ่ม " Clear " สำหรับระบุข้อมูลช่องต่างๆให้ว่างเปล่า

หมายเลข 17 ปุ่ม " ออก " สำหรับออกจากระบบงาน

หมายเลข 18 พื้นที่ แสดงการประมวลผลการคาดการณ์ข้อมูลส่วนบุคคล

3.4.4 หน้าจอการทำนายความเสี่ยงโดยกำหนดองค์ประกอบการเดินทาง

การทำนายความเสี่ยงโดยกำหนดองค์ประกอบการเดินทาง

ตารางข้อมูลพนักงานขับรถ

(1)

(2) (3)

ระบุการเดินทาง

(4) (5) (6) (7)

(8)

ภาพที่ 4.34 หน้าจอการทำนายความเสี่ยงโดยกำหนดองค์ประกอบการเดินทาง

ภาพที่ 4.34 อธิบายส่วนประกอบต่างๆของหน้าหัวข้อการทำนายความเสี่ยงโดยกำหนดองค์ประกอบการเดินทาง ได้แก่

หมายเลข 1 ตาราง " ข้อมูลส่วนบุคคล " เลือกข้อมูลผู้สมัคร, พนักงานขับรถบริษัทฯ

หมายเลข 2 ช่อง " เดือน พ.ศ." ระบุช่วง เดือน พ.ศ. ของข้อมูล

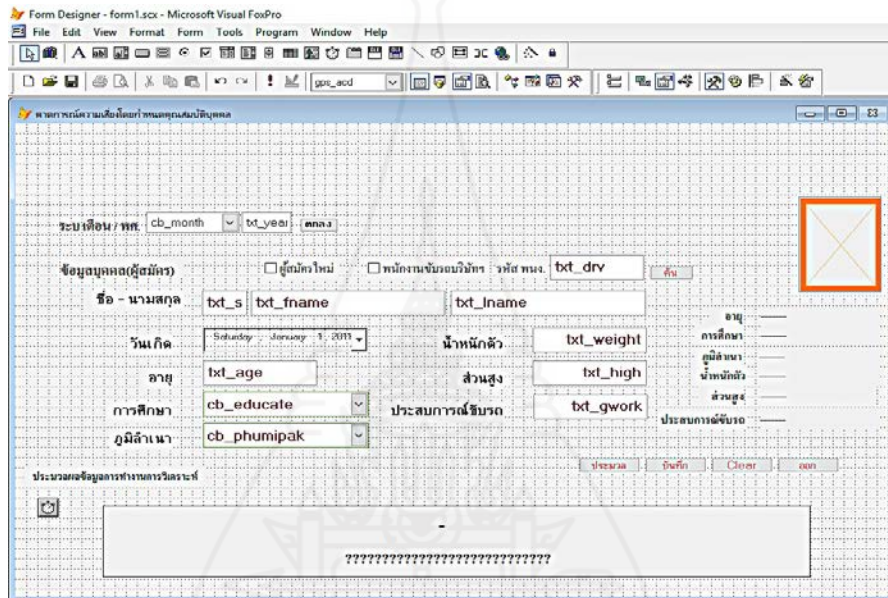
หมายเลข 3 ปุ่ม " ตกลง " แสดงข้อมูลผู้สมัคร, พนักงานขับรถบริษัทฯ

หมายเลข 4 ช่อง " องค์ประกอบการเดินทาง" มีส่วนประกอบของ ประเภทเส้นทางเดินทาง และ ประเภทรถ

- หมายเลข 5 ปุ่ม "ประมวลผล" แสดงประมวลผลการระบุข้อมูล
- หมายเลข 6 ปุ่ม "บันทึก" บันทึกข้อมูลการประมวลผลของผู้สมัคร
- หมายเลข 7 ปุ่ม "ออก" ออกจากระบบงาน
- หมายเลข 8 พื้นที่แสดงผลการประมวลผลการคาดการณ์

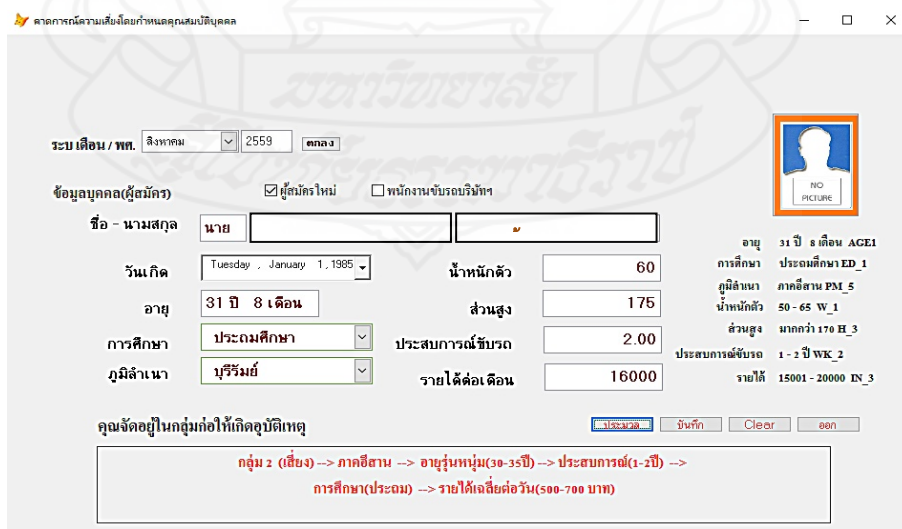
3.5 การออกแบบหน้าจอการทำงาน

3.5.1 ออกแบบหน้าจอการทำงานนายความเสี่ยงโดยกำหนดคุณสมบัติส่วนบุคคล



ภาพที่ 4.35 ออกแบบหน้าจอการทำงานนายความเสี่ยงโดยกำหนดคุณสมบัติส่วนบุคคล

3.5.2 แสดงผลการทำนายความเสี่ยงโดยกำหนดคุณสมบัติส่วนบุคคล



ภาพที่ 4.36 แสดงผลการทำนายความเสี่ยงโดยกำหนดคุณสมบัติส่วนบุคคล

การทำงาน

- 1) ระบุ เดือน - พ.ศ. สำหรับกำหนดช่วงการตรวจสอบคุณสมบัติบุคคล
- 2) คลิก ผู้สมัครใหม่ หรือ พนักงานขับรถบริษัท
 - เลือกคลิก ผู้สมัครใหม่ สิ่งที่ต้องระบุ ดังนี้ คำนำหน้า ชื่อ นามสกุล วันเกิด การศึกษา ภูมิลำเนา น้ำหนักตัว ส่วนสูง ประสบการณ์ทำงาน และรายได้ต่อเดือน
 - เลือกคลิก พนักงานขับรถบริษัท ให้ระบุรหัสพนักงานเท่านั้น ระบบจะดึงข้อมูลมาจากฐานข้อมูลบุคคล และระบุรายได้ล่าสุด
- 3) คลิกปุ่ม ประมวลผล สำหรับประมวลผลความเสี่ยงการเกิดอุบัติเหตุ
- 4) คลิกปุ่ม บันทึก สำหรับการบันทึกประมวลผล
- 5) คลิกปุ่ม Clear สำหรับล้างผลการประมวลผลการทำนายหรือคาดการณ์
- 6) คลิกปุ่ม ออก สำหรับออกจากโปรแกรม

3.5.3 รายงานผลการทำนายความเสี่ยงโดยกำหนดคุณสมบัติส่วนบุคคล

ลำดับ	สถานะ	ชื่อผู้สมัคร / พวงมาลัย	อายุ	กลุ่ม	ผลการทำนายความเสี่ยง
1.		นาย จาห์	37 ปี 10 เดือน	กลุ่ม 9	(ดีเยี่ยม) -> ประสบการณ์น้อยมาก(ในนี้ 1 ปี) -> ส่วนสูงเหมาะสม(161-170 ซม)
2.		นาย ชีระศักดิ์	54 ปี 4 เดือน	กลุ่ม 8	ผ่าน
3.	ผู้สมัคร	นาย สุรชาติ	51 ปี 0 เดือน	กลุ่ม 8	(ดีเยี่ยม) -> ภาวะสุขภาพดี -> อายุช่วงปกติ(30-35ปี) -> ประสบการณ์(มากกว่า 1ปี) -> การศึกษา(ประถม)
4.	ผู้สมัคร	นาย บุญมี	31 ปี 0 เดือน	กลุ่ม 2	(ดีเยี่ยม) -> ภาวะสุขภาพดี -> อายุช่วงหนุ่ม(30-35ปี) -> ประสบการณ์(1-2ปี) -> การศึกษา(ประถม) -> รายได้ต่อเดือน
5.	ผู้สมัคร	นาย สมการ	41 ปี 5 เดือน	กลุ่ม 4	(ดีเยี่ยม) -> ภาวะสุขภาพดี -> อายุช่วงหนุ่มใหญ่(41-45) -> ประสบการณ์(น้อยกว่า 1ปี) -> การศึกษา(ประถม) -> ร
6.	ผู้สมัคร	นาย สุรกัน	36 ปี 7 เดือน	กลุ่ม 8	ผ่าน
7.	ผู้สมัคร	นาย ครรช	40 ปี 6 เดือน	กลุ่ม 8	ผ่าน
8.	ผู้สมัคร	นาย สมชัย	50 ปี 0 เดือน	กลุ่ม 9	(ดีเยี่ยม) -> ส่วนสูงเหมาะสม(161-170 ซม) -> น้ำหนักตัวน้อย(50-65 กก)
9.	ผู้สมัคร	นาย ชนเชษฐ์	48 ปี 8 เดือน	กลุ่ม 3	(ดีเยี่ยม) -> ภาวะสุขภาพดี -> อายุช่วงหนุ่มใหญ่(46-50) -> ประสบการณ์(มากกว่า 1ปี) -> การศึกษา(ประถม) -> รายได้

ภาพที่ 4.37 รายงานผลการทำนายความเสี่ยงโดยกำหนดคุณสมบัติส่วนบุคคล

การทำงาน

- 1) ระบุ เดือน - พ.ศ. สำหรับระบุเดือนและปี พ.ศ.
- 2) คลิกปุ่ม ตกลง สำหรับแสดงรายงานผลการทำนายคาดการณ์ข้อมูลผู้สมัครขับรถ
- 3) คลิกปุ่ม พิมพ์ สำหรับพิมพ์รายงานผลรูปแบบกระดาษ
- 4) คลิกปุ่ม ออก สำหรับออกจากโปรแกรม

3.5.4 ออกแบบหน้าจอการทำนายความเสี่ยงโดยกำหนดองค์ประกอบการเดินรถ

จำนวน	สถานะ	ชื่อผู้สมัคร - พวง ใบรถ	อายุ	กลุ่มที่	ผลการทำนายบุคคล	ประเมิน
๑	๑	๑	๑	๑	๑	๑

ระบุ เดือน / พ.ศ. ปี

ประมวลผลข้อมูลการทำงานการวิเคราะห์

บุคคล

- กำหนดเส้นทาง -> 8 (เส้นทางตะวันออก) 100%
- ชั่วโมงการขับรถ -> มากกว่า 8 ชม.ต่อวัน 85.7%
- ประเภทรถ -> Golden dagon Benz Van
- สถานที่ -> ภายนอก สถานี บริษัท

ประมวลผล พิมพ์ ออก

ภาพที่ 4.38 ออกแบบหน้าจอการทำนายความเสี่ยงโดยกำหนดองค์ประกอบการเดินรถ

3.5.5 การแสดงผลการทำนายความเสี่ยงโดยกำหนดองค์ประกอบการเดินรถ

ภาพที่ 4.39 แสดงผลการทำนายความเสี่ยงโดยกำหนดองค์ประกอบการเดินรถ

ลำดับ	สถานะ	ชื่อผู้สมัคร - พนง. ขับรถ	อายุ	กลุ่มที่	ผลการทำนายผล	ประเภ
1.		นาย ชวติ	37 ปี 10 เดือน	กลุ่ม 9	(ดีเร็ว) -> ประสิทธิภาพน้อยมาก(ในเชิง 1 ปี) -> ส่วนสูงพอประมาณ(161-170 ซม)	
2.		นาย จิระศักดิ์	54 ปี 4 เดือน	กลุ่ม 0	ผ่าน	
3.		ผู้สมัคร นาย สุชาติ	51 ปี 8 เดือน	กลุ่ม 1	(เสีย) -> คุมทรงขับไม่คล -> อายุช่วงหนักหน่วง(54-55) -> ประสิทธิภาพ(มาก)	
4.		ผู้สมัคร นาย บุญชัย	31 ปี 8 เดือน	กลุ่ม 2	(เสีย) -> ควบคุมรถ -> อายุรุ่นหนุ่ม(30-35ปี) -> ประสิทธิภาพ(1-2ปี) -> การศึกษา	
5.		ผู้สมัคร นาย สมพร	41 ปี 5 เดือน	กลุ่ม 4	(เสีย) -> ควบคุมรถ -> อายุช่วงหนักหน่วง(41-45) -> ประสิทธิภาพ(น้อยกว่าปี)	
6.		ผู้สมัคร นาย สุรศักดิ์	36 ปี 7 เดือน	กลุ่ม 0	ผ่าน	
7.		ผู้สมัคร นาย สรรณ	46 ปี 6 เดือน	กลุ่ม 0	ผ่าน	
8.		ผู้สมัคร นาย สมชัย	56 ปี 8 เดือน	กลุ่ม 9	(ดีเร็ว) -> ส่วนสูงพอประมาณ(161-170 ซม) -> น้ำหนักตัวน้อย(50-65 กก)	
9.		ผู้สมัคร นาย ชนะศักดิ์	48 ปี 8 เดือน	กลุ่ม 3	(เสีย) -> ควบคุมรถ -> อายุช่วงหนักหน่วง(46-50) -> ประสิทธิภาพ(มากกว่าปี) ->	

ระบบ เดือน / พ.ศ. มิถุนายน 2559

ประมวลผลข้อมูลการทำงานการวิเคราะห์

(เสีย) -> ควบคุมรถ -> อายุช่วงหนักหน่วง(41-45) -> ประสิทธิภาพ(น้อยกว่าปี) -> การศึกษา(ประถม) -> รายได้เฉลี่ยต่อวัน(มากกว่า)

กำหนดเส้นทาง -> ---

ชั่วโมงการขับรถ -> ทำงานมากกว่า 8 ชั่วโมง ต่อวัน 63.6%

ประเภทรถ -> BENZ

สถานที่ -> สถานี 68.7% อื่นๆ 63.6%

ภาพที่ 4.39 แสดงผลการทำนายความเสี่ยงโดยกำหนดองค์ประกอบการเดินรถ

การทำงาน

- 1) ระบุ เดือน - พ.ศ. สำหรับกำหนดช่วงการตรวจสอบคุณสมบัติบุคคล
- 2) คลิกปุ่ม ตกลง สำหรับการแสดงข้อมูลผู้สมัครขับรถ
- 3) คลิก ตารางแสดงผลรายงานข้อมูล สำหรับเลือกข้อมูลผู้สมัครขับรถหรือพนักงานขับรถบริษัท(ทดสอบ)
- 4) คลิกปุ่ม ประมวลผล สำหรับประมวลผลความเสี่ยงการเกิดอุบัติเหตุ
- 5) คลิกปุ่ม บันทึก สำหรับการบันทึกประมวลผล
- 6) คลิกปุ่ม ออก สำหรับออกจากโปรแกรม

4. ประเมินผลโปรแกรมใช้งาน

กลุ่มผู้ใช้งาน ผู้บริหาร หมายถึง บุคลากรในองค์กร ได้แก่ ผู้จัดการฝ่ายบุคคล ผู้จัดการฝ่ายเดินรถ ผู้จัดการฝ่ายอุบัติเหตุ รวมถึงผู้บริหาร ที่ต้องการข้อมูลการบริการการจัดการเดินรถ ซึ่งสามารถใช้โปรแกรมของผู้วิจัยเพื่อทำการทำนายเกี่ยวกับบุคคล เช่น ผู้สมัครใหม่ พนักงานขับรถ เพื่อโยกย้ายตำแหน่งหรือเส้นทาง เพื่อให้ได้ข้อมูลมาประกอบ การตัดสินใจ

เจ้าหน้าที่ฝ่ายบุคคล หมายถึง เจ้าหน้าที่ทำการนำข้อมูลของผู้สมัครใหม่(ขับรถ) ป้อนเข้าสู่โปรแกรม หรือการนำพนักงานขับรถบริษัทมาทำการทำนาย ซึ่งข้อมูลส่วนบุคคลมีอยู่ในฐานข้อมูล สามารถที่จะระบุรหัสพนักงานของคนขับเพื่อทำการทำนายได้เช่นกัน และผู้ดูแลระบบ หมายถึง ผู้กำหนดสิทธิผู้ใช้งาน

4.1 แบบประเมินความพึงพอใจของผู้ใช้ระบบ

การสัมภาษณ์ใช้เครื่องมือ 2 ชนิด ได้แก่ แบบสอบถามเพื่อประเมินความพึงพอใจการใช้งานของระบบ และแบบสอบถามความคิดเห็นและข้อเสนอแนะเพิ่มเติมหลังการใช้งานระบบ

ตอนที่ 1 แบบสอบถามเพื่อประเมินความพึงพอใจในการใช้งานระบบ

การประเมินระบบใช้แบบสอบถามเป็นตัววัดผล ซึ่งใช้กลุ่มตัวอย่าง 2 กลุ่ม คือ กลุ่มผู้บริหาร และกลุ่มเจ้าหน้าที่

ดีมาก(ระดับ 5) หมายถึง ผู้ตอบแบบสอบถามมีความพึงพอใจต่อระบบมากที่สุด

ดี (ระดับ 4) หมายถึง ผู้ตอบแบบสอบถามมีความพึงพอใจต่อระบบมาก

ปานกลาง (ระดับ 3) หมายถึง ผู้ตอบแบบสอบถามมีความพึงพอใจต่อระบบปานกลาง

พอใช้ (ระดับ 2) หมายถึง ผู้ตอบแบบสอบถามมีความพึงพอใจต่อระบบน้อย

ปรับปรุง (ระดับ 1) หมายถึง ผู้ตอบแบบสอบถามมีความพึงพอใจต่อระบบน้อยที่สุด

จากข้อมูลที่ได้จากการตอบแบบสอบถามของกลุ่มผู้ใช้งานระบบ โดยใช้กับจำนวนเจ้าหน้าที่ หรือผู้ที่เกี่ยวข้อง ได้แก่

ฝ่ายบุคคล ผู้จัดการบุคคล และหัวหน้าแผนกบุคคล

ฝ่าย GPS เจ้าหน้าที่

ฝ่ายการเงินรถ ผู้จัดการฝ่ายการเงินรถ

ฝ่ายอุบัติเหตุ หัวหน้าแผนกอุบัติเหตุ และเจ้าหน้าที่

จากข้อมูลที่ได้จากการตอบแบบสอบถามของกลุ่มผู้ใช้งานระบบ มีจำนวนผู้ใช้ทั้งสิ้น

7 คน ซึ่งจะระบุจำนวนและร้อยละของผู้ที่ตอบในแต่ละคำถาม ตามตารางที่ 4.35

ตารางที่ 4.35 แสดงผลการประเมินความพึงพอใจการใช้งานระบบ

ลักษณะการใช้งานโปรแกรม ในด้านต่างๆ	ปาน					N	\bar{x}	SD
	ดีมาก	ดี	กลาง	พอใช้	ปรับปรุง			
	5	4	3	2	1			
1. ความสะดวกและง่ายในการใช้ระบบ	5	2				7	4.71	0.49
2. ความสมบูรณ์ข้อมูลพนักงานขับรถ	6	1				7	4.85	0.38
3. ความรวดเร็วในประมวลผลของ ระบบงาน	7					7	5	0.00
4. ความชัดเจนในการแบ่งกลุ่มบุคคล	5	2				7	4.71	0.49
5. ความแม่นยำในการทำนายผลข้อมูล การเกิดอุบัติเหตุการขับรถ	6	1				7	4.85	0.38
6. ข้อมูลข่าวสารตรงความต้องการ	5	1	1			7	4.57	0.79
7. ข้อมูลที่ได้ช่วยในการตัดสินใจ	6	1				7	4.85	0.38
8. ความเหมาะสมการโต้ตอบกับผู้ใช้	6	1				7	4.85	0.38
9. คำที่ผู้ใช้สามารถปฏิบัติตามได้ง่าย	7					7	5	0.00
รวม	53	9	1	-	-	63	4.84	0.18

ตารางที่ 4.36 เกณฑ์การกำหนดระดับความพึงพอใจ

ค่าเฉลี่ยระดับความพึงพอใจ	ระดับการแปลผล
5.0 - 4.5	ดีมาก
4.4 - 3.5	ดี
3.4 - 2.5	ปานกลาง
2.4 - 1.5	พอใช้
1.4 - 0	ปรับปรุง

ผลการวิเคราะห์ ข้อมูลจากแบบสอบถามทั้งหมด โดยทำการอ่านค่า และจัดทำผลการวิเคราะห์ได้ ตามตารางที่ 4.38

ตารางที่ 4.37 ผลการวิเคราะห์ข้อมูลจากการตอบแบบสอบถามเพื่อประเมินความพึงพอใจ

หัวข้อการประเมิน	ค่าเฉลี่ย	ระดับความพึงพอใจ
1. ความสะดวกและง่ายในการใช้ระบบ	4.7	ดีมาก
2. ความสมบูรณ์ข้อมูลพนักงานขับรถ	4.8	ดีมาก
3. ความรวดเร็วในประมวลผลของระบบงาน	5	ดีมาก
4. ความชัดเจนในการแบ่งกลุ่มบุคคล	4.7	ดีมาก
5. ความแม่นยำในการทำนายผลข้อมูลการเกิดอุบัติเหตุการขับรถ	4.8	ดีมาก
6. ข้อมูลข่าวสารตรงความต้องการ	4.5	ดีมาก
7. ข้อมูลที่ได้ช่วยในการตัดสินใจ	4.8	ดีมาก
8. ความเหมาะสมการโต้ตอบกับผู้ใช้	4.8	ดีมาก
9. คำศัพท์ที่ผู้ใช้สามารถปฏิบัติตามได้ง่าย	5	ดีมาก
ความพึงพอใจ โดยภาพรวม	4.8	ดีมาก

ตอนที่ 2 แบบสอบถามความคิดเห็นและข้อเสนอแนะเพิ่มเติมหลังการใช้งาน

หลังจากได้ทดสอบระบบด้วยการติดตั้งระบบเพื่อใช้งานจริงและการสัมภาษณ์ด้วยแบบสอบถามความคิดเห็นและข้อเสนอแนะสามารถสรุปผลได้ดังนี้

ข้อที่ 1. ท่านคิดว่าระบบสารสนเทศนี้ ช่วยท่านในเรื่องใดบ้าง

- ก. ลดขั้นตอนการทำงาน
- ข. การคัดเลือกพนักงานขับรถ
- ค. เป็นเครื่องมือช่วยในการตัดสินใจ
- ง. เพิ่มประสิทธิภาพในการทำงาน

ข้อที่ 2. ปัญหาและอุปสรรคที่พบจากการใช้งานในระบบนี้ คืออะไรบ้าง

- ก. มีตัวอักษรหน้าจอบางส่วนไม่สมบูรณ์
- ข. ข้อความยังมีความหมายไม่ตรงประเด็น
- ค. ระบบสามารถใช้งานภายในองค์กร
- ง. ต้องการนำข้อมูลประวัติพนักงานขับรถจากสถานีตำรวจมาใช้ร่วมกัน
- จ. เพิ่มความสามารถในการทำนาย พยากรณ์ในเส้นทางเดินรถอื่นๆ

ข้อที่ 3. ข้อเสนอแนะเพิ่มเติมจากการใช้งานระบบนี้

- ก. นำตารางการฝึกอบรมพนักงานขับรถกับเงื่อนไขการทำนาย
- ข. เพิ่มหัวข้อลักษณะบุคคลมากขึ้น เช่น ศาสนา จำนวนบุตร แต่งงาน เป็นต้น

4.2 แบบประเมินความเชื่อมั่นคำทำนายระบบ

คำชี้แจง

1. แบบประเมินชุดนี้สำหรับผู้ใช้งาน เพื่อประเมินความเชื่อมั่นในระบบการทำนายการเกิดอุบัติเหตุรถยนต์โดยสารสาธารณะ

2. การแสดงความคิดเห็นของผู้ใช้งานที่มีต่อระบบการทำนายการเกิดอุบัติเหตุรถยนต์โดยสารสาธารณะที่ได้พัฒนาขึ้นจะประกอบด้วยส่วนของรายการประเมินที่อยู่ด้านซ้ายมือ และมาตราส่วนประมาณค่าอยู่ด้านขวามือ จำนวน 5 ช่อง โปรดทำเครื่องหมาย ลงในช่องทางด้านขวามือที่ตรงกับความคิดเห็นของท่าน โดยกำหนดค่าความหมายดังนี้

- 5 หมายถึง ระดับความเชื่อมั่นดีมาก
- 4 หมายถึง ระดับความเชื่อมั่นดี
- 3 หมายถึง ระดับความเชื่อมั่นปานกลาง
- 2 หมายถึง ระดับความเชื่อมั่นพอใช้
- 1 หมายถึง ระดับความเชื่อมั่นปรับปรุง

รายละเอียดแบบสอบถาม

แบบสอบถามแบ่งเป็น 2 ตอน ดังนี้

ตอนที่ 1 ความเชื่อมั่นของระบบการทำนาย

ตารางที่ 4.38 แสดงผลการประเมินความเชื่อมั่นของระบบการทำนาย

หัวข้อการประเมิน	ระดับการประเมิน					N	\bar{x}	SD
	ดีมาก	ดี	ปานกลาง	พอใช้	ปรับปรุง			
	5	4	3	2	1			
1. ความสะดวกและง่ายต่อการใช้งาน	4	2	1			7	4.42	0.77
2. ความน่าเชื่อถือของผลการทำนาย	4	2	1			7	4.28	0.77
3. ผลการทำนายตรงกับความต้องการ	5	1	1			7	4.57	0.73
4. ช่วยประกอบการตัดสินใจ	4	3				7	4.57	0.50
รวม	17	8	3			28	4.35	1.33

ตารางที่ 4.39 ผลการวิเคราะห์ข้อมูลจากการประเมินความเชื่อมั่นของระบบการทำนาย

หัวข้อการประเมิน	ค่าเฉลี่ย	ระดับความพึงพอใจ
1. ความสะดวกและง่ายในการใช้ระบบ	4.42	ดี
2. ความน่าเชื่อถือของผลการทำนาย	4.28	ดี
3. ผลการทำนายตรงกับความต้องการ	4.57	ดีมาก
4. ช่วยประกอบการตัดสินใจ	4.57	ดีมาก
ความพึงพอใจโดยภาพรวม	4.35	ดี

ตอนที่ 2 แบบสอบถามความคิดเห็นและข้อเสนอแนะเพิ่มเติมหลังการใช้งานแสดงความคิดเห็นของผู้ใช้งานที่มีต่อระบบการทำนายการเกิดอุบัติเหตุ โดยสารสาธารณะหลังจากได้ทดสอบระบบเพื่อใช้งานจริงและการสัมภาษณ์ด้วยแบบสอบถามความคิดเห็นและข้อเสนอแนะสามารถสรุปผลได้ดังนี้

ข้อที่ 1. ท่านคิดว่าของผลการทำนายช่วยท่านในเรื่องใดบ้าง

ก. การตัดสินใจ

ข. เกิดความรอบคอบขึ้น

ค. การวางแผนด้านบุคลากร

ข้อที่ 2. ปัญหาที่พบจากการทำนาย คืออะไรบ้าง

ก. ความหมายต้องการแก้ไขบางคำ

ข. บุคลากรส่วนใหญ่การศึกษาน้อย

ค. เพิ่มความสามารถในการทำนายเส้นทางเดินรถอื่นๆ

สรุปผลการประเมิน

การวิเคราะห์ข้อมูลจากแบบสอบถามเพื่อประเมินความพึงพอใจต่อการใช้งานระบบของผู้ใช้งาน พบว่าค่าเฉลี่ยความพึงพอใจโดยภาพรวมต่อการใช้งานระบบอยู่ในระดับดีมาก โดยผู้มีความพึงพอใจในหัวข้อความสะดวกและง่ายในการใช้งาน ความถูกต้องของข้อมูลในการทำนายผลมีความใกล้เคียงความเป็นจริง ความรวดเร็วในการแสดงผลเร็ว และสามารถตอบคำถามส่วนต่างๆของผู้เกี่ยวข้องได้ และความพึงพอใจต่อการตอบสนองการใช้งานได้ดีเป็นตัวช่วยหรือเครื่องมือในการตัดสินใจอีกทางหนึ่ง โดยอ้างอิงจากข้อมูลในอดีต



บทที่ 5

สรุปการวิจัย และข้อเสนอแนะ

การศึกษาวิจัยเรื่อง การประยุกต์ระบบ GPS และการทำเหมืองข้อมูลเพื่อทำนาย และป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด ผู้วิจัยได้พัฒนากล้องข้อมูลระบบ GPS โดยใช้หลักการ Big Data รองรับข้อมูลขนาดใหญ่ เพิ่มประสิทธิภาพการสืบค้น การวิเคราะห์ข้อมูลที่รวดเร็ว และการทำเหมืองข้อมูลการเกิดอุบัติเหตุรถโดยสารสาธารณะเพื่อสร้างแบบจำลองทำนาย ได้ข้อสรุปผลการวิจัยการอภิปรายผล และข้อเสนอแนะดังนี้

1 สรุปผลการวิจัย

1.1 สร้างระบบจัดการข้อมูล GPS

ผู้วิจัยนำข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS มาสร้างคลังข้อมูล โดยใช้โปรแกรม MONGODB จัดการฐานข้อมูลให้สามารถรองรับกับข้อมูลขนาดใหญ่ และผ่านกระบวนการอีทีแอล (EXTRACT TRANSFORM AND LOAD : ETL) ซึ่งผู้วิจัยได้ออกแบบส่วนของการนำเข้าข้อมูลอยู่ในรูปแบบของเท็กซ์ไฟล์ (CSV) ทำการคัดเลือก คัดแยกข้อมูลที่ไม่เกี่ยวข้องให้ได้รูปแบบข้อมูลถูกต้อง แล้วใช้คำสั่ง COMMAND LINE ของโปรแกรม MongoDB เป็นเครื่องมือนำเข้าคลังข้อมูล GPS

1.2 สร้างแบบจำลองการทำนาย

ข้อมูลที่ใช้ในการทำเหมืองข้อมูลได้มาจากคลังข้อมูลระบบ GPS และคลังข้อมูลการเกิดอุบัติเหตุรถโดยสารสาธารณะที่สร้างขึ้นมา ซึ่งใช้เทคนิคการทำเหมืองข้อมูลมาสร้างระบบการวิเคราะห์ใช้อัตราความเร็วของรถโดยสารสาธารณะ และทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ ส่วนขั้นตอนการทำเหมืองข้อมูลใช้รูปแบบจำลองของ CRISP - DM Model โดยใช้ข้อมูลติดตามรถโดยสารสาธารณะ ข้อมูลสถานะของรถโดยสารสาธารณะ ข้อมูลประวัติรถโดยสาร ข้อมูลสถานะของรถโดยสาร ข้อมูลการจัดกลุ่มรถโดยสาร ข้อมูลการประเภทรถโดยสาร และข้อมูลตำแหน่งที่ตั้งต่างๆกำหนดความสัมพันธ์ระหว่างข้อมูลเพื่อนำเข้าคลังข้อมูลระบบ GPS และข้อมูลการเกิดอุบัติเหตุรถโดยสารสาธารณะระหว่างพ.ศ. 2556-2559 จำนวน 406 ระเบียบ ประกอบด้วย

ข้อมูลการเกิดอุบัติเหตุโดยสาธาธารณะ ข้อมูลความเสียหายของรถยนต์คู่กรณี ข้อมูลรายละเอียดรถคู่กรณี ข้อมูลประวัติพนักงาน ข้อมูลความเสียหายของบริษัท ข้อมูลขั้นตอนทำงาน และรายจ่าย ประวัติรถโดยสาธาธารณะ ข้อมูลประเภทรถโดยสาธาธารณะ กำหนดความสัมพันธ์ระหว่างข้อมูลเพื่อนำเข้าคลังข้อมูลการเกิดอุบัติเหตุโดยสาธาธารณะ

การสร้างตัวแบบสำหรับการวิเคราะห์ใช้อัตราความเร็วของรถโดยสาธาธารณะ และตัวแบบจำลองทำนายการเกิดอุบัติเหตุโดยสาธาธารณะ อธิบายได้ดังนี้

1.2.1 แบบจำลองสำหรับการวิเคราะห์ใช้อัตราความเร็วของรถโดยสาธาธารณะ

สามารถออกแบบจำลองการวิเคราะห์อัตราความเร็วรถโดยสาธาธารณะได้ 2 แบบ ได้แก่ การใช้ความเร็วรถของกลุ่มรถยนต์ และการใช้ความเร็วรถของตำแหน่งที่ตั้ง ซึ่งใช้เทคนิคเรียนรู้ต้นไม้ตัดสินใจ Decision Tree Learner ด้วยการกำหนดเป้าหมาย (TARGET) ของฟิลด์หมายเลขรถ (TRUCK_NAME) และระบุเงื่อนไขที่เพื่อการวิเคราะห์ใน INCLUDE BLOCK ของ Decision Tree Learner Node เช่น ฟิลด์ความเร็ว (R_SPEED) เพื่อทำนายแบบจำลองการวิเคราะห์ความเร็วรถของกลุ่มรถยนต์ หรือฟิลด์ความเร็ว (R_SPEED) กับฟิลด์ตำแหน่งที่ตั้ง (TAM_TNAME) เพื่อทำนายแบบจำลองวิเคราะห์การใช้ความเร็วรถของตำแหน่งที่ตั้ง และกำหนดคุณลักษณะการวัดค่าด้วย Quality Measure (QM) ในรูป Gain / Gain Ratio ผลการวิจัยพบว่าสามารถจำแนกกลุ่มรถโดยสาธาธารณะ และกลุ่มตำแหน่งที่ตั้งในรูปแบบภาพต้นไม้ตัดสินใจ ซึ่งสามารถอธิบายผลลัพธ์ที่ได้ เช่น การเปรียบเทียบร้อยละ กับจำนวนครั้งทั้งหมด เช่น ข้อมูลหมายเลขรถ 42 (158) ใช้ความเร็วมากกว่าหรือเท่ากับ 90 กม.ต่อชั่วโมง มีจำนวน 999 ครั้ง คิดเป็นร้อยละ 10.7 ของจำนวนครั้งรวม (TOTAL) 9379 ครั้ง และอีกหนึ่งตัวอย่าง เช่น หมายเลข 36 (36) ใช้อัตราความเร็วรถผ่านสถานที่บางไฟจำนวน 5 ครั้งคิดเป็นร้อยละ 50 ของจำนวนทั้งหมด 10 ครั้ง ที่ความเร็วมากกว่าหรือเท่ากับ 93 กม.ต่อชั่วโมง

1.2.2 แบบจำลองทำนายการเกิดอุบัติเหตุโดยสาธาธารณะ ใช้เทคนิคการจัด

กลุ่ม(Clustering) ภูควความสัมพันธ์ (Association) และ การจำแนก (Classification) โดยใช้ข้อมูลการเกิดอุบัติเหตุโดยสาธาธารณะ ซึ่งกระบวนการออกแบบจำลองเพื่อการทำนายต้องการจำแนกกลุ่มออกเป็นสองกลุ่ม คือ กลุ่มเสี่ยง (เป็นบุคคลที่ถูกจัดให้มีอัตราการก่อให้เกิดอุบัติเหตุการขับรถโดยสาธาธารณะมากที่สุด) และกลุ่มเฝ้าระวัง (เป็นบุคคลที่ถูกจัดว่าอาจจะก่อให้เกิดอุบัติเหตุการขับรถโดยสาธาธารณะ) แบบจำลองเพื่อการทำนายของกลุ่มเสี่ยง มี 2 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 การแบ่งกลุ่มเสี่ยง โดยใช้ข้อมูลในส่วนของคุณสมบัติบุคคล ได้แก่ อายุ การศึกษา ภูมิลำเนา น้ำหนักตัว ส่วนสูง ประสบการณ์ทำงาน รายได้ และแบ่งข้อมูลด้วยเทคนิค Split Test ได้แก่ Training Data และ Testing Data ทำการทดสอบแบบจำลองด้วยอัลกอริทึม Simple

K-Means เพื่อทำการแบ่งกลุ่มเสี่ยง ทำให้ได้ผลลัพธ์ของข้อมูลเหมือนกัน และได้กลุ่มเสี่ยงเป็นจำนวน 4 กลุ่ม ที่มีลักษณะเฉพาะกลุ่มเสี่ยงชัดเจน

ขั้นตอนที่ 2 นำกลุ่มเสี่ยงแต่ละกลุ่มมาทำการทดสอบร่วมกับข้อมูลองค์ประกอบของปัจจัยการเกิดอุบัติเหตุ ได้แก่ เส้นทางเดินรถ จำนวนชั่วโมงขับรถ ประเภทรถที่ใช้ และสถานที่เกิดเหตุ (บริษัท, สถานี, ภายนอก) ด้วยอัลกอริทึม Decision Tree Learner เพื่อหาองค์ประกอบปัจจัยการเกิดอุบัติเหตุในแต่ละกลุ่มเสี่ยงด้วย Decision Tree to Ruleset Node ให้ได้ผลลัพธ์การทำนายสามารถนำไปพัฒนาเป็นระบบการทำนายผลการเกิดอุบัติเหตุรถโดยสารสาธารณะ ดังตารางที่ 5.1

ตารางที่ 5.1 สรุปผลการทดสอบของกลุ่มเสี่ยง

กลุ่มเสี่ยง	ลักษณะเฉพาะกลุ่มบุคคล	การเกิดเหตุ	เกิดเหตุร้อยละ
กลุ่ม 1	ชาย -> กรุงเทพฯปริมณฑล ->	- ขับรถมากกว่า 8 ชั่วโมงต่อวัน	90.5
	อายุ 51-55 -> ประสบการณ์มากกว่า 7 ปี -	- ขับรถต่อวัน 7-8 ชั่วโมง รถตู้ V9905	50
	> การศึกษา(ประถม) -> รายได้เฉลี่ยต่อวัน 300-500 บาท	- ขับรถต่อวันน้อยกว่า 6 ชั่วโมง รถ BENZ เกิดเหตุที่สถานี	75
กลุ่ม 2	ชาย -> ภาคอีสาน -> อายุ 30-35 ปี ->	- ขับรถ 7-8 ชั่วโมงต่อวัน	74-75
	ประสบการณ์ 1-2 ปี -> การศึกษา (ประถม) -> รายได้ต่อวัน 500-700 บาท	- BENZ, GOLDEN DAGON, SUNLONG - สถานี, ภายนอก	73 - 83 74 - 75
กลุ่ม 3	ชาย -> ภาคกลาง -> อายุ 46--50 ปี ->	-เส้นทาง B389 เกิดเหตุภายในสถานี	100
	ประสบการณ์มากกว่า 7 ปี -> การศึกษา	- เส้นทาง B048 เกิดเหตุภายในสถานี	77.8
	ประถม -> รายได้ต่อวันมากกว่า 800 บาท	- ขับรถมากกว่า 8 ชั่วโมงต่อวันเส้นทาง B389	85
		- ขับรถมากกว่า 8 ชั่วโมงต่อวันเส้นทาง B048	58.5
กลุ่ม 4	ชาย -> ภาคตะวันออก -> อายุ 41-45 ->	- ขับรถน้อยกว่า 6 ชั่วโมงต่อวัน รถ BENZ	63.6
	ประสบการณ์น้อยกว่า 4 ปี -> การศึกษา (ประถม) -> รายได้เฉลี่ยต่อวันมากกว่า 800 บาท	- รถ BENZ เกิดเหตุภายในสถานี	68.7

1.2.3 แบบจำลองเพื่อการทำนายของกลุ่มเฟ้อระวัง เป็นการนำข้อมูลคุณสมบัติบุคคล ได้แก่ อายุ การศึกษา ภูมิลำเนา น้ำหนักตัว ส่วนสูง ประสบการณ์ทำงาน รายได้ มาทดสอบแบบจำลอง โดยใช้กฎความสัมพันธ์ (Association Rule) ด้วยวิธี Apriori เพื่อหาความสัมพันธ์ของคุณสมบัติบุคคล โดยกำหนดค่าความเชื่อมั่น Confidence ไม่ต่ำกว่าร้อยละ 60 และแสดงผลลัพธ์ของกฎ 50 จำนวน เพื่อให้ได้คุณลักษณะบุคคลที่แสดงค่าความเชื่อมั่น (Confidence) ของกฎความสัมพันธ์และค่าลิฟต์ (Lift) คือค่าสหสัมพันธ์ที่บ่งบอกถึงกฎความสัมพันธ์ที่มีค่ามากกว่า 1 เท่านั้น ทั้งนี้การกำหนดค่า Confidence ให้มีค่าระดับกลางๆ เพื่อให้ได้ผลลัพธ์ของข้อมูลจากกฎความสัมพันธ์อยู่ในระดับข้อมูลกลุ่มเฟ้อระวังและไม่มีข้อมูลถูกจัดไว้ในกลุ่มเสี่ยง ส่วนการกำหนดค่า Confidence ให้มีค่ามากหรือร้อยละ 80 ขึ้นไปผลลัพธ์ที่ได้มีจำนวนข้อมูลน้อย ซึ่งไม่ตรงกับความต้องการ จากผลการวิจัยพบว่า การหาคุณลักษณะบุคคลกลุ่มเฟ้อระวังด้วยกฎความสัมพันธ์ของอัลกอริทึม Apriori กำหนดค่าความเชื่อมั่น Confidence ไว้ที่ร้อยละ 60 และสามารถคัดเลือกค่าที่เหมาะสมมาใช้ในการพัฒนาโปรแกรมใช้งาน ดังตารางที่ 5.2

ตารางที่ 5.2 คุณลักษณะบุคคลกลุ่มเฟ้อระวัง

ลำดับ	ความสัมพันธ์นอกกลุ่มกับอุบัติเหตุ	Confidence	Lift
1	ภาคกลาง -> การศึกษาประถม	89	1.17
2	ประสบการณ์น้อย(1-2ปี) -> การศึกษาประถม	89	1.04
3	รายได้ต่อวัน(700-800บาท) ->การศึกษาประถม	78	1.03
4	อายุ 30-35 ปี -> น้ำหนักตัว 50 - 65 ก.ก.	72	1.15
5	ส่วนสูง 161-170 ซม. -> น้ำหนักตัว 50- 65 ก.ก.	72	1.15
6	ประสบการณ์ไม่ถึง 1 ปี -> ส่วนสูง161-170 ซม.	71	1.46
7	อายุ 45 - 50 ปี	66	1.05
8	การศึกษามัธยมต้น	66	1.04
9	รายได้เฉลี่ยต่อวัน 700 - 800บาท	66	1.04
10	ประสบการณ์น้อยมากไม่ถึง 1 ปี	66	1.04

1.3 สร้างโปรแกรมใช้งาน

การพัฒนาระบบทำนายการเกิดอุบัติเหตุของรถโดยสารสาธารณะ ผู้วิจัยใช้โปรแกรม Microsoft Visual FoxPro v9.0 ในการพัฒนาเพื่อความสามารถในสร้างและออกแบบการใช้งาน จากผลการประเมินความเชื่อมั่นของการทำนาย โดยผู้ใช้งานจำนวน 7 คน พบว่า ผลการประเมินความเชื่อมั่นโดยรวมมีค่าเฉลี่ยเท่ากับ 4.84 ถือว่าการทำนายการเกิดอุบัติเหตุของรถโดยสารสาธารณะ โดยรวมอยู่ในระดับดีมาก สามารถนำไปใช้ป็นเครื่องมือในการพิจารณาคัดเลือกพนักงานขับรถโดยสาร เพื่อลดความเสี่ยงต่อการเกิดอุบัติเหตุทางรถโดยสารสาธารณะ

2 การอภิปรายผล

2.1 สร้างระบบจัดการข้อมูล GPS

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาค้างข้อมูลระบบ GPS ซึ่งในการพัฒนาค้างข้อมูลใช้ซอฟต์แวร์มอองโกดีบี(MongoDB 2.6.7) มารองรับข้อมูลขนาดใหญ่ตามหลักการของ Big data จากแหล่งข้อมูลการติดตามรถโดยสารสาธารณะระบบ GPS เดิม ผู้วิจัยทำการรวบรวมข้อมูล ออกแบบคลังข้อมูล และสร้างกระบวนการ ETL ในการดึงข้อมูลเข้าสู่คลังข้อมูลที่พัฒนาขึ้นใหม่นั้น จำเป็นต้องใช้เครื่องมือติดต่อกับฐานข้อมูลเดิม และเปลี่ยนรูปแบบข้อมูลเป็นเท็กซ์ไฟล์ (CSV) ซึ่งในงานวิจัยใช้ด้วยโปรแกรม KNIME v2.12 เป็นตัวจัดการ ส่วนวิธีนำข้อมูลเข้าคลังข้อมูลที่พัฒนา จะใช้ฟังก์ชันของซอฟต์แวร์มอองโกดีบี (MongoDB v2.6.7) ผ่านคำสั่ง Command Line ผลที่ได้จากการพัฒนาค้างข้อมูล GPS มีประสิทธิภาพด้านการประมวลผลข้อมูล การสืบค้นข้อมูลที่รวดเร็ว และสามารถพัฒนาระบบสนับสนุนด้านการประมวลผลเชิงวิเคราะห์ การทำเหมืองข้อมูล และสร้างรายงาน โปรแกรม KNIME v2.12 เป็นเครื่องมือที่ช่วยออกแบบสำหรับกลุ่มผู้ใช้งานในระดับผู้บริหารที่ต้องใช้ข้อมูลเพื่อประกอบการตัดสินใจ ซึ่งสอดคล้องกับงานวิจัยของ ประกายมาศ ศรีสุขทักษิณ และคณะ(2557) ที่ทำการทดสอบประสิทธิภาพของฐานข้อมูลระหว่างฐานข้อมูลเชิงสัมพันธ์ MySQL กับฐานข้อมูลไม่สัมพันธ์ MongoDB ซึ่งผลของการวิจัยพบว่าปริมาณข้อมูลที่นำมาทดสอบมีขนาดใหญ่ ฐานข้อมูลไม่สัมพันธ์ ของ MongoDB สามารถประมวลผลข้อมูล เช่น การค้นหา การอ่าน การบันทึก และการลบข้อมูลได้เร็วกว่าหลายเท่ามาก เมื่อเทียบกับฐานข้อมูลเชิงสัมพันธ์ MySQL ซึ่งสอดคล้องกับงานวิจัย

2.2 สร้างแบบจำลองการทำนาย

งานวิจัยนี้ นำเสนอการใช้เทคนิคเหมืองข้อมูลในการสร้างแบบจำลอง Connolly&E.Begg (2002: 1232) กล่าวถึง "กระบวนการที่นำข้อมูลที่มีอยู่ในฐานข้อมูลขนาดใหญ่ มาทำการศึกษา วิเคราะห์ ทำความเข้าใจ นำผลลัพธ์ที่ได้จากการศึกษามาใช้ในการตัดสินใจทาง ธุรกิจ และการนำเทคนิคเหมืองข้อมูล เพื่อออกแบบจำลองจากข้อมูล" ที่แนะนำถึงเทคนิคและ ขั้นตอนการทำเหมืองข้อมูล ซึ่งสอดคล้องต่อการสร้างแบบจำลองเพื่อทำนายการเกิดอุบัติเหตุ โดยสารสาธารณะ ของกลุ่มเสี่ยง และกลุ่มเฝ้าระวัง ด้วยการค้นหาคุณลักษณะบุคคลและปัจจัยการ เกิดอุบัติเหตุโดยสารร่วมกับเทคนิคการทำเหมืองข้อมูล เพื่อใช้เป็นข้อมูลประกอบการตัดสินใจคัด กรองบุคคลเพื่อเลือกพนักงานขับรถ ซึ่งผลการวิจัยสอดคล้องกับงานวิจัยของ พล โกสุขวัฒน์ (2556: 81-82) สร้างแบบจำลองการแบ่งกลุ่มนักศึกษา เพื่อเห็นการจัดข้อมูลที่มีลักษณะประจำ เช่น อายุ เพศ ภูมิภาค ที่กระจกระบายให้อยู่เป็นกลุ่ม พบว่าการจัดกลุ่มนักศึกษาใช้เทคนิคเหมือง ข้อมูลแบบคลัสเตอร์ (Clustering) ด้วยวิธี Simple K-Means สามารถหากลุ่มที่มีความแตกต่าง และ จุดเด่นแต่ละกลุ่มได้อย่างชัดเจน และเสกสรรค์ วิสัยทัศน์ (2558: 78-89) การสร้างแบบพยากรณ์ ผลการเรียนรู้ระดับมัธยมศึกษา การจัดกลุ่มข้อมูลนักศึกษาโดยใช้คุณลักษณะของนักศึกษาที่เป็น ข้อมูลประวัติส่วนตัวมาจัดแบ่งกลุ่ม 2 แบบ คือ ไม่มีการคัดเลือกคุณลักษณะ และวิธีการคัดเลือก คุณลักษณะ ด้วยอัลกอริทึม K-Means ผลที่ได้พบว่าแบบไม่จัดกลุ่มนำมาคัดเลือกคุณลักษณะ ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ให้ค่าความถูกต้องร้อยละ 94.48 ซึ่ง นักเรียนสามารถนำผลการพยากรณ์ไปประกอบการตัดสินใจและวางแผนการเรียน และไพฑูริย์ จันทร (2550: 36-45) การพัฒนาระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียนของ นักศึกษาระดับปริญญาตรี ใช้เทคนิคต้นไม้ตัดสินใจด้วย อัลกอริทึม C5.0 ในการสร้างตัวแบบ พบว่าผลลัพธ์ผลลัพธ์แสดงภาพออกมาเป็นต้นไม้ตัดสินใจ และสามารถจำแนกกฎ (Rule set) มีค่า การทำนายถูกต้องเท่ากับร้อยละ 76.84 นำตัวแบบที่ได้ไปพัฒนาระบบเลือกแผนการเรียนระดับ มัธยมศึกษาตอนปลาย ใช้เป็นเครื่องมือช่วยตัดสินใจเลือกแผนการเรียนได้ และ พิจิตรา จอมศรี (2549: 39-41) การทำนายเนื้อหาของเว็บโดยใช้เหมืองข้อมูล ได้ใช้เทคนิคการค้นหากฎความสัมพันธ์ ด้วยวิธี Apriori พบว่าสามารถจำแนกกฎความสัมพันธ์เนื้อหาของเว็บ และแสดงผลในรูปการจำแนกกฎ ความสัมพันธ์ ระบุถึงค่าความเชื่อมั่นและค่าสนับสนุน และให้ค่าความถูกต้องของตัวแบบข้อมูลเท่ากับ ร้อยละ 66.67 ดังนั้นสามารถนำแบบจำลองที่ได้ไปใช้ในการทำนายเนื้อหาเว็บได้ ซึ่งสอดคล้องกับแนว ทางการวิจัยการหาคุณลักษณะบุคคลที่ไม่ได้จัดอยู่ในกลุ่ม

นอกจากงานวิจัยได้นำเทคนิคการทำเหมืองข้อมูลมาวิเคราะห์ข้อมูลการเกิดอุบัติเหตุ โดยสารของพนักงานขับรถ เพื่อผลลัพธ์การทำนายจากตัวแบบที่สร้างได้ให้ค่าความสัมพันธ์ของ

กลุ่มกับการเกิดอุบัติเหตุอยู่ในเกณฑ์สูง การประเมินผลจากผู้ใช้งานอยู่ในเกณฑ์ดีมาก การนำผลการ
ทำนายไปใช้ประกอบการตัดสินใจคัดกรองบุคคลเพื่อเลือกพนักงานขับรถ

3 ปัญหาและอุปสรรค

3.1 ข้อมูลไม่ได้ถูกออกแบบเพื่อใช้ในการทำเหมืองข้อมูล ความเหมาะสมหรือความ
สมบูรณ์จึงไม่เพียงพอ ต้องมีการแทนค่า หรือปรับเปลี่ยนค่าต่างๆของข้อมูลให้เหมาะสม ซึ่งอาจมี
ผลต่อการวิเคราะห์ข้อมูลผิดพลาดได้

3.2 การนำข้อมูลจากโปรแกรมระบบ GPS ใช้เวลานาน เนื่องจากข้อมูลมีขนาดใหญ่
การจัดเก็บข้อมูลติดตามรถยนต์ถูกแบ่งตารางเป็นรายคัน ความซับซ้อนของข้อมูลจึงเพิ่มมากขึ้น

3.3 ข้อมูลบางส่วนไม่สามารถเปิดเผยได้ ซึ่งจะกระทบต่อธุรกิจของผู้ประกอบการเดิน
รถ จึงต้องใช้ข้อความอื่นแทนในการสื่อถึงความหมายที่ต้องการอ้างอิง

3.4 ข้อมูลบางส่วนไม่เพียงพอต่อความต้องการ และขาดหายไปจากการระบุข้อความ
ของข้อมูลที่ไม่มีสมบูรณ์ หรือไม่ครบถ้วน เช่น ภูมิลำเนา อายุ ส่วนสูง รายได้ และน้ำหนัก เป็นต้น

3.5 ข้อมูลที่ใช้ในการทำตัวแบบทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ การ
อ้างอิงข้อมูลมาจากผู้ประกอบการเดินรถแหล่งเดียว และข้อมูลมีจำนวน 406 ระเบียบ

3.6 ความร่วมมือจากหลายหน่วยงานในการใช้ข้อมูล มีข้อจำกัด ภาวะเบียบ และไม่
สามารถเผยแพร่ได้ซึ่งอาจจะทำให้แบบจำลองมีผลคลาดเคลื่อนกับความเป็นจริงได้

3.7 โปรแกรม KNIME เป็นเครื่องมือที่ใช้ในการวิเคราะห์ และประมวลผลข้อมูล มี
ความสามารถในการรองรับข้อมูลที่หลากหลายรูปแบบ ฟังก์ชันที่ถูกออกแบบช่วยในการวิเคราะห์
ข้อมูลมีหลายชนิด และความหลากหลายของเครื่องอื่นๆ ผู้วิจัยมีเวลาจำกัดในการเรียนรู้ ซึ่งอาจทำ
ให้เกิดข้อผิดพลาดได้

4 ข้อเสนอแนะ

ในการวิจัยนี้ได้พัฒนาค้างข้อมูลระบบ GPS โดยใช้โปรแกรม MongoDB v2.6.7 เพื่อรองรับข้อมูลติดตามรถโดยสารสาธารณะระบบ GPS ผู้สนใจสามารถนำหลักการพัฒนาค้างข้อมูลนี้ไปพัฒนาโดยใช้โปรแกรมอื่น เช่น Hadoop, CouchDB, HBase และ Cassandra เป็นต้นเพื่อเปรียบเทียบประสิทธิภาพและการประมวลผลเชิงวิเคราะห์จากข้อมูล Big Data

และการสร้างแบบจำลองทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ ความสมบูรณ์ของแบบจำลองขึ้นอยู่กับข้อมูลที่นำมาใช้สร้างแบบจำลอง ควรมีข้อมูลเพิ่มเติมและเพิ่มข้อมูลการเกิดอุบัติเหตุโดยสารของผู้ประกอบการเดินรถเส้นทางอื่นๆ จะส่งผลให้เกิดความถูกต้องการทำนายดีมากยิ่งขึ้น



ภาคผนวก



ภาคผนวก ก
แบบประเมินความพึงพอใจของผู้ใช้ระบบ



แบบประเมินความพึงพอใจของผู้ใช้ระบบ

ชื่อวิทยานิพนธ์ การประยุกต์ระบบ GPS และการทำเหมืองข้อมูลเพื่อทำนาย และป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด

คำชี้แจง

แบบสอบถามความคิดเห็นชุดนี้ จัดทำขึ้นเพื่อใช้สอบถามความคิดเห็นเกี่ยวกับการใช้งานโปรแกรมคอมพิวเตอร์ มีวัตถุประสงค์เพื่อประเมินหาระดับความพึงพอใจของ “การประยุกต์ระบบจีพีเอส และการทำเหมืองข้อมูลเพื่อทำนาย และป้องกันการเกิดอุบัติเหตุของรถโดยสารสาธารณะบริการระหว่างจังหวัด” ที่ได้พัฒนาขึ้น ซึ่งแบบสอบถามนี้ ได้แบ่งออกเป็น 2 ตอน ประกอบด้วย

ตอนที่ 1 ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม

ตอนที่ 2 ข้อมูลความคิดเห็นเกี่ยวกับความพึงพอใจที่มีต่อระบบ

ในฐานะที่ท่านเป็นผู้หนึ่งที่มีความเกี่ยวข้องกับการใช้ระบบที่พัฒนาขึ้นนี้ ขอได้โปรดพิจารณาและกรุณาตอบคำถามให้ครบทุกข้อตามความเป็นจริง เพราะคำตอบของท่านจะเป็นประโยชน์อย่างยิ่งต่อการพัฒนาโปรแกรมในครั้งนี้ เพื่อที่ผู้พัฒนาจะได้นำข้อมูลไปวิเคราะห์และประเมินความพึงพอใจของโปรแกรมต่อไป

ขอขอบพระคุณเป็นอย่างยิ่งที่ท่านได้กรุณาให้ความร่วมมือในการตอบแบบสอบถามในครั้งนี้

2569600139

ชื่อ นาย เกรียงไกร เอี่ยมเลิศวงศ์ รหัสนักศึกษา

สาขาวิชา วิทยาศาสตร์และเทคโนโลยี

แขนงวิชา เทคโนโลยีสารสนเทศและการสื่อสาร

มหาวิทยาลัยสุโขทัยธรรมมาธิราช

อาจารย์ที่ปรึกษาหลัก รองศาสตราจารย์ ดร.วิภา เจริญภักดิ์

อาจารย์ที่ปรึกษาร่วม อาจารย์ ดร. ดวงดาว วิชาดากุล

ตอนที่ 1 ข้อมูลเกี่ยวกับระดับความพึงพอใจของผู้ใช้โปรแกรม

คำชี้แจง

1. แบบสอบถามความคิดเห็นตอนที่ 1 นี้ เป็นการสอบถามข้อมูลความคิดเห็นของผู้ตอบแบบสอบถามภายหลังจากที่ได้ทดลองใช้โปรแกรมที่พัฒนาขึ้น ซึ่งแบบสอบถามแบ่งออกเป็น 4 ด้าน คือ
 - 1.1 ด้านการตรงตามความต้องการของผู้ใช้ระบบ (Functional Requirement Test)
 - 1.2 ด้านการทำงานได้ตามฟังก์ชันงานของระบบ (Functional Test)
 - 1.3 ด้านความง่ายต่อการใช้งานระบบ (Usability Test)
 - 1.4 ด้านการรักษาความปลอดภัยของข้อมูลในระบบ (Security Test)
2. ในการตอบแบบสอบถามตอนที่ 1 นี้ ขอความกรุณาให้ท่านดำเนินการดังนี้

ทำเครื่องหมาย ✓ ลงในช่องในแบบสอบถามที่ตรงกับระดับความคิดเห็นของท่านมากที่สุด โดยตัวเลขของระดับความพึงพอใจแต่ละด้านมีความหมายดังนี้

 - 5 หมายถึง ความเหมาะสม/ความพึงพอใจในระดับมากที่สุด
 - 4 หมายถึง ความเหมาะสม/ความพึงพอใจในระดับมาก
 - 3 หมายถึง ความเหมาะสม/ความพึงพอใจในระดับปานกลาง
 - 2 หมายถึง ความเหมาะสม/ความพึงพอใจในระดับน้อย
 - 1 หมายถึง ความเหมาะสม/ความพึงพอใจในระดับน้อยที่สุด

ตัวอย่างการประเมิน

รายการประเมิน	ระดับความพึงพอใจ				
	5	4	3	2	1
การออกแบบหน้าจอมีความเป็นมาตรฐานเดียวกัน		✓			

แบบสอบถามความพึงพอใจด้านการตรงตามความต้องการผู้ใช้ระบบ (Functional Requirement Test)

รายการประเมิน	ระดับความพึงพอใจ				
	5	4	3	2	1
1. ความสะดวกและง่ายในการใช้ระบบ 2. ความสมบูรณ์ของข้อมูลพนักงานขับรถ 3. ความรวดเร็วในประมวลผลของระบบงาน 4. ความชัดเจนในการแบ่งกลุ่มบุคคล 5. ความแม่นยำในการทำนายผลข้อมูลการเกิดอุบัติเหตุการขับรถ 6. ข้อมูลข่าวสารตรงความต้องการ 7. ข้อมูลที่ได้ ช่วยในการตัดสินใจ 8. ความเหมาะสมการโต้ตอบกับผู้ใช้ 9. คำศัพท์ที่ผู้ใช้สามารถปฏิบัติตามได้ง่าย					

ตอนที่ 2 ความคิดเห็นและข้อเสนอแนะ

โปรดแสดงความคิดเห็นและข้อเสนอแนะเกี่ยวกับการพัฒนาระบบ

.....

.....

.....

.....

.....

.....

ขอขอบพระคุณเป็นอย่างสูงในการให้ข้อมูลที่เป็นจริงจากท่าน

ภาคผนวก ข

แบบประเมินความเชื่อมั่นการทำนายการเกิดอุบัติเหตุโดยสารสนเทศ



แบบประเมินความเชื่อมั่นการทำนายการเกิดอุบัติเหตุโดยสาธารณชน

คำชี้แจง

1. แบบประเมินชุดนี้สำหรับผู้ใช้งาน เพื่อประเมินความเชื่อมั่นในระบบการทำนายการเกิดอุบัติเหตุโดยสาธารณชน

2. การแสดงความคิดเห็นของผู้ใช้งานที่มีต่อระบบการทำนายการเกิดอุบัติเหตุโดยสาธารณชนที่ได้พัฒนาขึ้นจะประกอบด้วยส่วนของรายการประเมินที่อยู่ด้านซ้ายมือ และมาตราส่วนประมาณค่าอยู่ด้านขวามือ จำนวน 5 ช่อง โปรดทำเครื่องหมาย ลงในช่องทางด้านขวามือที่ตรงกับความคิดเห็นของท่าน โดยกำหนดค่าความหมายดังนี้

- | | |
|---|------------------------------------|
| 5 | หมายถึง ระดับความเชื่อมั่นดีมาก |
| 4 | หมายถึง ระดับความเชื่อมั่นมาก |
| 3 | หมายถึง ระดับความเชื่อมั่นปานกลาง |
| 2 | หมายถึง ระดับความเชื่อมั่นพอใช้ |
| 1 | หมายถึง ระดับความเชื่อมั่นปรับปรุง |

รายละเอียดแบบสอบถาม

แบบสอบถามแบ่งเป็น 2 ตอน ดังนี้

ตอนที่ 1 ความเชื่อมั่นของระบบการทำนาย

ตอนที่ 2 ความคิดเห็นและข้อเสนอแนะ



ตอนที่ 1 ความเชื่อมั่นของระบบการทำนาย

กลุ่มผู้ใช้งาน

[] ผู้บริหาร

[] พนักงาน

รายการประเมิน	ระดับการประเมิน				
	ดีมาก 5	มาก 4	ปานกลาง 3	พอใช้ 2	ปรับปรุง 1
1. ความสะดวกและง่ายต่อการใช้งาน					
2. ความน่าเชื่อถือของผลการทำนาย					
3. ผลการทำนายตรงกับความต้องการ					
4. ช่วยประกอบการตัดสินใจ					

ตอนที่ 2 ความคิดเห็นและข้อเสนอแนะ

.....

.....

.....

.....

.....

.....

.....

.....

*** ขอขอบคุณสำหรับการตอบแบบสอบถาม ***

ภาคผนวก ค
คู่มือการติดตั้ง โปรแกรม



คู่มือการติดตั้งโปรแกรม

1. การติดตั้งโปรแกรม KNIME v 2.12

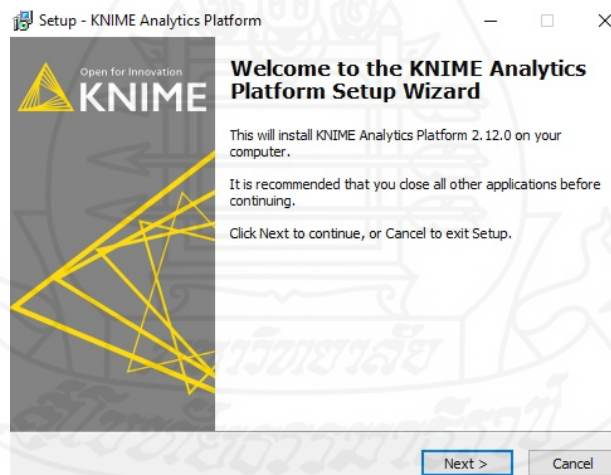
เครื่องมือพัฒนาการประยุกต์ระบบจีพีเอส และการทำเหมืองข้อมูลเพื่อทำนาย คือ โปรแกรม KNIME v2.12 มีขั้นตอนในการติดตั้งโปรแกรม มีดังนี้

- 1) ดาวน์โหลดไฟล์จากเว็บไซต์ของ KNIME.com



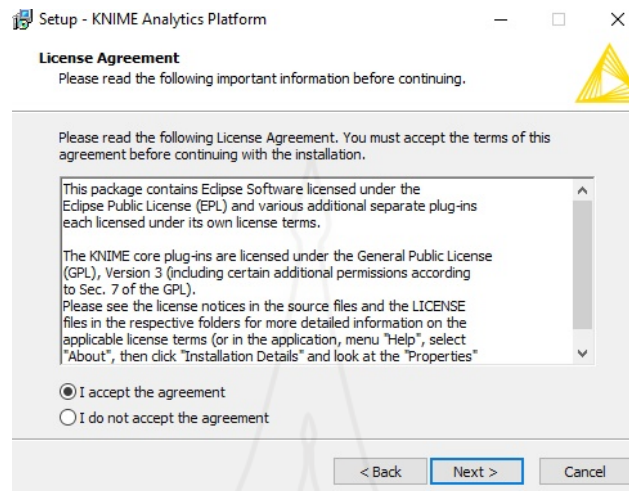
ภาพที่ 1.1 ไฟล์ติดตั้งโปรแกรม KNIME v2.12

- 2) เริ่มต้นไฟล์ติดตั้งโปรแกรม KNIME v2.12 โปรแกรมจะแสดงหน้า "Welcome to the KNIME Analytics Platform setup Wizard" กดปุ่ม "Next"



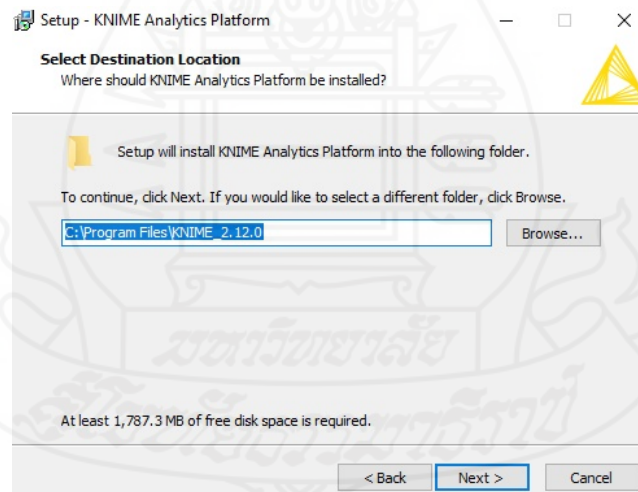
ภาพที่ 1.2 แสดงหน้าการติดตั้ง โปรแกรม KNIME v2.12

3) ยืนยันการยอมรับข้อตกลงระบบ คลิก "I accept the agreement" กดปุ่ม "Next"



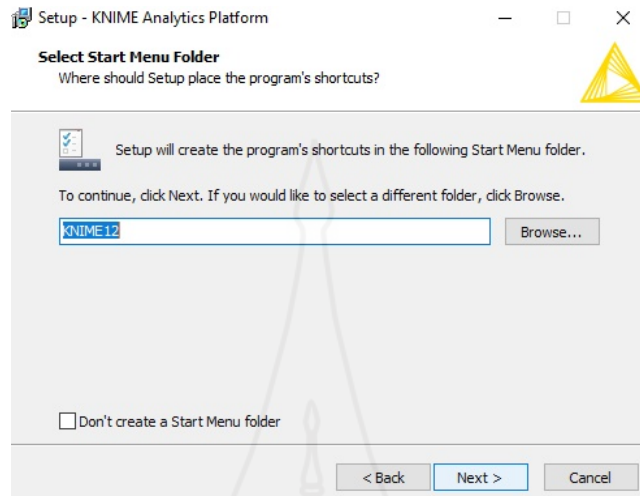
ภาพที่ 1.3 แสดงการยืนยันการยอมรับข้อตกลงโปรแกรมระบบ

4) กำหนดพื้นที่ติดตั้งโปรแกรม และพื้นที่ว่างไม่ต่ำกว่า 1,787.3 MB กดปุ่ม "Next"



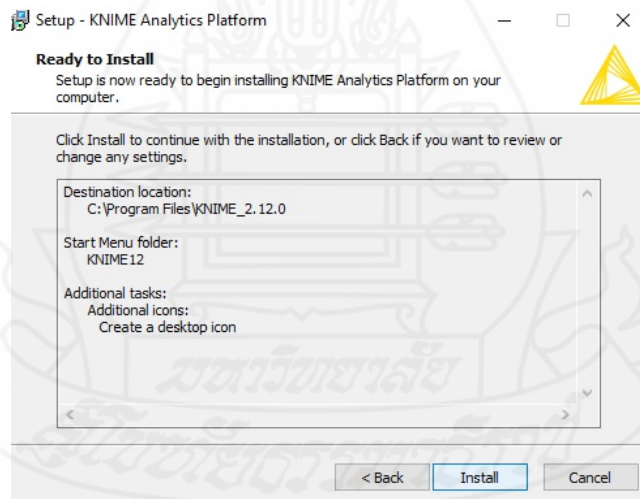
ภาพที่ 1.4 แสดงกำหนดพื้นที่การติดตั้งโปรแกรม

5) สร้างการเรียกใช้โปรแกรม KNIME กดปุ่ม “Next”



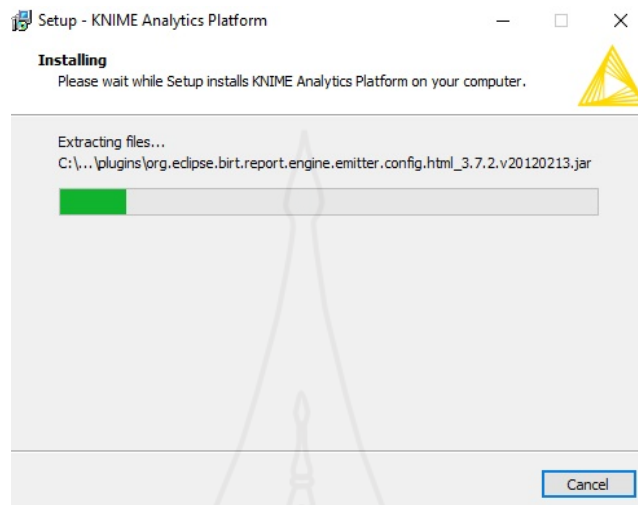
ภาพที่ 1.5 สร้างการเรียกใช้โปรแกรม KNIME

6) ยืนยันความพร้อมการติดตั้งโปรแกรม KNIME กดปุ่ม “Install”



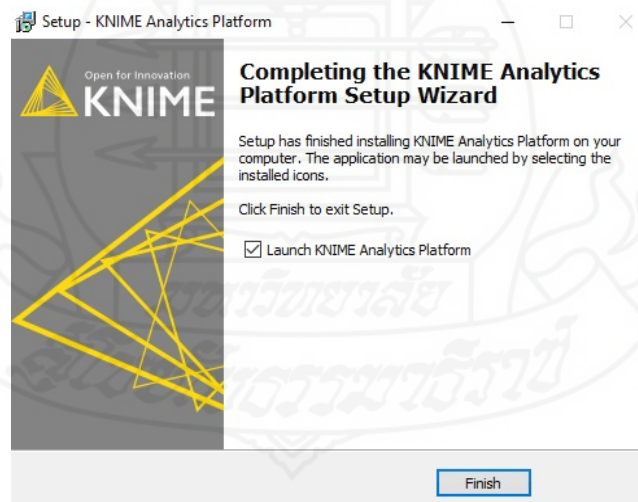
ภาพที่ 1.6 ยืนยันความพร้อมการติดตั้งโปรแกรม KNIME

7) แสดงกระบวนการติดตั้งโปรแกรม KNIME



ภาพที่ 1.7 แสดงกระบวนการติดตั้งโปรแกรม KNIME

8) แสดงความสำเร็จการติดตั้งโปรแกรม KNIME กดปุ่ม "Finish"

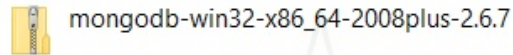


ภาพที่ 1.8 แสดงความสำเร็จติดตั้งโปรแกรม KNIME

2. การติดตั้งโปรแกรม MongoDB v 2.6.7

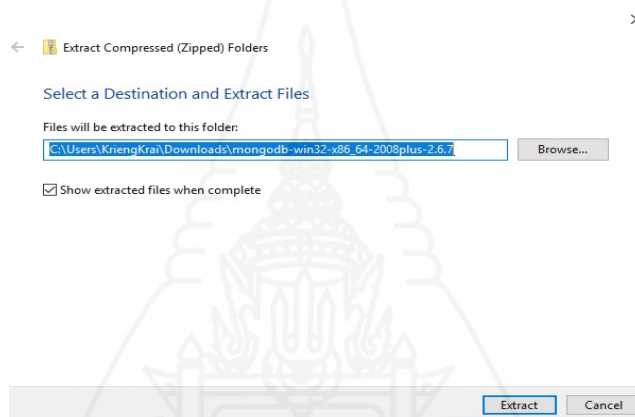
เครื่องมือพัฒนากล้องข้อมูลระบบ GPS โปรแกรม MongoDB v2.6.7 มีขั้นตอนในการติดตั้งโปรแกรม มีดังนี้

- 1) ดาวน์โหลดไฟล์จากเว็บไซต์ของ MongoDB.com

 mongodb-win32-x86_64-2008plus-2.6.7

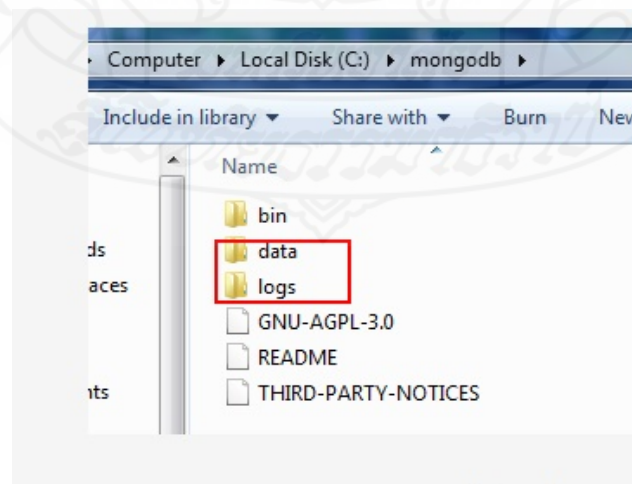
ภาพที่ 2.1 ไฟล์ติดตั้งโปรแกรม MongoDB v 2.6.7

- 2) ทำการแตกไฟล์ และระบุชื่อพื้นที่ที่ต้องการติดตั้งระบบกล้องข้อมูล กดปุ่ม "Extract"



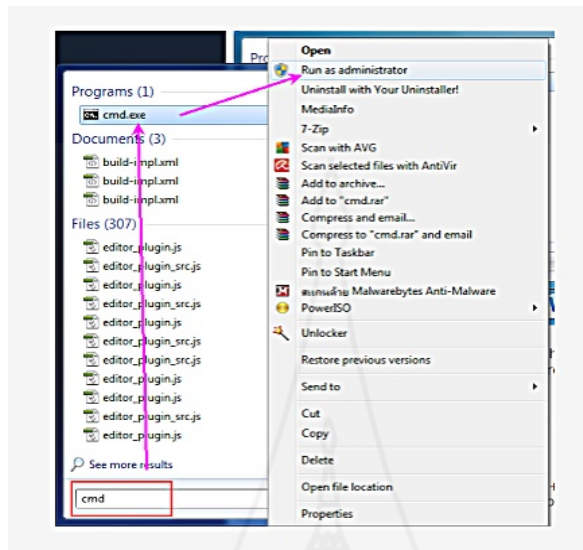
ภาพที่ 2.2 แสดงการแตกไฟล์ และระบุชื่อพื้นที่ MongoDB v 2.6.7

- 3) เปิดโฟลเดอร์ของ MongoDB สร้างโฟลเดอร์ data และ logs



ภาพที่ 2.3 แสดงสร้างโฟลเดอร์ data และ logs

4) กำหนดให้เปิด Service ให้กับ MongoDB เป็น Auto Start ทุกครั้งที่เปิดเครื่องใหม่



ภาพที่ 2.4 กำหนดให้เปิด Service ให้กับ MongoDB

5) เข้าไปใน Path ของโฟลเดอร์ MongoDB และไปที่ c:\mongodb\bin ระบุคำสั่งตาม

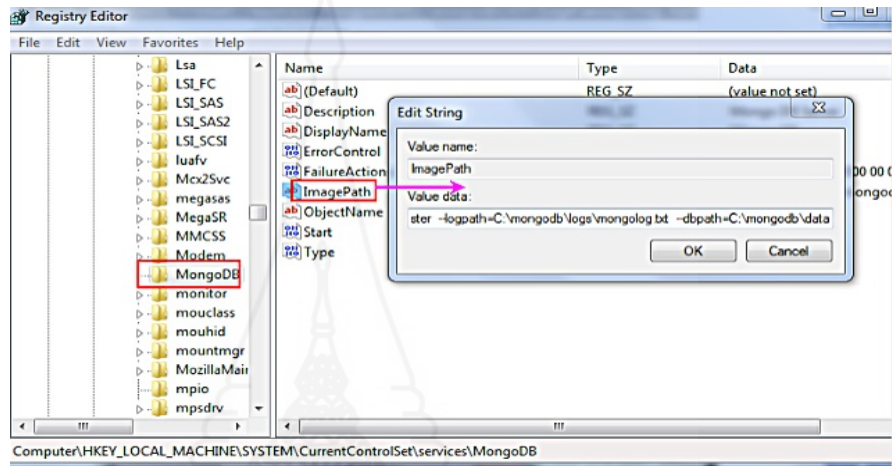
ภาพที่ 2.5

```
Administrator: C:\Windows\System32\cmd.exe
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Windows\system32>cd\
C:\>cd mongodb\bin
C:\mongodb\bin>mongod --install --rest --master --logpath=C:\mongodb\logs\mongolog.txt
```

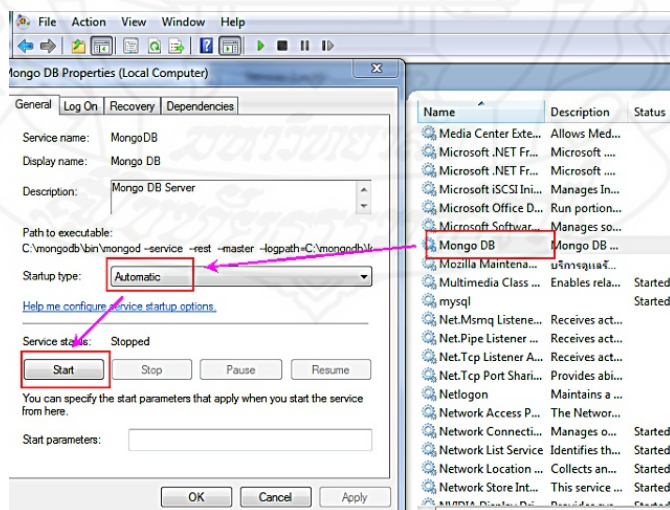
ภาพที่ 2.5 แสดงการกำหนดค่าใน โฟลเดอร์ "mongodb\bin"

6) ไปที่ Start เลือก Run พิมพ์ว่า "regedit" เปิด "Registry Editor" ไปที่ "HKEY_LOCAL_MACHINE -> SYSTEM->CurrentControlSet->services " คลิกที่ MongoDB และดับเบิลคลิกที่ ImagePath และพิมพ์ว่า C:\mongodb\bin\mongod --service --rest --master --logpath=C:\mongodb\logs\mongolog.txt --dbpath=C:\mongodb\data ในช่อง Value data



ภาพที่ 2.6 แสดงการกำหนดค่าใน ImagePath

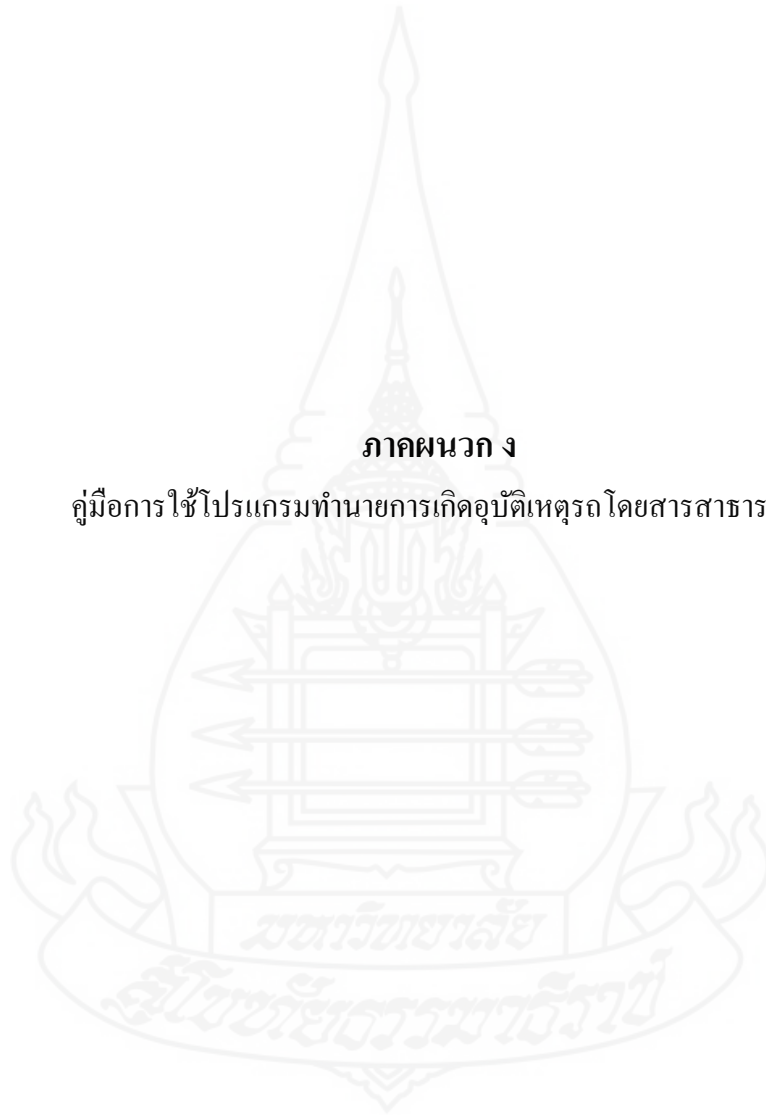
7) ไปที่ Start เลือก Run พิมพ์ว่า "service" เลือก "Component Service -> Service(local -> MongoDB)" กำหนด "Startup type" เป็น "Automatic" กดปุ่ม "Start"



ภาพที่ 2.7 แสดงการทำงานของ MongoDB

ภาคผนวก ง

คู่มือการใช้โปรแกรมทำนายการเกิดอุบัติเหตุโดยสารสนเทศ



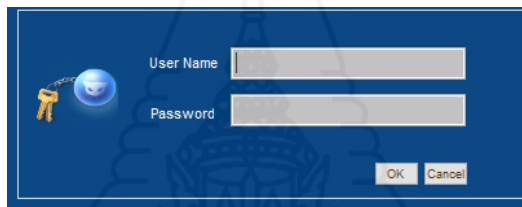
คู่มือการใช้โปรแกรมทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ

1. เปิดโปรแกรมทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ



ภาพที่ 1.1 โปรแกรมทำนายการเกิดอุบัติเหตุรถโดยสารสาธารณะ

2. ระบุ User Name และ Password ผู้ใช้งานระบบ



ภาพที่ 1.2 แสดงการระบุผู้ใช้ระบบงาน

3. เข้าสู่หน้าจอการเลือกข้อมูลเพื่อใช้ในการแสดงผลการทำงานการทำนายความเสี่ยงโดยกำหนดคุณสมบัติส่วนบุคคล

ข้อมูลบุคคล(ผู้ขับขี่)

ชื่อ - นามสกุล นาย [] []

วันเกิด วันพุธที่ 1 สิงหาคม 1, 1995

อายุ 31 ปี 8 เดือน

การศึกษา ประถมศึกษา

ภูมิลำเนา สุพรรณบุรี

นำพนักขับ 60

ส่วนสูง 175

ประสบการณ์ขับรถ 2.00

รายได้ต่อเดือน 16000

อายุ 31 ปี ระดับ ACB_1
การศึกษา ประถมศึกษา ED_1
ภูมิลำเนา ภาคอีสาน PRV_5
นำพนักขับ 50-65 PV_1
สัญชาติ มลฑลที่ 176 II_3
ประสบการณ์ขับรถ 1-2 ปี PCK_2
รายได้ 15001-20000 IN_3

คุณจัดอยู่ในกลุ่มค่าให้เกตุอุบัติเหตุ

กลุ่ม 2 (เสี่ยง) -> ภาคอีสาน -> อายุกลุ่ม(60-35ปี) -> ประสบการณ์(1-2ปี) -> การศึกษา(ประถม) -> รายได้เฉลี่ยต่อวัน(500-700 บาท)

ภาพที่ 1.3 โปรแกรมทำนายความเสี่ยงโดยกำหนดคุณสมบัติส่วนบุคคล

การทำงาน

1) ระบุ เดือน - พ.ศ. สำหรับกำหนดช่วงการตรวจสอบคุณสมบัติบุคคล

2) คลิก ผู้สมัครใหม่ หรือ พนักงานขับรถบริษัท

เลือกคลิก ผู้สมัครใหม่ สิ่งที่ต้องระบุ ดังนี้ คำนำหน้า ชื่อ นามสกุล วันเกิด การศึกษา ภูมิลำเนา น้ำหนักตัว ส่วนสูง ประสบการณ์ทำงาน และรายได้ต่อเดือน

เลือกคลิก พนักงานขับรถบริษัท ให้ระบุรหัสพนักงานเท่านั้น ระบบจะดึงข้อมูลมาจากฐานข้อมูลบุคคล และระบุรายได้ล่าสุด

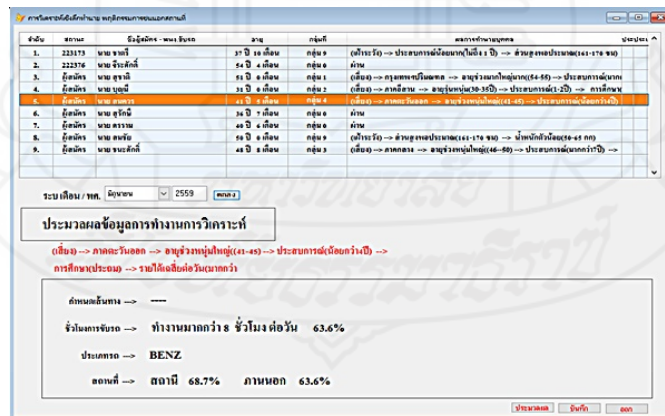
3) คลิกปุ่ม ประมวลผล สำหรับประมวลผลความเสี่ยงการเกิดอุบัติเหตุ

4) คลิกปุ่ม บันทึก สำหรับการบันทึกประมวลผล

5) คลิกปุ่ม Clear สำหรับล้างผลการประมวลผลการทำนายหรือคาดการณ์

6) คลิกปุ่ม ออก สำหรับออกจากโปรแกรม

4. เข้าสู่หน้าของการเลือกข้อมูลเพื่อใช้ในแสดงผลการทำงานการทำนายความเสี่ยงโดยกำหนดองค์ประกอบการเดินรถ



ภาพที่ 1.4 โปรแกรมทำนายความเสี่ยงโดยกำหนดองค์ประกอบการเดินรถ

การทำงาน

1) ระบุ เดือน - พ.ศ. สำหรับกำหนดช่วงการตรวจสอบคุณสมบัติบุคคล

2) คลิกปุ่ม ตกลง สำหรับการแสดงข้อมูลผู้สมัครขับรถ

3) คลิก ตารางแสดงผลรายงานข้อมูล สำหรับเลือกข้อมูลผู้สมัครขับรถหรือพนักงานขับรถบริษัทฯ(ทดสอบ)

4) คลิกปุ่ม ประมวลผล สำหรับประมวลผลความเสี่ยงการเกิดอุบัติเหตุ

5) คลิกปุ่ม บันทึก สำหรับการบันทึกประมวลผล

6) คลิกปุ่ม ออก สำหรับออกจากโปรแกรม

5. เข้าสู่หน้าของการเลือกข้อมูลเพื่อใช้ในรายงานผลการทำนายความเสี่ยงการเกิดอุบัติเหตุชุด โดยสาธารณะ

ลำดับ	สถานะ	ชื่อผู้สมัคร - พว. ยศ	อายุ	กลุ่ม	หมายเหตุ
1.		นาย ชวสิทธิ์	37 ปี 8 เดือน	กลุ่ม 9	(ตำรวจ) -> ประสบการณ์ขับรถบรรทุก(ในเชิง 5 ปี) -> ส่วนสูงของประมาณ(161-178 ซม)
2.		นาย ชวิษฐ์	34 ปี 4 เดือน	กลุ่ม 8	บ้าน
3.	ผู้สมัคร	นาย ชวสิทธิ์	31 ปี 8 เดือน	กลุ่ม 1	(ศิษย์) -> คุมขังขังบริเวณพัก -> อายุข้อมาใหม่(54-55) -> ประสบการณ์(มากกว่า 7 ปี) -> การศึกษา(ประถมศึกษา)
4.	ผู้สมัคร	นาย ชวสิทธิ์	31 ปี 8 เดือน	กลุ่ม 2	(ศิษย์) -> ภาคอีสาน -> อายุข้อมาใหม่(30-35 ปี) -> ประสบการณ์(1-2 ปี) -> การศึกษา(ประถมศึกษา) -> รายได้เฉลี่ยต่อปี
5.	ผู้สมัคร	นาย ชวสิทธิ์	41 ปี 8 เดือน	กลุ่ม 4	(ศิษย์) -> ภาคตะวันออก -> อายุข้อมาใหม่(41-45) -> ประสบการณ์(น้อยกว่า 1 ปี) -> การศึกษา(ประถมศึกษา) -> ร
6.	ผู้สมัคร	นาย ชวสิทธิ์	36 ปี 7 เดือน	กลุ่ม 8	บ้าน
7.	ผู้สมัคร	นาย ชวสิทธิ์	46 ปี 8 เดือน	กลุ่ม 8	บ้าน
8.	ผู้สมัคร	นาย ชวสิทธิ์	36 ปี 8 เดือน	กลุ่ม 9	(ตำรวจ) -> ส่วนสูงของประมาณ(161-178 ซม) -> น้ำหนักตัว(160-65 กก)
9.	ผู้สมัคร	นาย ชวสิทธิ์	46 ปี 8 เดือน	กลุ่ม 9	(ศิษย์) -> ภาคกลาง -> อายุข้อมาใหม่(46-50) -> ประสบการณ์(มากกว่า 7 ปี) -> การศึกษา(ประถมศึกษา) -> รายได้

ภาพที่ 1.5 โปรแกรมรายงานผลการทำนายความเสี่ยงการเกิดอุบัติเหตุชุด โดยสาธารณะ
การทำงาน

1) ระบุ เดือน - พ.ศ. สำหรับระบุเดือนและปี พ.ศ.

2) คลิกปุ่ม ตกลง แสดงรายงานผลการทำนายคาดการณ์ข้อมูลผู้สมัครขับรถ

3) คลิกปุ่ม พิมพ์ สำหรับพิมพ์รายงานผลรูปแบบกระดาษ

4) คลิกปุ่ม ออก สำหรับออกจากโปรแกรม

บรรณานุกรม



บรรณานุกรม

- พล โกสุภวัฒน์.(2556). การสร้างแบบจำลองและทำนายผลการเรียนของนักศึกษานิติศาสตร์ มหาวิทยาลัยธรรมศาสตร์ โดยการใช้เหมืองข้อมูล .(การค้นคว้าแบบอิสระปริญญา วิทยาศาสตร์มหาบัณฑิต).มหาวิทยาลัยธรรมศาสตร์,กรุงเทพฯ.
- สุรเดช บุญถวิล.(2550).ระบบบริการผู้โดยสารอัจฉริยะผ่านเทคโนโลยีตรวจสอบพิกัดจากดาวเทียม. ภาควิชาวิทยาการคอมพิวเตอร์,คณะเทคโนโลยีสารสนเทศ,วิทยาลัยนอร์ทกรุงเทพ, กรุงเทพฯ.
- วรพล ปัญจศิริประการ.(2553). ปัจจัยการยอมรับการนำระบบติดตามรถยนต์ GPS มาใช้ร่วมบริษัท ประกันภัย.(การค้นคว้าแบบอิสระปริญญาวิทยาศาสตรมหาบัณฑิต).มหาวิทยาลัย ธรรมศาสตร์,กรุงเทพฯ.
- ซัดชัย แก้วตา และ อัจฉรา มหาวีวัฒน์. (2553). การวินิจฉัยคดีด้วยเทคนิคต้นไม้ตัดสินใจ.(การ ค้นคว้าแบบอิสระปริญญาวิทยาศาสตรมหาบัณฑิต).มหาวิทยาลัยอุบลราชธานี, อุบลราชธานี.
- บุญมา เฟ่งชวน.(2548).การใช้เทคนิคเหมืองข้อมูลเพื่อพัฒนาระบบสนับสนุนการตัดสินใจ ด้าน การผลิตบัณฑิตระดับปริญญาตรี. (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต). มหาวิทยาลัยศิลปากร,กรุงเทพฯ.
- พิจิตรา จอมศรี. (2549).การทำนายเนื้อหาของเว็บโดยใช้เทคนิคเหมืองข้อมูล กรณีศึกษามหาวิทยาลัย ศิลปากร.(วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต).มหาวิทยาลัย ศิลปากร, กรุงเทพฯ.
- พวงทิพย์ แทนแสง.(2550). การทดสอบประสิทธิภาพการทำงานของอัลกอริธึมการไม่นิ่งกฎ สำหรับจำแนก.(การค้นคว้าแบบอิสระปริญญาวิทยาศาสตรมหาบัณฑิต).สถาบัน เทคโนโลยีพระจอมเกล้าพระนครเหนือ,กรุงเทพฯ.
- ศิริรัตน์ ช่านาญรบ และบุญธรรม พรเจริญ.(2555). การพัฒนาระบบสารสนเทศเพื่อส่งเสริม กิจกรรมชุมชน กรณีศึกษา ชุมชนนางเลิ้ง กรุงเทพมหานคร.มหาวิทยาลัย เทคโนโลยี ราชมนักลพระนคร,กรุงเทพฯ.
- ไพฑูรย์ จันทร์เรือง. (2550). ระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียนของนักศึกษาระดับ ปริญญาตรี.(การค้นคว้าแบบอิสระปริญญาวิทยาศาสตรมหาบัณฑิต).มหาวิทยาลัย เทคโนโลยีพระจอมเกล้าพระนครเหนือ,กรุงเทพฯ.

- ประกายมาศ ศรีสุขทักษิณ และคณะ.(2557). *การเปรียบเทียบความเร็วในการประมวลผลระหว่างฐานข้อมูลเชิงสัมพันธ์และฐานข้อมูลไม่สัมพันธ์แบบเอกสาร*. (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต). มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.
- นิรุทธ์ รวยรื่น.(2557). *การใช้แมพรีดิวซ์เชื่อมคอลเลกชันของฐานข้อมูลมอดกอดีบี*. (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต). มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี, กรุงเทพฯ.
- เสกสรรค์ วิลัยลักษณ์.(2558). *การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน*. (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต). มหาวิทยาลัยสุโขทัยธรรมาธิราช , นนทบุรี.
- โสภณ พินิจกิจเจริญกุล.(2559). *ทำความเข้าใจกับคำว่า Big Data กัน*. มหาวิทยาลัยราชภัฏพิบูลสงคราม.
- Joseph O.Chanh,(2013), "An Architecture for Big Data Analytics", Roosevelt University, USA.
- Wei, Z., Pierre, G. and Chi.(2012). "Scalable Join Queries in Cloud Data Stores", CCGRID '12 Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID 2012), USA.
- Calil, A. and Mello, R.S.(2012). "SimpleSQL: A Relational Layer for SimpleDB", ADBIS'12: Proceedings of the 16th East European conference on Advances in Databases and Information Systems, September 2012, Germany.
- Connolly, Thomas & Begg, Carolyn. (2002). *Database Systems :A Practical Approach to Design, implementation and management*. Third Edition. Addison Wesley, England.
- Best, John W. (1963). *Research in Education*. New Jersey: Prentice-Hall.
- Sadalage, P.J. and Fowler, M., (2013), *NoSQL distilled : a brief guide to the emerging world of polyglot*, 2nded., Pearson Education Inc, United States of America, pp. 9-28.
- Islam, R., (2011), *PHP and MongoDB Web Development Beginner's Guide*, 1sted., Packt Publishing, Birmingham, pp. 8-12.
- MongoDB Inc, Who Uses MongoDB? [Online], Available : <http://www.mongodb.com/who-uses-mongodb/>. [2014, September 21].

MongoDB Org, SQL to MongoDB Mapping Chart [Online], Available :

<http://docs.mongodb.org/manual/reference/sql-comparison>. [2013, September 2].

MongoDB Org, Introduction to MongoDB [Online], Available :

<http://www.mongodb.org/about/introduction/>. [2015, June 7].

Parth Chandarana and Vijayalakshmi, (2014), "Big data analytics frameworks", Department of Information Technology, India.



ประวัติผู้วิจัย

ชื่อ	นาย เกรียงไกร เอี่ยมเลิศวงศ์
วัน เดือน ปีเกิด	15 ธันวาคม 2508
สถานที่เกิด	เขตป้อมปราบศัตรูพ่าย จังหวัดกรุงเทพมหานคร
ประวัติการศึกษา	อุตสาหกรรมศาสตรบัณฑิต (อศ.บ.) สาขาวิชาวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีมหานคร พ.ศ. 2538
สถานที่ทำงาน	บริษัท รถรุ่งเรือง จำกัด เขตบางนา จังหวัดกรุงเทพมหานคร
ตำแหน่ง	หัวหน้าฝ่ายคอมพิวเตอร์

