การพยากรณ์การใช้สิทธิของลูกจ้างในสถานประกอบการเพื่อเพิ่มประสิทธิภาพการบริหารจัดการกองทุนเงินทดแทน สำนักงานประกันสังคม

นายอิทธิ บรรลือศักดิ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
เขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
พ.ศ. 2559

Claim Prediction of Enterprise Employees to Increase Management Efficiency for Workmen's Compensation Fund of Social Security Office

Mr. Itti Banlursak

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 การพยากรณ์การใช้สิทธิของลูกจ้างในสถานประกอบการเพื่อเพิ่มประสิทธิภาพการบริหารจัดการกองทุนเงินทดแทน สำนักงานประกันสังคม

ชื่อและนามสกุล นายอิทธิ บรรลือศักดิ์

แขนงวิชา เทคโนโลยีสารสนเทศและการสื่อสาร

สาขาวิชา วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

อาจารย์ที่ปรึกษา 1. รองศาสตราจารย์ ดร. วิภา เจริญภัณฑารักษ์

2. อาจารย์ ดร. ดวงดาว วิชาดากุล

วิทยานิพนธ์นี้ ได้รับความเห็นชอบให้เป็นส่วนหนึ่งของการศึกษา ตามหลักสูตรระดับปริญญาโท เมื่อวันที่ 21 ตุลาคม 2559

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. ประจวบ วานิชชัชวาล)

..... กรรมการ

(รองศาสตราจารย์ ดร. วิภา เจริญภัณฑารักษ์)

..... กรรมการ

(อาจารย์ ดร. ดวงดาว วิชาดากุล)

..... ประธานกรรมการบัณฑิตศึกษา

(รองศาสตราจารย์รสลิน ศิริยะพันธุ์)

ชื่อวิทยานิพนธ์ การพยากรณ์การใช้สิทธิของลูกจ้างในสถานประกอบการเพื่อเพิ่มประสิทธิภาพการบริหารจัดการกองทุนเงินทดแทน สำนักงานประกันสังคม

ผู้วิจัย นายอิทธิ บรรลือศักดิ์ **รหัสนักศึกษา** 2579600020 **ปริญญา** วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีสารสนเทศและการสื่อสาร) **อาจารย์ที่ปรึกษา** (1) รองศาสตราจารย์ ดร. วิภา เจริญภัณฑารักษ์ (2) อาจารย์ ดร. ดวงดาว วิชาดากุล **ปีการศึกษา** 2559

บทคัดย่อ

การวิจัยนี้ทำการพยากรณ์การใช้สิทธิของลูกจ้างในสถานประกอบการ เพื่อเพิ่มประสิทธิภาพการบริหารจัดการกองทุนเงินทดแทน สำนักงานประกันสังคม มีวัตถุประสงค์เพื่อศึกษา 1) การพัฒนาคลังข้อมูลเพื่อวิเคราะห์ข้อมูล และจัดทำรายงาน ตามหลักการธุรกิจอัจฉริยะ 2) การพยากรณ์การใช้สิทธิของลูกจ้างที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคม ด้วยวิธีการจำแนกประเภท และ 3) การประเมินผลการใช้งานคลังข้อมูล

การพัฒนาคลังข้อมูลโดยใช้รูปแบบโครงสร้างสตาร์ ซึ่งใช้ข้อมูลการใช้สิทธิของลูกจ้างที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 1-12 ปี พ.ศ.2556- 2557 ด้วยโปรแกรม Microsoft SQL Server 2008 R2 และทำกระบวนการอีทีแอล ด้วยเครื่องมือ SQL Server Integration Services และนำเสนอรายงานผ่านโปรแกรม Microsoft Excel 2007 จากนั้นใช้ข้อมูลการใช้สิทธิของลูกจ้างที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 3 จำนวน 3,636 ชุดข้อมูล สร้างตัวแบบพยากรณ์โดยการเลือกใช้ข้อมูล 2 แบบ คือ ชุดข้อมูลแบบไม่จัดกลุ่ม และชุดข้อมูลแบบจัดกลุ่ม นำไปผ่านกระบวนการคัดเลือกคุณลักษณะ ด้วยวิธีเลือกคุณสมบัติตามความสัมพันธ์ แล้วทำการทดสอบประสิทธิภาพแบบ 10-fold Cross Validation ด้วยโปรแกรม WEKA Version 3.8.0 โดยใช้อัลกอริทึม Multilayer Perceptron, Naïve Bayes และต้นไม้ตัดสินใจ J48 พบว่าชุดข้อมูลแบบไม่จัดกลุ่มที่ใช้อัลกอริทึม J48 ให้ค่าความถูกต้องสูงที่ร้อยละ 100 Multilayer Perceptron ให้ค่าความถูกต้องสูงที่ร้อยละ 99.67 และ Naïve Bayes ให้ค่าความถูกต้องสูงที่ร้อยละ 94.63 ส่วนชุดข้อมูลแบบจัดกลุ่มที่ใช้อัลกอริทึม J48 ให้ค่าความถูกต้องสูงที่ร้อยละ 100 Multilayer Perceptron ให้ค่าความถูกต้องสูงที่ร้อยละ 99.67 และ Naïve Bayes ให้ค่าความถูกต้องสูงที่ร้อยละ 94.00 และการประเมินผลการใช้งานคลังข้อมูล

ผลการวิจัยพบว่า คลังข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน ที่ใช้งานโดย ผู้บริหาร หัวหน้าส่วนงานกองทุนเงินทดแทน และเจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทน มีความพึงพอใจการใช้งานคลังข้อมูลอยู่ในระดับดี และในการทำเหมืองข้อมูลพยากรณ์ พบว่า ชุดข้อมูลแบบไม่จัดกลุ่มและจัดกลุ่ม โดยใช้เทคนิคตัวแบบพยากรณ์ต้นไม้ตัดสินใจ J48 ให้ค่าความถูกต้องสูงที่สุดที่ร้อยละ 100 และมีค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดที่ 0 เหมาะสม สำหรับนำไปสร้างระบบพยากรณ์กองทุนเงินทดแทน

คำสำคัญ คลังข้อมูล การจำแนกแบบเบย์ ต้นไม้ตัดสินใจ กองทุนเงินทดแทน สำนักงานประกันสังคม

Thesis title: Claim Prediction of Enterprise Employees to Increase Management Efficiency for Workmen's Compensation Fund of Social Security Office
Researcher: Mr. Itti Banlursak; **ID:** 2579600020;
Degree: Master of Science (Information and Communication Technology);
Thesis advisors: (1) Dr. Vipa Jaroenpuntaruk, Associate Professor;
(2) Dr. Duangdao Wichadakul; **Academic year:** 2016

Abstract

This research was about claim prediction of enterprise employees to increase management efficiency for workmen's compensation fund of social security office. The objectives of this study were as follows. 1) Developing data warehouse for analyzing and reporting with Business Intelligence. 2) Predicting the claim of workmen's compensation fund of social security office by classification technique, and 3) The evaluation of data warehouse.

Data warehouse was developed using Star Schema for the claim of workmen's compensation fund in Bangkok social security office, covering of the 1st to 12th area between 2013 and 2014. The tools were Microsoft SQL server 2008 R2 for data warehouse development, SQL Server Integration Service for ETL process and Microsoft Excel 2007 for multidimensional data analysis. Data for the model prediction were drawn from the claim of employee in compensation fund of social security office in Bangkok area 3 with 3,636 datasets. The sample data were divided into two groups (original datasets and clustering datasets) by feature selection process with Correlation-based Feature Selection method using 10-fold Cross Validation. Predictive modeling uses WEKA Version 3.8.0. The algorithm for prediction were Multilayer Perceptron, Naïve Bayes, and J48 Decision Tree. The results of original data tested by J48 algorithm had accuracy level at 100 percent: tested by Multilayer Perceptron algorithm had accuracy level at 99.67 percent: and tested by Naïve Bayes algorithm had accuracy level at 94.63 percent. While the results of clustering data tested by J48 algorithm had accuracy level at 100 percent; results tested by Multilayer Perceptron algorithm had accuracy level at 99.67 percent, and tested by Naïve Bayes algorithm had accuracy level at 94.00 percent. The evaluation of data warehouse was good level.

The evaluation results of data warehouse usage showed that the administrators, the director of the workmen's compensation fund, and the social security technical officers were satisfied with the claim of workmen's compensation fund' data warehouse at a good level. Data mining prediction with the original data using Decision Tree prediction model of J48 showed accuracy of the highest level at 100 percent and the square root of minimal error at 0 which was suitable to construct the workmen's compensation prediction system.

Keywords: Data Warehouse, Naïve Bayes, J48 Decision Tree, Workmen's Compensation fund, Social Security Office

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จลุล่วงได้ด้วยการสนับสนุนและความช่วยเหลือจากคณาจารย์หลายท่าน โดยเฉพาะท่านอาจารย์ รองศาสตราจารย์ ดร.วิภา เจริญภัณฑารักษ์ อาจารย์ที่ปรึกษาหลัก และท่านอาจารย์ ดร.ดวงดาว วิชาดากุล อาจารย์ที่ปรึกษาร่วมวิทยานิพนธ์ฉบับนี้ ซึ่งได้กรุณาเสียสละเวลาในการให้คำปรึกษา ให้ข้อเสนอแนะต่างๆ ในการจัดทำวิทยานิพนธ์ตลอดจนแนะนำแนวทางในการศึกษาค้นคว้างานวิจัยและข้อมูลที่นำมาปรับปรุงแก้ไขข้อบกพร่องต่างๆ จนกระทั่งวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดี ผู้วิจัยขอขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

นอกจากนี้ผู้วิจัยขอขอบพระคุณท่านอาจารย์กุลเทพ จิรลักษ์ และคณาจารย์ ประจำสาขาวิชาวิทยาสาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช ตลอดจนผู้ที่มีส่วนเกี่ยวข้องในการทำวิทยานิพนธ์ครั้งนี้ที่ได้กรุณาให้ความช่วยเหลือ คำแนะนำ และให้กำลังใจตลอดจนกระทั่งการทำวิทยานิพนธ์ฉบับนี้สำเสร็จได้ด้วยดี ผู้วิจัยขอขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

สุดท้ายนี้ผู้วิจัยขอขอบคุณสำนักงานประกันสังคม ที่ได้ให้โอกาสในการศึกษาต่อ และให้ความอนุเคราะห์ข้อมูลของสำนักงานประกันสังคม รวมทั้งขอขอบคุณผู้บริหาร และหัวหน้าส่วนงานที่ช่วยให้คำแนะนำ และความช่วยเหลือต่างๆ หากมีข้อบกพร่องหรือผิดพลาดประการใด ผู้วิจัยขออภัยในข้อบกพร่องและความผิดพลาดนั้นไว้ ณ ที่นี้

อิทธิ บรรลือศักดิ์

ตุลาคม 2559

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญตาราง ..... ฌ
สารบัญภาพ ..... ฎ
บทที่ 1 บทนำ 1
ความเป็นมาและความสำคัญของปัญหา 1
วัตถุประสงค์การวิจัย 2
ขอบเขตของการวิจัย 2
นิยามศัพท์เฉพาะ 4
ประโยชน์ที่คาดว่าจะได้รับ 6
บทที่ 2 วรรณกรรมที่เกี่ยวข้อง 7
สภาพทั่วไปของสำนักงานประกันสังคม 7
กฎหมายสำนักงานประกันสังคม กระทรวงแรงงาน 9
คลังข้อมูล 10
การทำเหมืองข้อมูล 15
ธุระกิจอัจฉริยะ 23
งานวิจัยที่เกี่ยวข้อง 24
บทที่ 3 วิธีดำเนินการวิจัย 29
ประชากรและกลุ่มตัวอย่าง 29
เครื่องมือที่ใช้ในการวิจัย 30
ขั้นตอนการดำเนินงาน 30
บทที่ 4 ผลการวิเคราะห์ข้อมูล 36
การพัฒนาคลังข้อมูล 36
การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ 46
การทำเหมืองข้อมูล 53
การประเมินความพึงพอใจต่อการใช้งานคลังข้อมูล 85

สารบัญ (ต่อ)

หน้า

บทที่ 5 สรุปการวิจัย อภิปรายผล และข้อเสนอแนะ 89
สรุปการวิจัย 89
อภิปรายผล 91
ปัญหาและอุปสรรค 93
ข้อเสนอแนะ 94
บรรณานุกรม 95
ภาคผนวก 99
ประวัติผู้วิจัย 102

สารบัญตาราง

หน้า

ตารางที่ 4.1 ตารางมุมมองมิติการใช้สิทธิลูกจ้างของสถานประกอบการ 38
ตารางที่ 4.2 ตารางข้อเท็จจริง FT_ALL_Injury ใช้สำหรับเก็บข้อมูลการใช้สิทธิลูกจ้าง 39
ตารางที่ 4.3 ตารางมิติ DT_Area ใช้สำหรับเก็บข้อมูลสำนักงานประกันสังคมพื้นที่ 40
ตารางที่ 4.4 ตารางมิติ DT_Position ใช้สำหรับเก็บข้อมูลตำแหน่ง 40
ตารางที่ 4.5 ตารางมิติ DT_Enterprise ใช้สำหรับเก็บข้อมูลสถานประกอบการ 40
ตารางที่ 4.6 ตารางมิติ DT_BusinessType ใช้สำหรับเก็บข้อมูลประเภทกิจการ 40
ตารางที่ 4.7 ตารางมิติ DT_Organ ใช้สำหรับเก็บข้อมูลอวัยวะ 41
ตารางที่ 4.8 ตารางมิติ DT_CauseInjury ใช้สำหรับเก็บข้อมูลสาเหตุการประสบอันตราย 41
ตารางที่ 4.9 ตารางมิติ DT_Period ใช้สำหรับเก็บข้อมูลระยะเวลา 41
ตารางที่ 4.10 ตารางมิติ DT_Age ใช้สำหรับเก็บข้อมูลอายุ 41
ตารางที่ 4.11 ตารางชั่วคราว Temp_Data ใช้สำหรับเก็บข้อมูลการใช้สิทธิลูกจ้าง
จากไฟล์เอ็กเซล (Excel File) 42
ตารางที่ 4.12 คุณลักษณะที่ใช้ในการสร้างตัวแบบพยากรณ์ 55
ตารางที่ 4.13 การแปลงข้อมูลที่เป็นผลลัพธ์ 56
ตารางที่ 4.14 การแปลงคุณลักษณะ 56
ตารางที่ 4.15 ผลการคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection 59
ตารางที่ 4.16 ผลการจัดกลุ่มข้อมูลด้วยอัลกอริทึม k-means 66
ตารางที่ 4.17 ผลการวัดประสิทธิภาพโดยใช้เทคนิคโครงข่ายประสาทเทียม
มัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม 71
ตารางที่ 4.18 ผลการวัดประสิทธิภาพโดยใช้เทคนิคเหมืองข้อมูลการจำแนกแบบเบย์
ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม 72
ตารางที่ 4.19 ผลการวัดประสิทธิภาพโดยใช้เทคนิคเหมืองข้อมูลต้นไม้ตัดสินใจ
โดยใช้อัลกอริทึม J48 ด้วยชุดข้อมูล แบบไม่จัดกลุ่ม 72
ตารางที่ 4.20 การเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม 73
ตารางที่ 4.21 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์
ด้วยชุดข้อมูลแบบจัดกลุ่ม 75
ตารางที่ 4.22 การเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ 77

สารบัญตาราง (ต่อ)

หน้า

ตารางที่ 4.23 ค่าร้อยละข้อทั่วไปของผู้ใช้งานคลังข้อมูล 86
ตารางที่ 4.24 ค่าเฉลี่ยของการประเมินความพึงพอใจระบบคลังข้อมูล 87
ตารางที่ 4.25 ผลการประเมินความพึงพอใจระบบคลังข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนเงินทดแทน สำนักงานประกันสังคม 88

สารบัญภาพ

หน้า

ภาพที่ 2.1 โครงสร้างสำนักงานประกันสังคม 8
ภาพที่ 2.2 สถาปัตยกรรมคลังข้อมูล 11
ภาพที่ 2.3 กระบวนการ Extract Transform Load 13
ภาพที่ 2.4 ลักษณะของคิวบ์ (Club) 14
ภาพที่ 2.5 การค้นหาความรู้ด้วยการทำเหมืองข้อมูล 15
ภาพที่ 2.6 กระบวนการ CRISP-DM Modelv 17
ภาพที่ 2.7 ตัวอย่างเทคนิคการทำเหมืองข้อมูลด้วยวิธีการจำแนกประเภท 19
ภาพที่ 2.8 ภาพต้นไม้ตัดสินใจ 20
ภาพที่ 2.9 ขั้นตอนการจัดกลุ่มข้อมูลด้วย k-means 21
ภาพที่ 2.10 ตัวอย่างการแบ่งข้อมูลด้วยวิธี K – fold Cross validation 22
ภาพที่ 2.11 สถาปัตยกรรมของ Business Intelligence และองค์ประกอบต่างๆ 24
ภาพที่ 2.12 โครงสร้างการทำคลังข้อมูลและเหมืองข้อมูลของโรงเรียน 27
ภาพที่ 3.1 ขั้นตอนการสร้างคลังข้อมูล 32
ภาพที่ 3.2 โมเดลขั้นตอนการสร้างตัวแบบการพยากรณ์ 33
ภาพที่ 4.1 โครงสร้างคลังข้อมูลแบบสตาร์ (Star Schema) 37
ภาพที่ 4.2 การนำข้อมูลเข้าสู่คลัง 43
ภาพที่ 4.3 การทำกระบวนการอีทีแอลด้วยเครื่องมือ SQL Server Integration Services (SSIS) 44
ภาพที่ 4.4 ขั้นตอนของการทำ (Extract Transform and Load : ETL) 45
ภาพที่ 4.5 การแสดงผลรวมของข้อมูลผ่านกระบวนการ (Extract Transform and Load : ETL) 45
ภาพที่ 4.6 การแสดงผลรวมของข้อมูลในคลังข้อมูล 46
ภาพที่ 4.7 Data Source View 47
ภาพที่ 4.8 การประมวลผลเชิงวิเคราะห์แบบออนไลน์ 48
ภาพที่ 4.9 รายงานประจำปีสำหรับผู้บริหาร 49
ภาพที่ 4.10 รายงานสำหรับหัวหน้าส่วนงานกองทุนเงินทดแทน 51
ภาพที่ 4.11 รายงานสำหรับเจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทน 52

สารบัญภาพ (ต่อ)

หน้า

ภาพที่ 4.12 ข้อมูลจากคลังข้อมูลในรูปแบบของไฟล์เอ็กเซล (Excel File) 54
ภาพที่ 4.13 ข้อมูลที่แปลงในรูปแบบของไฟล์ CSV (Comma delimited) 57
ภาพที่ 4.14 ชุดข้อมูลแบบไม่ผ่านการคัดเลือกคุณลักษณะ (Feature Selection) 58
ภาพที่ 4.15 การคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection 59
ภาพที่ 4.16 เทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม 60
ภาพที่ 4.17 ผลกฎการพยากรณ์เทคนิคโครงข่ายประสาทเทียม แบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) 61
ภาพที่ 4.18 ผลของการใช้เทคนิคการจำแนกแบบเบย์ (Naïve Bayes) ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม 62
ภาพที่ 4.19 ผลของการใช้เทคนิคต้นไม้ตัดสินใจโดยใช้อัลกอริทึม J48 ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม 63
ภาพที่ 4.20 เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม J48 64
ภาพที่ 4.21 เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม J48 65
ภาพที่ 4.22 การจัดกลุ่มข้อมูล (Clustering Data) ด้วยอัลกอริทึม k-means 66
ภาพที่ 4.23 ผลที่ได้จากการจัดกลุ่มข้อมูล (Clustering Data) 67
ภาพที่ 4.24 ผลของการใช้เทคนิคโครงข่ายประสาทเทียมแบบ มัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ด้วยชุดข้อมูลแบบจัดกลุ่ม 68
ภาพที่ 4.25 ผลการใช้เทคนิคการจำแนกแบบเบย์ (Naïve Bayes) ด้วยชุดข้อมูลแบบจัดกลุ่ม 69
ภาพที่ 4.26 ผลการใช้เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 ด้วยชุดข้อมูลแบบจัดกลุ่ม 70
ภาพที่ 4.27 กราฟเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม 74
ภาพที่ 4.28 กราฟเปรียบเทียบค่า Precision, Recall และ RMSE ของตัวแบบพยากรณ์ ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม 74
ภาพที่ 4.29 กราฟเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบจัดกลุ่ม 76
ภาพที่ 4.30 กราฟเปรียบเทียบค่า Precision, Recall และ RMSE ของตัวแบบพยากรณ์ ด้วยชุดข้อมูลแบบจัดกลุ่ม 76

สารบัญภาพ (ต่อ)

หน้า

ภาพที่ 4.31 กราฟเปรียบเทียบประสิทธิภาพค่าความถูกต้องของตัวแบบพยากรณ์ 78
ภาพที่ 4.32 แผนภาพบริบทของระบบพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน สำนักงานประกันสังคม 79
ภาพที่ 4.33 แผนภาพกระแสข้อมูลที่ระดับ 0 (Data Flow Diagram Level 0) 80
ภาพที่ 4.34 หน้าจอหลักระบบพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน สำนักงานประกันสังคม 81
ภาพที่ 4.35 หน้าจอการทำงานระบบพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน สำนักงานประกันสังคม 82
ภาพที่ 4.36 แสดงผลการพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการโครงสร้างที่ 1 82
ภาพที่ 4.37 แสดงผลการพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการโครงสร้างที่ 2 83
ภาพที่ 4.38 แสดงผลการพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการโครงสร้างที่ 3 84

บทที่ 1
บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

สำนักงานประกันสังคม เป็นหน่วยงานรัฐบาล ซึ่งให้ความคุ้มครองให้กับนายจ้าง ลูกจ้าง และผู้ประกันตน ตลอดระยะเวลา 25 ปีที่ผ่านมาสำนักงานประกันสังคมได้บริหารจัดการกองทุนประกันสังคมตามพระราชบัญญัติประกันสังคม พ.ศ. 2533 และกองทุนเงินทดแทนตามพระราชบัญญัติกองทุนเงินทดแทน พ.ศ. 2537 อีกทั้งยังคงพัฒนาสิทธิประโยชน์ควบคู่กันอย่างต่อเนื่อง และกาลเวลาที่มีการเปลี่ยนแปลงในด้านเศรษฐกิจ สังคม และการเมือง พร้องทั้งสำนักงานประกันสังคม ภายใต้ประเทศไทย ได้เข้าสู่ประชาคมอาเซียน (ASEAN Economic Community: AEC) ซึ่งได้เพิ่มความคุ้มครองสถานประกอบการขนาดเล็ก และขนาดใหญ่ โดยครอบคลุมแรงงานทั้งสิ้น 1.39 ล้านคน ปัจจุบันกองทุนเงินทดแทนได้ให้ความคุ้มครองลูกจ้างที่ประสบอันตรายที่เกิดจากการปฏิบัติงานให้นายจ้าง ซึ่งได้แก่กรณีประสบอันตราย เจ็บป่วย สูญหาย สูญเสียสมรรถภาพ เงินทดแทน ค่าทดแทน ค่ารักษาพยาบาล ค่าฟื้นฟูสมรรถภาพในการทำงาน การฟื้นฟูสมรรถภาพในการทำงาน และค่าทำศพ (สำนักงานประกันสังคม กระทรวงแรงงาน, 2537)

สำนักงานประกันสังคม มีภารกิจในการบริหารจัดการกองทุนประกันสังคมและกองทุนเงินทดแทน เพื่อให้ลูกจ้าง และผู้ประกันตนมีหลักประกันในการดำรงชีวิตที่มั่นคง การจัดเก็บเงินสมทบ ตามพระราชบัญญัติกองทุนเงินทดแทน พ.ศ. 2537 โดยจะจัดเก็บจากนายจ้างเพียงฝ่ายเดียว โดยให้ความคุ้มครองแก่ลูกจ้างทั้งประเทศ จำนวน 1.39 ล้านคน ซึ่งข้อมูลมีความซ้ำซ้อนและผิดพลาดบ่อย การสร้างคลังข้อมูลช่วยให้การสอบถามข้อมูลและสร้างรายงาน (Rifaie, Blas, and Muhsen, 2008) และการทำเหมืองข้อมูล (Data Mining) ซึ่งเป็นวิธีการหนึ่งในการวิเคราะห์ข้อมูลที่มีปริมาณมาก (Han, and Kamber, 2006) จากการทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้อง ปรากฏว่ามีนักวิจัยได้นำเทคนิคการทำเหมืองข้อมูลมาใช้ในการพยากรณ์ เช่นงานวิจัยของ (เสกสรรค์ วิลัยลักษณ์ วิภา เจริญภัณฑารักษ์ และดวงดาว วิชาดากุล, 2558) ได้ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (Multilayer Perceptron) มาเปรียบเทียบสร้างตัวแบบพยากรณ์ผลการเรียนของนักเรียน ให้ค่าความถูกต้องสูงที่สุดมีค่าร้อยละ 94.48 ต่อมา

(ปรีชา ลิ้มตระกูล วิภา เจริญภัณฑารักษ์ และวิทยา พรพัชรพงศ์, 2559) ได้ใช้เทคนิคการต้นไม้การตัดสินใจ (Decision Tree) มาทำการทดสอบตัวแบบพยากรณ์ผลผลิตมันสำปะหลัง ซึ่งให้ค่าความถูกต้องสูงถึงร้อยละ 80.12

งานวิจัยนี้จึงได้ศึกษาการพัฒนาคลังข้อมูลและจัดทำรายงานตามหลักการธุรกิจอัจฉริยะ รวมทั้งการพยากรณ์การใช้สิทธิของลูกจ้างในสถานประกอบการเพื่อเพิ่มประสิทธิภาพการบริหารการจัดการกองทุนเงินทดแทน ของสำนักงานประกันสังคม โดยใช้เทคนิคการทำเหมืองข้อมูลด้วยวิธีการจำแนกประเภท (Classification) เพื่อสร้างตัวแบบพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน เพื่อช่วยลดอัตราการประสบอันตรายของลูกจ้าง และใช้จัดตั้งงบประมาณประจำปีของกองทุนเงินทดแทน สำนักงานประกันสังคม ต่อไป

2. วัตถุประสงค์การวิจัย

2.1 การพัฒนาคลังข้อมูลเพื่อวิเคราะห์ข้อมูล และจัดทำรายงาน ตามหลักการธุรกิจอัจฉริยะ

2.2 การพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคม ด้วยอัลกอริทึมการจำแนกประเภท (Classification)

2.3 การประเมินผลการใช้งานคลังข้อมูล

3. ขอบเขตของการวิจัย

3.1 ขอบเขตด้านข้อมูล

3.1.1 การวิจัยในครั้งนี้ นำข้อมูลการใช้สิทธิของลูกจ้างในสถานประกอบการกองทุนเงินทดแทน สำนักงานประกันสังคม เพื่อจัดทำคลังข้อมูล โดยใช้ข้อมูลสำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 1 – 12 ระหว่างปี 2556 – 2557 จำนวน 61,924 ชุดข้อมูล

3.1.2 การวิจัยในครั้งนี้ นำข้อมูลการใช้สิทธิของลูกจ้างในสถานประกอบการกองทุนเงินทดแทน สำนักงานประกันสังคม เพื่อสร้างตัวแบบการพยากรณ์ โดยใช้ข้อมูลเป็น Data Set ของสำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 3 ปี 2557 จำนวน 3,636 ชุดข้อมูล

3.2 ขอบเขตด้านการพัฒนาคลังข้อมูลและธุรกิจอัจฉริยะ

3.2.1 การจัดทำการพัฒนาคลังข้อมูล เพื่อจัดทำคลังข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคม โดยเลือกใช้ข้อมูลการ

ใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคม กรุงเทพมหานครพื้นที่ 3 จำนวน 61,924 ชุดข้อมูล ใช้รูปแบบโครงสร้างสตาร์ (Star Schema) ด้วยโปรแกรม Microsoft SQL Server 2008 R2 และทำกระบวนการอีทีแอล

3.2.2 การจัดทำรายงานรูปแบบหลายมิติ ให้ผู้ใช้งาน ได้แก่ ผู้บริหาร หัวหน้าส่วนงานกองทุนเงินทดแทน เจ้าหน้าที่ปฏิบัติงานด้านกองทุนทดแทน โดยใช้ข้อมูลสำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 1- 12 จำนวน 61,924 ชุดข้อมูล และนำเสนอรายงานผ่าน Pivot Table ด้วยโปรแกรม Microsoft Excel 2007 โดยเชื่อมจากคลังข้อมูล

3.3 ขอบเขตการสร้างตัวแบบพยากรณ์

3.3.1 การสร้างตัวแบบพยากรณ์ เพื่อทำการพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคม ซึ่งใช้ข้อมูลแบบไม่จัดกลุ่ม (Original Data) และชุดข้อมูลแบบจัดกลุ่ม (Clustering Data) แล้วนำไปผ่านกระบวนการคัดเลือกคุณลักษณะ (Feature Selection) ด้วยวิธี Correlation-based Feature Selection และใช้เทคนิคเหมืองข้อมูลตัวแบบการพยากรณ์ 3 เทคนิค ได้แก่ ตัวแบบพยากรณ์โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ตัวแบบพยากรณ์จำแนกแบบเบย์ (Naïve Bayes) และตัวแบบพยากรณ์ต้นไม้ตัดสินใจ (Decision Tree) โดยใช้อัลกอริทึม J48

3.3.2 การวัดประสิทธิภาพการตัวแบบพยากรณ์ ด้วยการทดสอบประสิทธิภาพแบบ 10-fold Cross Validation ด้วยโปรแกรม WEKA 3.8.0

3.5 ขอบเขตด้านการประเมินผลการใช้งานคลังข้อมูล

การประเมินผลการใช้งานระบบคลังข้อมูล เพื่อให้ทราบถึงความพึงพอใจของผู้ใช้งานระบบคลังข้อมูล ได้แก่ ผู้บริหาร หัวหน้าส่วนงานกองทุนเงินทดแทน เจ้าหน้าที่ปฏิบัติงานด้านกองทุนทดแทน โดยใช้แบบสอบถามในการประเมินผลการใช้งานคลังข้อมูล ในการประเมินทั้งหมด 5 ด้าน ได้แก่ 1) มีความสะดวกและง่ายต่อการใช้งาน 2) การออกแบบรายงานเข้าใจได้ง่าย 3) ความถูกต้องของข้อมูล 4) ช่วยลดขั้นตอนในการทำงาน และ 5) ข้อมูลตรงกับความต้องการผู้ใช้งาน

4. นิยามศัพท์เฉพาะ

4.1 คลังข้อมูล (data warehouse) หมายความว่า ฐานข้อมูลขนาดใหญ่ ทีเก็บรวบรวมฐานข้อมูลจากหลายแหล่งหลายช่วงเวลา ซึ่งอาจมีโครงสร้าง (schema) ที่แตกต่างกัน มารวมไว้ในที่เดียวกัน

4.2 เหมืองข้อมูล (data mining) หมายความว่า การค้นหาความรู้ในฐานข้อมูล (Knowledge Discovery in Databases - KDD) เป็นเทคนิคเพื่อค้นหารูปแบบ (pattern) จากข้อมูลจำนวนมหาศาลโดยอัตโนมัติ โดยใช้ขั้นตอนวิธีจากวิชาสถิติ การเรียนรู้ของเครื่อง หรือในอีกนิยามหนึ่ง การทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับข้อมูล (โดยส่วนใหญ่จะมีจำนวนมาก) เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ

4.3 ต้นไม้ตัดสินใจ (Decision Tree) เป็นโครงสร้างข้อมูลชนิดลำดับชั้น (hierarchy) ใช้สนับสนุนการตัดสินใจ โดยจะมีลักษณะคล้ายต้นไม้จริงกลับหัวที่มีโหนดรากอยู่ด้านบนสุดและโหนดใบอยู่ล่างสุดของต้นไม้ ภายในต้นไม้จะประกอบไปด้วยโหนด (node) ซึ่งแต่ละโหนดจะมีคุณลักษณะ (attribute) เป็นตัวทดสอบกิ่งของต้นไม้ (branch) แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบ และใบ (leaf) ซึ่งเป็นสิ่งที่อยู่ล่างสุดของต้นไม้ตัดสินใจแสดงถึงกลุ่มของข้อมูล (class) หรือนั่นก็คือผลลัพธ์ที่ได้จากการทำนาย โหนดที่อยู่บนสุดของต้นไม้เรียกว่าโหนดราก (root node)

4.4 การจำแนกประเภทข้อมูล (Data classification) หากฎเพื่อระบุประเภทของวัตถุจากคุณสมบัติของวัตถุ เช่น หาความสัมพันธ์ระหว่างผลการเรียนต่าง ๆ เพื่อดูว่าผลการเรียนของนักเรียนมีแนวโน้มระดับใด และสามารถเลือกเรียนสายวิชาไหนกับผลคะแนน ส่วนในทางธุรกิจจะใช้เพื่อดูคุณสมบัติของผู้ที่จะก่อหนี้ดีหรือหนี้เสีย เพื่อประกอบการพิจารณาการอนุมัติเงินกู้

4.5 ธุรกิจอัจฉริยะ (Business Intelligence: BI) หมายความว่า เป็นเทคนิคที่ใช้คอมพิวเตอร์ในการวิเคราะห์ ดึงข้อมูลที่โดดเด่น ตัวอย่างเช่น ยอดขาย หรือ สินค้า จากหน่วยใดหน่วยหนึ่งมาแสดงผล โดยใช้รูปแบบการวิเคราะห์ข้อมูล และนำเสนอข้อมูลที่จะช่วยให้ผู้บริหารสามารถตัดสินใจได้เร็วขึ้น มีประสิทธิภาพมากขึ้น

4.6 อัลกอริทึม (Algorithm) หมายความว่า กระบวนการแก้ปัญหาที่สามารถอธิบายออกมาเป็นขั้นตอนที่ชัดเจน เมื่อนำเข้าอะไร แล้วจะต้องได้ผลลัพธ์เช่นไร กระบวนการนี้ประกอบด้วยจะประกอบด้วย วิธีการเป็นขั้นๆ และมีส่วนที่ต้องทำแบบวนซ้ำอีก จนกระทั่งเสร็จสิ้นการทำงาน อัลกอริทึม (Algorithm) ไม่ใช่คำตอบแต่เป็นชุดคำสั่งที่ทำให้ได้คำตอบ

4.7 สถานประกอบการ หมายถึง สถานที่ที่ประกอบธุรกิจ อุตสาหกรรม และบริการที่มีลูกจ้างตั้งแต่ 1 คนขึ้นไป ณ วันขึ้นทะเบียนประกันสังคม ยกเว้นกิจการเพาะปลูก ประมง ป่าไม้ และเลี้ยงสัตว์ซึ่งมิใช่ลูกจ้างตลอดปีและไม่มีงานลักษณะอื่นรวมอยู่ด้วย จำนวนสถานประกอบการจะนับรวมสำนักงานใหญ่และสำนักงานสาขา (กรณีที่แยกเป็นรายจังหวัดจะคำนึงถึงสถานที่ตั้งของสถานประกอบการเป็นหลัก)

4.8 นายจ้าง หมายถึง ผู้ซึ่งรับลูกจ้างเข้าทำงานโดยจ่าค่าจ้าง และให้หมายความรวมถึงผู้ซึ่งได้รับมอบหมายให้ทำงานแทนนายจ้างในกรณีที่นายจ้างเป็นนิติบุคคลให้หมายความ รวมถึงผู้มีอำนาจกระทำการแทนนิติบุคคล และผู้ซึ่งได้รับมอบหมายจากผู้มีอำนาจกระทำการแทนนิติบุคคลให้ทำการแทนด้วย

4.9 ลูกจ้าง หมายความว่า ผู้ซึ่งทำงานให้นายจ้างโดยรับค่าจ้าง

4.10 เงินสมทบ หมายความว่า เงินที่นายจ้างจ่ายสมทบเข้ากองทุนเงินทดแทนเพื่อใช้เป็นเงินทดแทนให้แก่ลูกจ้าง ตามพระราชบัญญัติประกันสังคม พ.ศ. 2537

4.11 เงินทดแทน หมายความว่า เงินที่จ่ายเป็นค่าทดแทน ค่ารักษาพยาบาล ค่าฟื้นฟูสมรรถภาพในการทำงาน และค่าทำศพ

4.12 ค่าทดแทน หมายความว่า เงินที่จ่ายให้ลูกจ้างหรือผู้มีสิทธิตามมาตรา 20 สำหรับการประสบอันตรายหรือเจ็บป่วยหรือสูญหายของลูกจ้างตามพระราชบัญญัติกองทุนเงินทดแทน พ.ศ. 2537

4.13 ค่ารักษาพยาบาล หมายความว่า ค่าใช้จ่ายเกี่ยวกับการตรวจ การรักษาการพยาบาล และค่าใช้จ่ายอื่นที่จำเป็นเพื่อให้ผลของการประสบอันตรายหรือการเจ็บป่วยบรรเทาหรือหมดสิ้นไป และหมายความรวมถึงค่าใช้จ่ายเกี่ยวกับอุปกรณ์ เครื่องใช้ หรือวัตถุที่ใช้แทนหรือทำหน้าที่แทนหรือช่วยอวัยวะที่ประสบอันตรายด้วย

4.14 ค่าฟื้นฟูสมรรถภาพในการทำงาน หมายความว่า ค่าใช้จ่ายที่จำเป็นเกี่ยวกับการฟื้นฟูสมรรถภาพในการทำงาน

4.15 ค่าทำศพ หมายความว่า ค่าใช้จ่ายเกี่ยวกับการจัดการศพของลูกจ้างตามประเพณีทางศาสนาของลูกจ้างหรือตามประเพณีแห่งท้องถิ่น ในกรณีที่ลูกจ้างถึงแก่ความตายเนื่องจากประสบอันตรายหรือเจ็บป่วยหรือสูญหาย

5. ประโยชน์ที่คาดว่าจะได้รับ

5.1 ทำให้มีระบบคลังข้อมูลสำหรับการจัดเก็บ วิเคราะห์ข้อมูลกองทุนเงินทดแทน สำนักงานประกันสังคม ที่เป็นระบบในการบริหารจัดการและปฏิบัติงาน

5.2 ทำให้มีตัวแบบพยากรณ์การใช้สิทธิของลูกจ้างที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคม ซึ่งช่วยให้ประมาณการจัดตั้งงบประมาณในการชดเชยให้กับลูกจ้างของสถานประกอบการ

5.3 ช่วยลดขั้นตอนในการปฏิบัติงานของผู้ที่เกี่ยวข้องด้านกองทุนเงินทดแทน สำนักงานประกันสังคม

5.4 ช่วยลดค่าใช้จ่ายและทรัพยากรในการปฏิบัติงานด้านกองทุนเงินทดแทน สำนักงานประกันสังคม อีกทั้งข้อมูลยังมีความถูกต้องตรงตามความต้องการของผู้ใช้งาน

บทที่ 2
วรรณกรรมที่เกี่ยวข้อง

การวิจัยครั้งนี้ผู้วิจัย ได้ทำการศึกษาและค้นคว้าข้อมูลเกี่ยวกับทฤษฎี และงานวิจัยที่เกี่ยวข้องจากเอกสาร บทความวิชาการต่างๆ เพื่อกำหนดกรอบแนวคิดที่จะใช้ในการศึกษา ดังนี้

1. สภาพทั่วไปของสำนักงานประกันสังคม
2. กฎหมายสำนักงานประกันสังคม กระทรวงแรงงาน
3. คลังข้อมูล
4. การทำเหมืองข้อมูล
5. ธุรกิจอัจฉริยะ
6. งานวิจัยที่เกี่ยวข้อง

1. สภาพทั่วไปของสำนักงานประกันสังคม

ยุทธศาสตร์สำนักงานประกันสังคม ระยะ 5 ปี พ.ศ. 2558-2562

วิสัยทัศน์ (ปี 2558-2562)

" ก้าวสู่องค์กรประกันสังคมชั้นนำในภูมิภาคอาเซียน "

พันธกิจ

ภารกิจหลัก ปรากฏตามกฎกระทรวง “การบริหารการประกันสังคมและเงินทดแทน โดย การจัดการที่มีประสิทธิภาพ เพื่อให้แรงงานมีหลักประกันการดำรงชีวิตที่มั่นคง ”

ประเด็นยุทธศาสตร์

ยุทธศาสตร์ที่ 1 การยกระดับการให้บริการสู่สากล ในการสนองความต้องการ และความคาดหวังของผู้มีส่วนได้ส่วนเสียในทุกกลุ่ม ควบคู่กับสร้างการรับรู้ และเข้าใจในคุณค่าของการประกันสังคม บนรากฐานของความเป็นธรรมและสมเหตุผลที่ทุกภาคส่วนร่วมเป็นเจ้าของ

ยุทธศาสตร์ที่ 2 การมุ่งพัฒนา สิทธิประโยชน์ ประโยชน์ทดแทน ให้มีความเป็นสากล รองรับการก้าวสู่การประกันสังคมนานาชาติ บนบรรทัดฐานของการพัฒนา ระบบเงินสมทบและระบบการลงทุน ที่มีศักยภาพสูง เพื่อให้สามารถรองรับ กับพัฒนาการได้อย่างมั่นคง มีเสถียรภาพ และเป็นที่เชื่อถือ เชื่อมั่นได้อย่างแท้จริง

ยุทธศาสตร์ที่ 3 การรื้อปรับและยกระดับระบบเทคโนโลยีสารสนเทศ IT และการบริหารสารสนเทศ MIT ของสำนักงานประกันสังคมสู่องค์กรที่มีการดำเนินการและการบริการที่ล้ำสมัย

ยุทธศาตร์ที่ 4 การพัฒนาการบริหารจัดการ “ทุนมนุษย์” ได้อย่างเป็นรูปธรรม และสอดคล้องกับ ทิศทางของโลกในปัจจุบันและอนาคต

ยุทธศาตร์ที่ 5 การพัฒนาองค์กรสู่องค์กรที่มีประสิทธิภาพสูง (HPO) ทันสมัย คล่องตัว โปร่งใส และมีธรรมาภิบาล

ยุทธศาสตร์ ASEAN ระบบประกันสังคมไทยร่วมเป็นส่วนสำคัญของการวางรากฐานระบบประกันสังคมอาเซียนได้อย่างโดดเด่น

โครงสร้างสำนักงานประกันสังคม

ภาพที่ 2.1 โครงสร้างสำนักงานประกันสังคม

สำนักงานประกันสังคม มีภารกิจที่สำคัญในการบริหารกองทุนเงินทดแทนและกองทุนประกันสังคมให้มีเสถียรภาพ โดยการจัดการที่มีประสิทธิภาพ ควบคู่ไปกับการบริการที่เป็นเลิศ เพื่อให้แรงงานมีหลักประกันการดำรงชีวิตที่มั่นคง ซึ่งภาระหน้าที่ คือ การเพิ่มประสิทธิภาพและ

ศักยภาพใน การจัดเก็บเงินสมทบและการมุ่งพัฒนาสิทธิประโยชน์ให้แก่ผู้ประกันตน ลูกจ้าง และนายจ้าง โดยกำหนดให้สำนักงานประกันสังคม มีภารกิจเกี่ยวกับการบริหารการประกันสังคมและเงินทดแทน โดยการจัดการที่มีประสิทธิภาพ ซึ่งแบ่งส่วนราชการเป็น ราชการบริหารส่วนกลาง และราชการบริหารส่วนภูมิภาค

ราชการบริหารส่วนกลาง ประกอบด้วย กลุ่มตรวจสอบภายใน กลุ่มพัฒนาระบบบริหาร สำนักเงินสมทบ สำนักจัดระบบบริการทางการแพทย์ กองบริหารการเงินและบัญชี สำนักเทคโนโลยีสารสนเทศ สำนักสิทธิประโยชน์ สำนักเลขานุการกรม กองบริหารทรัพยากรบุคคล กองกฎหมาย กองนโยบายและแผน กองคลัง กองบริหารการลงทุน สำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 1-12 สำนักเสริมสร้างความมั่นคงแรงงานนอกระบบ ศูนย์ฟื้นฟูสมรรถภาพคนงานประจำภาค 1-5 โดยการบริหารงานภายใต้ เลขานุการสำนักงานประกันสังคม และราชการบริหารส่วนภูมิภาค ประกอบด้วย สำนักงานประกันสังคมจังหวัด (76 จังหวัด) สำนักงานประกันสังคมจังหวัดสาขา (49 สาขา) โดยการบริหารงานภายใต้ ผู้ว่าราชการจังหวัด

2. กฎหมายสำนักงานประกันสังคม กระทรวงแรงงาน

2.1 พระราชบัญญัติเงินทดแทน พ.ศ. 2537

มาตรา 4 พระราชบัญญัตินี้ไม่ใช้บังคับแก่

(1) ราชการส่วนกลาง ราชการส่วนภูมิภาค และราชการส่วนท้องถิ่น

(2) รัฐวิสาหกิจตามกฎหมายว่าด้วยพนักงานรัฐวิสาหกิจสัมพันธ์

(3) นายจ้างซึ่งประกอบธุรกิจโรงเรียนเอกชนตามกฎหมายว่าด้วยโรงเรียนเอกชน เฉพาะในส่วนที่เกี่ยวกับครูหรือครูใหญ่

(4) นายจ้างซึ่งดำเนินกิจการที่มิได้มีวัตถุประสงค์เพื่อแสวงหากำไรในทางเศรษฐกิจ

(5) นายจ้างอื่นตามที่กำหนดในกฎกระทรวง

มาตรา 5 ในพระราชบัญญัตินี้

“นายจ้าง” หมายความว่า ผู้ซึ่งตกลงรับลูกจ้างเข้าทำงานโดยจ่ายค่าจ้างให้ และหมายความรวมถึงผู้ซึ่งได้รับมอบหมายให้ทำงานแทนนายจ้าง ในกรณีที่นายจ้างเป็นนิติบุคคล ให้หมายความรวมถึง ผู้มีอำนาจกระทำการแทนนิติบุคคลและผู้ซึ่งได้รับมอบหมายจากผู้มีอำนาจกระทำการแทนนิติบุคคลให้ทำการแทนด้วย

“ลูกจ้าง” หมายความว่า ผู้ซึ่งทำงานให้นายจ้างโดยรับค่าจ้างไม่ว่าจะเรียกชื่ออย่างไร แต่ไม่รวมถึงลูกจ้างซึ่งทำงานเกี่ยวกับงานบ้านอันมิได้มีการประกอบธุรกิจรวมอยู่ด้วย

“ค่าจ้าง” หมายความว่า เงินทุกประเภทที่นายจ้างให้แก่ลูกจ้างเป็นค่าตอบแทนการทำงานในวันและเวลาทำงานปกติไม่ว่าจะคำนวณตามระยะเวลา หรือคำนวณตามผลงานที่ลูกจ้างทำได้และให้หมายความรวมถึงเงินที่นายจ้างจ่ายให้แก่ลูกจ้างในวันหยุดและวันลาซึ่งลูกจ้างไม่ได้ทำงานด้วยทั้งนี้ ไม่ว่าจะกำหนด คำนวณ หรือจ่ายในลักษณะใด หรือโดยวิธีใด และไม่ว่าจะเรียกชื่ออย่างไร

“เงินสมทบ” หมายความว่า เงินที่นายจ้างจ่ายสมทบเข้ากองทุนเงินทดแทนเพื่อใช้เป็นเงินทดแทนให้แก่ลูกจ้าง

“กองทุน” หมายความว่า กองทุนเงินทดแทน

3. คลังข้อมูล

3.1 ทฤษฎีคลังข้อมูล

คลังข้อมูล เป็นระบบข้อมูลขนาดใหญ่ที่เอื้ออำนวยให้ผู้ใช้งานสามารถใช้ข้อมูลได้อย่างมีประสิทธิภาพ เป็น ที่เก็บรวบรวมข้อมูลจากหลายแหล่งที่มา หลายช่วงเวลา มีคุณลักษณะของข้อมูลแตกต่างกัน นำมารวมไว้ ณ ที่เดียวกันใช้เพื่อการวิเคราะห์ (กิตติพงษ์ กลมกล่อม, 2552: 5)

คลังข้อมูล (Data warehouse) เป็นการเก็บข้อมูลเชิงหัวข้อ โดยทำการรวบรวมข้อมูลจากแหล่งที่มาต่างๆ ซึ่งจะเก็บข้อมูลเป็นระยะเวลานาน โดยข้อมูลจะไม่เปลี่ยนแปลงบ่อย และใช้ข้อมูลช่วยในการบริหารการสนับสนุนการตัดสินใจการดำเนินงาน (Roiger and Geatz, 2003: 148)

3.2 สถาปัตยกรรมคลังข้อมูล

สถาปัตยกรรมคลังข้อมูล เป็นโครงสร้างมาตาฐานที่ใช้อธิบายแนวความคิดและกระบวนการของคลังข้อมูล โดยทั่วไปแล้วคลังข้อมูลแต่ละระบบจะมีความแตกต่างกันเพื่อปรับให้เหมาะสมกับวัตถุประสงค์ขององค์กรแต่ละแห่ง (เลิศ เลิศศิริโสภณ : 2541) ทั้งนี้ยังมีส่วนประกอบภายในสถาปัตยกรรม ได้แก่

3.2.1 Operational database ทำหน้าที่จัดเก็บข้อมูลภายในระบบปฏิบัติการนอกองค์กร

3.2.2 Information assess layer ผู้ใช้ปลายทางสามารถติดต่อผ่านโดยตรงด้วยฮาร์ดแวร์ ที่ใช้ในการประมวลผล โดยมีเครื่องมือเป็นตัวกลางคอยช่วยติดต่อ ซึ่งสามารถวิเคราะห์ข้อมูลที่มีความซ้ำซ้อน และแสดงข้อมูลในมิติมุมมองได้หลากหลาย

3.2.3 Data access layer เป็นการประสานในส่วนของ Information assess layer กับ Operational layer

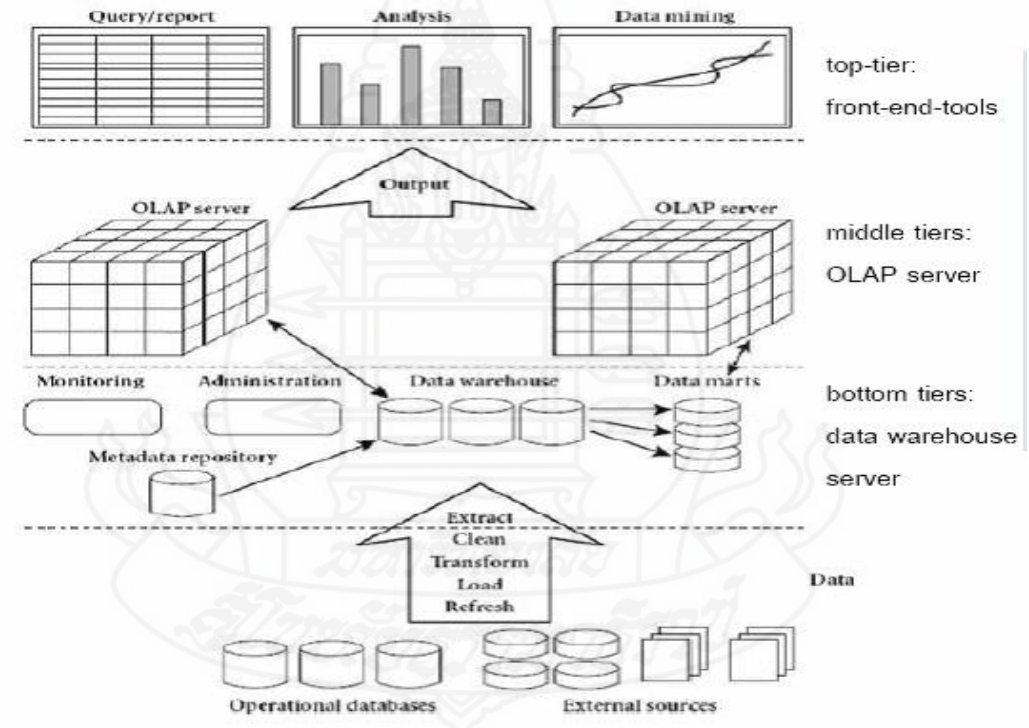
3.2.4 Data director layer (metadata) ส่วนของการเพิ่มความเร็วในการเรียกใช้ข้อมูลจากคลังข้อมูล

3.2.5 Process management layer ส่วนของการจัดการกระบวนการทำงาน

3.2.6 Application messaging layer ทำหน้าที่ของการส่งข้อมูลทางเครือข่ายภายในองค์กร

3.2.7 Data ware house layer (physical) การเข้าถึงข้อมูลจากแหล่งเก็บข้อมูลจาก Information data และ External data ให้สามารถใช้งานได้ง่าย

3.2.8 Data staging layer กระบวนการแก้ไขข้อมูลและดึงข้อมูลจาก External data



ภาพที่ 2.2 สถาปัตยกรรมคลังข้อมูล

ที่มา : http://www.scribd.com/doc/23295033/3-Tier-Achtecture Retrieved August 29, 2016.

3.3 กระบวนการในการพัฒนาคลังข้อมูล

เนื่องจากคลังข้อมูลประกอบด้วยข้อมูลจำนวนมาก และจะต้องมีการออกแบบคลังข้อมูลเพื่อรวบรวม ในการวิเคราะห์ข้อมูลและสิ่งที่สำคัญที่ต้องคำนึงถึง คือปัจจัยในการจัดทำคลังข้อมูล (สุวรรณี อัศวกุลชัย : 2555) ทั้งนี้เนื่องจากในการพัฒนาคลังข้อมูล ประกอบด้วย 7 ขั้นตอนหลักๆ ดังนี้

3.3.1 การศึกษากระบวนการทางธุรกิจ (business process) นับเป็นขั้นตอนแรกของการพัฒนาคลังข้อมูล เพื่อให้ทราบถึงการทำงานของระบบงานปัจจุบัน ทำให้สามารถกำหนดได้ว่าจะเริ่มสร้างระบบงานขั้นตอนใดก่อน โดยการศึกษาการไหลของกระบวนการ (process flow) และการไหลของข้อมูล (data flow) จากนั้นจึงพัฒนาเป็นแบบจำลอง เสมือนเป็นการเตรียมข้อมูลก่อนการจัดเก็บลงคลังข้อมูล กระบวนการรวบรวมเพื่อทำเป็นคลังข้อมูล

3.3.2 การศึกษาความต้องการของผู้ใช้ (user requirement) ซึ่งเป็นหัวใจสำคัญในการพัฒนาคลังข้อมูล ทั้งนี้เนื่องจากถ้าพัฒนาระบบขึ้นมาแล้วแต่ไม่ตรงกับความต้องการของผู้ใช้และไม่มีใครใช้ ก็ถือว่าการพัฒนาระบบไม่ประสบความสำเร็จ

3.3.3 การพัฒนาแบบจำลองข้อมูล (data modeling) ซึ่งถูกจัดเก็บในรูปแบบของฐานข้อมูลเชิงสัมพันธ์ (relational database) ซึ่งมีแหล่งข้อมูลจากหลายแหล่ง รวมถึงการจัดเตรียมข้อมูลจากฐานข้อมูลระบบปฏิบัติการทั่วไปที่ใช้ประจำวัน หรือ ข้อมูลจากทรานแซกชั่นต่างๆ (transection systems) เพื่อเตรียมการจัดเก็บลงในคลังข้อมูล สำหรับแบบจำลองหลายมิติของคลังข้อมูล (multi-dimensional of data warehouse) เป็นข้อมูลที่จัดเก็บในฐานข้อมูลหลายมิติอาจมาจากฐานข้อมูลเชิงสัมพันธ์ เพื่อให้สามารถสืบค้นข้อมูลได้หลากหลายรูปแบบ ประกอบด้วยสตาร์สกีมา (star schema) สโนแฟลกสกีมา (snowflake schema) และแบบหลายตารางข้อเท็จจริง (fact constellations) ซึ่งจะนำข้อมูลจากแบบจำลองเหล่านี้ไปสร้างเป็นคิวบ์ (cube)

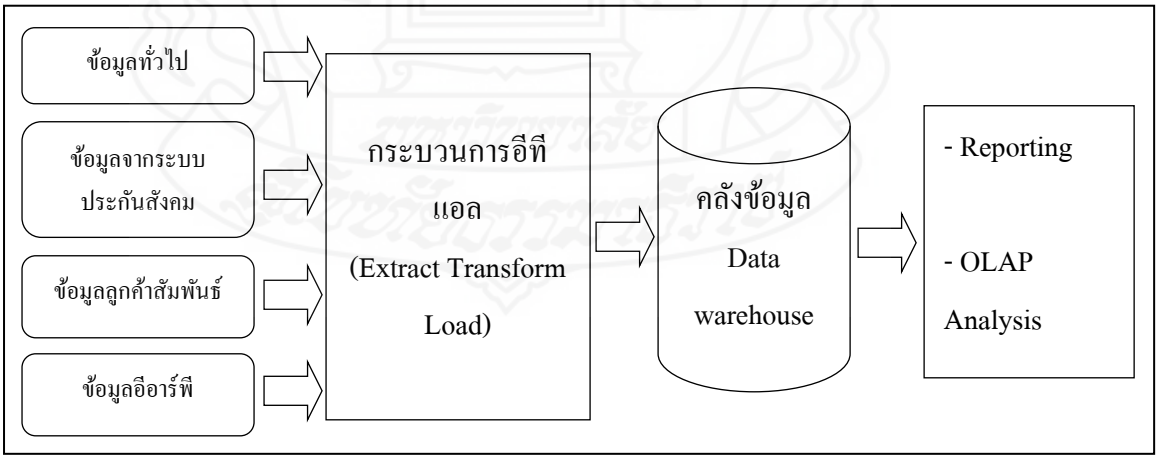
3.3.4 กระบวนการออกแบบการดึงข้อมูลจากแหล่งข้อมูลต่างๆ (data source) การเตรียมพื้นที่พักข้อมูล (staging area) สำหรับข้อมูลที่ดึงออกมาจากฐานข้อมูลระบบปฏิบัติการทั่วไปเพื่อเข้าสู่คลังข้อมูล (data warehouse) แบ่งเป็น 3 ขั้นตอนหลัก ดังนี้ (ชัยเลิศ พิชิตพรชัย, 2555,น. 38)

1) การคัดแยกหรือสกัดข้อมูล (Extract - E) เป็นการนำข้อมูลออกมาจากฐานข้อมูลที่มาจากแหล่งข้อมูลที่แตกต่างกันตามที่ต้องการมาเก็บไว้ในคลังข้อมูล เช่น ฐานข้อมูลปฏิบัติการ ฐานข้อมูลภายนอกองค์กร ขั้นตอนนี้มีการความสำคัญมากเนื่องจากเป็นขั้นตอนที่ดึงข้อมูลที่ถูกต้องมีผลต่อการดำเนินการในขั้นตอนต่อไปให้มีความถูกต้องด้วย การพัฒนาคลังข้อมูลส่วนใหญ่จะดำเนินการโดยการคัดแยกข้อมูลจากฐานข้อมูลปฏิบัติการหลายๆ ระบบที่แตกต่างกัน ซึ่งมักจะมีการจัดรูปแบบข้อมูลที่แตกต่างกัน วัตถุประสงค์หลักของขั้นตอนนี้คือ การดึงข้อมูล

ออกมาจากแหล่งข้อมูลโดยจัดให้อยู่ในรูปแบบที่เหมาะสมสำหรับการจัดรูปแบบ (transform) ต่อไป โดยมากในขั้นตอนนี้มักมีกระบวนการคัดกรองที่มีรูปแบบ (pattern) หรือโครงสร้าง (structure) ตรงตามที่ได้กำหนดไว้ ไม่เช่นนั้นข้อมูลจะถูกตัดทิ้งไป

2) การแปลงข้อมูล (Transform - T) เป็นการนำข้อมูลที่ได้มาจากการคัดแยกมาจัดรูปแบบ หรือเปลี่ยนแปลงรูปแบบของข้อมูลให้ถูกต้องสอดคล้องกันตามโครงสร้างของคลังข้อมูล เช่น การจับคู่ข้อมูล (data mapping) การทำความสะอาดข้อมูล (data cleansing) การตรวจสอบและแก้ไขข้อมูลให้ถูกต้อง การกำจัดข้อมูลที่ผิดพลาดออก การแปลงข้อมูลยังรวมถึงการปรับปรุงรูปแบบของข้อมูลที่สามารถนำไปวิเคราะห์ได้ การแปลงข้อมูลเริ่มจากการวิเคราะห์แหล่งข้อมูล กำหนดการส่งข้อมูลรวบรวมหรือสร้างข้อมูลภายนอก วางแผนและการกำหนดงานหลักของการแปลงข้อมูล และตรวจสอบความถูกต้องของข้อมูล

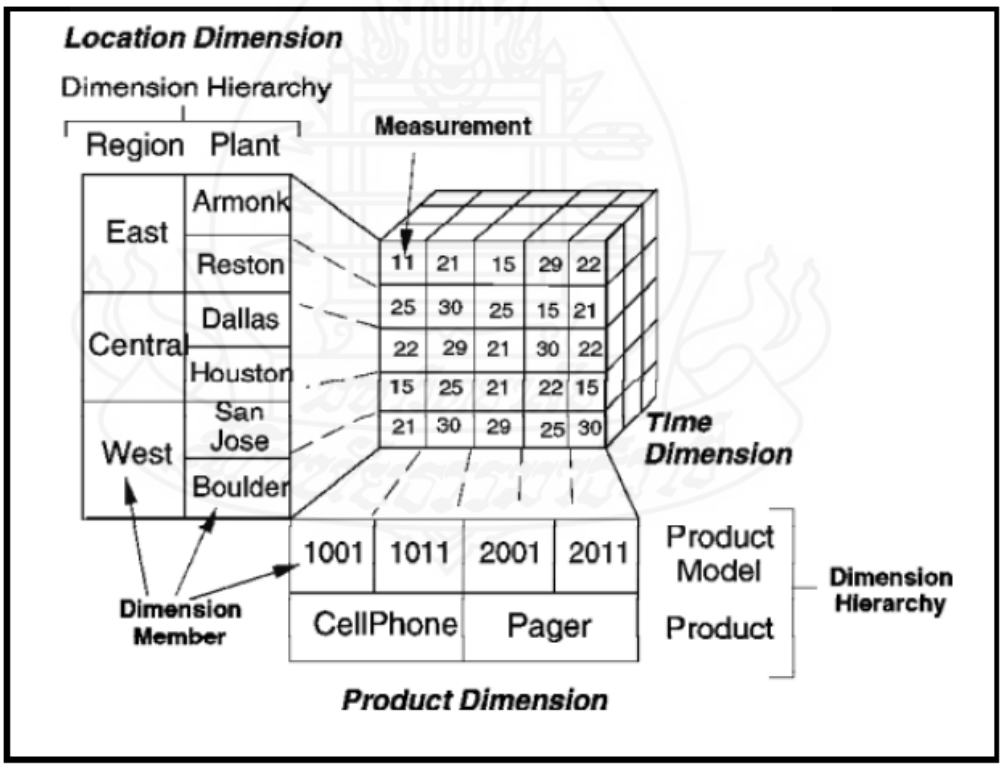
3) การโหลดหรือการนำเข้าข้อมูล (Load - L) เป็นการจัดเก็บข้อมูล หรือการนำข้อมูลที่ผ่านการแปลงข้อมูลให้อยู่ในรูปแบบที่ถูกต้องเข้าสู่คลังข้อมูล (data warehouse) ซึ่งอาจมีการกำหนดช่วงเวลา ความถี่ในการโหลดข้อมูล ในบางคลังข้อมูลจะโหลดข้อมูลทับข้อมูลเดิม ความถี่ในการโหลดข้อมูลอาจกำหนดเป็นทุกชั่วโมง ทุกวัน ทุกสัปดาห์ ทุกเดือน ทุกไตรมาส ในบางครั้งคลังข้อมูลอาจจะโหลดข้อมูลแบบต่อท้ายเก็บข้อมูลเดิมไว้เป็นประวัติข้อมูล ซึ่งอาจกำหนดให้โหลดทุกชั่วโมง ในขณะที่กำลังทำการโหลดข้อมูลจะมีการติดต่อกับฐานข้อมูลของคลังข้อมูล ข้อจำกัด (Constraints) จะกำหนดในสกีมาของฐานข้อมูล หรือกำหนดที่ทริกเกอร์ (trigger) ที่กำหนดไว้ในขณะที่มีการโหลด ซึ่งจะมีผลต่อคุณภาพของข้อมูล



ภาพที่ 2.3 กระบวนการ Extract Transform Load

3.3.5 การศึกษาวิเคราะห์ ออกแบบ และจัดทำระบบคลังข้อมูลกลาง (data warehouse) รวมทั้งนำข้อมูลเข้า เพื่อจัดเก็บข้อมูลต่างๆ ประกอบด้วยหลายๆ ดาต้ามาร์ท (data mart) จากฐานข้อมูลของหน่วยงานเป็นประโยชน์ต่อการวิเคราะห์ตัดสินใจ และการใช้ข้อมูลต่างๆ โดยกำหนดส่วนย่อยที่สุดในคลังข้อมูล (granularity) การออกแบบคุณลักษณะของตารางมิติ (dimension table) รวมถึงการออกแบบตารางข้อเท็จจริง (fact table)

3.3.6 การแสดงรายงานการประเมินผลเชิงวิเคราะห์ออนไลน์ (Online Analytic Process : OLAP) จากเหมืองข้อมูล (data mining) เพื่อแสดงในรูปแบบข้อมูลที่เห็นเป็นภาพ (data visualization) หรือในรูปแบบธุรกิจอัจฉริยะ (Business Intelligence – BI) เครื่องมือที่ช่วยดึงและนำเสนอข้อมูลในหลายมิติ (Multidimensional) ที่ประกอบไปด้วยมิติ (Dimension) และตารางข้อเท็จจริง (Fact Table) ในตารางข้อเท็จจริงก็จะประกอบไปด้วยเมเซอร์ (Measure) เป็นข้อมูลที่ต้องการวัดในเชิงปริมาณ กับข้อมูลที่ใช้เชื่อมความสัมพันธ์กับมิติ โดยที่ข้อมูลจะถูกเก็บอยู่ในรูปแบบของลูกบาศก์ที่มีหลายมิติเรียกว่า คิวบ์ (Cube) เพื่อช่วยประกอบการตัดสินใจในระดับสูง ดังภาพที่ 2.4



ภาพที่ 2.4 ลักษณะของคิวบ์ (Club)

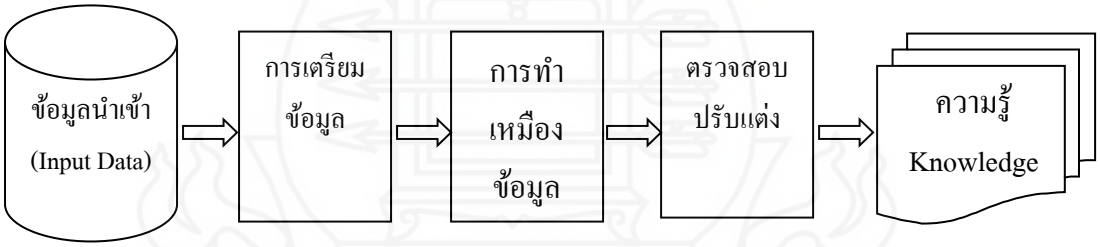
ที่มา: Ballard and others (1998, pp. 44)

3.3.7 การตรวจสอบคลังข้อมูล (data warehouse testing) เป็นขั้นตอนสำคัญ ถ้าไม่มีการตรวจสอบก็อาจทำให้เกิดความผิดพลาดเกิดขึ้น เช่น ได้คำตอบที่ไม่ถูกต้อง ที่นำไปสู่การสูญเสียทางธุรกิจอย่างมากนั่นเอง ดังนั้นการตรวจสอบคลังข้อมูลจึงจำเป็นต้องดำเนินการอย่างมีประสิทธิภาพ ซึ่งประกอบด้วยเทคโนโลยี กระบวนการ และบุคคลกรที่ถูกต้อง

4. การทำเหมืองข้อมูล

เทคโนโลยีสำหรับช่วยในการวิเคราะห์ที่มีประสิทธิภาพสูง ถูกใช้อย่างแพร่หลายในองค์กร ซึ่งระบบจะทำหน้าที่ค้นหาลักษณะที่มีความน่าสนใจและคล้ายกัน ตลอดจนวิเคราะห์ความสัมพันธ์ของข้อมูล และได้ความสัมพันธ์ในแต่ละกลุ่มที่เชื่อมโยงกัน (Nicholson : 2003)

การทำเหมืองข้อมูลต้องอาศัยวิธีการทางคอมพิวเตอร์ร่วมกับวิธีการทางสถิติ ในการสร้างตัวแบบพยากรณ์หรือแบบจำลองโมเดล (Modelling) ด้วยวิธีการจัดกลุ่ม (Cluster) การค้นหาความสัมพันธ์ (Association) การพยากรณ์ (Classification) การสร้างแบบจำลองต้องมีกฎเกณฑ์ในการดำเนินการเพื่อได้แบบพยากรณ์หรือแบบจำลองที่ง่ายและไม่ซับซ้อนต่อการใช้งาน โดยจะนำข้อมูลขนาดใหญ่มาทำการค้นหาความรู้ใหม่ ด้วยอัลกอริทึมต่างๆ ที่มีให้เลือกใช้หลากหลาย ดังภาพที่ 2.5



ภาพที่ 2.5 การค้นหาความรู้ด้วยการทำเหมืองข้อมูล

4.1 ประเภทของการทำเหมืองข้อมูล

การค้นหารูปแบบของข้อมูลเกิดขึ้น โดยแบ่งตามลักษณะงาน ได้ดังนี้ (วิภา เจริญภัณฑรักษ์, น. 2555)

4.1.1 การหาลักษณะของข้อมูลที่ตรงกัน (Data Characterization) การค้นหาลักษณะจากข้อมูลที่เป็นกลุ่มเป้าหมาย เพื่อหาลักษณะโดยดูจากพฤติกรรมของลูกค้า

4.1.2 การค้นหาความสัมพันธ์ (Association) การหาความสัมพันธ์ของข้อมูลที่ต้องการเพื่อให้ได้ความสัมพันธ์ที่เกิดขึ้นคู่กัน โดยอาศัยการสร้างกฎความสัมพันธ์ที่เกิดขึ้น

4.1.3 การจำแนกประเภทและการพยากรณ์ ในการทำเหมืองข้อมูลประเภทนี้สามารถแบ่งได้ดังนี้

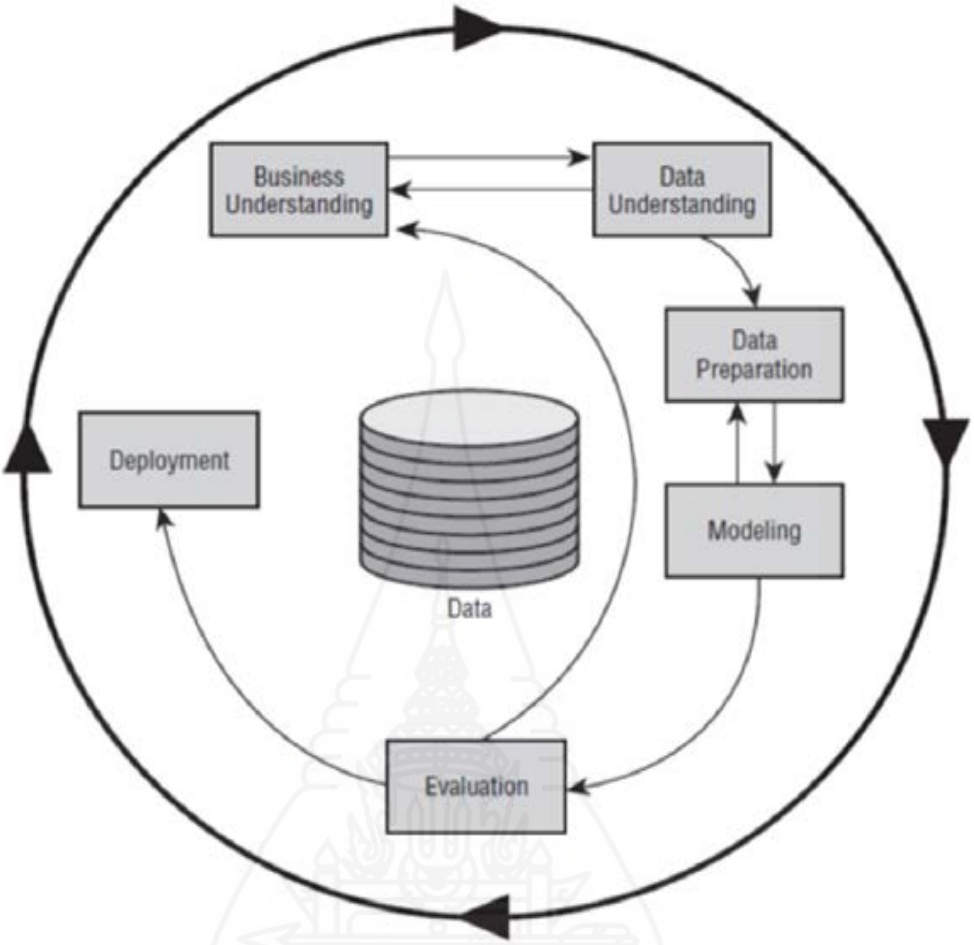
1) การจำแนกประเภท (Classification) เป็นการจำแนกข้อมูลที่ต้องการว่าอยู่ประเภทไหน โดยอาศัยแบบจำลองหรือตัวแบบพยากรณ์ที่สร้างขึ้น และใช้ข้อมูลในการทดสอบตัวแบบพยากรณ์ที่ได้สร้างขึ้น เพื่อใช้เป็นข้อมูลในการตัดสินใจ

2) การพยากรณ์ (Prediction) เป็นข้อมูลประเภทที่มีค่าต่อเนื่องหรือข้อมูลที่มีการลำดับการเกิดของข้อมูล เช่น ข้อมูลอายุ เงินเดือน เวลา เป็นต้น แล้วนำมาทำการสร้างตัวแบบพยากรณ์และทดสอบประสิทธิภาพ

4.1.4 การจัดกลุ่ม คือ การทำเหมืองประเภทข้อมูลแบบจัดกลุ่ม (Clustering) เป็นเทคนิคการจัดกลุ่มข้อมูลที่มีความคล้ายคลึงกันมากที่สุด โดยไม่ได้กำหนดกลุ่มไว้ล่วงหน้า เพื่อจำแนกข้อมูลในแต่ละหน่วยชุดข้อมูล โดยจะมีค่าความห่างจากจุดศูนย์กลางและการวัดค่าเบี่ยงเบนเป็นการรวมหน่วยที่คล้ายคลึงกันมากที่สุดไว้เป็นกลุ่มเดียวกัน

4.2 ขั้นตอนการทำเหมืองข้อมูล

กระบวนการในการทำเหมืองข้อมูลมีกระบวนการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล ด้วยกระบวนการมาตรฐานที่เรียกว่า “Cross-Industry Standard Process for Data Mining” หรือเรียกย่อๆ ว่า “CRISP-DM Model” ซึ่งมีกระบวนการทำงานดังภาพที่ 2.6



ภาพที่ 2.6 กระบวนการ CRISP-DM Model

ที่มา: Umair Shafique and Haseeb Qaiser (2014: pp.219)

4.2.1 การวิเคราะห์ปัญหาและโอกาสทางธุรกิจ (Business Understanding) เป็นขั้นตอนการทำความเข้าใจ ระบุปัญหาหรือโอกาสในเชิงธุรกิจ เพื่อหารูปแบบที่เหมาะสมต่อการนำมาวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูล

4.2.2 การทำความเข้าใจข้อมูล (Data Understanding) เป็นขั้นตอนการรวบรวมข้อมูล ที่เกี่ยวข้อง จากแหล่งข้อมูลที่ถูกต้องน่าเชื่อถือ ซึ่งข้อมูลที่ได้มีปริมาณมากพอ และเป็นข้อมูลที่เหมาะสมเพียงพอต่อการนำไปใช้ในวิเคราะห์ข้อมูล

4.2.3 ขั้นตอนการเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนที่สำคัญและใช้เวลามากที่สุด การทำไมน์นิ่งจะได้ผลดีหรือไม่ขึ้นอยู่กับขั้นตอนนี้การเตรียมข้อมูล แบ่งการทำงานออกเป็น 3 ขั้นตอน ได้แก่

4.2.4 การสร้างโมเดล (Modeling) เป็นขั้นตอนการนำเทคนิคเหมืองข้อมูลมาใช้ในการวิเคราะห์ข้อมูล เพื่อสร้างโมเดลให้ได้ผลลัพธ์ที่ดีที่สุด

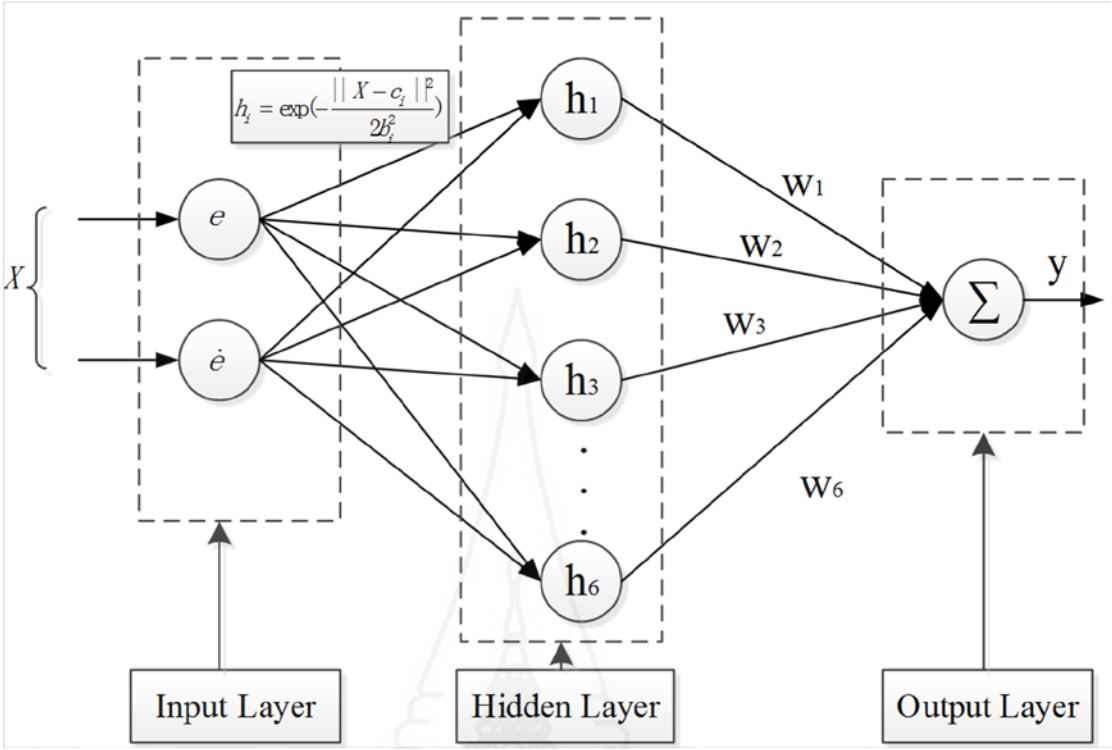
4.2.5 การวัดประสิทธิภาพ (Evaluation) เป็นขั้นตอนการประเมินหรือวัดประสิทธิภาพของโมเดลวิเคราะห์ข้อมูลในขั้นตอนก่อนหน้านั้น เพื่อวัดความน่าเชื่อถือของโมเดล

4.2.6 การนำโมเดลที่พัฒนาไปใช้งาน (Deployment) เป็นขั้นตอนในการนำโมเดลที่ได้จากการวิเคราะห์ข้อมูลไปใช้งานจริง เพื่อช่วยในการตัดสินใจ

4.3 เทคนิคการทำเหมืองข้อมูล

4.3.1 เทคนิคโครงข่ายประสาทเทียม (Neural Networks) มีลักษณะคล้ายกับการส่งสัญญาณประสาทของสมองมนุษย์ โครงข่ายประสาทเทียมจะประกอบด้วยเซลล์ประสาทเทียมหรือโหนด (Node) จำนวนมากเชื่อมต่อกัน ในการเชื่อมต่อนั้นแบ่งออกเป็นกลุ่มย่อย เรียกว่า ชั้น (Layer) ชั้นแรกเป็นชั้นข้อมูลนำเข้าบางที่เรียกว่า ชั้นรับข้อมูล (Input layer) ส่วนชั้นสุดท้าย เรียกว่า ชั้นส่งข้อมูลออก (Output layer) และชั้นที่อยู่ระหว่างชั้นรับข้อมูลและขั้นส่งข้อมูลออก เรียกว่า ชั้นซ่อนหรือแอบแฝง (Hidden layer) โครงข่ายประสาทเทียมที่มีโครงข่ายแบบหลายชั้น (Multilayer) จะเป็นโครงข่ายที่มีชั้นแอบแฝงตั้งแต่หนึ่งชั้นขึ้นไป โครงข่ายแบบหลายชั้นจะใช้เมื่อมีปัญหาความซับซ้อน เพื่อเพิ่มจำนวนโหนดที่มีการคำนวณ ลักษณะของโครงข่ายแบบหลายชั้นเป็น ดังภาพที่ 2.7

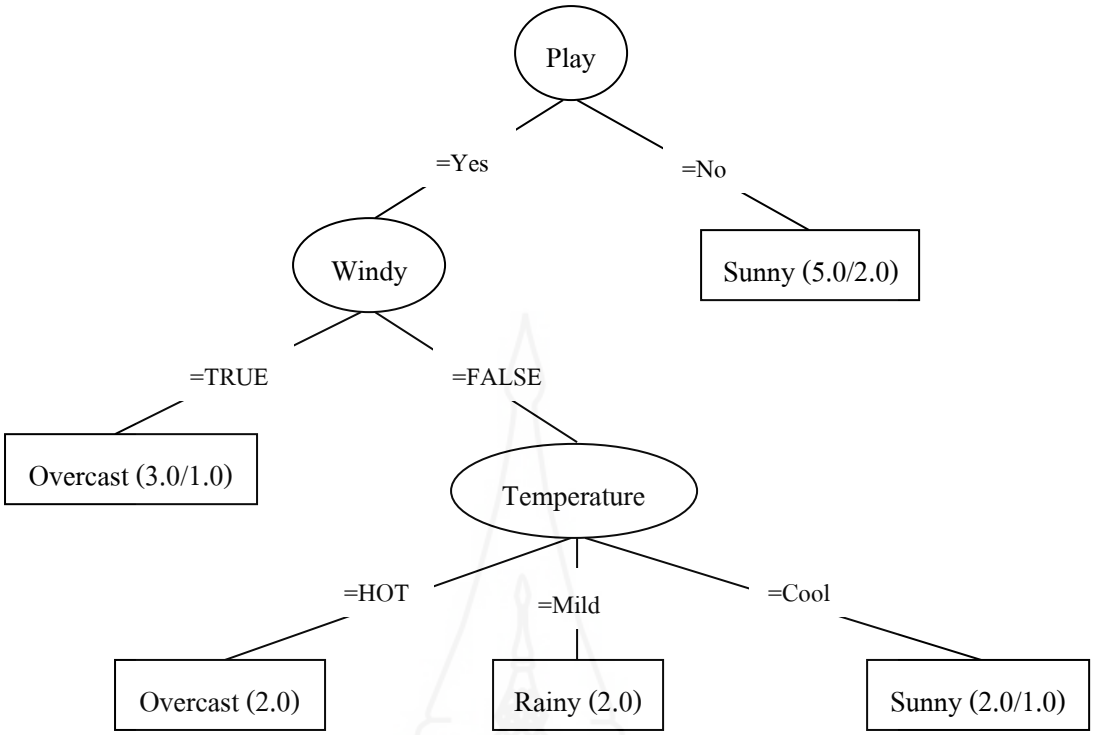
โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (Multilayer Perceptron: MLP) เป็นรูปแบบหนึ่งของโครงข่ายประสาทเทียมที่มีโครงสร้างเป็นแบบชั้นใช้สำหรับงานที่มีความซับซ้อนซึ่งได้ผลเป็นอย่างดี โดยมีกระบวนการฝึกฝนเป็นแบบ Supervise และใช้ขั้นตอนการส่งค่าย้อนกลับ (Back propagation) สำหรับในการฝึกฝนจะประกอบด้วย 2 ส่วนย่อยคือ การส่งผ่านไปข้างหน้า (Forward Pass) และการส่งผ่านย้อนกลับ (Backward Pass) สำหรับการส่งผ่านไปข้างหน้าข้อมูลจะผ่านเข้าชั้นของข้อมูลเข้า และจะส่งผ่านจากอีกชั้นหนึ่งไปสู่อีกชั้นหนึ่งจนกระทั่งถึงชั้นข้อมูลออก ส่วนการส่งย้อนกลับค่าน้ำหนักการเชื่อมต่อจะถูกปรับเปลี่ยนให้สอดคล้องกับกฎการแก้ข้อผิดพลาด คือ ผลต่างของผลตอบที่แท้จริงกับผลตอบเป้าหมายเกิดเป็นสัญญาณผิดพลาด ซึ่งจะถูกส่งย้อนกลับเข้าสู่โครงข่ายประสาทเทียมในทิศทางตรงกันข้ามกับการเชื่อมต่อ ค่าน้ำหนักการเชื่อมต่อจะถูกปรับจนกระทั่งผลตอบที่แท้จริงเข้าใกล้ผลตอบเป้าหมาย



ภาพที่ 2.7 ตัวอย่างเทคนิคการทำเหมืองข้อมูลด้วยวิธีการจำแนกประเภท

ที่มา: Han et al., (2006)

4.3.2 เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคต้นไม้ตัดสินใจมีโครงสร้างคล้ายต้นไม้ จะประกอบด้วยโหนดในการตัดสินใจที่แสดงถึงข้อกำหนดของตัวแปรและเชื่อมต่อกันด้วยกิ่งก้าน (Branch) โดยมีโครงสร้างความสัมพันธ์เริ่มจากโหนดราก (Root node) อยู่ด้านบนสุดของแผนภูมิต้นไม้ และเชื่อมต่อโหนดในการตัดสินใจโดยเลือกแอตทริบิวต์ที่มีผลต่อการตัดสินใจแบ่งประเภทของข้อมูลไปยังกิ่งก้านต่างๆ เพื่อเชื่อมไปยังโหนดใบ (Leaf node) หรือเรียกว่า โหนดลูกที่จะแสดงคลาสของข้อมูล ดังในภาพที่ 2.8

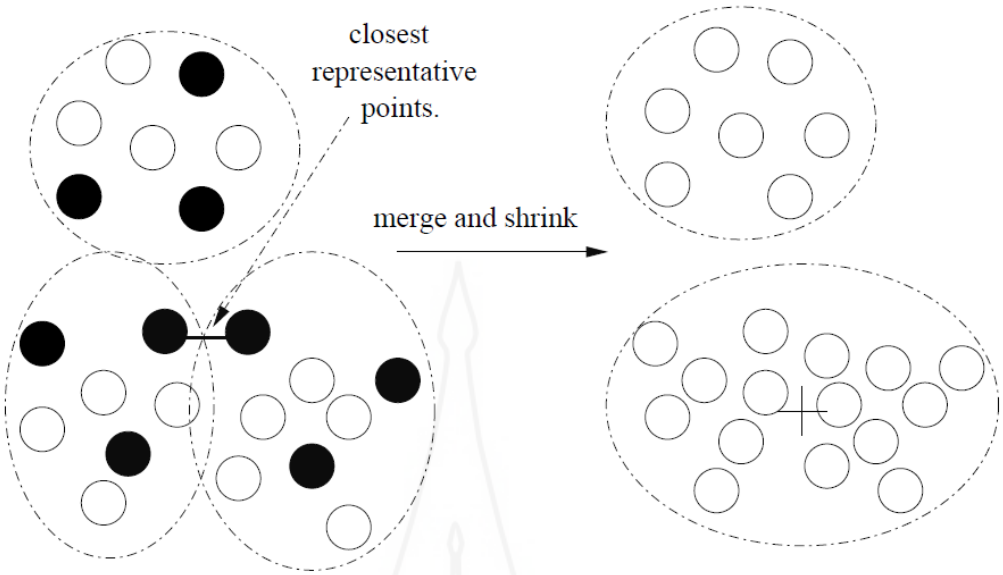


ภาพที่ 2.8 ภาพต้นไม้ตัดสินใจ

ที่มา : http://www.cs.waikato.ac.nz/~ml/weka/ Retrieved August 29, 2016.

4.3.3 เทคนิคการจัดกลุ่ม (Clustering) เป็นการทำเหมืองข้อมูลในรูปแบบของการจัดกลุ่ม (Clustering) ไม่ได้กำหนดหรือแบ่งกลุ่มไว้ล่วงหน้าเป็นเทคนิคจัดกลุ่มที่มีลักษณะคล้ายคลึงกันมากที่สุด (Nearest neighbor) เพื่อจำแนกข้อมูลในแต่ละชุดข้อมูล โดยวัดค่าความห่างจากจุดศูนย์กลาง (Distance measurement) การวัดค่าเบี่ยงเบน (Deviation measurement) เป็นการรวมลักษณะที่คล้ายกันมากเข้าเป็นกลุ่มเดียวกัน

ในการวิจัยนี้จะใช้เทคนิคการจัดกลุ่มข้อมูลด้วยแบบเคมีน (k-means) ซึ่งเป็นที่นิยมไว้ใช้ในการแบ่งกลุ่มข้อมูลเป็นการเรียนรู้แบบไม่มีผู้สอนมีประสิทธิภาพในด้านของการเข้าใจง่าย และสามารถใช้ได้กับข้อมูลหลายประเภท โดยจะทำการแบ่งกลุ่มข้อมูลออกเป็น k กลุ่มตามที่ต้องการ (Han, et al., 2013) โดยในแต่ละกลุ่มจะมีข้อมูลจุดศูนย์กลางประจำกลุ่ม (Centroid) ซึ่งได้จากค่าเฉลี่ยของกลุ่ม และในการพิจารณาว่าข้อมูลจะอยู่กลุ่มใดจะดูจากระยะห่างระหว่างข้อมูลกับจุดศูนย์กลางประจำกลุ่มในแต่ละกลุ่ม ข้อมูลที่มีระยะห่างจากจุดศูนย์กลางประจำกลุ่มใดน้อยที่สุดถูกจัดอยู่ในกลุ่มนั้นๆ เมื่อข้อมูลทุกตัวถูกพิจารณาจัดกลุ่มแล้ว ก็จะทำการคำนวณหาค่าเฉลี่ยจุดศูนย์กลางประจำกลุ่มใหม่ หากจุดศูนย์กลางของแต่ละกลุ่มมีการเปลี่ยนตำแหน่งก็จะทำการวนซ้ำเพื่อจัดกลุ่มข้อมูลตามจุดศูนย์กลางกลุ่มที่เปลี่ยนไป ดังภาพที่ 2.9



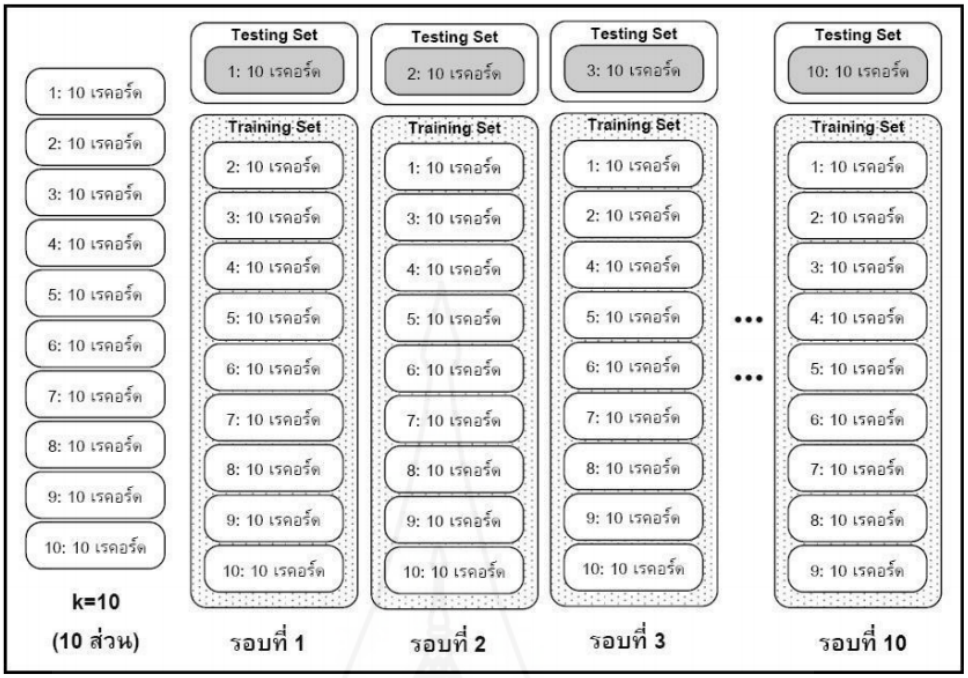
ภาพที่ 2.9 ขั้นตอนการจัดกลุ่มข้อมูลด้วย k-means

ที่มา: Han Kamber, (2006)

4.4 การทดสอบประสิทธิภาพแบบ k-fold Cross Validation

การวัดประสิทธิภาพของแบบจำลองในวิจัยนี้ จะใช้การวัดค่าความถูกต้องของข้อมูล (Accuracy) และค่ารากที่สองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) ส่วนการทดสอบทำการเลือกข้อมูลชุด (Training data) และข้อมูลชุดทดสอบ (Testing data) ซึ่งใช้วิธีการทดสอบประสิทธิภาพแบบ k – fold Cross validation

การจำแนกข้อมูลโดยใช้เทคนิคการทำเหมืองข้อมูล ด้วยต้นไม้ตัดสินใจ (Decision Tree) โครงข่ายประสาทเทียม (Neural Networks) และการจำแนกแบบเบย์ (Naïve Bayes) นั้น จะต้องมีการแบ่งข้อมูลออกเป็นข้อมูลชุด และข้อมูลชุดทดสอบ มาทำการทดสอบตัวแบบที่ได้สร้างขึ้น และการใช้วิธี k – fold Cross validation ซึ่งเหมาะสำหรับชุดข้อมูลจำนวนไม่มาก โดยจะแบ่งข้อมูลออกเป็น k ชุด จำนวนเท่าๆ กัน ตามที่ได้เลือกใช้เพื่อทำการวัดและทดสอบประสิทธิภาพตัวแบบพยากรณ์ (วรารัตน์ รุ่งวรวุฒิ, 2554) ซึ่งเป็นวิธีการที่ให้ค่าความแม่นยำ ด้วยการแบ่งข้อมูลออกเป็น 10 ส่วน โดยแบ่งข้อมูลการสร้างตัวแบบเป็น 9 ส่วน และตัวทดสอบ 1 ส่วน แล้วสลับสับเปลี่ยนข้อมูลในการสร้างตัวแบบและทดสอบ ใช้วิธีการวัดค่าความถูกต้อง (Accuracy) จากการสร้างตัวแบบพยากรณ์ ดังภาพที่ 2.10



ภาพที่ 2.10 ตัวอย่างการแบ่งข้อมูลด้วยวิธี K – fold Cross validation

ที่มา: Yoshua Bengio, Yves Grandvalet (2015, น. 75-95). Bias in Estimating the Variance of K-Fold Cross-Validation

4.4.1 การวัดค่าความถูกต้อง (Accuracy) ในการสร้างตัวแบบพยากรณ์ สามารถคำนวณได้ โดยใช้ค่าของชุดข้อมูลที่ได้นำมาพยากรณ์ (Total Number of Instances) หารด้วยจำนวนตัวอย่างที่นำมาทำการพยากรณ์ (Correctly Classified Instance) แล้วคูณด้วย 100 จะได้ค่าความถูกต้อง (Accuracy) ออกมาดังตัวอย่าง

$$Accuracy = \frac{TotalNumberofIns\tan ces}{CorrectyClassifiedIns\tan ce} x100$$

โดย Correctly Classified Instance คือ จำนวนตัวอย่างที่ทำนายถูกต้อง

Total Number of Instances คือ จำนวนตัวอย่างทั้งหมด

ที่มา: C. Lakshmi Devasena (2013, น. 133)

4.4.2 ค่ารากที่สองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) คือ การวัดค่าความแตกต่างระหว่างค่าจริงกับค่าที่ได้จากแบบจำลอง หาก RSME มีค่าน้อย แสดงว่าแบบจำลองใกล้เคียงกับค่าจริง สามารถคำนวณได้จาก

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{'} - Y_i\right)^2}$$

โดย Y_i แทน ค่าของข้อมูลที่เกิดขึ้นจริง

$Y_i^{'}$ แทน ค่าที่ได้จากการพยากรณ์

n แทน จำนวนนำเข้าทั้งหมด n

ที่มา: Hye Rin Kim (2009, pp. 4)

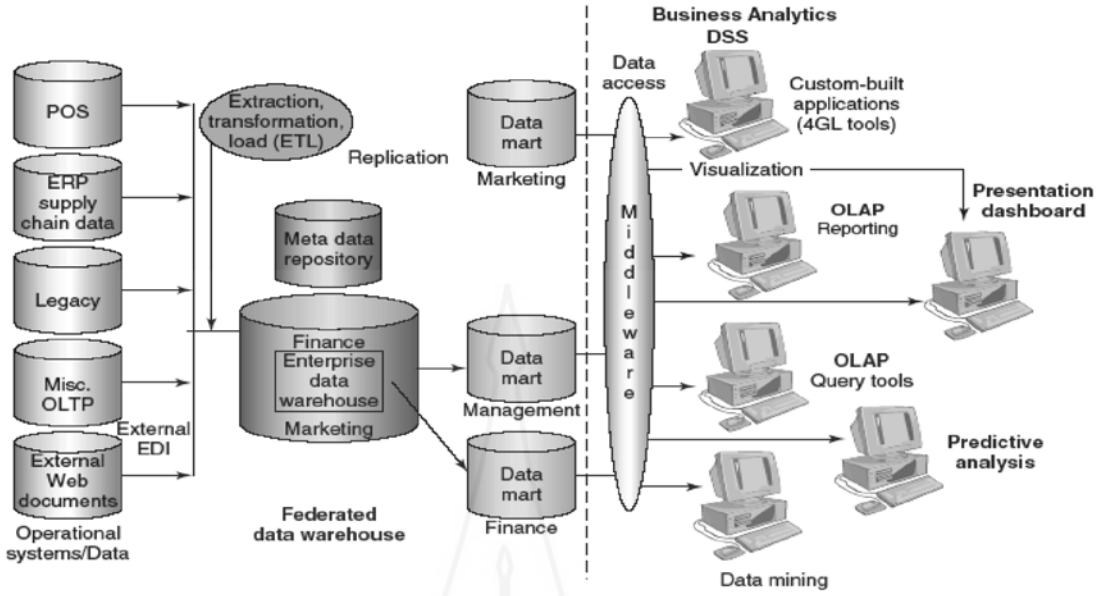
5. ธุระกิจอัจฉริยะ

ธุรกิจอัจฉริยะ เป็นการเข้าถึงข้อมูล การเลือกใช้ข้อมูล การวิเคราะห์และรวมถึงการพัฒนาต่อยอดข้อมูลเพื่อให้เข้าใจในสิ่งที่ซ้อนอยู่ในข้อมูล ที่จะช่วยให้เกิดการตัดสินใจที่มีประสิทธิภาพ อีกทั้งผู้บริหารยังใช้ข้อมูลที่ทันสมัยตลอดเวลา มีการปรับปรุงข้อมูลเพื่อใช้ในการเพิ่มยอดขาย บริการเพื่อรักษาลูกค้า พยากรณ์การเงินในอนาคต และยังช่วยลดความเสี่ยงในการดำเนินงาน

ธุรกิจอัจฉริยะ ยังสามารถช่วยให้เกิดประโยชน์ในด้านต่างๆ เช่น การกำหนดวิสัยทัศน์และพันธกิจขององค์กร กำหนดปัจจัยสู่ความสำเร็จ กำหนดเป้าหมายเชิงกลยุทธ์ กำหนดกลยุทธ์ในการวัดผล และกำหนดแผนงานที่จะต้องดำเนินงาน หากผู้ใช้สามรถนำข้อมูลไปใช้งานได้อย่างมีประสิทธิภาพ ก็จะเกิดประโยชน์ต่อองค์กรและทำให้องค์กรสามารถดำเนินกิจการได้อย่างมีประสิทธิภาพมากยิ่งขึ้นไป

Stackowiak et al. (2007) ได้กล่าวไว้ว่า Business Intelligence หรือธุรกิจอัจฉริยะเป็นกระบวนการรวบรวมข้อมูลขนาดใหญ่ และนำข้อมูลเหล่านี้มาทำการวิเคราะห์ รวมถึงการทำรายงานที่เป็นผลจากการวิเคราะห์เพื่อการช่วยในการตัดสินใจในการทำงานประจำวัน

Cui et al. (2007) ได้กล่าวไว้ว่า ธุรกิจอัจฉริยะเป็นวิธีการปรับปรุงประสิทธิภาพในการดำเนินธุรกิจขององค์กร โดยการให้ความช่วยเหลือที่มีประสิทธิภาพสำหรับการตัดสินใจของผู้บริหารเพื่อให้ผู้บริหารมีข้อมูลในการดำเนินการต่อองค์กร



ภาพที่ 2.11 สถาปัตยกรรมของ Business Intelligence และองค์ประกอบต่างๆ

ที่มา : Turban et al., (2007)

6. งานวิจัยที่เกี่ยวข้อง

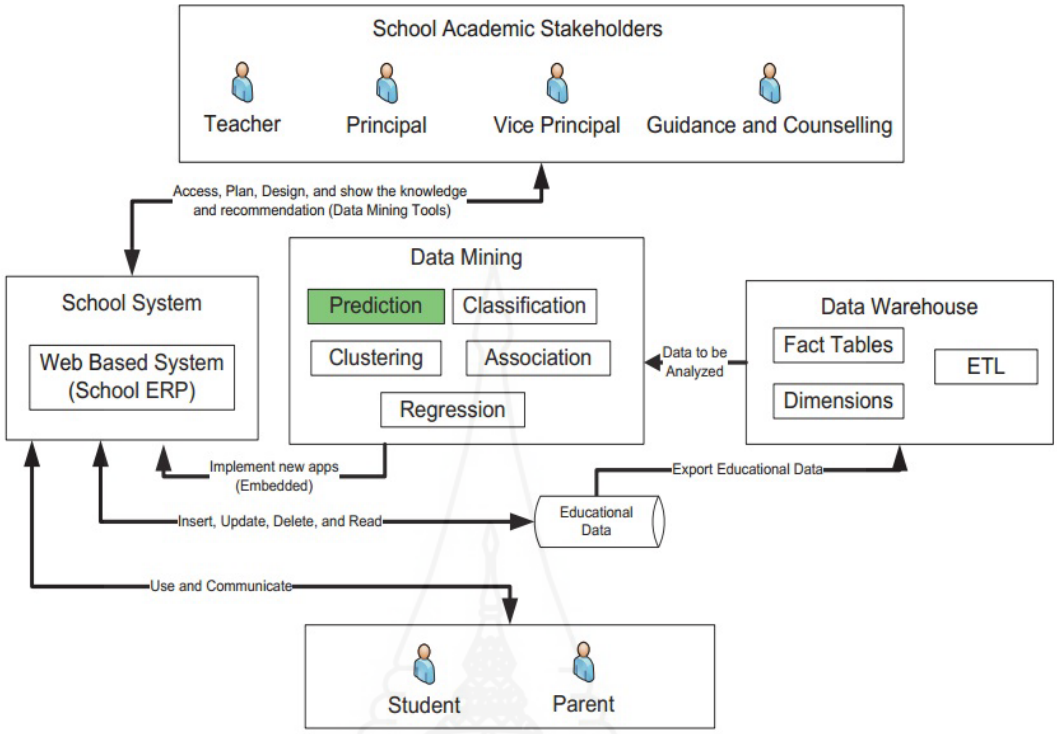
นายเสกสรรค์ วิลัยลักษณ์ (2558) ได้ทำการวิจัย เรื่องการใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา เพื่อพัฒนาคลังข้อมูล และสร้างตัวแบบพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา โดยใช้ข้อมูลนักเรียนระดับมัธยมศึกษาระหว่างปีการศึกษา 2548 – 2556 เพื่อทำการพัฒนาคลังข้อมูลโดยใช้โครงสร้างแบบสโนว์เฟลก สกีมา พร้อมนำเสนอรายงานด้วยเครื่องมือ Qlikview จากนั้นใช้ข้อมูลนักเรียนระดับมัธยมศึกษาปีที่ 4 ระหว่าง ปีการศึกษา 2553 – 2556 จำนวน 525 ระเบียน ประกอบด้วย 16 คุณลักษณะ มาสร้างตัวแบบพยากรณ์ผลการเรียนโดยใช้ชุดข้อมูล 2 แบบ คือ ข้อมูลแบบไม่จัดกลุ่ม (Original Data) และข้อมูลแบบจัดกลุ่ม (Data Cluster) และนำผ่านกระบวนการคัดเลือกคุณลักษณะ (Feature Selection) ซึ่งใช้วิธี Correlation-based Feature Selection (CFS) และวิธี Information Gain (IG) แล้วใช้เทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ซัพพอร์ตเวกเตอร์แมชชีน (SVM) และต้นไม้ตัดสินใจ (Decision Tree) มาสร้างตัวแบบพยากรณ์และเปรียบเทียบตัวแบบ ด้วยการทดสอบประสิทธิภาพแบบ 10-fold Cross Validation ผลพบว่า คลังข้อมูลนักเรียนระดับมัธยมศึกษาที่ใช้

งานโดยผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้นมีความพึงพอใจต่อการใช้งานคลังข้อมูลอยู่ในระดับดี และการทำเหมืองข้อมูล พบว่า ชุดข้อมูลแบบไม่จัดกลุ่มนำมาผ่านการคัดเลือกคุณลักษณะด้วยวิธี Correlation-based Feature Selection (CFS) ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ให้ค่าความถูกต้องสูงที่สุดที่ร้อยละ 94.48 และมีค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดที่ 0.1880 เหมาะสมสาหรับการสร้างระบบพยากรณ์ผลการเรียนเฉลี่ยของนักเรียน ซึ่งนักเรียนสามารถนาผลการพยากรณ์ไปประกอบการตัดสินใจและวางแผนการเรียน

ชฎิภกรน์ ทรายหมอ (2557) ได้ทำการวิจัยเรื่อง การพัฒนาแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่น ในผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล :กรณีโรงพยาบาลปทุมธานี วัตถุประสงค์เพื่อสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นในผู้ป่วยโรคเบาหวาน และทำการทดสอบแบบจำลองโดยใช้กับกลุ่มงานอายุรกรรม ห้องตรวจผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี โดยใช้ประวัติผลการตรวจรักษาผู้ป่วยที่เป็นโรคเบาหวาน โรงพยาบาลปทุมธานี ปี 2553 – ปี 2556 จำนวน 17,043 ชุดข้อมูล จากนั้นทำความเข้าใจปัญหาที่ดำเนินงานการตรวจรักษาของแพทย์ต่อผู้ป่วย และโครงสร้างข้อมูล เพื่อใช้สร้างแบบจำลองให้ตรงตามความต้องการของผู้ใช้ เพื่อให้แบบจำลองมีความแม่นยำ ทั้งนี้ข้อมูลที่ใช้ในการสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน คือข้อมูลผลตรวจการวินิจฉัยโรคเบาหวานตั้งแต่ที่ผู้ป่วยเป็นโรคเบาหวาน แล้วนำข้อมูลที่ได้ไปใช้ในการสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ทำการสร้างแบบจำลองด้วยการคัดเลือกตัวแปรที่สำคัญได้แก่ คือรหัสโรคที่ลงผลการตรวจวินิจฉัยของโรคเบาหวาน ทำการเปรียบเทียบตัวแบบการพยากรณ์ โดยวิธีการหาความสัมพันธ์ของข้อมูลระหว่าง Apriori Algorithm และ FPGrowth Algorithm พยากรณ์ภาวะแทรกซ้อนโดยใช้โปรแกรม WEKA เพื่อใช้วิเคราะห์ข้อมูลสำหรับการพยากรณ์ภาวะแทรกซ้อนโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน MicroSoft Visual Foxpro เพื่อจัดเก็บข้อมูลผลการตรวจวินิจฉัยโรคเบาหวาน Microsoft Execl 2010 สำหรับสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน จากนั้นทดสอบแบบจำลองกับสถานการณ์จริง โดยพัฒนาระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน โดยใช้ข้อมูลตั้งแต่ ปี 2557 – ปี 2557 กับกลุ่มงานอายุรกรรม ห้องตรวจผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี ข้อมูลที่นำเสนอมีประวัติผู้ป่วย ผลการพยากรณ์ภาวะแทรกซ้อนของโรคเบาหวาน อาการตามกลุ่มโรค ระดับความเสี่ยงต่อการเป็นอัมพฤกษ์ อัมพาต ระดับความเสี่ยงต่อการเสียชีวิต ข้อควรปฏิบัติและคำแนะนำจากแพทย์ ผลการพยากรณ์ความสัมพันธ์ภาวะแทรกซ้อนของโรคเบาหวานหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ในผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี พบว่าผู้ป่วยที่เป็นโรค 1) เบาหวานชนิดที่

ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา (รหัสโรค E103) แล้วจะเป็นโรคแทรกซ้อนดังนี้ 1.1) โรคจอตาจากโรคเบาหวานชนิดไม่งอกขยายหรือเอ็นพีดีอาร์ (รหัสโรค H3600) คิดเป็น 65 % 1.2) โรคต้อกระจกในผู้สูงอายุ (รหัสโรค H251) คิดเป็น 46 % 1.3) โรคจอตาจากโรคเบาหวานชนิดงอกขยายหรือพีดีอาร์ (รหัสโรค H3602) คิดเป็น 44 % 2) โรคเบาหวานชนิดที่ไม่ต้องพึ่งอินซูลินร่วมกับภาวะแทรกซ้อนทางตา (รหัสโรค E113) แล้วจะเป็นโรคแทรกซ้อน 2.1) โรคจอตาจากโรคเบาหวานชนิดไม่งอกขยาย (รหัสโรค H3600) คิดเป็น 77 % 2.2) โรคต้อกระจกในผู้สูงอายุ (รหัสโรค H251) คิดเป็น49 % 2.3) โรคจอตาจากโรคเบาหวานชนิดงอกขยาย (รหัสโรค H3602) คิดเป็น 36 % 2.4) โรคไตวายเรื้อรังที่มิได้ระบุรายละเอียด (รหัสโรค N189) คิดเป็น 20 % 3) โรคเบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางประสาท (รหัสโรค E114) แล้วจะเป็นโรคแทรกซ้อน 3.1) โรคต้อกระจกในผู้สูงอายุ (รหัสโรค H251) คิดเป็น 24 % 3.2) โรคหัวใจขาดเลือดเรื้อรังที่มิได้ ระบุรายละเอียด (รหัสโรค I251) คิดเป็น 20 %

Kurniawan and Halim (2013) ได้ศึกษาเรื่อง การพัฒนาคลังข้อมูลและการใช้เทคนิคเหมืองข้อมูลในการวิเคราะห์และพยากรณ์ผลการเรียน มีวัตถุประสงค์เพื่อช่วยให้นักเรียนมีผลการเรียนที่ดีขึ้น ผลการวิจัยพบว่า การสร้างคลังข้อมูลออกแบบโดยใช้โครงสร้างแบบสตาร์ (Star schema) และนำเสนอข้อมูลในรูปแบบของเว็บไซต์ (Web Application) ทำการดึงข้อมูลจากคลังข้อมูล เพื่อสร้างตัวแบบพยากรณ์ด้วยการทำเทคนิคเหมืองข้อมูล ซึ่งใช้วิธี การจำแนกประเภทข้อมูล (Classification) การหาความสัมพันธ์ข้อมูล (Association) และการจัดกลุ่มข้อมูล (Clustering) ผลจากการสร้างตัวแบบพยากรณ์ จะทำการแสดงบนหน้าเว็บไซต์ของโรงเรียน ซึ่งแสดงให้เห็นว่าการทำคลังข้อมูลจะช่วยเพิ่มประสิทธิภาพด้านข้อมูลให้มีความถูกต้องตามความต้องการของผู้ใช้ อีกทั้งยังช่วยลดขั้นตอนการทำงาน และการทำเหมืองข้อมูลสามารถช่วยพยากรณ์ผลการเรียนของนักเรียนได้อย่างมีประสิทธิภาพ ดังภาพที่ 2.12



ภาพที่ 2.12 โครงสร้างการทำคลังข้อมูลและเหมืองข้อมูลของโรงเรียน

ที่มา: Kurniawan and Halim (2013: 102)

นายอภิชาติ ปัญญา (2557) ได้ทำการวิจัยเรื่อง การทำเหมืองข้อมูลสำหรับการพัฒนาเว็บไซต์ กรณีศึกษาเว็บไซต์มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ลำปาง มีวัตถุประสงค์เพื่อสร้างคลังข้อมูลเข้าใช้งานเว็บไซต์ ทำเหมืองข้อมูลวิเคราะห์จำแนกกลุ่มการใช้งานเว็บใช้โดยใช้อัลกอริทึมจัดกลุ่ม และทำเหมืองข้อมูลพยากรณ์จำนวนผู้ใช้งานเว็บไซต์โดยใช้อัลกอริทึมอนุกรมเวลา โดยใช้โปรแกรม SQL Server 2008 และ SQL Service Analysis กระบวนการเริ่มจากการรวบรวมล็อกไฟล์การเข้าใช้งานเว็บไซต์ตั้งแต่วันที่ 1 มกราคม 2556 ถึง 31 ธันวาคม 2556 และข้อมูลเว็บไซต์หน่วยงานต่างๆ ของมหาวิทยาลัยฯ มาสร้างคลังข้อมูลด้วยกระบวนการอีทีแอล จากนั้นนำข้อมูลจากคลังข้อมูลมาสร้างคิวบ์ เพื่อนำเสนอรายงานจากการประมวลผลข้อมูลเชิงวิเคราะห์หรือโอแลป และนำข้อมูลจากคลังข้อมูลมาทำเหมืองข้อมูลโดยใช้แบบจำลองคริสพ์-ดีเอ็ม ผลจากการทำเหมืองข้อมูลพบว่าการจัดกลุ่มของการเข้าใช้เว็บไซต์แบ่งออกเป็น 6 กลุ่ม อนุกรมเวลาทำให้สามารถพยากรณ์ปริมาณผู้ใช้บริการเว็บไซต์ได้ ส่วนกฎความสัมพันธ์ที่เกิดขึ้นทั้งหมด 16 กฎ และการจัดกลุ่มโดยใช้ลำดับ พบว่าจัดกลุ่มออกเป็น 66 กลุ่มซึ่งกลุ่มที่น่าสนใจมีทั้งหมด 6 กลุ่ม การทำเหมืองข้อมูลสามารถทำให้ได้องค์ความรู้ใหม่ ซึ่งสามารถนำมาใช้ในการวางแผนพัฒนา

ปรับปรุงและดูแลเว็บไซต์ได้ และยังสามารถทำให้ผู้บริหารได้รับรายงานหลายมิติที่สามารถมาประกอบการตัดสินใจ

นายคเชนทร์ อึ่งสกุล (2552) ได้ศึกษาเรื่อง การพัฒนาระบบคลังข้อมูลด้านแรงงานของสำนักงานสถิติจังหวัดเชียงใหม่ มีวัตถุประสงค์เพื่อพัฒนาเครื่องมือสำหรับการวิเคราะห์ข้อมูลและแสดงผลในรูปแบบรายงานหลายมิติ เพื่อเป็นประโยชน์ในการวางแผน และใช้เป็นข้อมูลประกอบการตัดสินใจของผู้บริหาร ในการพัฒนาระบบคลังข้อมูลนี้ผู้เขียนได้ดำเนินการพัฒนาโปรแกรมแปลงและถ่ายโอนข้อมูล โดยใช้โปรแกรม Microsoft Visual Basic ในการพัฒนาโปรแกรม ซึ่งช่วยในการแปลงและถ่ายโอนข้อมูลต่างๆ ไปยังคลังข้อมูลที่ได้ออกแบบคลังข้อมูลที่จัดการด้วย Microsoft SQL Server 2005 ส่วนการพัฒนาเครื่องมือผลิตรายงานหลายมิติที่มีความยืดหยุ่นตามความต้องการของผู้บริหาร ซึ่งใช้โปรแกรมโปรแคลริติเดสทอป โปรเฟชชันนอล พัฒนาส่วนต่อประสานกับผู้ใช้ผลจากการนำระบบไปใช้จริง โดยผู้บริหารและเจ้าหน้าที่ของสำนักงานสถิติจังหวัดเชียงใหม่ ในข้อมูลด้านโครงสร้างกำลังแรงงานของแต่ละจังหวัด ภาวะการมีงานทำของประชากรและภาวะการว่างงานของประชากร สามารถนำเสนอผลในรูปแบบแผนภูมิต่างๆ ได้หลากหลายและมีรูปแบบที่ทันสมัย ใช้งานได้ง่าย สามารถใช้ระบบนี้เป็นเครื่องมือสำหรับการวิเคราะห์ข้อมูลให้อยู่ในรูปแบบรายงานหลายมิติได้ และสามารถนำรายงานที่ได้ไปใช้ในการวางแผนและกำหนดทิศทางในการพัฒนาองค์กรได้อย่างมีประสิทธิภาพ

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยนี้เป็นการวิจัยเชิงพัฒนา (Development research) โดยการพยากรณ์การใช้สิทธิของลูกจ้างในสถานประกอบการเพื่อเพิ่มประสิทธิภาพการบริหารการจัดการกองทุนเงินทดแทน สำนักงานประกันสังคม เพื่อนำผลการวิเคราะห์มาใช้จัดตั้งงบประมาณ สำรองการเบิกจ่ายในปีงบประมาณ และเพื่อนำมาป้องกันการเกิดอุบัติหตุ โดยการอบรมจากเจ้าหน้าที่ด้านความปลอดภัย โดยมีรายละเอียดการดำเนินการวิจัย ดังนี้

1. ประชากรและกลุ่มตัวอย่าง
2. เครื่องมือที่ใช้ในการดำเนินการวิจัย
3. ขั้นตอนการดำเนินงาน

1. ประชากรและกลุ่มตัวอย่าง

1.1 ประชากร

ข้อมูลของการประสบอันตรายจากการทำงานของลูกจ้างในสถานประกอบการเป็นข้อมูลการเบิกสิทธิประโยชน์ของลูกจ้างและสถานประกอบการของกองทุนเงินทดแทน โดยใช้ข้อมูลจำนวน 12 เขตพื้นที่ในกรุงเทพมหานคร ได้แก่ สำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 1-12 ระหว่างปี 2556 – 2557 โดยเลือกใช้ข้อมูลด้านกองทุนเงินทดแทนมาใช้ในการทำวิจัยครั้งนี้

1.2 กลุ่มตัวอย่าง

1.2.1 การจัดทำคลังข้อมูล ผู้วิจัยได้ใช้ข้อมูลการใช้สิทธิในการเบิกเงินของลูกจ้างที่เกิดอุบัติเหตุเนื่องจากการทำงานให้นายจ้าง ของกองทุนเงินทดแทน สำนักงานประกันสังคม โดยเลือกข้อมูลเฉพาะพื้นที่ในกรุงเทพมหานคร ทั้งหมด 12 เขตพื้นที่ ประกอบกด้วย สำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 1 – 12 ซึ่งจะเก็บรวบรวมข้อมูลในรูปแบบเอ็กเซล (Excel) ช่วงเวลาตั้งแต่วันที่ 1 มกราคม 2556 ถึง วันที่ 31 ธันวาคม 2557 มีจำนวนข้อมูลทั้งหมด 61,924 ชุดข้อมูล

1.2.2 การสร้างตัวแบบพยากรณ์ ได้นำข้อมูลจากคลังข้อมูลมาทำ Data Set ซึ่งจะเลือกใช้ข้อมูลการใช้สิทธิในการเบิกเงินของลูกจ้างที่เกิดอุบัติเหตุเนื่องจากการทำงาน สำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 3 ปี 2557 จำนวนข้อมูล 3,636 ชุดข้อมูล

2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการดำเนินการวิจัย มีดังนี้

2.1 Windows 8 คอมพิวเตอร์ตั้งโต๊ะ จำนวน 1 เครื่อง

2.2 Microsoft SQL Server 2008 R2

2.2.1 เครื่องมือ SQL Server Integration Services (SSIS) ใช้ในการจัดทำคลังข้อมูล

2.2.2 เครื่องมือ SQL Server Analysis Service (SSAS) ใช้ในการสร้างคิวบ์ข้อมูล

2.3 Microsoft Excel 2007 โดยใช้เทคนิค Pivot Table ในการนำเสนอรูปแบบรายงานสำหรับผู้บริหาร หัวหน้าส่วนงานกองทุนเงินทดแทน และเจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทน

2.4 WEKA Version 3.8.0 ใช้สร้างตัวแบบพยากรณ์ โดยใช้ข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินดองทุนเงินทดแทน ที่ได้จากการทำคลังข้อมูล

2.5 แบบสอบถาม ใช้ในการการประเมินผลการใช้งานคลังข้อมูล

3. ขั้นตอนการดำเนินงาน

3.1 การพัฒนาคลังข้อมูล

3.1.1 การวิเคราะห์ความต้องการของผู้ใช้งาน

ผู้วิจัยได้วิเคราะห์ความต้องการของผู้ใช้งานจากผู้บริหาร หัวหน้าส่วนงานกองทุนเงินทดแทน และเจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทน ได้ดังนี้

1) ระบบงานที่ใช้อยู่ในปัจจุบันเป็นระบบงานที่ใช้งานบนเว็บไซต์ (Web Application) ซึ่งระบบมีปัญหาที่ผิดพลาดบ่อยครั้งที่เกิดขึ้นและไม่สเถียรภาพในการใช้งาน

2) ผู้บริหารไม่สามารถดูข้อมูลผลของการประสบอันตรายในรูปแบบรายงานได้ทันทีต้องรอให้เจ้าหน้าที่สั่งพิมพ์รายงานข้อมูลออกมาให้เท่านั้น

3.1.2 การออกแบบคลังข้อมูล

ผู้วิจัยได้ออกแบบคลังข้อมูล โดยใช้ข้อมูลที่ได้จากฐานข้อมูลงานกองทุนเงินทดแทน และจัดทำตารางข้อเท็จจริง (Fact Table) ตารางมิติ (Dimension Table) และตารางชั่วคราว (Temp Table) ซึ่งใช้โครงสร้างแบบสตาร์ (Star Schema)

3.1.3 การสร้างคลังข้อมูล

กระบวนการสร้างคลังข้อมูล นำข้อมูลเข้าสู่คลังข้อมูลด้วยกระบวนการอีทีแอล (Extract Transform and Load : ETL) โดยใช้เครื่องมือ SQL Server Integration Services (SSIS) ของโปรแกรม Microsoft SQL Server 2008 R2

3.1.4 การนำคลังข้อมูลไปใช้งาน

1) การทำรายงาน ผู้วิจัยได้นำรายงานจากการทำการประมวลผลเชิงวิเคราะห์ออนไลน์ (Online Analytical Processing : OLAP) นำเสนอด้วยโปรแกรม Microsoft Excel 2007

2) การนำข้อมูลจาก Web Application เป็นข้อมูลจากการปฏิบัติงานของหน่วยงานเข้าสู่คลังข้อมูล โดยนำข้อมูลจากฐานข้อมูล (Database) ดึงข้อมูลออกมาให้อยู่ในรูปแบบไฟล์เอ็กเซล (Excel File) ซึ่งข้อมูลจะเป็นช่วงเวลาตั้งแต่วันที่ 01 มกราคม 2556 - 31 ธันวาคม 2557

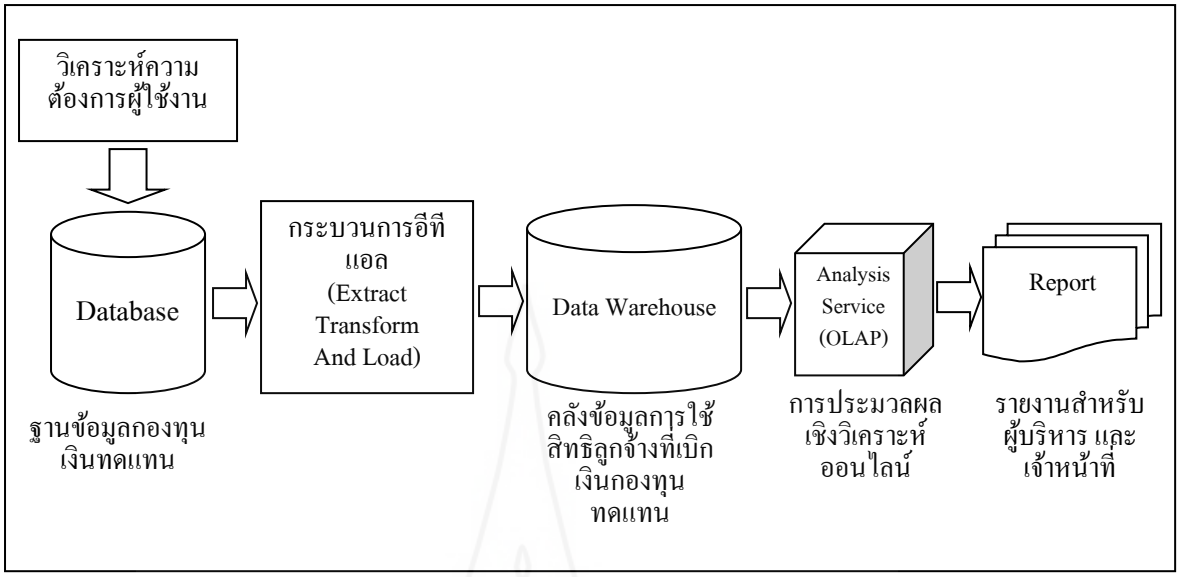
3.2 การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ

3.2.1 การสร้างคิวบ์ (Club)

ผู้วิจัยได้สร้างคิวบ์ (Club) ด้วยโปรแกรม SQL Server Analysis Service (SSAS) ซึ่งสามารถดูข้อมูลที่เป็นการประมวลผลเชิงวิเคราะห์ออนไลน์ในลักษณะหลายมิติ (Online Analytical Processing : OLAP)

3.2.2 ระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติตามหลักธุรกิจอัจฉริยะ (Online Analytical Processing)

การนำเสนอรายงานให้ผู้ใช้งานทั้ง 3 กลุ่มได้แก่ ผู้บริหาร หัวหน้าส่วนงานกองทุนเงินทดแทน และเจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทน ด้วยมิติมุมมองต่างๆ โดยใช้เครื่องมือ Pivot Table ของโปรแกรม Microsoft Excel 2007



ภาพที่ 3.1 ขั้นตอนการสร้างคลังข้อมูล

3.3 การทำเหมืองข้อมูล

งานวิจัยนี้ดำเนินการทำเหมืองข้อมูลตามแบบจำลองของคริสพ์ – ดีเอ็ม (CRISP – DM Model) ซึ่งประกอบด้วย 6 ขั้นตอน ดังนี้

3.3.1 ทำความเข้าใจปัญหาและโอกาสทางธุรกิจ (Business Understanding)

วัตถุประสงค์ของงานวิจัยนี้เพื่อใช้เทคนิคเหมืองข้อมูลในการสร้างตัวแบบพยากรณ์การใช้สิทธิลูกจ้างที่เบิกเงินกองทุนเงินทดแทน เพื่อใช้ในการจัดตั้งงบประมาณประจำปี และป้องกันการลดการประสบอันตราย โดยการจัดอบรมให้ความรู้ป้องกันการประสบอันตรายจากการทำงานจากเจ้าหน้าที่วิชาชีพ (จป.)

3.3.2 ทำความเข้าใจข้อมูล (Data Understanding)

ในการทำเหมืองข้อมูลใช้ข้อมูลที่ได้มาจากคลังข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน แล้วผ่านกระบวนการอีทีแอล (Extract Transform and Load: ETL) โดยใช้เครื่องมือ SQL Server Integration Services (SSIS) ของโปรแกรม Microsoft SQL Server 2008 R2 และนำข้อมูลจากฐานข้อมูลออกมาในรูปแบบของ Excel File

3.3.3 เตรียมข้อมูล (Data Preparation)

นำข้อมูลที่จากคลังข้อมูลในรูปแบบของไฟล์เอ็กเซล นำไปทำขั้นตอนการเตรียมข้อมูล (Data Preparation) เพื่อใช้ในการทำเหมืองข้อมูล ดังต่อไปนี้

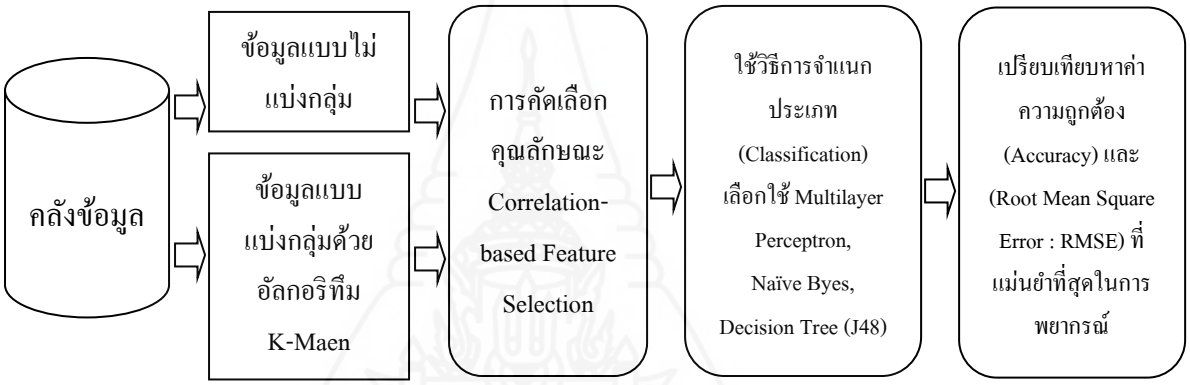
1) การคัดกรองข้อมูล (Data Cleaning)

2) การเลือกข้อมูล (Data Selection)

3) การแปลงข้อมูล (Data Transformation)

3.3.4 สร้างแบบพยากรณ์ (Modeling)

การสร้างตัวแบบพยากรณ์ โดยข้อมูลออกเป็น 2 แบบ ได้แก่ ชุดข้อมูลแบบไม่จัดกลุ่ม (Original Data) และชุดข้อมูลแบบจัดกลุ่ม (Clustering Data) นำมาผ่านการจัดกลุ่มข้อมูลด้วยอัลกอริทึม k-means โดยใช้โปรแกรม WEKA Version 3.8.0 แล้วใช้เทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP), เทคนิคการจำแนกแบบเบย์ (Naïve Bayes) และเทคนิคต้นไม้ตัดสินใจ (Decision Tree) โดยเลือกใช้อัลกอริทึม J48 ดังภาพที่ 3.1



ภาพที่ 3.2 โมเดลขั้นตอนการสร้างตัวแบบการพยากรณ์

3.3.5 ประเมินความถูกต้องหรือความแม่นยำของตัวแบบพยากรณ์

1) การประเมินความถูต้องของตัวแบบพยากรณ์ด้วย ค่าความถูกต้อง (Accuracy) และ ค่าความคลาดเคลื่อน (RMSE)

การประเมินประสิทธิภาพตัวแบบพยากรณ์ ใช้การประเมินประสิทธิภาพโดยเลือกใช้วิธีการวัดค่าความถูกต้อง (Accuracy) ในการสร้างตัวแบบพยากรณ์ และค่าความคลาดเคลื่อน (Root Mean Square Error : RMSE) ด้วยโปรแกรม WEKA Version 3.8.0

2) การสร้างตัวแบบพยากรณ์ด้วย K-fold Cross Validation

การประเมินประสิทธิภาพตัวแบบพยากรณ์ ใช้การประเมินประสิทธิภาพของแบบ K-fold Cross Validation โดยกำหนดค่า k เท่ากับ 10 เพื่อแบ่งชุดข้อมูลสำหรับการทดสอบด้วยโปรแกรม WEKA Version 3.8.0

3) การเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม (Original Data) กับชุดข้อมูลแบบจัดกลุ่ม (Clustering Data)

การทดสอบประสิทธิภาพตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม (Original Data) ที่มีค่าความถูกต้องสูงที่สุดมาทำการเปรียบเทียบกับ ตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบจัดกลุ่ม (Clustering Data) ที่มีค่าความถูกต้องสูงที่สุด

3.3.6 การนำโมเดลไปใช้งาน (Model Deployment)

เป็นการนำตัวแบบพยากรณ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลมาทำการพัฒนาและสร้างระบบพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน สำนักงานประกันสังคม เพื่อนำไปใช้กับผู้บริหาร เพื่อใช้ประกอบการตัดสินใจในการจัดตั้งงบประมาณประจำปีของกองทุนเงินทดแทน สำนักงานประกันสังคม และช่วยในการวิเคราะห์ลดอัตราการเกิดอุบัติเหตุ เพื่อการบริหารการจัดการกองทุนเงินทดแทนที่ดียิ่งขึ้น

3.4 การประเมินความพึงพอใจต่อการใช้งานคลังข้อมูล

3.4.1 การออกแบบเครื่องมือการประเมินความพึงพอใจของผู้ใช้ระบบคลังข้อมูลของการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน โดยมีผู้ประเมินประกอบด้วย จำนวน 9 ท่าน และประเมินทั้งหมด 5 ด้านได้แก่ 1) มีความสะดวกและง่ายต่อการใช้งาน 2) การออกแบบรายงานเข้าใจได้ง่าย 3) ความถูกต้องของข้อมูล 4) ช่วยลดขั้นตอนในการทำงาน และ 5) ข้อมูลตรงกับความต้องการผู้ใช้งาน

3.4.2 การประเมินผลความพึงพอใจระบบคลังข้อมูล เป็นการประเมินผลของการดำเนินงานหลังจากนำแบบสอบถามที่ได้จากการประเมินของกลุ่มผู้ใช้งาน มาทำการวิเคราะห์และแปลผล โดยใช้ค่าสถิติการวัดหาค่ากลางของข้อมูล คือ ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) และการประเมินผลได้กำหนดเกณฑ์ระดับคบความคิดเห็น แบ่งเป็น 5 ระดับ คือ 5 = มากที่สุด 4 = มาก 3 = ปานกลาง 2 = น้อย 1= น้อยมากที่สุด

3.4.3 นำค่าเฉลี่ยที่ได้จากประเมิน มาทำการเปรียบเทียบด้วยค่าสถิติ คือ ค่าเฉลี่ยเลขคณิต (x) และค่าส่วนเบี่ยงเบนมาตรฐาน S.D.

$$\frac{\sum x}{n}$$

$\bar{X}$ คือ ค่าเฉลี่ยเลขคณิต

$\sum x$ คือ ผลบวกของข้อมูลทุกค่า

n คือ จำนวนข้อมูลทั้งหมด

$$S.D. = \sqrt{\frac{(x-\bar{x})^2}{n-1}}$$

X คือ ข้อมูล

$\overline{X}$ คือ ค่าเฉลี่ยเลขคณิต

n คือ จำนวนข้อมูลทั้งหมด

เบสท์ (Best.,1963: 204-208)

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การศึกษาการวิจัยเรื่อง การพยากรณ์การใช้สิทธิของลูกจ้างในสถานประกอบการเพื่อเพิ่มประสิทธิภาพการบริหารการจัดการกองทุนเงินทดแทน สำนักงานประกันสังคม ผู้วิจัยได้ทำการพัฒนาระบบคลังข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทนในการนำเสนอรายงาน และนำมาข้อมูลมาวิเคราะห์หาความรู้ใหม่โดยเทคนิคการทำเหมืองข้อมูลตามขั้นตอนของแบบจำลอง คริสพ์ – ดีเอ็ม (CRISP – DM Model) ซึ่งผลการวิจัยและการวิเคราะห์ข้อมูล ประกอบด้วย 5 ส่วน ได้แก่

1. การพัฒนาคลังข้อมูล
2. การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ
3. การพัฒนาตัวแบบพยากรณ์
4. การนำโมเดลไปใช้งาน
5. การประเมินผลการใช้งานคลังข้อมูล

1. การพัฒนาคลังข้อมูล

1.1 การวิเคราะห์ความต้องการของผู้ใช้งาน

จากการสอบถามข้อมูลจากผู้บริหาร หัวหน้าส่วนงานกองทุนเงินทดแทน เจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทน และผู้วิจัยที่เป็นผู้ใช้งานข้อมูลระบบงานกองทุนเงินทดแทนอยู่เป็นประจำ พบว่าปัญหาและความต้องการของผู้ใช้งานซึ่งสามารถสรุปได้ดังนี้

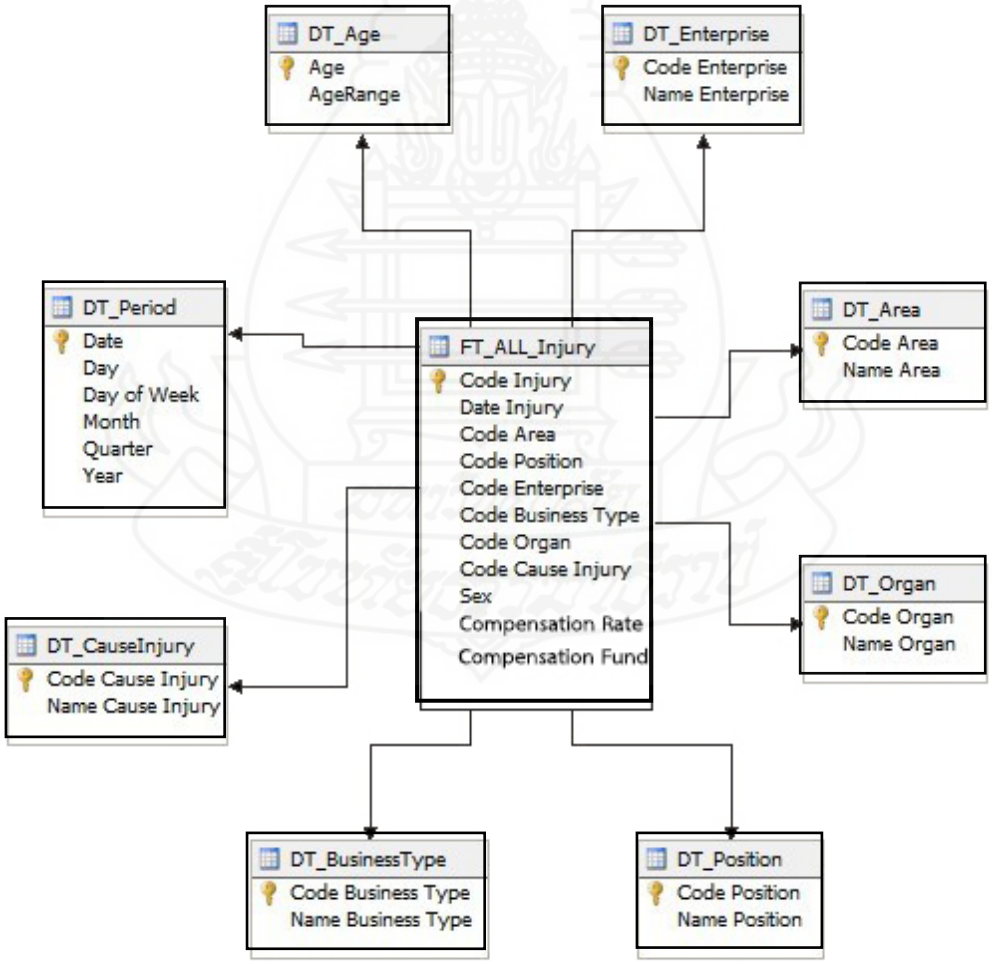
1.1.1 ระบบงานที่ใช้อยู่ในปัจจุบันเป็นระบบงานที่ใช้งานบนเว็บไซต์ (Web Application) ซึ่งในปัจจุบันการทำงานของระบบงานกองทุนเงินทดแทน มีขั้นตอนที่ยุ่งยากในการใช้งาน และปัญหาของข้อมูลที่ผิดพลาดบ่อยครั้งที่เกิดขึ้น ระบบไม่เสถียรภาพพอต่อการใช้งาน

1.1.2 ผู้บริหารสามารถดูข้อมูลผลของการประสบอันตรายในรูปแบบรายงาน ต้องรอให้เจ้าหน้าที่สั่งพิมพ์รายงานข้อมูลออกมาให้เท่านั้น ซึ่งความต้องการของผู้บริหารต้องการดูรายงานที่สามารถเรียกดูได้จากคอมพิวเตอร์ และเป็นผลสรุปที่สามารถนำไปประกอบการจัดตั้งงบประมาณได้

จากผลที่ได้ผู้วิจัยได้ทำการศึกษาและวิเคราะห์ความต้องการ และศึกษาจากงานวิจัยต่างๆ ที่เกี่ยวข้อง จากงานวิจัยของ (เสกสรร วิลัยลักษณ์: 2558) ได้พัฒนาคลังข้อมูลสำหรับผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา เพื่อสร้างคลังข้อมูลผลการเรียนในการนำไปสร้างตัวแบบการพยากรณ์ และทดสอบเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ และนำตัวแบบพยากรณ์ที่ดีที่สุดนำไปพัฒนาระบบผลการเรียน เพื่อใช้ในการตัดสินใจของนักเรียนในการเลือกเรียนสายวิชาต่างๆ

1.2 การออกแบบคลังข้อมูล

1.2.1 ในการออกแบบคลังข้อมูล ผู้วิจัยได้ออกแบบโครงสร้างคลังข้อมูลแบบโครงสร้างสตาร์ (Star Schema) โดยนำข้อมูลที่ได้จากฐานข้อมูลงานกองทุนเงินทดแทน สำนักงานประกันสังคม ด้วยเครื่องมือ SQL Server Management Studio ของโปรแกรม Microsoft SQL Server 2008 R2



ภาพที่ 4.1 โครงสร้างคลังข้อมูลแบบสตาร์ (Star Schema)

1.2.2 จัดการข้อมูลจากฐานข้อมูลที่ได้ให้เรียบร้อย ก่อนนำข้อมูลเข้าสู่คลังข้อมูล

ในส่วนของข้อมูลในฐานข้อมูลเดิมและความต้องการในการวิเคราะห์ข้อมูล ซึ่งจำเป็นจะต้องทราบถึงลักษณะของข้อมูลที่มีอยู่ และทำการแบ่งกลุ่มข้อมูลที่จัดเก็บไว้ในคลังข้อมูล ให้สามารถแบ่งกลุ่มข้อมูลได้ดังนี้

1) ข้อมูลรายละเอียดการประสบอันตราย เช่น เลขที่การประสบอันตราย วันที่ประสบอันตราย เพศ อายุ สาเหตุที่ประสบอันตราย อวัยวะ

2) ข้อมูลสถานประกอบการ เช่น เลขที่บัญชีสถานประกอบการ ชื่อสถานประกอบการ รหัสกิจการ

3) ข้อมูลสำนักงานประประกันสังคม เช่น รหัสสำนักงาน ชื่อสำนักงาน

4) ข้อมูลตำแหน่งหน้าที่ เช่น รหัสตำแหน่ง ชื่อตำแหน่ง

5) ข้อมูลอวัยวะ เช่น รหัสอวัยวะ ชื่ออวัยวะ

จากข้อมูลที่ได้จากฐานข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการจึงทำการกำหนดมุมมองหรือมิติ เพื่อใช้ในการออกแบบตารางข้อเท็จจริงและตารางมิติ ดังนี้

ตารางที่ 4.1 ตารางมุมมองมิติการใช้สิทธิลูกจ้างของสถานประกอบการ

ลำดับ	ชื่อ	ตัวอย่างข้อมูล	ประเภท
1	รายละเอียดการประสบอันตราย	เลขที่ประสบอันตราย 100356/00001/01 วันที่ประสบอันตราย 02 มีนาคม 2556 เพศชาย อายุ 35 ปี สาเหตุการประสบอันตราย ขณะทำงานตกนั่งร้าน อวัยวะ ขาซ้ายหัก	มุมมองหรือมิติ
2	รายละเอียดสถานประกอบการ	เลขที่บัญชีสถานประกอบการ 10-0001870-1 บริษัท ช.ช้าง จำกัด (มหาชน)	มุมมองหรือมิติ
3	รายละเอียดสำนักงานประกันสังคม	สำนักงานประกันสังคมกรุงเทพมหานคร พื้นที่ 3 รหัส 1003	มุมมองหรือมิติ
4	รายละเอียดตำแหน่ง	พนักงานส่งเอกสาร รหัส 915	มุมมองหรือมิติ
5	รายละเอียดอวัยวะ	ศีรษะ รหัส 0110	มุมมองหรือมิติ
6	รายละเอียดสาเหตุการประสบอันตราย	อุบัติเหตุรถยนต์ รหัส 11	มุมมองหรือมิติ

จากตารางที่ 4.1 ทำการกำหนดมุมมองหรือมิติ ค่าที่ต้องการวัด เพื่อใช้ในการออกแบบตารางข้อเท็จจริง (Fact Table) และตารางมิติ (Dimension Table) ซึ่งมีตารางข้อเท็จจริง 1 ตาราง คือ FT_ALL_Injury ตารางมิติ 8 ตาราง คือ DT_Area, DT_Position, DT_Enterprise, DT_BusinessType, DT_Organ, DT_CauseInjury, DT_ Period, DT_Age และตารางชั่วคราว 1 ตาราง คือ Temp_Data

ตารางที่ 4.2 ตารางข้อเท็จจริง FT_ALL_Injury ใช้สำหรับเก็บข้อมูลการใช้สิทธิลูกจ้าง

ลำดับ	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	Code Injury	Varchar	เลขที่ประสบอันตราย	Primary Key
2	Date Injury	Date time	วันที่ประสบอันตราย	Foreign Key (DT_Period)
3	Code Area	Varchar	สำนักงานประกันสังคมพื้นที่	Foreign Key (DT_Area)
4	Code Position	Varchar	รหัสตำแหน่ง	Foreign Key (DT_Position)
5	Code Enterprise	Varchar	รหัสสถานประกอบการ	Foreign Key (DT_Enterprise)
6	Code Business Type	Varchar	รหัสประเภทกิจการ	Foreign Key (DT_BusinessType)
7	Code Organ	Varchar	รหัสอวัยวะ	Foreign Key (DT_Organ)
8	Code Cause Injury	Varchar	รหัสสาเหตุการประสบอันตราย	Foreign Key (DT_CauseInjury)
9	Sex	Varchar	เพศ	
10	Age	Int	อายุ	Foreign Key (DT_Age)
11	Compensation Rate	Int	การจ่ายค่ารักษาพยาบาลและค่าทดแทน	
12	Compensation Fund	Varchar	การวินิจฉัยจ่ายค่าทดแทน	

ตารางที่ 4.3 ตารางมิติ DT_Area ใช้สำหรับเก็บข้อมูลสำนักงานประกันสังคมพื้นที่

ลำดับ	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	Code Area	Varchar	รหัสสำนักงานประกันสังคม	Primary Key
2	Name Area	Varchar	ชื่อสำนักงานประกันสังคม	

ตารางที่ 4.4 ตารางมิติ DT_Position ใช้สำหรับเก็บข้อมูลตำแหน่ง

ลำดับ	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	Code Position	Varchar	รหัสตำแหน่ง	Primary Key
2	Name Position	Varchar	ชื่อตำแหน่ง	

ตารางที่ 4.5 ตารางมิติ DT_Enterprise ใช้สำหรับเก็บข้อมูลสถานประกอบการ

ลำดับ	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	Code Enterprsie	Varchar	รหัสสถานประกอบการ	Primary Key
2	Name Enterprise	Varchar	ชื่อสถานประกอบการ	

ตารางที่ 4.6 ตารางมิติ DT_BusinessType ใช้สำหรับเก็บข้อมูลประเภทกิจการ

ลำดับ	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	Code Business Type	Varchar	รหัสประเภทกิจการ	Primary Key
2	Name Business Type	Varchar	ชื่อประเภทกิจการ	

ตารางที่ 4.7 ตารางมิติ DT_Organ ใช้สำหรับเก็บข้อมูลอวัยวะ

ลำดับ	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	Code Organ	Varchar	รหัสอวัยวะ	Primary Key
2	Name Organ	Varchar	ชื่ออวัยวะ	

ตารางที่ 4.8 ตารางมิติ DT_CauseInjury ใช้สำหรับเก็บข้อมูลสาเหตุการประสบอันตราย

ลำดับ	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	Code Cause Injury	Varchar	รหัสาเหตุการประสบอันตราย	Primary Key
2	Name Cause Injury	Varchar	ชื่อสาเหตุการประสบอันตราย	

ตารางที่ 4.9 ตารางมิติ DT_Period ใช้สำหรับเก็บข้อมูลระยะเวลา

ลำดับ	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	Date	Date time	ระยะเวลา	Primary Key
2	Day	Varchar	วันที่	
3	Day of Week	Varchar	วันในแต่ละสัปดาห์ เช่น จันทร์ อังคาร	
4	Month	Varchar	เดือน	
5	Quarter	Varchar	ไตรมาส	
6	Year	Varchar	ปี	

ตารางที่ 4.10 ตารางมิติ DT_Age ใช้สำหรับเก็บข้อมูลอายุ

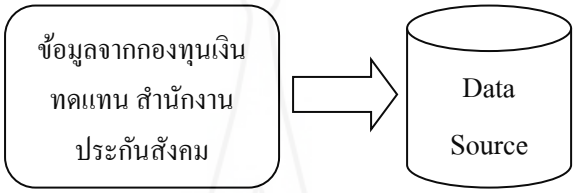
ลำดับ	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	Age	Int	อายุ	Primary Key
2	AgeRange	Varchar	ช่วงอายุ	

ตารางที่ 4.11 ตารางชั่วคราว Temp_Data ใช้สำหรับเก็บข้อมูลการใช้สิทธิลูกจ้างจาก Excel File

ลำดับ	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	Code Injury	Varchar	เลขที่ประสบอันตราย	Primary Key
2	Date1	Date time	วันที่รับเรื่องประสบอันตราย	
3	Code Position	Varchar	รหัสตำแหน่ง	
4	Name Position	Varchar	ชื่อตำแหน่ง	
5	Sex	Varchar	เพศ	
6	Age	Int	อายุ	
7	Code Enterprise	Varchar	รหัสสถานประกอบการ	
8	Name Enterprise	Varchar	ชื่อสถานประกอบการ	
9	Code Business Type	Varchar	รหัสประเภทกิจการ	
10	Date Injury	Date Time	วันที่ประสบอันตราย	
11	Code Organ	Varchar	รหัสอวัยวะ	
12	ผล	Varchar	รหัสผลของการประสบอันตราย	
13	Code Cause Injury	Varchar	รหัสสาเหตุการประสบอันตราย	
14	สิ่ง	Varchar	เพศ	
15	Name Cause Injury	Varchar	ชื่อสาเหตุการประสบอันตราย	
16	Medical	Varchar	ค่ารักษาพยาบาล	

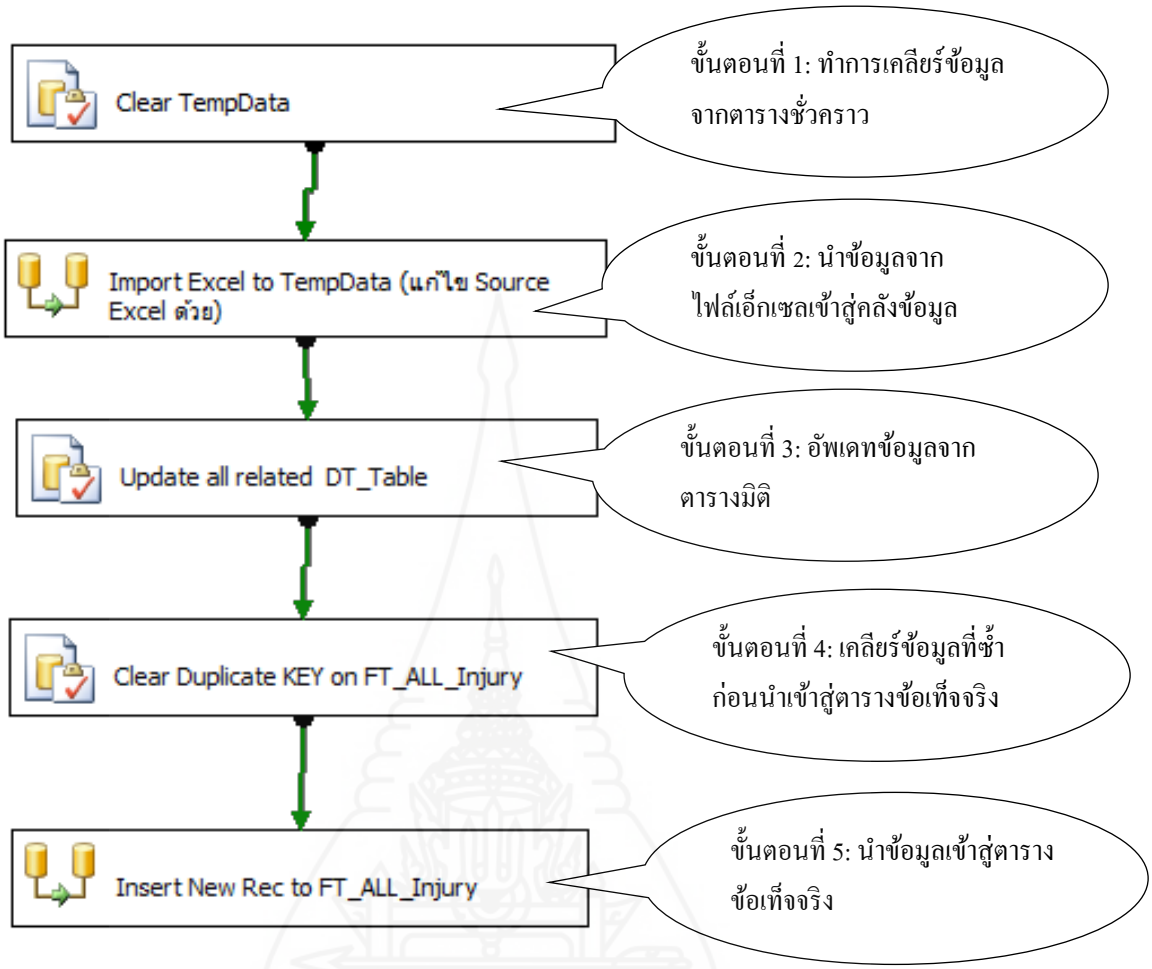
1.3 การสร้างคลังข้อมูล

1.3.1 การนำข้อมูลจากฐานข้อมูลเข้าสู่คลังข้อมูล ด้วยข้อมูลจากไฟล์เอ็กเซล โดยเป็นไฟล์ข้อมูลการใช้สิทธิของลูกจ้างที่เบิกเงินกองทุนทดแทน ชื่อไฟล์ คือ ALL_Injury 2007.xlsx โดยใช้เครื่องมือ SQL Server Integration Services (SSIS) ของโปรแกรม Microsoft SQL Server 2008 R2 สร้างแพ็คเกจขึ้นมา โดยใช้ชื่อว่า ETL Transection



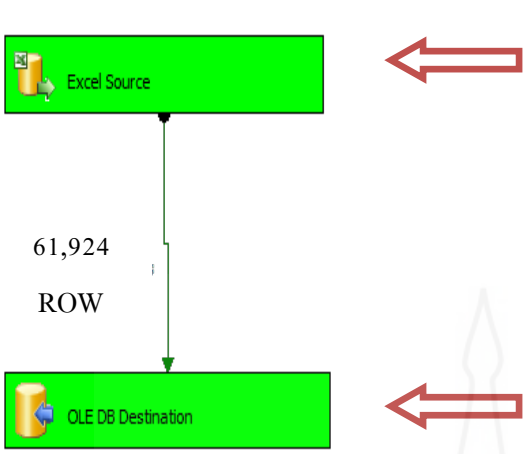
ภาพที่ 4.2 การนำข้อมูลเข้าสู่คลัง

1.3.2 การสร้างขั้นตอนกระบวนการทำงานอีทีแอล (Extract Transform and Load: ETL) โดยมีขั้นตอนการทำกระบวนการ ดังนี้ 1) ทำการเคลียร์ข้อมูลจากตารางชั่วครว (Temp Data) 2) ดึงข้อมูลจากไฟล์เอ็กเซล (Excel File) เข้าสู่คลังข้อมูล 3) ใช้คำสั่งเพื่อทำการอัพเดทข้อมูลจากตารางมิติทั้งหมด 4) เคลียร์ข้อมูลที่มีค่าซ้ำซ้อนให้มีความถูกต้องก่อนจะนำข้อมูลเข้าสู่ตารางข้อเท็จจริงในคลังข้อมูล 5) นำข้อมูลทั้งหมดที่ผ่านการเคลียร์ข้อมูลและทำความสะอาดข้อมูลให้มีความถูกต้องแล้วเข้าสู่ตารางข้อเท็จจริง และพร้อมใช้งานในการจัดทำคิวบ์ข้อมูล เพื่อแสดงการประมวลผลวิเคราะห์เชิงออนไลน์ ดังภาพที่ 4.3



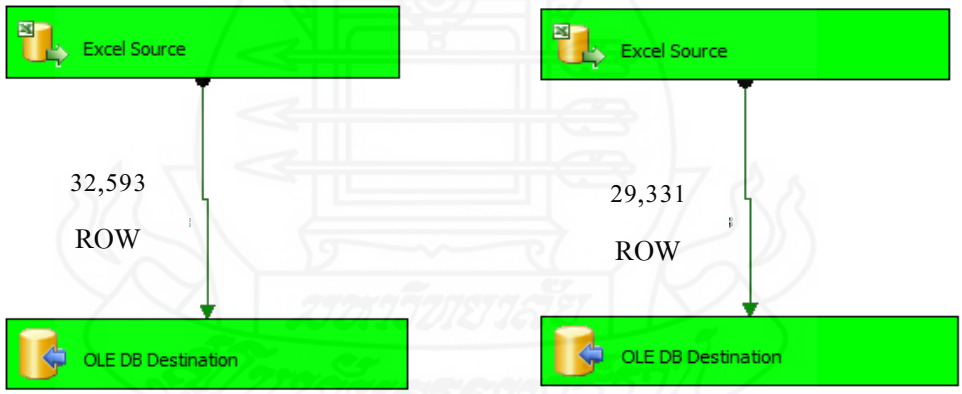
ภาพที่ 4.3 การทำกระบวนการอีทีแอลด้วยเครื่องมือ SQL Server Integration Services (SSIS)

ในกระบวนการนำเข้าข้อมูลแบ่งออกเป็น การถ่ายโอนข้อมูลนำเข้าสู่ที่พักข้อมูล จากนั้นทำการคัดลอกข้อมูลจากที่พักข้อมูลเข้าสู่ตารางข้อเท็จจริง และข้อมูลตารางมิติหรือมุมมองที่ออกแบบไว้นั้น และขั้นสุดท้ายเป็นการล้างข้อมูลออกจากที่พักข้อมูล ดังภาพที่ 4.4



ภาพที่ 4.4 ขั้นตอนของการทำ (Extract Transform and Load : ETL)

ผลที่ได้จากการทำกระบวนการอีทีแอล (Extract Transform and Load : ETL) ด้วยเครื่องมือ SQL Server Integration Services (SSIS) โดยโปรแกรม Microsoft SQL Server 2008 R2 จะได้ข้อมูลการใช้สิทธิของลูกจ้างที่เบิกเงินกองทุนทดแทน สำนักงานประกันสังคม ทั้งหมด 2 ปี คือ ปี 2556 – 2557 ดังภาพที่ 4.5



ภาพที่ 4.5 การแสดงผลรวมของข้อมูลผ่านกระบวนการ (Extract Transform and Load : ETL)

หลังจากได้นำข้อมูลเข้าสู่คลังข้อมูลแล้ว พบว่า มีจำนวนข้อมูลในตารางรวมการใช้สิทธิของลูกจ้างสถานประกอบการที่ประสบอันตราย (FT_ALL_Injury) ในปี 2556 และปี 2557 ทั้งหมดจำนวนข้อมูล 61,924 ชุดข้อมูล ดังภาพที่ 4.6

Results | Messages

	Code Injury	Date Injury	Code Area	Code Position	Code Enterprise	Code Business Type	Code Organ	Code Cause Injury	Sex	Age
61910	101257/02545/03	2014-11-12 00:00:00.000	1012	915	1000402088	1616	0700	07	M	16
61911	101257/02546/04	2014-11-07 00:00:00.000	1012	931	1000096467	1301	0123	18	M	32
61912	101257/02547/01	2014-10-03 00:00:00.000	1012	931	1000096467	1301	0651	07	M	20
61913	101257/02548/02	2014-10-09 00:00:00.000	1012	931	1000096467	1301	0541	07	M	38
61914	101257/02549/03	2014-10-03 00:00:00.000	1012	931	1000096467	1301	0652	07	M	51
61915	101257/02550/04	2014-10-02 00:00:00.000	1012	931	1000096467	1301	0121	08	F	41
61916	101257/02551/01	2014-10-14 00:00:00.000	1012	931	1000096467	1301	0121	08	F	45
61917	101257/02552/02	2014-11-11 00:00:00.000	1012	931	1000096467	1301	0522	01	F	49
61918	101257/02553/03	2014-11-05 00:00:00.000	1012	931	1000096467	1301	0512	04	F	21
61919	101257/02554/04	2014-11-03 00:00:00.000	1012	513	1000081877	1602	0662	05	F	20
61920	101257/02555/01	2014-10-30 00:00:00.000	1012	824	1000500462	0405	0661	04	M	39
61921	101257/02556/02	2014-10-29 00:00:00.000	1012	341	1000140164	1606	0551	07	M	34
61922	101257/02557/02	2014-11-27 00:00:00.000	1012	522	1000562611	1503	0652	07	M	28
61923	101257/02558/03	2014-10-28 00:00:00.000	1012	513	1000081877	1602	0662	04	M	18
61924	101257/02559/04	2014-11-06 00:00:00.000	1012	915	1001616481	1616	0522	04	F	45

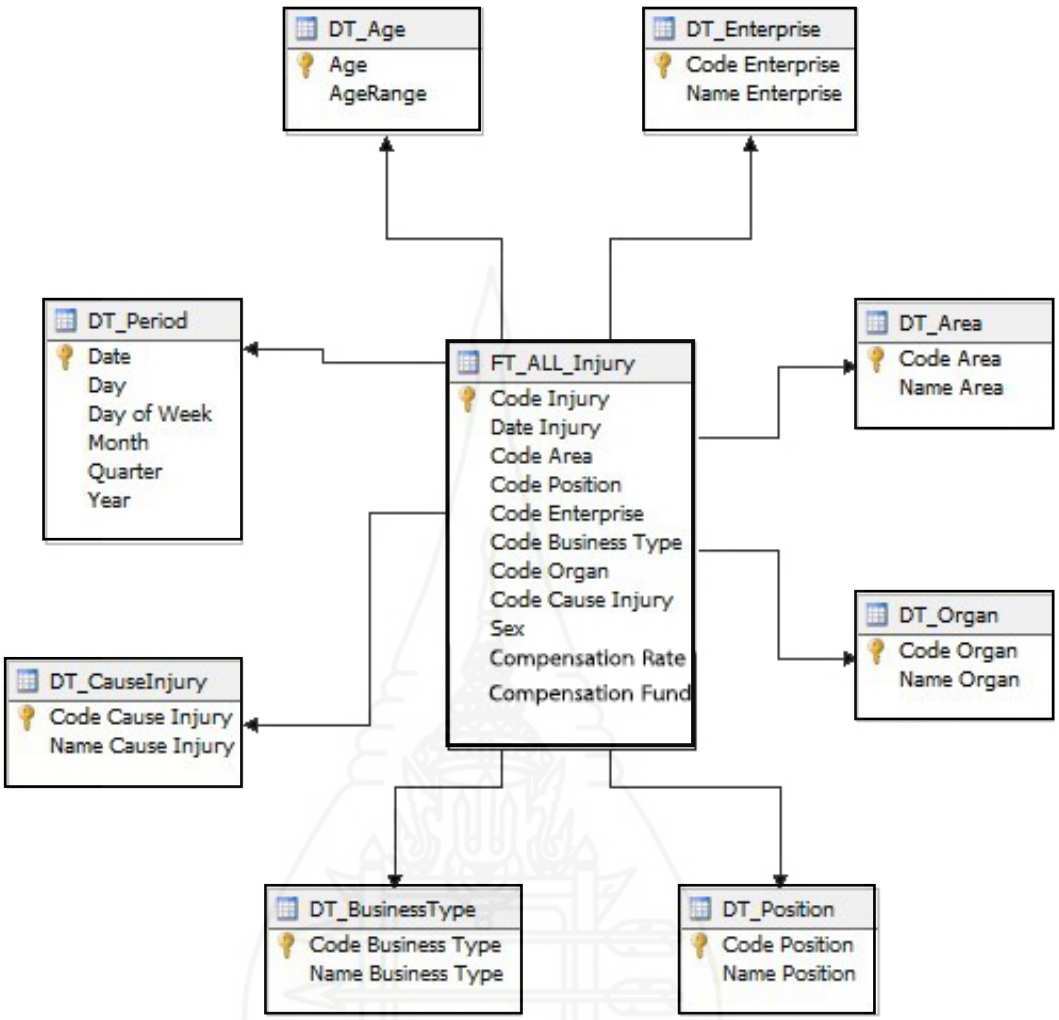
Query executed successfully. | STUDIO (10.50 RTM) | StudiO\StudiQ (54) | master | 00:00:01 | 61924 rows

ภาพที่ 4.6 การแสดงผลรวมของข้อมูลในคลังข้อมูล

2. การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ

2.1 การสร้างคิวบ์ (Club)

ผู้วิจัยทำการสร้างคิวบ์ (Cube) ใช้ข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทนจากคลังข้อมูล จำนวน 61,294 ชุดข้อมูล ด้วยเครื่องมือ SQL Server Analysis Service (SSAS) ซึ่งได้เชื่อมต่อกับคลังข้อมูล ของโปรแกรม Microsoft SQL Server 2008 R2 จากนั้นทำการสร้าง Data Source View ซึ่งประกอบด้วยมิติมุมมอง (Dimension) ได้แก่ มิติสำนักงานประกันสังคมพื้นที่ (DT_Area) , มิติตำแหน่ง (DT_Position), มิติสถานประกอบการ (DT_Enterprise), มิติประเภทกิจการ (DT_BusinessType) มิติอวัยวะ (DT_Organ) มิติสาเหตุการประสบอันตราย (DT_CauseInjury) มิติระยะเวลา (DT_Period) และมิติอายุ (DT_Age) โดยมีเมเชอร์หรือค่าที่ต้องการวัด (Measure) คือ จำนวนรายการประสบอันตราย (FT_ALL_Injury) ดังภาพที่ 4.7



ภาพที่ 4.7 Data Source View

จากนั้นสร้างคิวบ์ (Club) จะสามารถดูข้อมูลที่เป็นการประมวลผลเชิงวิเคราะห์ออนไลน์ในลักษณะหลายมิติ (Online Analytical Processing : OLP) ซึ่งผู้ใช้งานสามารถเปลี่ยนมุมมอง (Dimension) ได้แก่ มิติปีที่ประสบอันตราย มิติตำแหน่ง มิติสถานประกอบการ มิติประเภทกิจการ มิติอวัยวะ มิติสาเหตุการประสบอันตราย มิติเพศ และมิติอายุ และเมเชอร์หรือค่าที่ต้องการวัด (Measure) คือ จำนวนรายการประสบอันตราย ดังภาพที่ 4.8

		Year							
		2550	2552	2553	2554	2555	2556	2557	Grand Total
Name Business Type	Age Range	จำนวนรายการ	จำนวนรายการ	จำนวนรายการ	จำนวนรายการ	จำนวนรายการ	จำนวนรายการ	จำนวนรายการ	จำนวนรายการ
⊟ การก่อสร้าง	15-21				3.00	76.00	386.00	370.00	835.00
	22-25				1.00	86.00	556.00	485.00	1,128.00
	26-30				2.00	101.00	637.00	591.00	1,331.00
	31-40			1.00	3.00	219.00	1,377.00	1,133.00	2,733.00
	41-50			2.00	4.00	193.00	1,214.00	939.00	2,352.00
	51-60				2.00	93.00	497.00	386.00	978.00
	61-70					5.00	29.00	20.00	54.00
	71-80							1.00	1.00
	Total			3.00	15.00	773.00	4,696.00	3,925.00	9,412.00
⊟ การค้าวัสดุก่อสร้าง	15-21					12.00	106.00	98.00	216.00
	22-25					19.00	97.00	56.00	172.00
	26-30					32.00	167.00	138.00	337.00
	31-40				1.00	45.00	383.00	269.00	698.00
	41-50					29.00	240.00	180.00	449.00
	51-60					6.00	87.00	64.00	157.00
	61-70						16.00	9.00	25.00
	Total				1.00	143.00	1,096.00	814.00	2,054.00
⊟ ขนส่งสินค้า ผู้โดยสารทางรถยนต์	15-21					7.00	54.00	44.00	105.00
	22-25					20.00	78.00	48.00	146.00
	26-30					24.00	109.00	116.00	249.00
	31-40				2.00	47.00	305.00	239.00	593.00
	41-50	1.00		3.00	2.00	37.00	225.00	171.00	439.00
	51-60				2.00	13.00	93.00	77.00	185.00
	61-70				1.00	3.00	11.00	10.00	25.00
	Total	1.00		3.00	7.00	151.00	875.00	705.00	1,742.00

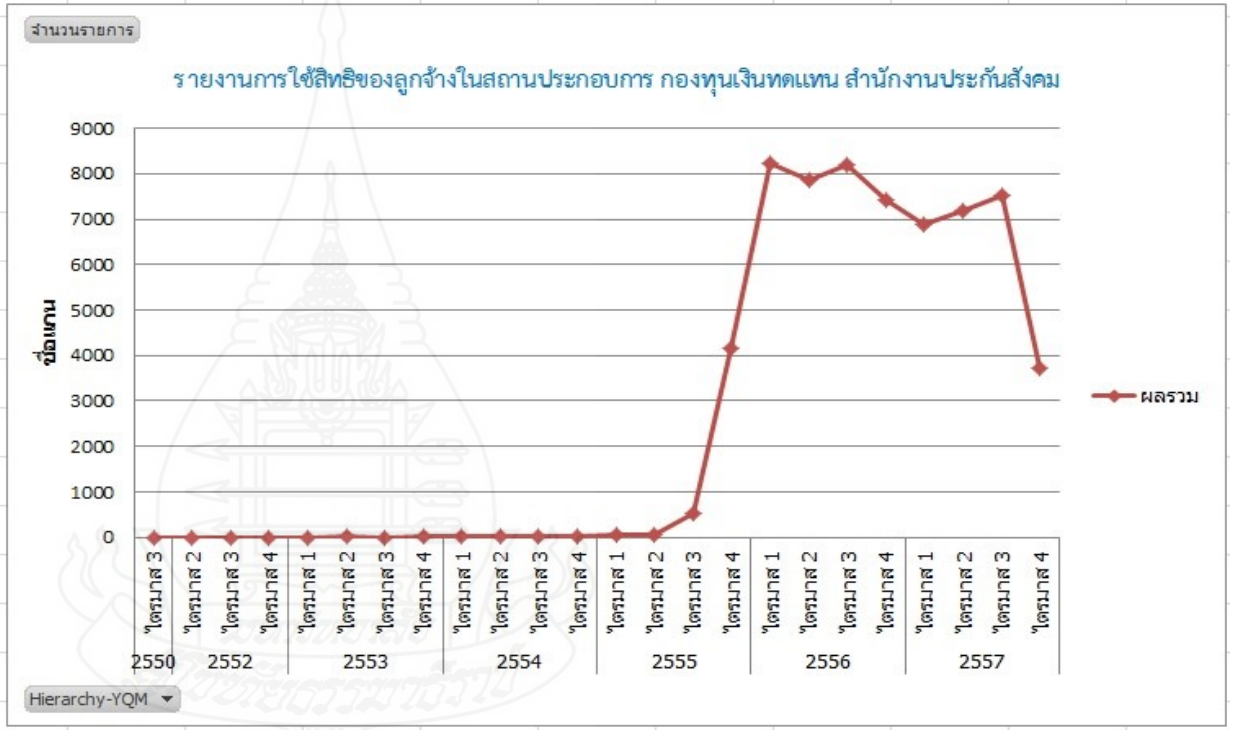
ภาพที่ 4.8 การประมวลผลเชิงวิเคราะห์แบบออนไลน์

จากภาพที่ 4.8 แสดงให้เห็นถึงข้อมูลในหลายมิติ ด้านบนจะเป็นข้อมูลของปีการประสบอันตรายที่มีการเบิกเงินกองทุนทดแทน ส่วนด้านซ้ายจะเป็นข้อมูลของประเภทกิจการและช่วงอายุที่มีการประสบอันตรายที่มีการเบิกเงินกองทุนทดแทน แสดงให้เห็นถึงผลของการประสบอันตรายแต่ละประเภทกิจการ เช่น สถานประกอบการที่ทำประเภทกิจการก่อสร้าง จะมีการประสบอันตรายมากที่สุด โดยจะมีตั้งแต่ปี 2553 – 2557 โดยรวมแล้วมีถึง 9,412 ราย เป็นต้น ซึ่งในการแสดงผลข้อมูลหลายมิติสามารถปรับเปลี่ยนข้อมูลหรือมิติในการดูข้อมูลได้หลากหลาย ดังภาพที่ 4.8 และถ้าหากต้องการดูข้อมูลลึกลงไปอีกก็สามารถปรับเปลี่ยนรูปแบบการแสดงผลได้อีกเช่นกัน

2.2 ระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติตามหลักธุรกิจอัจฉริยะ (Online Analytical Processing)

การแสดงรายงานด้วยโปรแกรม Microsoft Excel 2007 ในส่วนของ Pivot Table ในการนำเสนอรายงานให้ผู้ใช้งานทั้ง 3 กลุ่มได้แก่ ผู้บริหาร หัวหน้าส่วนงานกองทุนเงินทดแทน และเจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทน ตามมุมมองต่าง ๆ ผู้บริหารใช้ดูภาพรวมทั้งหมดของสำนักงานประกันสังคม รูปแบบของการนำเสนอรายงานนั้นเป็นข้อมูลจำนวนการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน ในแต่ละปีงบประมาณ ดังภาพที่ 4.9

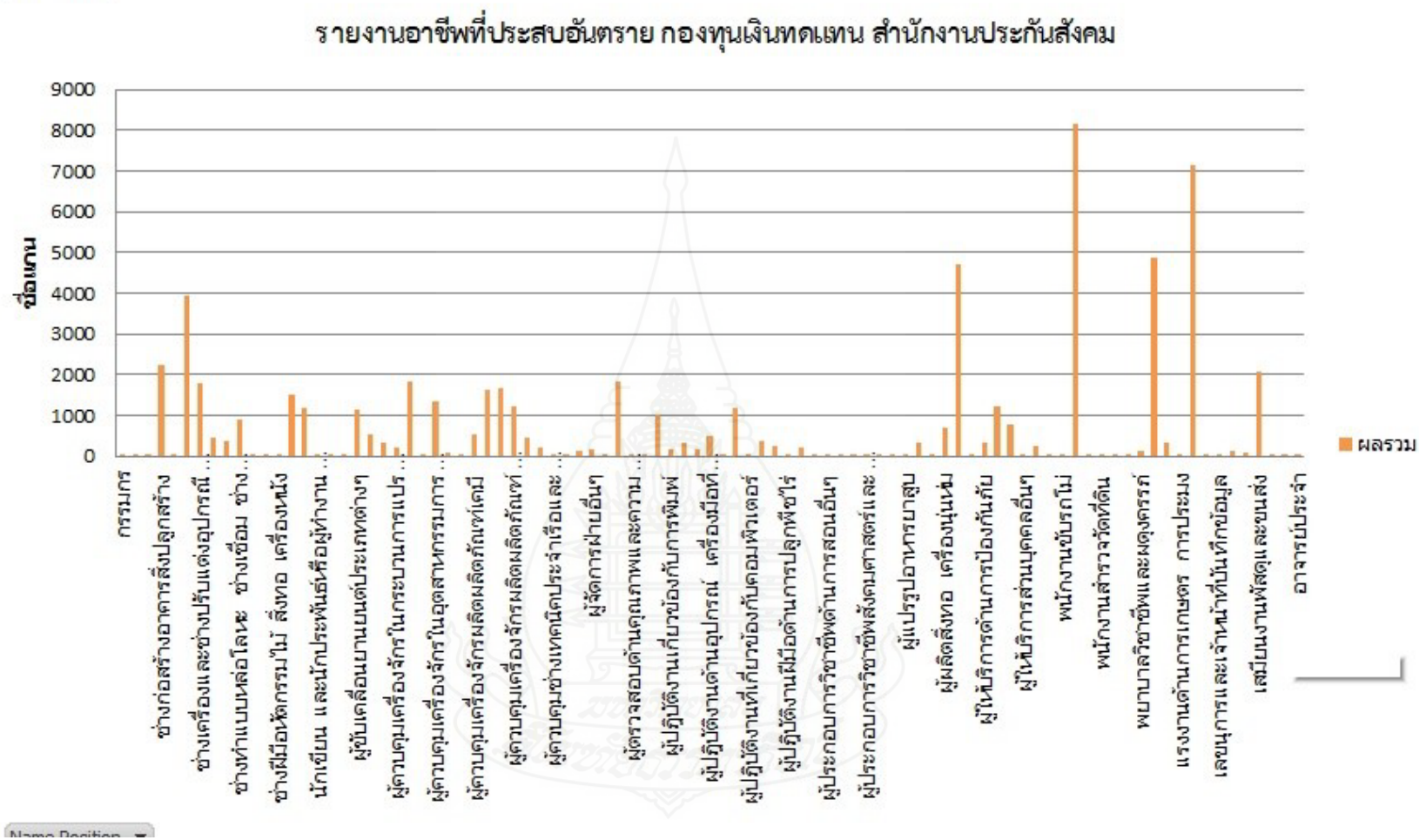
ป้ายชื่อแถว	จำนวนรายการ
⊞ไตรมาส 2	18
⊞ไตรมาส 3	18
⊞ไตรมาส 4	20
⊟2555	4815
⊞ไตรมาส 1	50
⊞ไตรมาส 2	71
⊞ไตรมาส 3	523
⊞ไตรมาส 4	4171
⊟2556	31690
⊞ไตรมาส 1	8221
⊞ไตรมาส 2	7875
⊞ไตรมาส 3	8179
⊞ไตรมาส 4	7415
⊟2557	25320
⊞ไตรมาส 1	6888
⊞ไตรมาส 2	7184
⊞ไตรมาส 3	7509
⊞ไตรมาส 4	3739
ผลรวมทั้งหมด	61924



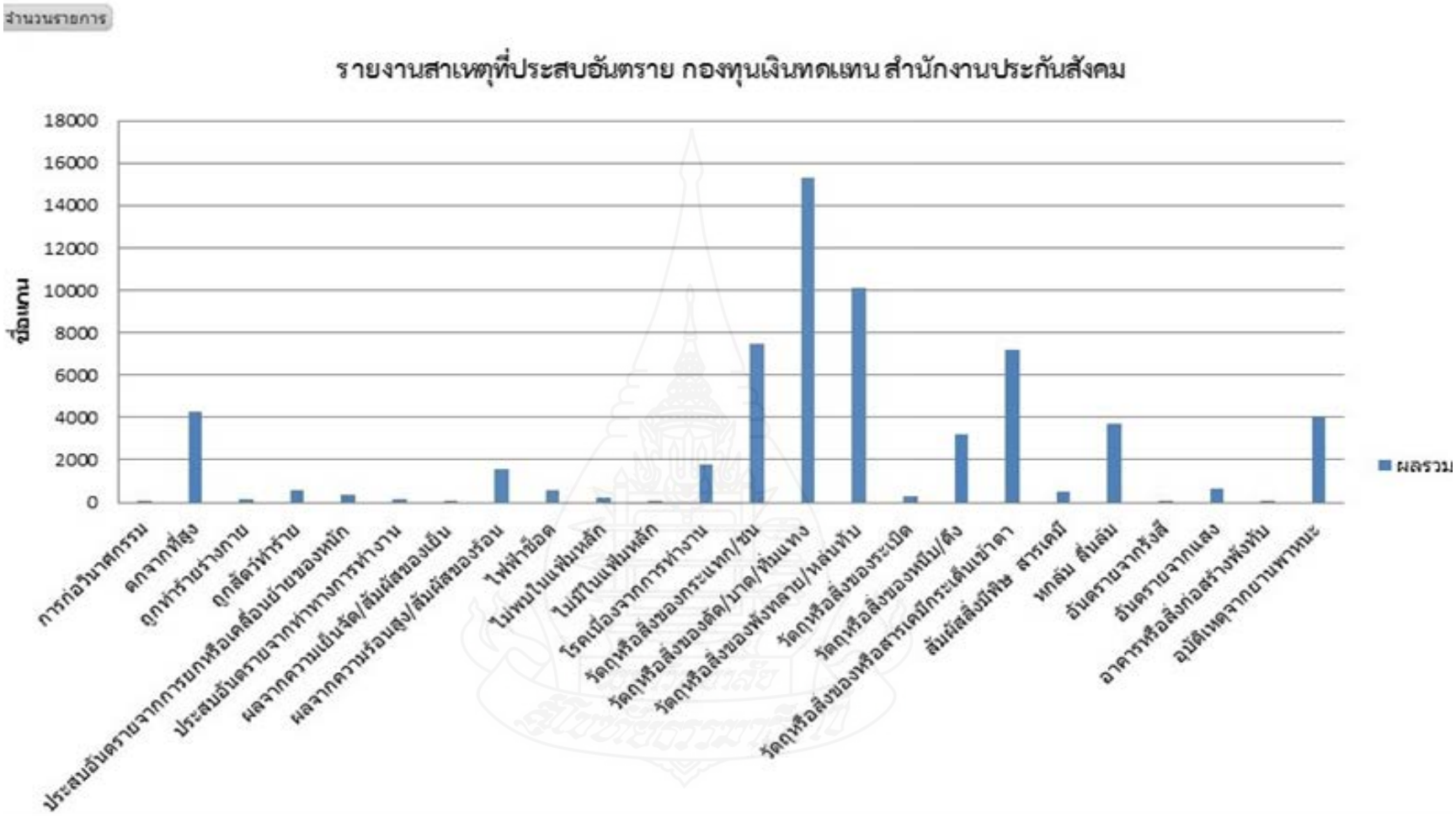
ภาพที่ 4.9 รายงานประจำปีสำหรับผู้บริหาร

หัวหน้าส่วนงานกองทุนเงินทดแทนใช้ดูข้อมูลในส่วนของการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน รูปแบบของการนำเสนอรายงานนั้นข้อมูลเป็นในส่วนของตำแหน่งที่ประสบอันตราย สถานประกอบการที่ประสบอันตราย ประเภทกิจการที่ประสบอันตราย และช่วงอายุที่ประสบอันตราย ดังภาพที่ 4.10

เจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทนใช้ดูข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน รูปแบบของการนำเสนอรายงานนั้นเป็นข้อมูลรายบุคคล เพื่อนำไปวิเคราะห์การจ่ายค่ารักษาพยาบาลและค่าทดแทนต่างๆ ดังภาพที่ 4.11



ภาพที่ 4.10 รายงานสำหรับหัวหน้าส่วนงานกองทุนเงินทดแทน



ภาพที่ 4.11 รายงานสำหรับเจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทน

3. การทำเหมืองข้อมูล

งานวิจัยนี้ดำเนินการทำเหมืองข้อมูลตามแบบจำลองของคริสพ์ – ดีเอ็ม (CRISP – DM Model) ซึ่งประกอบด้วย 6 ขั้นตอน ดังนี้

3.1 ทำความเข้าใจปัญหาและโอกาสทางธุรกิจ (Business Understanding)

ผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้อง และได้ศึกษาข้อมูลเกี่ยวกับการพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน และวิเคราะห์ปัญหาที่เกิดขึ้นกับสำนักงานประกันสังคม พบว่า ข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทนมีความซ้ำซ้อนและผิดพลาดบ่อย ผู้บริหารไม่สามารถดูรายงานข้อมูลมาประกอบการตัดสินใจในการจัดตั้งงบประมาณประจำปีได้ทันที และยังพบว่ามีการเกิดการประสบอันตรายขึ้นบ่อยครั้ง ซึ่งทำให้ยากต่อการวิเคราะห์ในการป้องกันการประสบอันตราย หากมีเครื่องมือที่สามารถทำการพยากรณ์ และวิเคราะห์ข้อมูลให้ผู้บริหารสามารถตัดสินใจจัดตั้งงบประมาณได้ และยังใช้ในการป้องกันการประสบอันตรายได้อีกด้วย

3.2 ทำความเข้าใจข้อมูล (Data Understanding)

ผู้วิจัยทำการรวบรวมข้อมูลที่เกี่ยวข้องมาทำการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล ผู้วิจัยจึงได้สร้างคลังข้อมูลดังกล่าวขึ้นมา ซึ่งเป็นข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทนโดยใช้ข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 1-12 ปี 2556 - 2557 จำนวนข้อมูล 61,924 ชุดข้อมูล ในการทำเหมืองข้อมูลใช้ข้อมูลที่มาจากคลังข้อมูลที่ได้สร้างขึ้น และผ่านกระบวนการอีทีแอล โดยใช้เครื่องมือ SQL Server Integration Services (SSIS) ของโปรแกรม Microsoft SQL Server 2008 R2 และนำข้อมูลที่ได้ออกมาให้อยู่ในรูปแบบของไฟล์เอ็กเซล (Excel File) ซึ่งประกอบด้วยข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน ปี 2557 จำนวนข้อมูล 3,636 ชุดข้อมูล มาใช้เป็น (Data Set) ใช้ในการสร้างตัวแบบพยากรณ์ โดยจะประกอบด้วยชุดข้อมูล ดังนี้ รหัสประสบอันตราย (Code Injury), วันที่ประสบอันตราย (Date Injury), รหัสพื้นที่สำนักงานประกันสังคม (Code Area), รหัสตำแหน่งหน้าที่ (Code Position), รหัสสถานประกอบการ (Code Enterprise), รหัสประเภทกิจการ (Code Business Type), รหัสอวัยวะ (Code Organ), รหัสสาเหตุการประสบอันตราย (Code Cause Injury), เพศ (Sex), และ อายุ (Age) ดังภาพที่ 4.12

	A	B	C	D	E	F	G	H	I	J	K
1	Code Injury	Date Injury	Code Area	Code Position	Code Enterprise	Code Business Type	Code Organ	Code Cause Injury	Sex	Age	
2	100357/00001/01	9/10/2013 0:00	1003	832	1002383501	1402	410	1	M	45	
3	100357/00002/01	29/10/2013 0:00	1003	832	1001290437	1402	121	11	M	30	
4	100357/00003/01	9/12/2013 0:00	1003	832	1000118681	1402	521	11	M	33	
5	100357/00004/01	16/11/2013 0:00	1003	931	1000018741	1301	552	7	M	39	
6	100357/00005/01	1/12/2013 0:00	1003	931	1000069648	1301	110	1	M	48	
7	100357/00006/01	6/12/2013 0:00	1003	931	1000018741	1301	561	7	M	36	
8	100357/00007/01	2/12/2013 0:00	1003	931	1000018741	1301	310	4	M	44	
9	100357/00008/01	13/10/2013 0:00	1003	931	1000018741	1301	110	1	M	40	
10	100357/00009/01	10/12/2013 0:00	1003	712	1000087352	1301	700	1	M	41	
11	100357/00010/02	21/9/2013 0:00	1003	712	1000045269	1301	632	11	M	39	
12	100357/00011/02	17/5/2013 0:00	1003	931	1000018741	1301	542	5	F	49	
13	100357/00012/02	19/10/2013 0:00	1003	721	1000125980	1301	410	1	M	35	
14	100357/00013/02	2/11/2013 0:00	1003	931	1000018741	1301	562	6	M	28	
15	100357/00014/02	19/11/2013 0:00	1003	931	1000018741	1301	121	8	M	30	
16	100357/00015/02	1/10/2013 0:00	1003	342	1002713579	1604	122	8	M	43	
17	100357/00016/02	17/11/2013 0:00	1003	931	1000018741	1301	541	14	M	52	
18	100357/00017/02	5/11/2013 0:00	1003	341	1002178584	1601	122	8	F	23	
19	100357/00018/03	16/11/2013 0:00	1003	341	1000063305	1601	631	11	M	32	
20	100357/00019/03	4/12/2013 0:00	1003	931	1001299124	1301	562	4	M	58	
21	100357/00020/03	3/12/2013 0:00	1003	931	1000018741	1301	651	7	M	29	
22	100357/00021/03	15/11/2013 0:00	1003	931	1000018741	1301	651	4	M	25	
23	100357/00022/03	4/12/2013 0:00	1003	712	1001299124	1301	700	4	F	34	
24	100357/00023/03	10/11/2013 0:00	1003	931	1000018741	1301	110	5	M	40	
25	100357/00024/03	21/11/2013 0:00	1003	724	1000018741	1301	160	14	M	30	
26	100357/00025/03	23/11/2013 0:00	1003	931	1000018741	1301	542	7	M	30	
27	100357/00026/04	26/11/2013 0:00	1003	522	1000689069	1501	541	5	M	37	
28	100357/00027/04	17/11/2013 0:00	1003	931	1000174492	1301	632	7	M	38	
29	100357/00028/04	21/11/2013 0:00	1003	522	1000336972	1501	551	4	M	40	
30	100357/00029/04	28/8/2013 0:00	1003	712	1000018741	1301	542	5	M	55	
31	100357/00030/04	11/11/2013 0:00	1003	522	1000288251	1501	122	7	F	38	
32	100357/00031/04	29/8/2013 0:00	1003	723	1000425185	1610	123	18	M	36	

ภาพที่ 4.12 ข้อมูลจากคลังข้อมูลในรูปแบบของไฟล์เอ็กเซล (Excel File)

3.3 การเตรียมข้อมูล (Data Preparation)

นำข้อมูลที่จากคลังข้อมูลออกมาในรูปแบบของไฟล์ Excel โดยข้อมูลที่ได้ออกมายังไม่สามารถนำไปทำการวิเคราะห์ด้วยเทคนิคการทำเหมืองข้อมูลได้ เพื่อให้ได้ผลการวิจัยที่มีประสิทธิภาพ ต้องทำขั้นตอนการเตรียมข้อมูล (Data Preparation) เพื่อใช้ในการทำเหมืองข้อมูลดังต่อไปนี้

3.3.1 การคัดกรองข้อมูล (Data Cleaning) คัดกรองข้อมูลที่ไม่เกี่ยวข้องหรือข้อมูลที่ผิดปกติออก โดยเลือกใช้ข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 3 ปี 2557 และนำมาทำ Data Set จำนวน 3,636 ชุดข้อมูล

3.3.2 การเลือกข้อมูล (Data Selection) ในการเลือกข้อมูลที่ใช้ในการวิเคราะห์ผู้วิจัยได้เลือกใช้ข้อมูลที่จะนำมาทำการพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน โดยใช้คุณลักษณะของลูกจ้างที่มีการใช้สิทธิ ซึ่งจะมีคุณลักษณะคือ เลขที่ประสบอันตราย, วันที่ประสบอันตราย, สำนักงานประกันสังคมเขตพื้นที่, รหัสตำแหน่ง, เลขที่บัญชีสถานประกอบการ, รหัสกิจการ, รหัสอวัยวะ, รหัสสาเหตุการประสบอันตราย, เพศ และอายุ ดังตารางที่ 4.12

ตารางที่ 4.12 คุณลักษณะที่ใช้ในการสร้างตัวแบบพยากรณ์

ข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน	
ชื่อข้อมูล	**ความหมาย**
Code Injury	เลขที่ประสบอันตราย
Code Position	รหัสตำแหน่ง
Code Enterprise	เลขที่บัญชีสถานประกอบการ
Code Business Type	รหัสกิจการ
Code Organ	รหัสอวัยวะ
Code Cause Injury	รหัสสาเหตุการประสบอันตราย
Sex	เพศ
Age	อายุ
Group Age	ช่วงอายุ
Compensation Rate	การจ่ายค่ารักษาพยาบาลและค่าทดแทน
Compensation Fund	การวินิจฉัยจ่ายค่าทดแทน

3.3.3 การแปลงข้อมูล (Data Transformation)

ในขั้นตอนการแปลงข้อมูลใช้โปรแกรม Microsoft Excel 2007 ในการแปลงข้อมูล ซึ่งข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน มีความละเอียดมากเกินไปผู้วิจัยจึงต้องทำการจัดกลุ่มผลการจ่ายค่ารักษาพยาบาลและค่าทดแทนเฉลี่ยใหม่ โดยแบ่งการจ่ายค่ารักษาพยาบาลและค่าทดแทน ออกเป็น 3 กลุ่ม เพื่อใช้ในการสร้างตัวแบบพยากรณ์ (Model) ซึ่งช่วยลดการกระจายของข้อมูลดังนี้

High คือ มีการจ่ายค่ารักษาพยาบาลและค่าทดแทน ตั้งแต่ 300,000 บาท

Medium คือ มีการจ่ายค่ารักษาพยาบาลและค่าทดแทน ตั้งแต่ 110,000 บาท

Low คือ มีการจ่ายค่ารักษาพยาบาลและค่าทดแทน ตั้งแต่ 45,000 บาท

รูปแบบของการแปลงข้อมูลที่เป็นผลลัพธ์ที่ต้องการ (Output) ดังตารางที่ 4.13 และข้อมูลที่เป็นคุณลักษณะ (Input) มีรายละเอียดของการแปลงข้อมูลดังตารางที่ 4.14

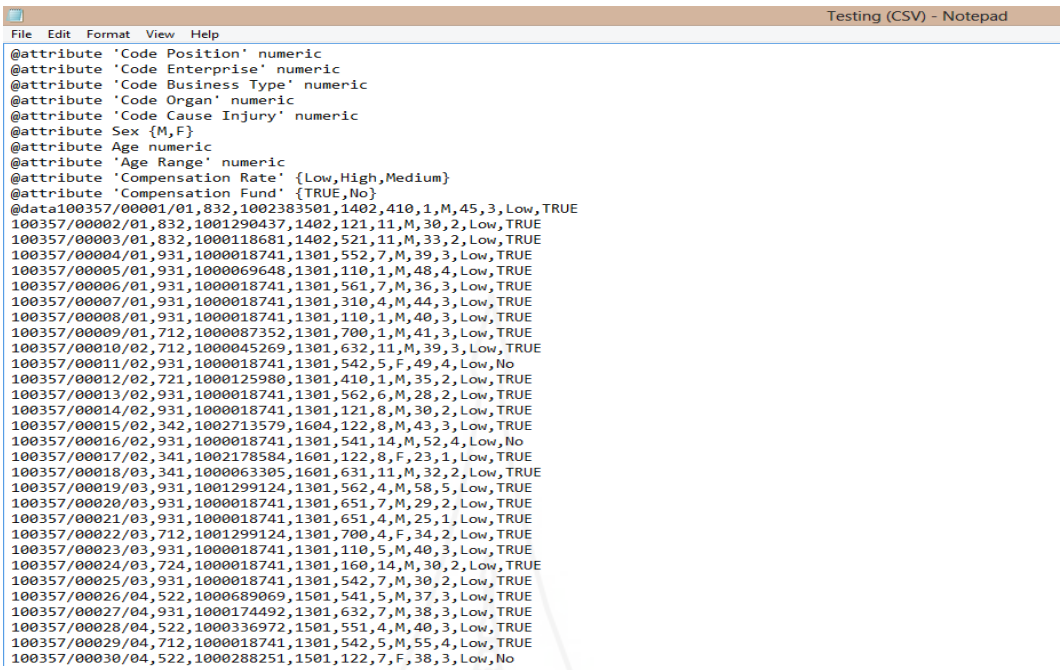
ตารางที่ 4.13 การแปลงข้อมูลที่เป็นผลลัพธ์

คุณลักษณะ	คำอธิบาย	ข้อมูลที่ถูกแปลง
Compensation Rate	การจ่ายค่ารักษาพยาบาลและค่าทดแทน	High = จ่ายค่ารักษาพยาบาลและค่าทดแทน ตั้งแต่ 300,000 บาท
		Medium = จ่ายค่ารักษาพยาบาลและค่าทดแทน ตั้งแต่ 110,000 บาท
		Low = จ่ายค่ารักษาพยาบาลและค่าทดแทน ตั้งแต่ 45,000 บาท

ตารางที่ 4.14 การแปลงคุณลักษณะ

ลำดับ	คุณลักษณะ	คำอธิบาย	ข้อมูลที่ถูกแปลง
1	Compensation Fund	หลักเกณฑ์การวินิจฉัยจ่ายเงิน	Yes = วินิจฉัยว่ามีสิทธิได้รับค่ารักาพยาบาลและค่าทดแทน No = วินิจฉัยว่าไม่มีสิทธิได้รับค่ารักาพยาบาลและค่าทดแทน
2	Age Range	ช่วงอายุ	1 = อายุระหว่าง 15 – 25 2 = อายุระหว่าง 26 – 35 3 = อายุระหว่าง 36 – 45 4 = อายุระหว่าง 46 – 60 5 = อายุระหว่าง 61 – 68

นำข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน สำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 3 ปี 2557 จากคลังข้อมูล จำนวน 3,636 ชุดข้อมูล มาทำการแปลงข้อมูลด้วยโปรแกรม Microsoft Excel 2007 แล้วบันทึกข้อมูลให้อยู่ในรูปแบบของไฟล์ CSV (Comma delimited) เพื่อให้สามารถใช้งานกับโปรแกรม WEKA ได้ และนำไฟล์เข้าโปรแกรม WEKA เพื่อทำการบันทึกให้อยู่ในรูปแบบไฟล์ ARFF เพื่อให้สามารถใช้ในการวิเคราะห์ข้อมูลได้ ดังภาพที่ 4.13



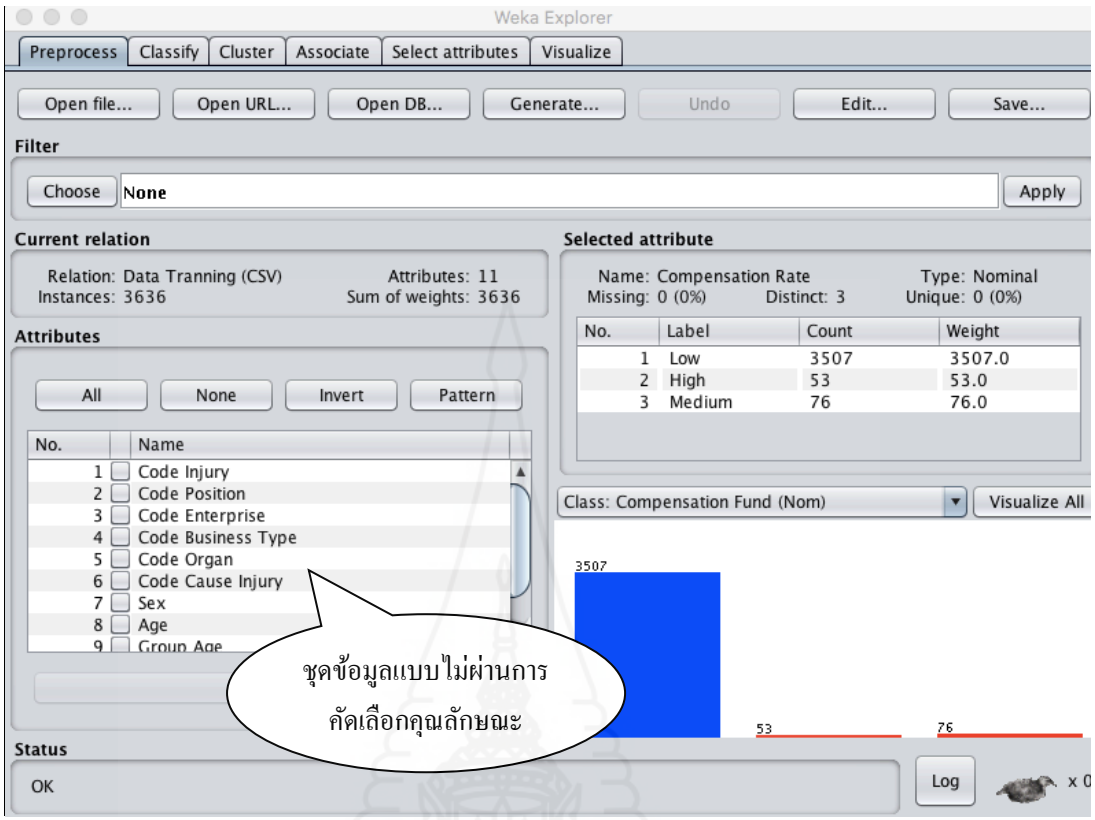
ภาพที่ 4.13 ข้อมูลที่แปลงในรูปแบบของไฟล์ CSV (Comma delimited)

3.4 การสร้างตัวแบบพยากรณ์ (Modeling)

การสร้างตัวแบบพยากรณ์ ใช้ข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนเงินทดแทน สำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 3 ปี 2557 ที่ผ่านการคัดเลือกข้อมูลแล้วนำมาทำ Data Set ซึ่งมีจำนวน 3,636 ชุดข้อมูล โดยการเลือกใช้ข้อมูล 2 แบบ คือ ชุดข้อมูลแบบไม่จัดกลุ่ม (Original Data) และชุดข้อมูลแบบจัดกลุ่ม (Clustering Data) ด้วยอัลกอริทึม k-means และผลที่ได้จากการวิจัยเป็นดังนี้

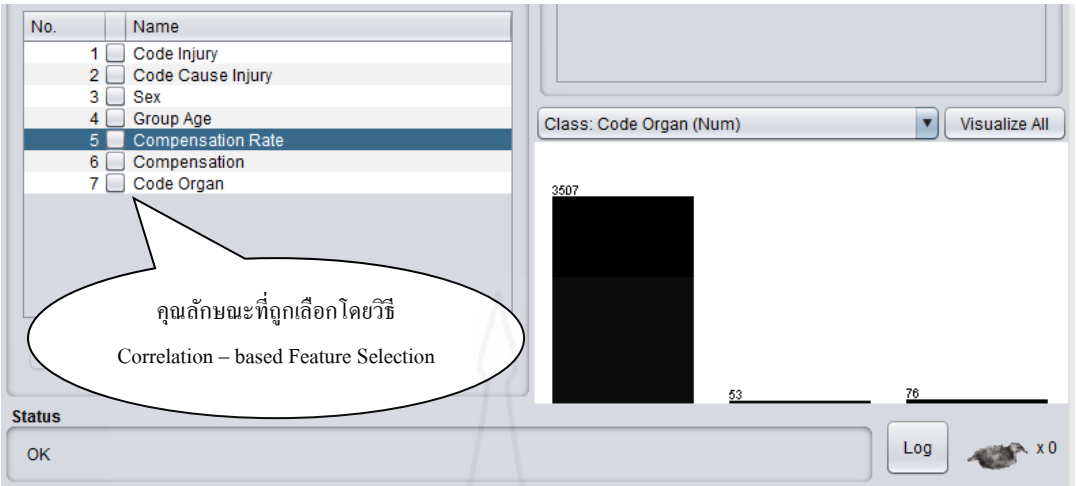
3.4.1 การสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม

1) ผู้วิจัยทำการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม โดยใช้ชุดข้อมูล (Data set) แบบไม่ผ่านการคัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะ (Feature Selection) ด้วยวิธี Correlation – based Feature Selection ซึ่งชุดข้อมูลที่ไม่มีการคัดเลือกคุณลักษณะดังภาพที่ 4.14



ภาพที่ 4.14 ชุดข้อมูลแบบไม่ผ่านการคัดเลือกคุณลักษณะ (Feature Selection)

จากการคัดเลือกคุณลักษณะ (Feature Selection) ด้วยวิธี Correlation – based Feature Selection (CFS) โดยใช้โปรแกรม WEKA พบว่า คุณลักษณะที่ถูกคัดเลือกจาก 11 แอตทริบิวต์ เหลือเพียง 6 แอตทริบิวต์ ดังภาพที่ 4.15 และได้ผลของการคัดเลือกคุณลักษณะ ด้วยวิธี Correlation – based Feature Selection ดังตารางที่ 4.15

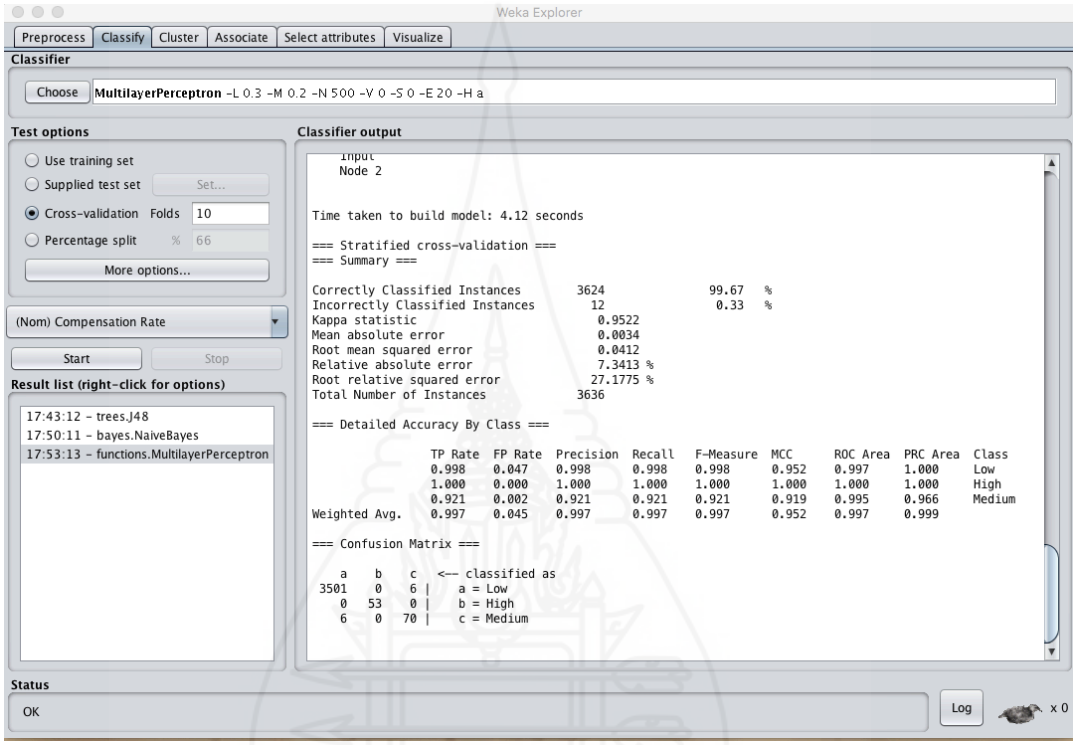


ภาพที่ 4.15 การคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection

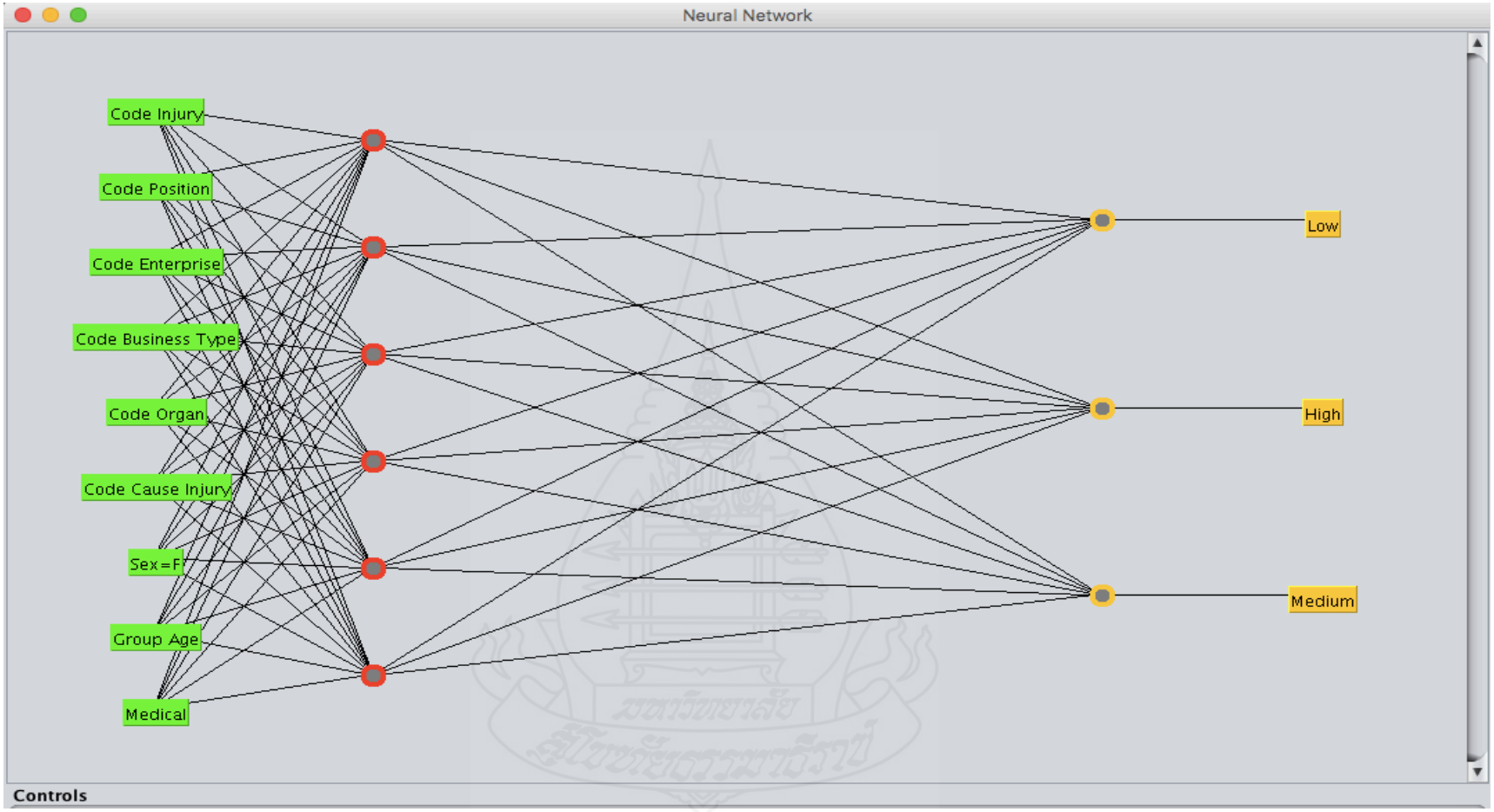
ตารางที่ 4.15 ผลการคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection

ลำดับที่	คุณลักษณะ	คำอธิบาย
1	Code Injury	เลขที่ประสบอันตราย
2	Code Cause Injury	รหัสสาเหตุการประสบอันตราย
3	Sex	เพศ
4	Group Age	ช่วงอายุ
5	Compensation Rate	การจ่ายค่ารักษาพยาบาลและค่าทดแทน
6	Compensation	การวินิจฉัยจ่ายค่าทดแทน

2) การใช้เทคนิคเหมืองข้อมูลโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ร่วมกับวิธีที่ไม่ผ่านการคัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection ผลการพยากรณ์ให้ค่าความถูกต้อง (Accuracy) เท่ากับร้อยละ 99.67 และมีค่าความคลาดเคลื่อน (RMSE) เท่ากับ 0.0412 ดังภาพที่ 4.16

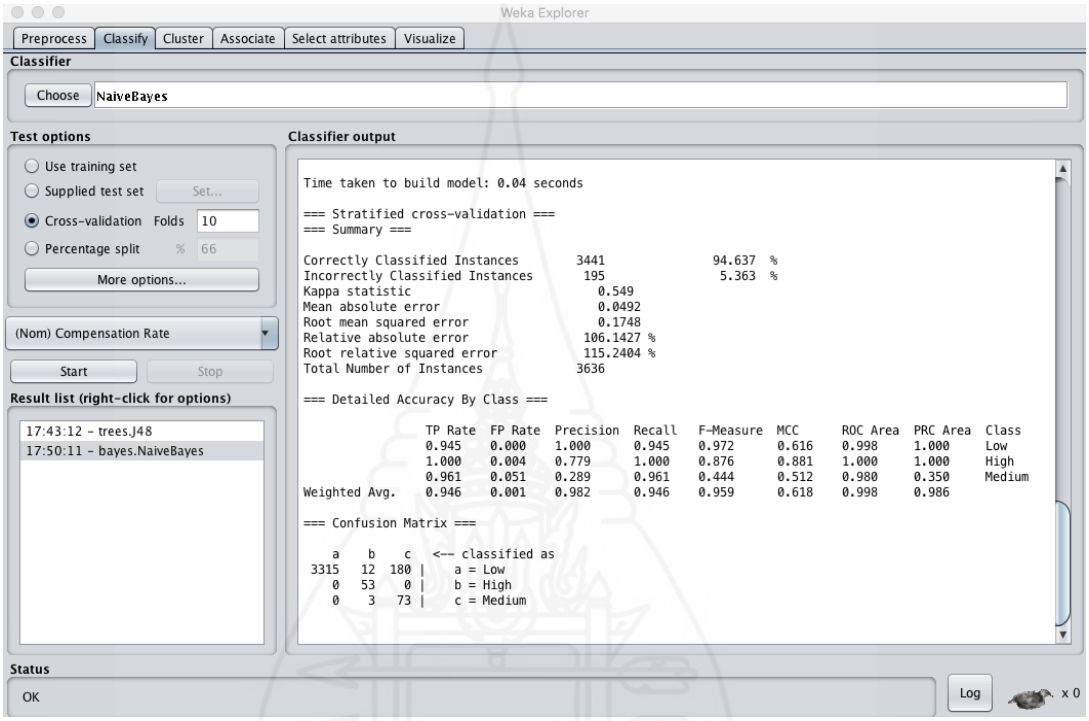


ภาพที่ 4.16 ผลของการใช้เทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม



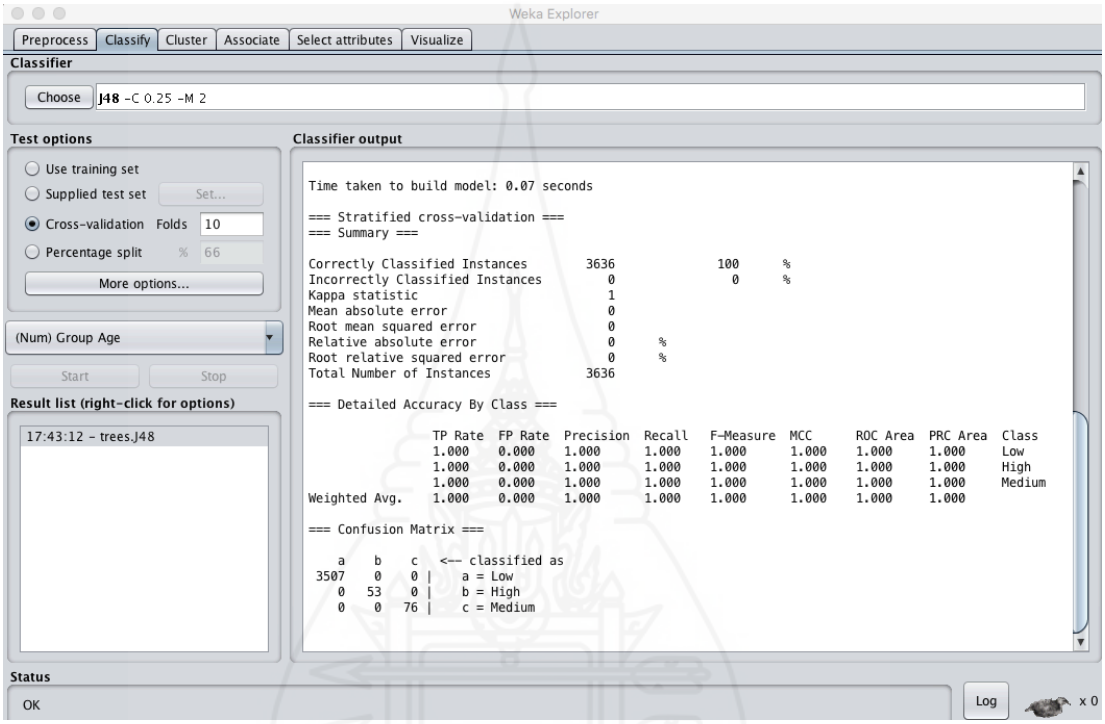
ภาพที่ 4.17 เทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP)

3) การใช้เทคนิคการทำเหมืองข้อมูลการจำแนกแบบเบย์ (Naïve Bayes) ร่วมกับวิธีที่ไม่ผ่านการคัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection ผลการพยากรณ์ให้ค่าความถูกต้อง (Accuracy) เท่ากับร้อยละ 94.63 และมีค่าความคลาดเคลื่อน (RMSE) เท่ากับ 0.1748 ดังภาพที่ 4.18



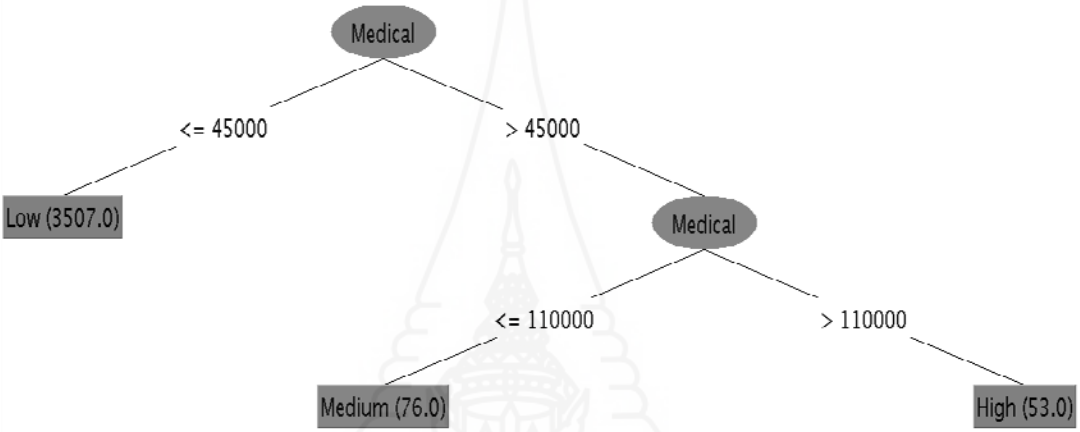
ภาพที่ 4.18 ผลของการใช้เทคนิคการจำแนกแบบเบย์ (Naïve Bayes) ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม

4) การใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึมแบบ J48 ในการสร้างตัวแบบพยากรณ์ ร่วมกับวิธีที่ไม่ผ่านการคัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะ ด้วยวิธี Correlation – based Feature Selection ผลการพยากรณ์ให้ค่าความถูกต้อง (Accuracy) เท่ากับร้อยละ 100 และมีค่าความคลาดเคลื่อน (RMSE) เท่ากับ 0 ดังภาพที่ 4.19

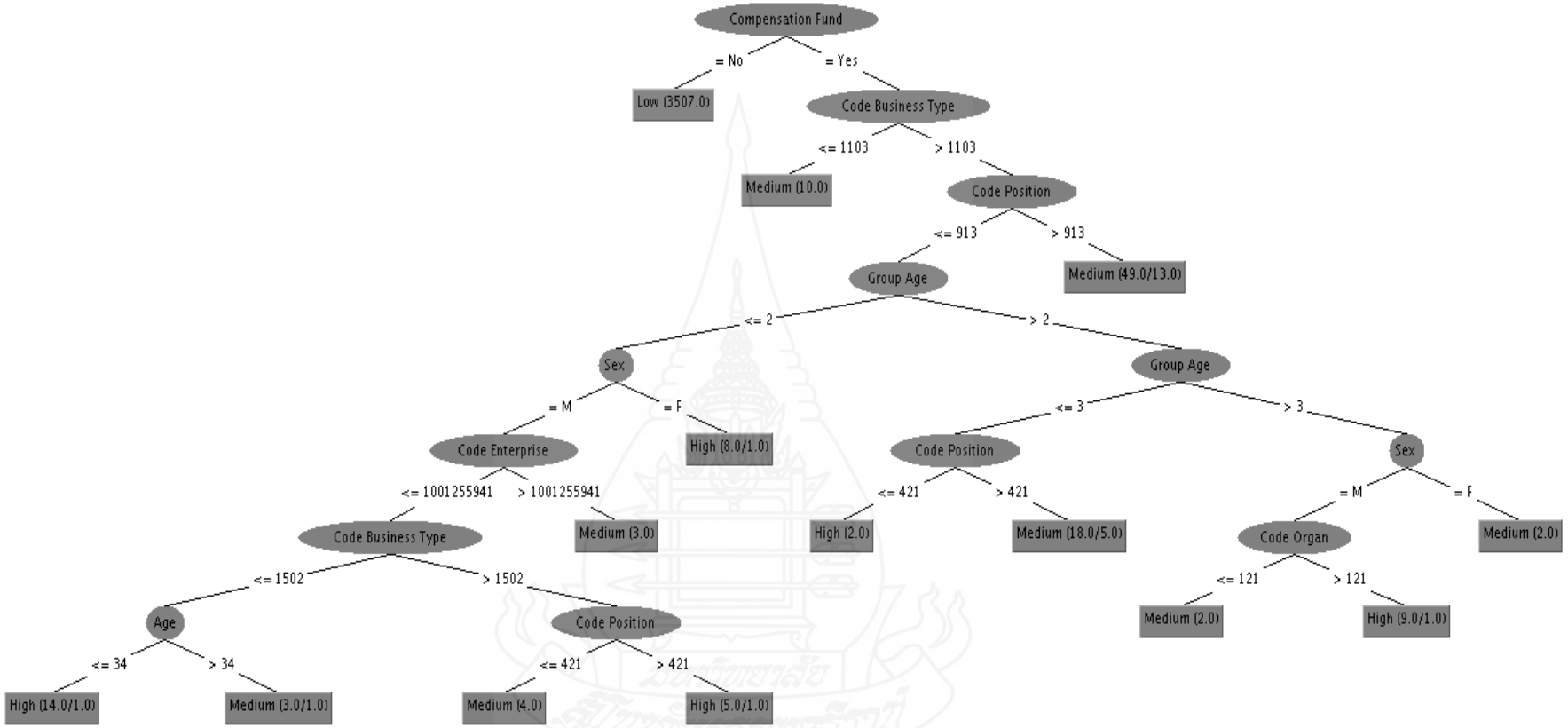


ภาพที่ 4.19 ผลของการใช้เทคนิคต้นไม้ตัดสินใจโดยใช้อัลกอริทึม J48 ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม

และในส่วนของกฎการพยากรณ์ที่ได้สามารถอธิบายได้ว่า กรณีที่ลูกจ้างเบิกค่ารักษาพยาบาลไม่เกิน 45,000 จะอยู่ในเกณฑ์ที่การใช้สิทธิเบิกกองทุนทดแทนต่ำ หากลูกจ้างเบิกค่ารักษาพยาบาลตั้งแต่ 45,001 – 110,000 จะอยู่ในเกณฑ์ที่การใช้สิทธิเบิกกองทุนทดแทนปานกลาง และลูกจ้างเบิกค่ารักษาพยาบาลมากกว่า 110,000 จะมีความเสี่ยงสูงในการใช้สิทธิเบิกกองทุนทดแทน ดังภาพที่ 4.20



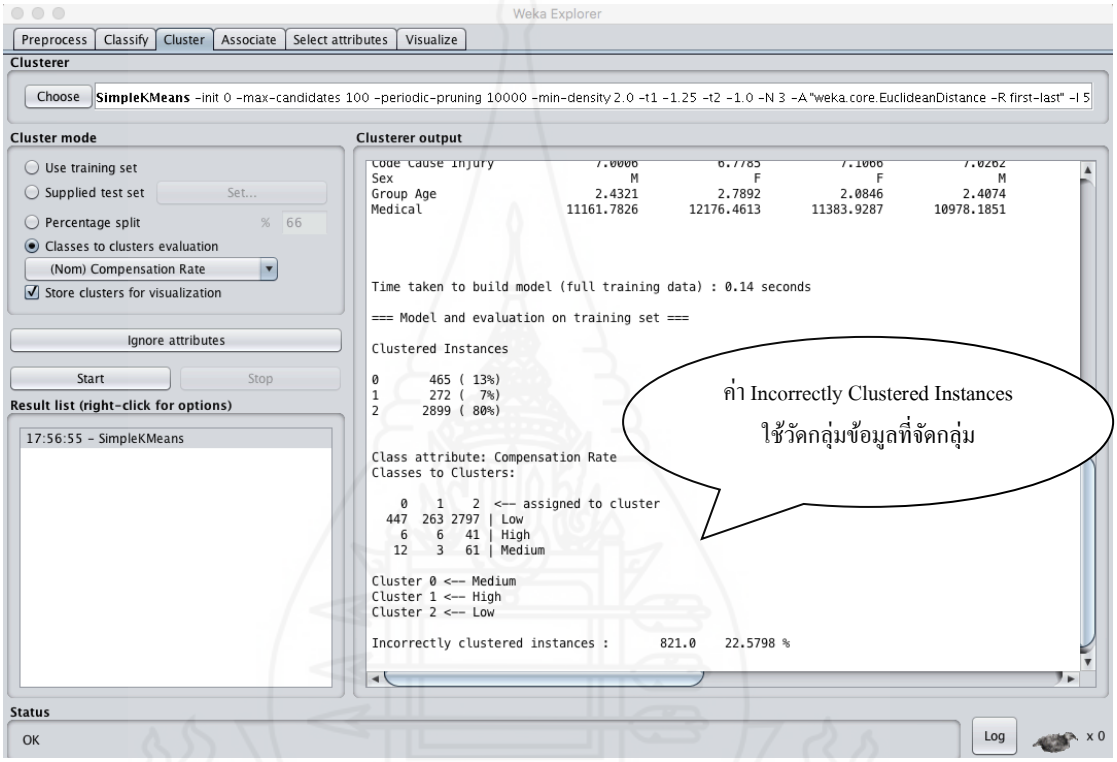
ภาพที่ 4.20 เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม J48



ภาพที่ 4.21 เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม J48

3.4.2 การสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม

1) การใช้เทคนิคเหมืองข้อมูลในการจัดกลุ่มข้อมูล (Clustering Data) ด้วยอัลกอริทึม k-means ด้วยโปรแกรม WEKA ดังภาพที่ 4.22 เพื่อจัดกลุ่มข้อมูลที่ใช้ในการสร้างตัวแบบพยากรณ์ โดยได้กำหนดค่า Num Cluster ระหว่าง 2 - 5 ซึ่งเป็นค่าที่ใช้ในการแบ่งกลุ่มข้อมูลผลที่ได้ ดังตารางที่ 4.16

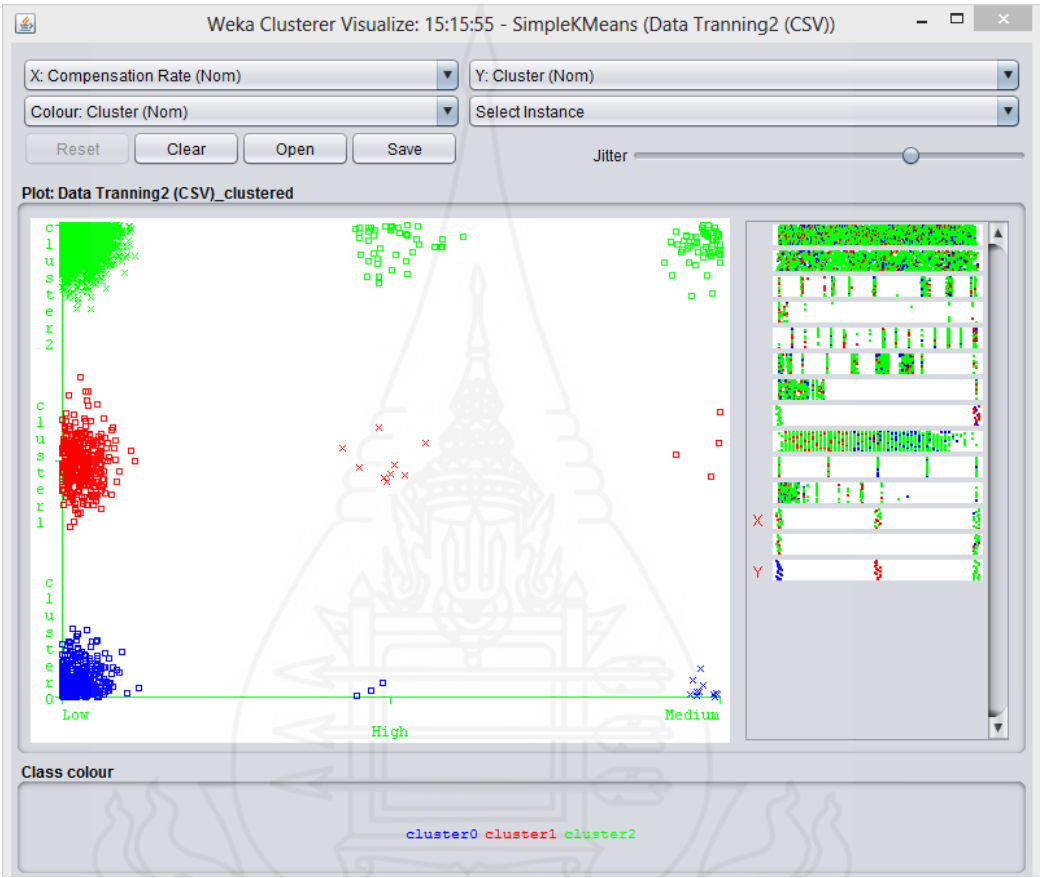


ภาพที่ 4.22 การจัดกลุ่มข้อมูล (Clustering Data) ด้วยอัลกอริทึม k-means

ตารางที่ 4.16 ผลการจัดกลุ่มข้อมูลด้วยอัลกอริทึม k-means

ลำดับ	Num Cluster	Class	Incorrectly Clustered Instances
1	2	2	44.77%
2	**3**	**3**	**22.57%**
3	4	4	43.70%
4	5	5	66.96%

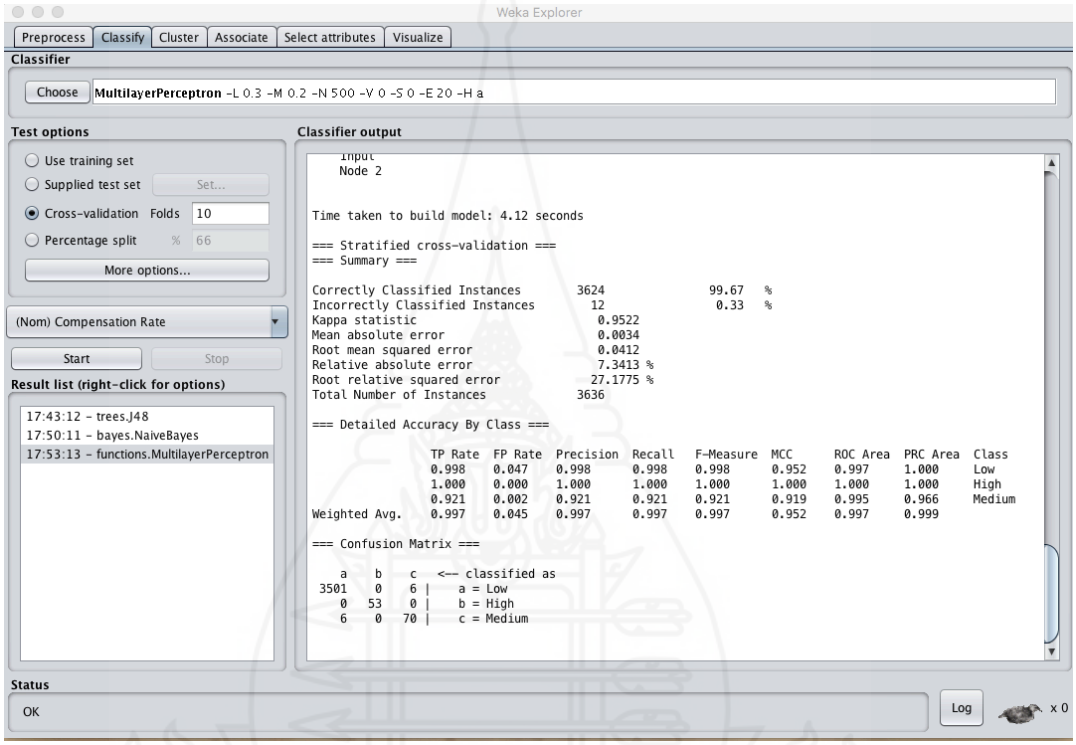
จากตารางที่ 4.16 เมื่อพิจารณาค่าความผิดพลาดในการจำแนกกลุ่ม (Incorrectly Clustered Instances: ICI) พบว่า การจำแนกกลุ่มโดยใช้ค่า Num Cluster เท่ากับ 3 มีค่าความผิดพลาดในการจำแนกข้อมูลน้อยที่สุด เท่ากับ 22.57% และนำค่าที่ได้มาทำการการจำแนกกลุ่มข้อมูลได้ 3 กลุ่ม โดยผลที่ได้จากการจัดกลุ่มข้อมูล (Clustering Data) ดังภาพที่ 4.23



ภาพที่ 4.23 ผลที่ได้จากการจัดกลุ่มข้อมูล (Clustering Data)

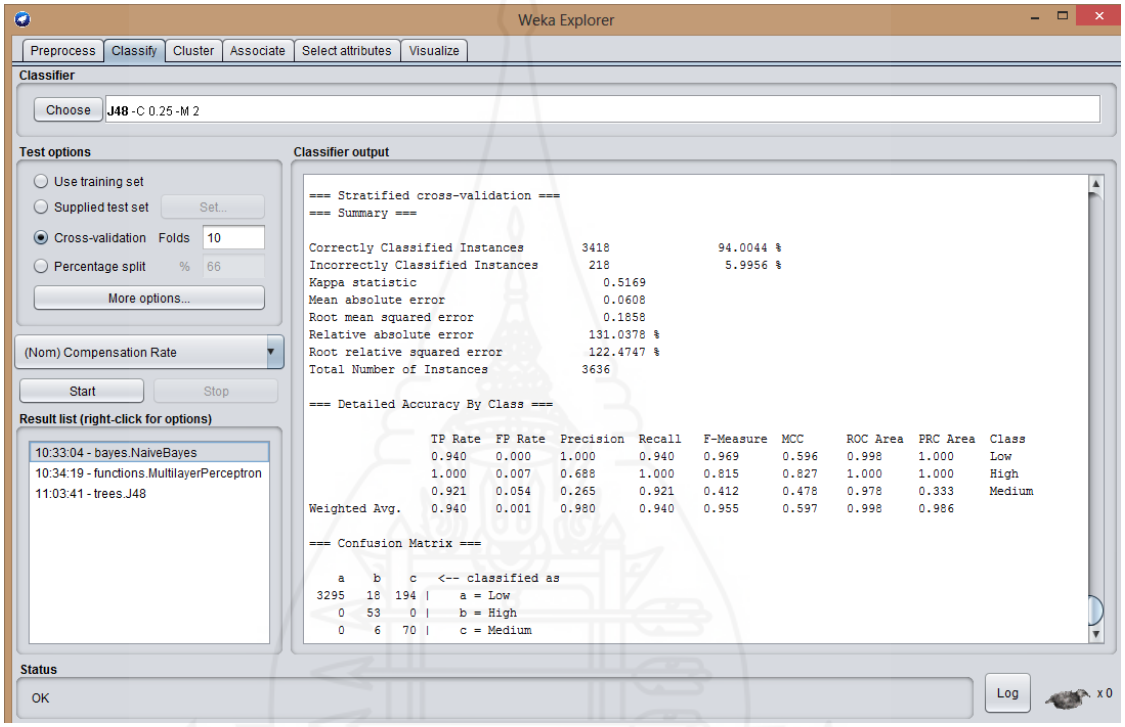
ผลการจำแนกกลุ่มข้อมูลที่ได้คือ Cluster0 มีจำนวนข้อมูล 322 ชุดข้อมูล Cluster1 มีจำนวนข้อมูล 415 ชุดข้อมูล และ Cluster2 มีจำนวนข้อมูล 2,899 ชุดข้อมูล

2) การใช้เทคนิคเหมืองข้อมูลโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ในการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบจัดกลุ่ม (Clustering Data) ร่วมกับวิธีที่ไม่ผ่านการคัดเลือกคุณลักษณะ และวิธีที่ผ่านการคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection ผลการพยากรณ์ให้ค่าความถูกต้อง (Accuracy) เท่ากับร้อยละ 73.81 และมีค่าความคลาดเคลื่อน (RMSE) เท่ากับ 0.4359 ดังภาพที่ 4.24



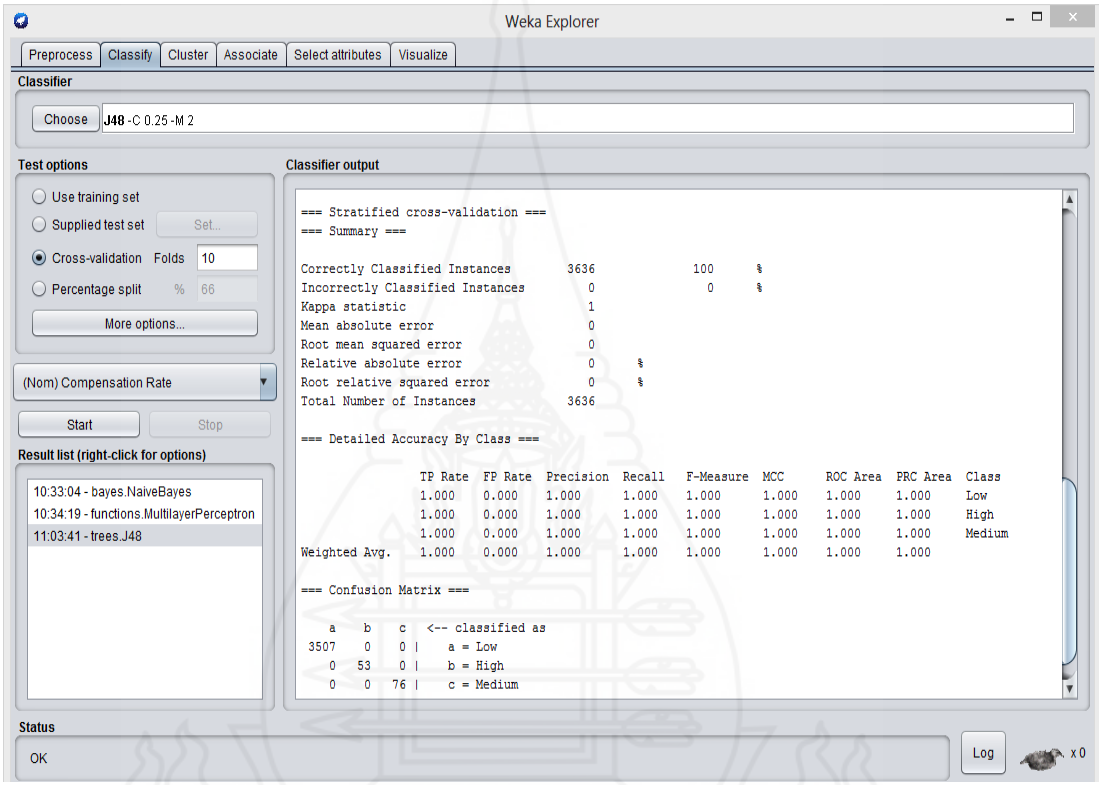
ภาพที่ 4.24 ผลของการใช้เทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ด้วยชุดข้อมูลแบบจัดกลุ่ม

3) การใช้เทคนิคเหมืองข้อมูลโครงข่ายประสาทเทียมการจำแนกแบบเบย์ (Naïve Bayes) ในการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบจัดกลุ่ม (Clustering Data) ร่วมกับวิธีที่ไม่ผ่านการคัดเลือกคุณลักษณะ และวิธีที่ผ่านการคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection ผลการพยากรณ์ให้ค่าความถูกต้อง (Accuracy) เท่ากับร้อยละ 94.00 และมีค่าความคลาดเคลื่อน (RMSE) เท่ากับ 0.1858 ดังภาพที่ 4.25



ภาพที่ 4.25 ผลการใช้เทคนิคการจำแนกแบบเบย์ (Naïve Bayes) ด้วยชุดข้อมูลแบบจัดกลุ่ม

4) การใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึมแบบ J48 ในการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบจัดกลุ่ม (Clustering Data) ร่วมกับวิธีที่ไม่ผ่านการคัดเลือกคุณลักษณะ และวิธีผ่านการคัดเลือกคุณลักษณะด้วย Correlation – base Feature Selection ผลการพยากรณ์ให้ค่าความถูกต้อง (Accuracy) เท่ากับร้อยละ 100 และมีค่าความคลาดเคลื่อน (RMSE) เท่ากับ 0 ดังภาพที่ 4.26



ภาพที่ 4.26 ผลการใช้เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม J48 ด้วยชุดข้อมูลแบบจัดกลุ่ม

3.5 ประเมินความถูกต้องหรือความแม่นยำของตัวแบบพยากรณ์

3.5.1 การสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม (Original Data)

จากการทดสอบประสิทธิภาพด้วยวิธี K-fold Cross Validation โดยกำหนดค่า K เท่ากับ 10 ซึ่งเป็นวิธีการที่ให้ค่าความแม่นยำ ด้วยการแบ่งข้อมูลออกเป็น 10 ส่วน โดยแบ่งข้อมูลการสร้างตัวแบบเป็น 9 ส่วน และตัวทดสอบ 1 ส่วน แล้วสลับสับเปลี่ยนข้อมูลในการสร้างตัวแบบพยากรณ์และทดสอบ โดยใช้วิธีการวัดค่าความถูกต้อง (Accuracy) จากการสร้างตัวแบบพยากรณ์สามารถคำนวณได้จาก

$$Accuracy = \frac{TotalNumberofIns\tan ces}{CorrectyClassifiedIns\tan ce} x100$$

และค่ารากที่สองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) สามารถคำนวณได้จาก

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{'} - Y_i\right)^2}$$

ผลทดสอบประสิทธิภาพตัวแบบการพยากรณ์โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ในการสร้างตัวแบบพยากรณ์ด้วยข้อมูลแบบไม่จัดกลุ่ม (Original Data) ร่วมกับวิธีที่ไม่ผ่านการคัดเลือกคุณลักษณะ และวิธีผ่านการคัดเลือกคุณลักษณะด้วย Correlation – based Feature Selection ดังตารางที่ 4.17

ตารางที่ 4.17 ผลการวัดประสิทธิภาพโดยใช้เทคนิคโครงข่ายประสาทเทียมมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม

Feature Selection	**Precision**	**Recall**	**Accuracy**	**Root Mean Square Error**
Original	0.997	0.997	99.67	0.4385
Correlation – based Feature Selection	0.930	0.959	95.92	0.1653

ผลทดสอบประสิทธิภาพตัวแบบการพยากรณ์การจำแนกแบบเบย์ (Naïve Bayes) ในการสร้างตัวแบบพยากรณ์ด้วยข้อมูลแบบไม่จัดกลุ่ม (Original Data) ร่วมกับวิธีที่ไม่ผ่านการคัดเลือกคุณลักษณะ และวิธีผ่านการคัดเลือกคุณลักษณะด้วย Correlation – based Feature Selection ดังตารางที่ 4.18

ตารางที่ 4.18 ผลการวัดประสิทธิภาพโดยใช้เทคนิคเหมืองข้อมูลการจำแนกแบบเบย์ (Naïve Bayes) ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม

Feature Selection	**Precision**	**Recall**	**Accuracy**	**Root Mean Square Error**
Original	0.982	0.946	94.63	0.1748
Correlation – based Feature Selection	0.964	0.995	95.92	0.1653

ผลทดสอบประสิทธิภาพตัวแบบการพยากรณ์ต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม J48 ในการสร้างตัวแบบพยากรณ์ด้วยข้อมูลแบบไม่จัดกลุ่ม (Original Data) ร่วมกับวิธีที่ไม่ผ่านการคัดเลือกคุณลักษณะ และวิธีผ่านการคัดเลือกคุณลักษณะด้วย Correlation – based Feature Selection ดังตารางที่ 4.19

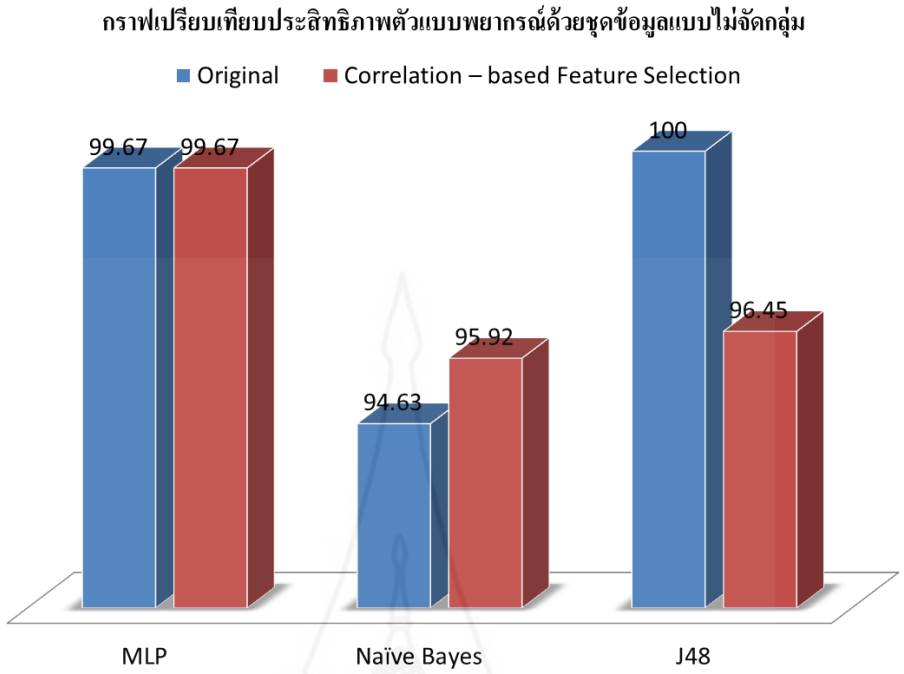
ตารางที่ 4.19 ผลการวัดประสิทธิภาพโดยใช้เทคนิคเหมืองข้อมูลต้นไม้ตัดสินใจ โดยใช้อัลกอริทึม J48 ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม

Feature Selection	**Precision**	**Recall**	**Accuracy**	**Root Mean Square Error**
Original	1.000	1.000	100	0
Correlation – based Feature Selection	0.930	0.965	96.45	0.1517

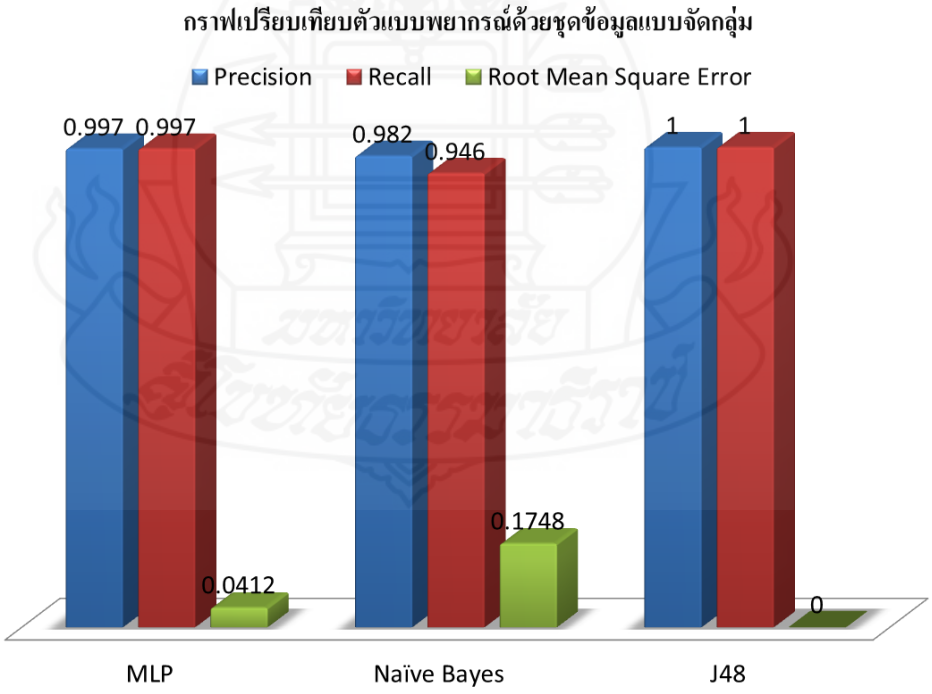
จากผลการทดสอบประสิทธิภาพด้วยวิธี 10 fold Cross-Validation พบว่าการจำแนกข้อมูลที่ไม่ผ่านวิธีการคัดเลือกคุณลักษณะ (Feature Selection) โดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม J48 ให้ค่าความถูกต้อง เท่ากับร้อยละ 100 ซึ่งให้ค่าความถูกต้องสูงและให้ค่าความคลาดเคลื่อนน้อยที่สุด (Root Mean Squared Error) เท่ากับร้อยละ 0 ซึ่งให้ผลการพยากรณ์ดีกว่าโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) และเทคนิคการจำแนกแบบเบย์ (Naïve Bayes) ดังตารางที่ 4.20

ตารางที่ 4.20 การเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม

Algorithm	Feature Selection	Precision	Recall	Accuracy	Root Mean Square Error
MLP	Original	0.997	0.997	99.67	0.0412
	Correlation – based Feature Selection	0.997	0.997	99.67	0.0412
Naïve Bayes	Original	0.982	0.946	94.63	0.1748
	Correlation – based Feature Selection	0.964	0.995	95.92	0.1653
J48	Original	1.000	1.000	100	0
	Correlation – based Feature Selection	0.930	0.965	96.45	0.1517



ภาพที่ 4.27 กราฟเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม



ภาพที่ 4.28 กราฟเปรียบเทียบค่า Precision, Recall และ RMSE ของตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม

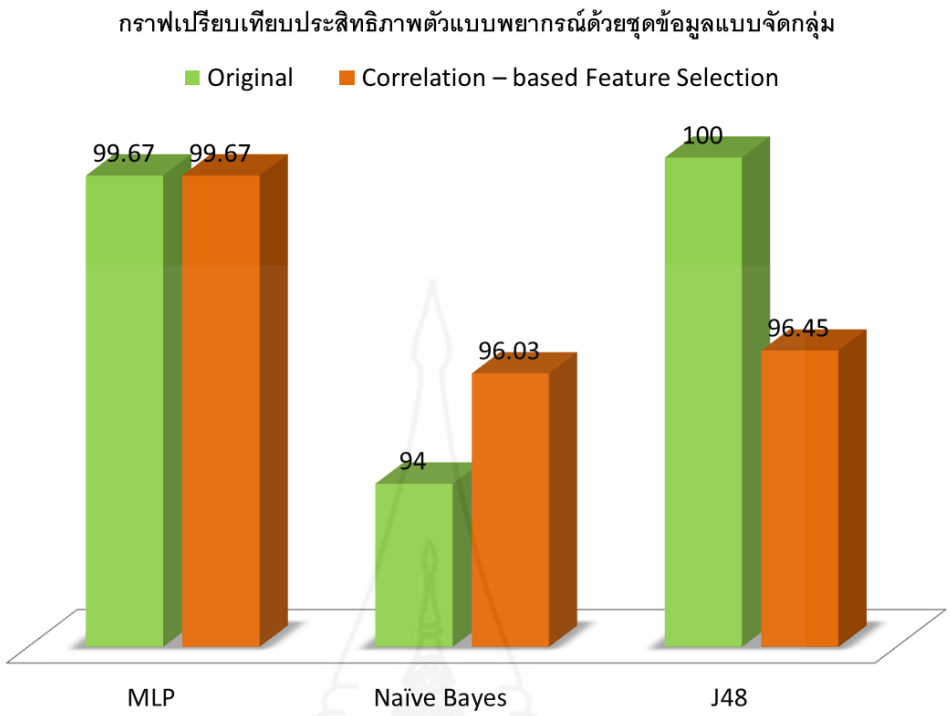
3.5.2 การสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบจัดกลุ่ม (Clustering Data)

จากการทดสอบประสิทธิภาพตัวแบบการพยากรณ์โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ตัวแบบพยากรณ์การจำแนกแบบเบย์ (Naïve Bayes) และเทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม J48 ร่วมกับวิธีแบบไม่ผ่านการคัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วยข้อมูลแบบจัดกลุ่ม ผลที่ได้ ดังตารางที่ 4.21

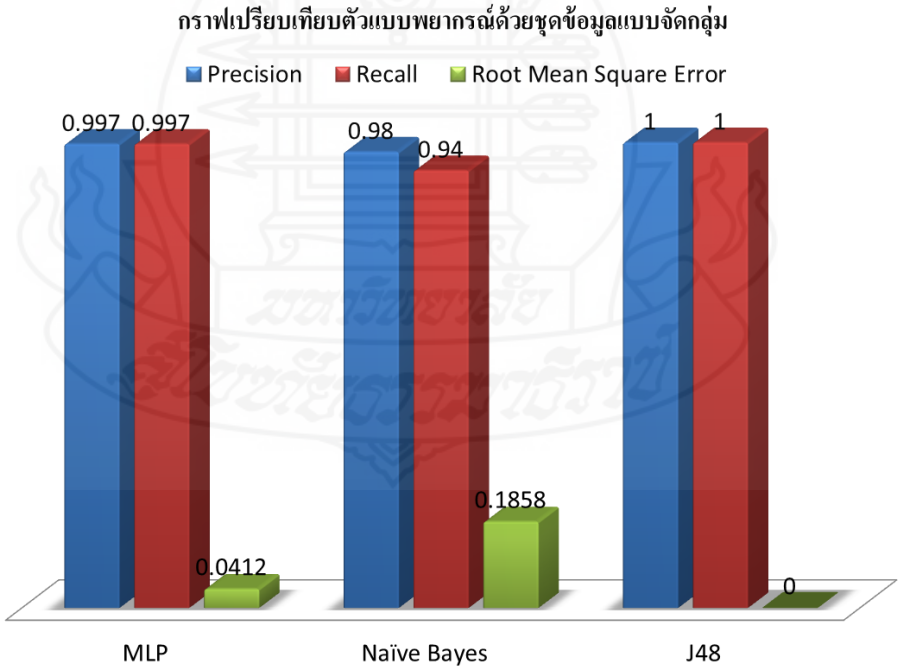
ตารางที่ 4.21 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบจัดกลุ่ม

Data Set	Algorithm	Feature Selection	Attribute	Precision	Recall	Accuracy	Root Mean Square Error
Clustering	MLP	Original	14	0.997	0.997	99.67	0.0412
		CFS	7	0.997	0.997	99.67	0.0412
	Naïve Bayes	Original	14	0.980	0.940	94.00	0.1858
		CFS	7	0.930	0.960	96.03	0.1622
	J48	Original	14	1.000	1.000	100	0
		CFS	7	0.930	0.965	96.45	0.1517

จากตารางที่ 4.21 ผลการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบจัดกลุ่ม (Clustering Data) พบว่าตัวแบบพยากรณ์ต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม J48 โดยไม่ผ่านวิธีการคัดเลือกคุณลักษณะ และร่วมกับวิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection ให้ค่าความถูกต้อง เท่ากับร้อยละ 100 ซึ่งให้ค่าความถูกต้องสูงและให้ค่าความคลาดเคลื่อนน้อยที่สุด (Root Mean Squared Error) เท่ากับร้อยละ 0 ซึ่งให้ผลการพยากรณ์ดีกว่าโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) และเทคนิคการจำแนกแบบเบย์ (Naïve Bayes) ดังภาพที่ 4.29 และภาพที่ 4.30



ภาพที่ 4.29 กราฟเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบจัดกลุ่ม



ภาพที่ 4.30 กราฟเปรียบเทียบค่า Precision, Recall และ RMSE ของตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบจัดกลุ่ม

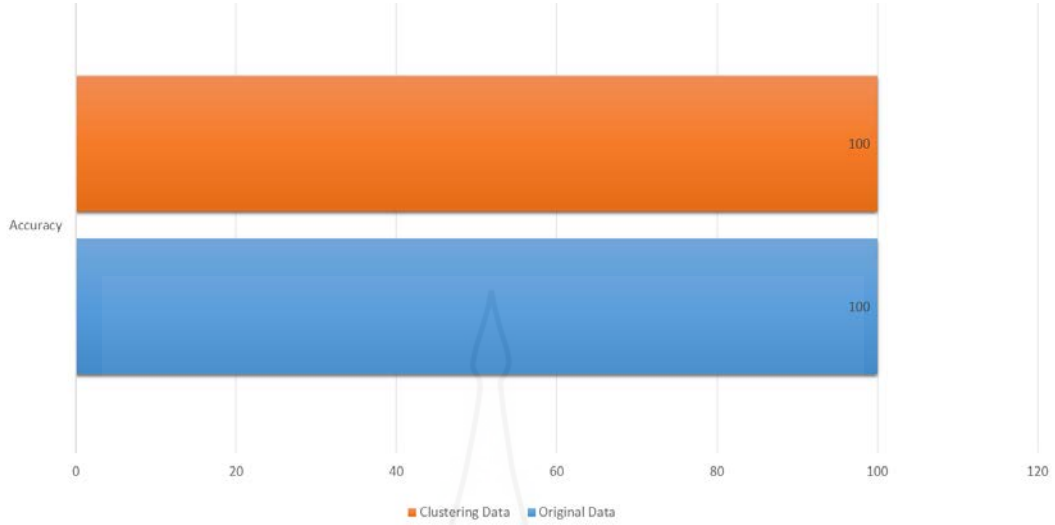
3.5.3 การเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม (Original Data) กับชุดข้อมูลแบบจัดกลุ่ม (Clustering Data)

นำผลที่ได้จากตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม (Original Data) ที่มีค่าความถูกต้องสูงที่สุดมาเปรียบเทียบกับตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบจัดกลุ่ม (Clustering Data) ที่มีค่าความถูกต้องสูงที่สุดในแต่ละกลุ่ม ผลการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ ดังตารางที่ 4.22

ตารางที่ 4.22 การเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์

Prediction Model	Algorithm	Feature Selection	Attribute	Precision	Recall	Accuracy	Root Mean Square Error
Original Data	J48	Original	12	1.000	1.000	100	0
Clustering Data	J48	Original	14	1.000	1.000	100	0

จากตารางที่ 4.22 การเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์พบว่าตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม (Original Data) โดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม J48 ซึ่งคัดเลือกคุณลักษณะไว้ 12 แอตทริบิวต์ และชุดข้อมูลแบบจัดกลุ่ม (Clustering) ซึ่งคัดเลือกคุณลักษณะไว้ 14 แอตทริบิวต์ มีค่าความถูกต้อง (Accuracy) ร้อยละ 100 มีค่าความแม่นยำ (Precision) และค่า (Recall) เท่ากับร้อยละ 1.000 ส่วนค่าความคลาดเคลื่อนน้อย (Root Mean Squared Error) เท่ากับ 0 ดังภาพที่ 4.31



ภาพที่ 4.31 กราฟเปรียบเทียบประสิทธิภาพค่าความถูกต้องของตัวแบบพยากรณ์

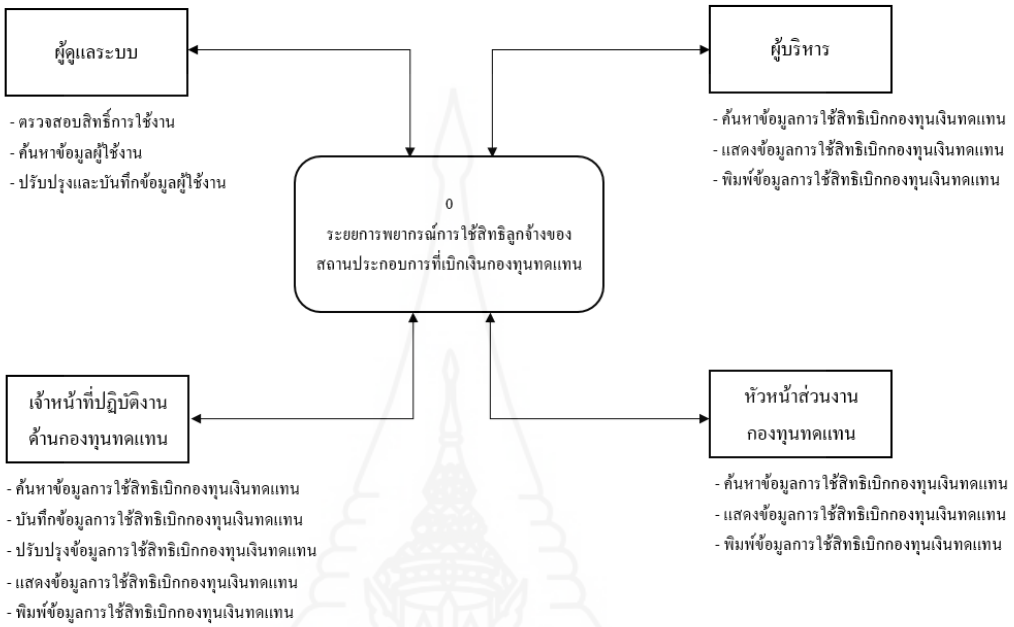
ดังนั้นผู้วิจัยจึงเลือกใช้ตัวแบบพยากรณ์ที่ใช้ชุดข้อมูลแบบไม่จัดกลุ่ม (Original Data) โดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม J48 ซึ่งเหมาะสมที่สุดในการนำพัฒนาเป็นระบบการพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคม เนื่องจากให้ค่าความถูกต้องของการพยากรณ์ดีที่สุด และค่าความคลาดเคลื่อนน้อยที่สุด

3.6 การนำโมเดลไปใช้งาน (Model Deployment)

ระบบพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน สำนักงานประกันสังคม ผู้วิจัยได้ใช้รูปแบบการแสดงผลบนเว็บไซต์ (Web Application) โดยใช้โปรแกรมภาษา PHP ร่วมกับโปรแกรม Adobe Dreamweaver CS 6.0 ในการพัฒนาเพื่อเพิ่มความสะดวกในการใช้งาน ซึ่งเป็นการนำตัวแบบพยากรณ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล โดยใช้ตัวแปรในการวิเคราะห์ข้อมูล คือ Code Injury, Code Enterprise, Business Type, Compensation Rate และจะทำการส่งออกข้อมูลผลการวิเคราะห์ ให้สามารถใช้งานได้จริง เพื่อประกอบการตัดสินใจในการใช้จัดตั้งงบประมาณประจำปี และใช้ในการวางแผนลดอัตราการเกิดประสบอันตรายจากการทำงาน สำหรับผู้บริหาร หัวหน้าส่วนงานกองทุนทดแทน และเจ้าหน้าที่ปฏิบัติงานกองทุนเงินทดแทน

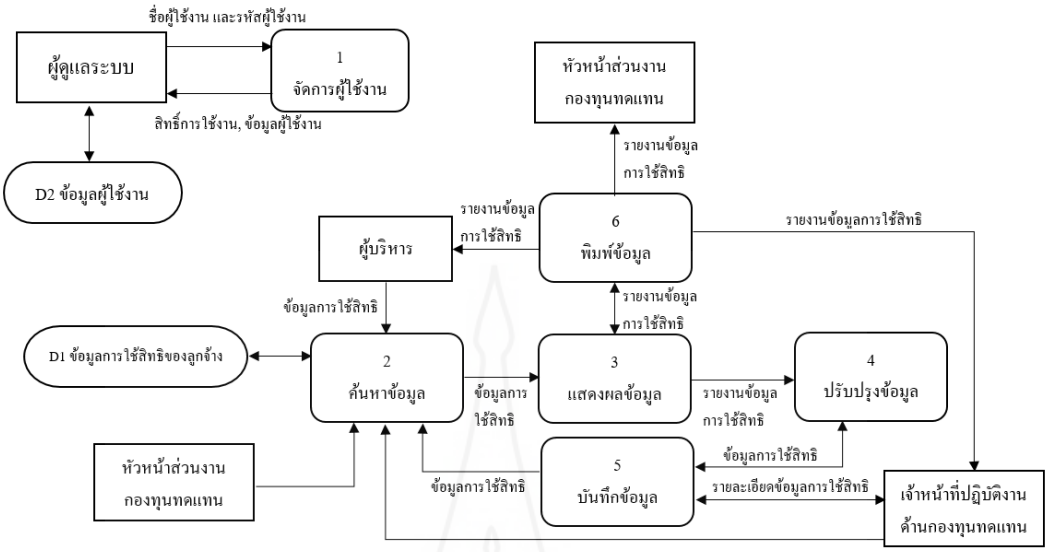
3.6.1 ทำการออกแบบระบบโดยใช้แผนภาพกระแสข้อมูล (Data Flow Diagram : DFD) เป็นเครื่องมือในการวิเคราะห์ระบบ โดยแสดงการไหลของข้อมูล และความสัมพันธ์ของข้อมูลที่ได้จากการวิเคราะห์ระบบเพื่อแสดงการนำเข้าข้อมูล (Input) และแสดงผลลัพธ์ (Output)

ของระบบในการทำงานแต่ละส่วน โดยเส้นทางการไหลของข้อมูลดังกล่าว ซึ่งลำดับภาพรวมของกระบวนการทำงาน (Context Diagram) ดังภาพที่ 4.32



ภาพที่ 4.32 แผนภาพบริบทของระบบพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน สำนักงานประกันสังคม

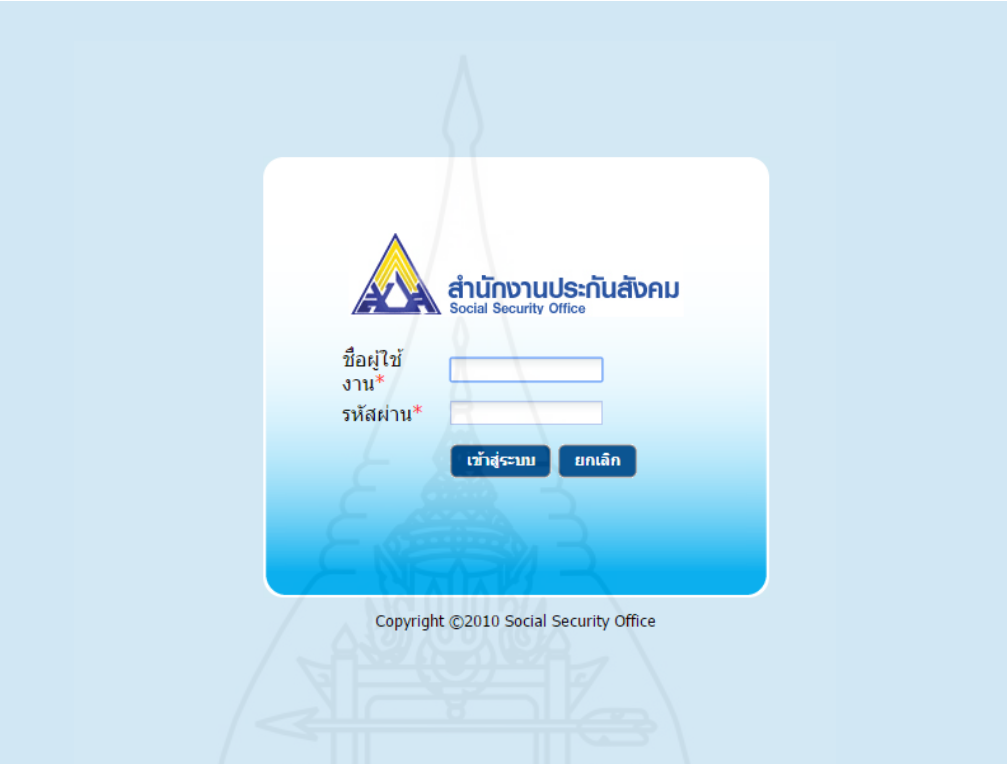
จากภาพที่ 4.32 แสดงเส้นทางของข้อมูลที่เข้าและออกจากแหล่งข้อมูลที่มีผลกระทบต่อระบบโดยมีผู้ที่เกี่ยวข้องกับการใช้งาน ได้แก่ ผู้บริหาร หัวหน้าส่วนงานกองทุนทดแทน เจ้าหน้าที่ปฏิบัติงานด้านกองทุนทดแทน และผู้ดูแลระบบ



ภาพที่ 4.33 แผนภาพกระแสข้อมูลที่ระดับ 0 (Data Flow Diagram Level 0)

จากภาพที่ 4.33 แสดงกระแสข้อมูลในส่วนของการจัดการผู้ใช้งานระบบ ได้แก่ ผู้บริหาร หัวหน้าส่วนงานกองทุนทดแทน เจ้าหน้าที่ปฏิบัติงานด้านกองทุนทดแทน และผู้ดูแลระบบ ซึ่งสามารถค้นหาข้อมูล แสดงการพยากรณ์ และสั่งพิมพ์ข้อมูลการพยากรณ์ได้

3.6.2 หน้าจอหลักระบบพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน สำนักงานประกันสังคม สำหรับผู้ใช้งานต้องทำการใส่ชื่อผู้ใช้งานและรหัสผ่านเพื่อเข้าสู่ระบบ ดังภาพที่ 4.34



ภาพที่ 4.34 หน้าจอหลักระบบพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิก เงินกองทุนทดแทน สำนักงานประกันสังคม

3.6.3 หน้าจอการทำงานระบบพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน สำนักงานประกันสังคม ประกอบด้วยขั้นตอน ดังภาพที่ 4.35

1) ค้นหาตามหมายเลขการประสบอันตราย (Code Injury) ผู้ใช้จะต้องทำการกรอกหมายเลขการประสบอันตราย

2) เมื่อทำการกรอกหมายเลขการประสบอันตรายแล้ว จะปรากฏชื่อสถานประกอบการในช่อง (Enterprise) ประเภทกิจการ (Business Type) และในส่วนของการเบิกเงินจากกองทุนเงินทดแทน (Claim Medical)

3) ทำการกดปุ่ม Prediction เพื่อหาผลการพยากรณ์

ภาพที่ 4.35 หน้าจอการทำงานระบบพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน สำนักงานประกันสังคม

จากภาพที่ 4.35 ระบบจะทำการพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน โดยการค้นหาตามหมายเลขการประสบอันตราย (Code Injury) ซึ่งการวิเคราะห์ข้อมูลจากโปรแกรม WEKA 3.8.0 จะใช้ตัวแปร Code Injury, Code Enterprise, Business Type, Compensation Rate และจะแสดงผลการพยากรณ์ ดังภาพที่ 4.36

ภาพที่ 4.36 แสดงผลการพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการโครงสร้างที่ 1

จากภาพที่ 4.36 ผลการพยากรณ์ที่ได้ของการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิก เงินกองทุนทดแทน สำนักงานประกันสังคมอยู่ในระดับต่ำ (Low) ซึ่งหมายความว่า ถ้าลูกจ้างของสถานประกอบการเบิกค่ารักษาพยาบาลและค่าทดแทน ไม่เกินจำนวนเงิน 45,000 บาท จะมีผลต่อการเรียกเก็บอัตราการประสบอันตราย (ความเสี่ยง) อยู่ในเกณฑ์ต่ำ

ภาพที่ 4.37 แสดงผลการพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการโครงสร้างที่ 2

จากภาพที่ 4.37 ผลการพยากรณ์ที่ได้ของการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิก เงินกองทุนทดแทน สำนักงานประกันสังคมอยู่ในระดับต่ำ (Medium) ซึ่งหมายความว่า ถ้าลูกจ้างของสถานประกอบการเบิกค่ารักษาพยาบาลและค่าทดแทน จำนวนเงิน 45,001 – 110,00 บาท จะมีผลต่อการเรียกเก็บอัตราการประสบอันตราย (ความเสี่ยง) อยู่ในเกณฑ์ปานกลาง

ภาพที่ 4.38 แสดงผลการพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการโครงสร้างที่ 3

จากภาพที่ 4.38 ผลการพยากรณ์ที่ได้ของการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิก เงินกองทุนทดแทน สำนักงานประกันสังคมอยู่ในระดับต่ำ (High) ซึ่งหมายความว่า ถ้าลูกจ้างของสถานประกอบการเบิกค่ารักษาพยาบาลและค่าทดแทน เกินจำนวนเงิน 110,000 บาทขึ้นไป จะมีผลต่อการเรียกเก็บอัตราการประสบอันตราย (ความเสี่ยง) อยู่ในเกณฑ์สูง

4. การประเมินความพึงพอใจต่อการใช้งานคลังข้อมูล

4.1 การออกแบบเครื่องมือการประเมินความพึงพอใจคลังข้อมูล

การออกแบบเครื่องมือการประเมินความพึงพอใจของผู้ใช้ระบบคลังข้อมูลของการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคม โดยใช้เครื่องมือเป็นแบบประเมินความพึงพอใจ ซึ่งจะให้ผู้ประเมินความพึงพอใจต่อคลังข้อมูลเป็นผู้ที่เกี่ยวข้องในการทำงานกองทุนเงินทดแทน สำนักงานประกันสังคม ได้แก่ ผู้บริหาร หัวหน้าส่วนงานกองทุนทดแทน และเจ้าหน้าที่ปฏิบัติงานด้านกองทุนทดแทน จำนวน 9 คน และแบบประเมินความพึงพอใจต่อคลังข้อมูล จะมีการประเมิน 2 ตอน คือ ตอนที่ 1 ข้อมูลทั่วไป และตอนที่ 2 ความพึงพอใจการใช้งาน ซึ่งจะประกอบด้วยทั้งหมด 5 ด้าน ได้แก่ 1) มีความสะดวกและง่ายต่อการใช้งาน 2) การออกแบบรายงานเข้าใจได้ง่าย 3) ความถูกต้องของข้อมูล 4) ช่วยลดขั้นตอนในการทำงาน และ 5) ข้อมูลตรงกับความต้องการผู้ใช้งาน

4.2 ประเมินความพึงพอใจระบบคลังข้อมูล

การประเมินความพึงพอใจระบบคลังข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนเงินทดแทน สำนักงานประกันสังคม โดยใช้แบบสอบถามเป็นเครื่องมือในการประเมิน ซึ่งมีมาตรวัดอัตราส่วน 5 ระดับ (Rating Scale) ได้แก่

มากที่สุด	(ระดับ 5) หมายถึง ระดับความพึงพอใจมากที่สุด
มาก	(ระดับ 4) หมายถึง ระดับความพึงพอใจมาก
ปานกลาง	(ระดับ 3) หมายถึง ระดับความพึงพอใจปานกลาง
น้อย	(ระดับ 2) หมายถึง ระดับความพึงพอใจน้อย
น้อยที่สุด	(ระดับ 1) หมายถึง ระดับความพึงพอใจน้อยที่สุด

ตอนที่ 1 ข้อมูลทั่วไป

จากการตอบแบบสอบถามของกลุ่มผู้ใช้งานระบบคลังข้อมูล มีจำนวนทั้งหมด 9 ท่าน แบ่งเป็น ผู้บริหาร 3 ท่าน หัวหน้าส่วนงานกองทุนเงินทดแทน จำนวน 2 ท่าน และเจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทน จำนวน 4 ท่าน และจากตัวแปรนำมาหาจำนวนร้อยละของผู้ตอบในตอนที่ 1 ดังตารางที่ 4.23

ตารางที่ 4.23 ค่าร้อยละข้อทั่วไปของผู้ใช้งานคลังข้อมูล

ตัวแปร	จำนวน (คน)	ร้อยละ
1.เพศ	**9**	**100**
1.1 ชาย	0	0
1.2 หญิง	9	100
2. อายุ	**9**	**100**
2.1 อายุ 15 – 30 ปี	0	0
2.2 อายุ 31 – 45 ปี	3	33.33
2.3 อายุ 46 – 60 ปี	6	66.67
3.สถานภาพ	**9**	**100**
3.1 ผู้บริหาร	3	33.33
3.2 หัวหน้าส่วนงานกองทุนเงินทดแทน	2	22.22
3.3 เจ้าหน้าที่ปฏิบัติงานกองทุนเงินทดแทน	4	44.45

ตอนที่ 2 ความพึงพอใจการใช้งาน

จากข้อมูลที่ได้จากการตอบแบบสอบถามของกลุ่มผู้ใช้งาน จำนวน 9 ท่าน ซึ่งแบ่งเป็นผู้บริหาร 3 ท่าน หัวหน้าส่วนงานกองทุนเงินทดแทน จำนวน 2 ท่าน และเจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทน จำนวน 4 ท่าน นำมาหาจำนวนค่าเฉลี่ยของผู้ตอบแต่ละข้อคำถาม ในตอนที่ 2 ดังตารางที่ 4.24

ตารางที่ 4.24 ค่าเฉลี่ยของการประเมินความพึงพอใจระบบคลังข้อมูล

รายการประเมิน	ดีมาก (5)	ดี (4)	ปานกลาง (3)	น้อย (2)	น้อยที่สุด (1)	จำนวนแบบสอบถาม (N)	ค่าเฉลี่ย (X)	ค่าเบี่ยงเบนมาตรฐาน (SD)
1. มีความสะดวกและง่ายต่อการใช้งาน	3	4	2	0	0	9	4.11	0.74
2. การออกแบบรายงานเข้าใจได้ง่าย	2	4	3	0	0	9	3.88	0.78
3. ความถูกต้องของข้อมูล	3	4	2	0	0	9	4.11	0.74
4. ช่วยลดขั้นตอนในการทำงาน	4	3	2	0	0	9	4.22	0.79
5. ข้อมูลตรงกับความต้องการผู้ใช้งาน	3	3	3	0	0	9	4.00	0.81
รวม	**15**	**18**	**12**	**0**	**0**	**45**	**4.06**	**0.80**

จากตารางที่ 4.24 นำค่าเฉลี่ยที่ได้มาเปรียบเทียบกับเกณฑ์ของเบสท์ (Best.1981, p. 82) ดังนี้

ค่าเฉลี่ย 4.50 – 5.00 หมายถึง ความคิดเห็นอยู่ในระดับดีมาก
ค่าเฉลี่ย 3.50 – 4.49 หมายถึง ความคิดเห็นอยู่ในระดับดี
ค่าเฉลี่ย 2.50 – 3.49 หมายถึง ความคิดเห็นอยู่ในระดับปานกลาง
ค่าเฉลี่ย 1.50 – 2.49 หมายถึง ความคิดเห็นอยู่ในระดับน้อย
ค่าเฉลี่ย 1.00 – 1.49 หมายถึง ความคิดเห็นอยู่ในระดับน้อยที่สุด

ผลที่ได้จากการวิเคราะห์ข้อมูลจากแบบสอบถาม จำนวน 9 ชุด นำมาทำการแปลผลเทียบกับเกณฑ์ ดังตารางที่ 4.25

ตารางที่ 4.25 ผลการประเมินความพึงพอใจระบบคลังข้อมูลการใช้สิทธิลูกจ้างของสถาน ประกอบการที่เบิกเงินกองทุนเงินทดแทน สำนักงานประกันสังคม

รายการประเมิน	ค่าเฉลี่ย (x)	การแปลผล
1. มีความสะดวกและง่ายต่อการใช้งาน	4.11	ระดับดี
2. การออกแบบรายงานเข้าใจได้ง่าย	3.88	ระดับดี
3. ความถูกต้องของข้อมูล	4.11	ระดับดี
4. ช่วยลดขั้นตอนในการทำงาน	4.22	ระดับดี
5. ข้อมูลตรงกับความต้องการผู้ใช้งาน	4.00	ระดับดี
ความพึงพอใจโดยภาพรวม	**4.06**	**ระดับดี**

จากตารางที่ 4.25 ผลของการประเมินระดับความพึงพอใจระบบคลังข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน พบว่า ระบบคลังข้อมูลมีความสะดวกสบายต่อการใช้งานมีค่าเฉลี่ยเท่ากับ 4.11 มีการออกแบบรายงานเข้าใจได้ง่ายมีค่าเฉลี่ยเท่ากับ 3.88 มีความถูกต้องของข้อมูลมีค่าเฉลี่ยเท่ากับ 4.11 ช่วยลดขั้นตอนในการทำงานมีค่าเฉลี่ยเท่ากับ 4.22 และข้อมูลตรงกับความต้องการของผู้ใช้งานมีค่าเฉลี่ยเท่ากับ 4.00 ผลการประเมินความพึงพอใจของผู้ใช้งานมีค่าเฉลี่ยรวมเท่ากับ 4.06 ซึ่งเมื่อแปลผลมีระดับความพึงพอใจอยู่ในระดับดี ระบบคลังข้อมูลมีประสิทธิภาพสามารถนำไปใช้งานได้

บทที่ 5
สรุปการวิจัย อภิปรายผล และข้อเสนอแนะ

การศึกษาการวิจัยเรื่อง การพยากรณ์การใช้สิทธิของลูกจ้างในสถานประกอบการเพื่อเพิ่มประสิทธิภาพการบริหารการจัดการกองทุนเงินทดแทน สำนักงานประกันสังคม ผู้วิจัยได้ทำการพัฒนาระบบคลังข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทนในการนำเสนอรายงาน และนำข้อมูลมาวิเคราะห์หาความรู้ใหม่โดยเทคนิคการทำเหมืองข้อมูล ได้ข้อสรุปผลการวิจัย การอภิปรายผล และข้อเสนอแนะดังนี้

1. สรุปการวิจัย

1.1 การพัฒนาคลังข้อมูล

งานวิจัยครั้งนี้ผู้วิจัยได้นำข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคม ซึ่งข้อมูลที่ได้นำมาทำการพัฒนาคลังข้อมูลเป็นข้อมูลรูปแบบไฟล์เอ็กเซล (Excel File) และทำการสร้างคลังข้อมูลด้วยโปรแกรม Microsoft SQL Server 2008 R2 โดยใช้โครงสร้างแบบสตาร์ (Star Schema) และใช้เครื่องมือ SQL Business Intelligence Development Studio ผ่านกระบวนการอีทีแอล (Extract Transform and Load) ซึ่งผู้วิจัยได้ออกแบบส่วนของการนำเข้าข้อมูลอยู่ในรูปแบบของไฟล์เอ็กเซล (Excel File) โดยทำการคัดแยกข้อมูลที่ไม่เกี่ยวข้องออก และทำความสะอาดข้อมูลให้มีความถูกต้องและสมบูรณ์สามารถนำไปประมวลผลเชิงวิเคราะห์ออนไลน์ได้

1.2 การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ

ผู้วิจัยได้ใช้เครื่องมือ Microsoft SQL Server Analysis Service (SSAS) ในการสร้าง Analysis Service Project สำหรับการสร้างคิวบ์ข้อมูล (Data Cube) ที่เป็นฐานข้อมูลแบบหลายมิติ และกำหนดข้อมูลจากตารางข้อมูลที่มีอยู่ในคลังข้อมูลที่ได้ออกแบบโครงสร้างคลังข้อมูลแบบสตาร์ (Star Schema) ซึ่งในการประมวลผลเชิงวิเคราะห์ออนไลน์ (Online Analytical Processing) จะประกอบด้วยมิติมุมมอง (Dimension) ต่างๆ ได้ดังนี้ มิติสำนักงานประกันสังคมพื้นที่ (DT_Area) , มิติตำแหน่ง (DT_Position), มิติสถานประกอบการ (DT_Enterprise), มิติประเภทกิจการ (DT_BusinessType) มิติอวัยวะ (DT_Organ) มิติสาเหตุการประสบอันตราย (DT_CauseInjury)

มิติระยะเวลา (DT_Period) และมิติอายุ (DT_Age) โดยมีเมเชอร์หรือค่าที่ต้องการวัด (Measure) คือ จำนวนรายการประสบอันตราย (FT_ALL_Injury) ซึ่งสามารถนำข้อมูลมาแสดงแบบหลายมิติได้ โดยใช้เครื่องมือ Pivot Table ของโปรแกรม Microsoft Excel 2007 ซึ่งจะทำการเชื่อมต่อกับคิวบ์ที่ได้สร้างไว้ มาสร้างเป็นมุมมองของกลุ่มผู้ใช้งาน โดยแบ่งเป็น 3 กลุ่ม คือ ผู้บริหารจะแสดงรายงานการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน ในแต่ละปีงบประมาณ หัวหน้าส่วนงานกองทุนเงินทดแทนใช้ดูข้อมูลสถานประกอบการที่ประสบอันตราย ประเภทกิจการที่ประสบอันตราย เพื่อวิเคราะห์การเรียกเก็บอัตราเงินสมทบ และเจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทนใช้ดูข้อมูลรายบุคคล เพื่อนำไปวิเคราะห์การจ่ายค่ารักษาพยาบาล จากผลการประเมินความพึงพอใจคลังข้อมูล โดยผู้ใช้งานจำนวน 9 ท่าน คือ ผู้บริหาร จำนวน 3 ท่าน หัวหน้าส่วนงานกองทุนเงินทดแทน จำนวน 2 ท่าน และเจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทน จำนวน 4 ท่าน พบว่า ผลการประเมินความความพึงพอใจโดยรวมมีค่าเฉลี่ยรวมเท่ากับ 4.06 เมื่อเทียบกับเกณฑ์อยู่ในระดับที่ดี ระบบคลังข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทนนั้น มีประสิทธิภาพมากสามารถนำไปใช้งานได้

1.3 การสร้างตัวแบบพยากรณ์

ข้อมูลที่นำมาใช้ในการทำเหมืองข้อมูลได้มาจากคลังข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทนที่ได้สร้างขึ้นมา ซึ่งใช้เทคนิคการทำเหมืองข้อมูลมาสร้างตัวแบบพยากรณ์ โดยผู้วิจัยดำเนินงานวิจัยในการทำเหมืองข้อมูลตามแบบจำลองของคริสพ์ - ดีเอ็ม (CRISP - DM Model) ซึ่งใช้ข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน สำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 3 ปี 2557 จำนวน 3,636 ชุดข้อมูล ประกอบด้วยข้อมูลจำนวน 11 คุณลักษณะ ได้แก่ Code Injury, Code Position, Code Enterprise, Code Business Type, Code Organ, Code Cause Injury, Sex, Age, Group Age, Compensation Rate, Compensation Fund

การสร้างตัวแบบพยากรณ์ใช้ข้อมูลเป็น 2 แบบ คือ ชุดข้อมูลแบบไม่จัดกลุ่ม (Original Data) เป็นข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน โดยเลือกใช้ Data set สำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 3 ปี 2557 จำนวนข้อมูล 3,636 ชุดข้อมูล มาทำการคัดเลือกคุณลักษณะ (Feature Selection) ด้วยวิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection และได้ตัวแปรทั้งหมด 6 คุณลักษณะ และชุดข้อมูลแบบจัดกลุ่ม (Clustering Data) โดยใช้อัลกอริทึม k-means จำนวน Cluster เท่ากับ 3 มีค่าความผิดพลาดในการจำแนกข้อมูลน้อยที่สุด เท่ากับ 22.57% ซึ่งผลการแบ่งข้อมูลตามกลุ่มที่ได้คือ Cluster0 มีจำนวนข้อมูล 322 ชุดข้อมูล Cluster1 มีจำนวนข้อมูล 415 ชุดข้อมูล และ Cluster2 มีจำนวน

ข้อมูล 2,899 ชุดข้อมูล มาทำการคัดเลือกคุณลักษณะ (Feature Selection) ด้วยวิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection และได้ตัวแปรทั้งหมด 7 คุณลักษณะ แล้วใช้เทคนิคโครงข่ายประสาทเทียมมัลติเลเยอร์เพอร์เซฟตรอน (Multilayer Perceptron) เทคนิคการจำแนกแบบเบย์ (Naïve Bayes) และเทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม J48 ผลการวิจัยพบว่า ตัวแบบพยากรณ์ด้วยชุดข้อมูลแบบไม่จัดกลุ่ม (Original Data) ร่วมกับเทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม J48 ซึ่งให้ค่าความถูกต้อง (Accuracy) ที่ร้อยละ 100 มีค่าความแม่นยำ (Precision) และค่า (Recall) เท่ากับร้อยละ 1.000 ส่วนค่ารากที่สองของความคลาดเคลื่อน (Root Mean Squared Error) เท่ากับ 0 ซึ่งเหมาะสมที่สุดกับข้อมูลที่ใช้ในการสร้างตัวแบบพยากรณ์ในงานวิจัยนี้มากที่สุด และยังสามารถนำไปใช้ในการวิเคราะห์การจัดตั้งงบประมาณของผู้บริหาร สำหรับการบริหารงานกองทุนเงินทดแทน สำนักงานประกันสังคม และใช้เป็นการประเมินผลการจัดฝึกอบรมการป้องการเกิดการประสบอันตรายจากการทำงาน ตามนโยบายและพันธกิจของหน่วยงาน ที่ให้ความคุ้มครองดูแลลูกจ้าง นายจ้าง และผู้ประกันตนอย่างต่อเนื่องไป

การพัฒนาระบบพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน สำนักงานประกันสังคม ผู้วิจัยได้ใช้รูปแบบการแสดงผลบนเว็บไซต์ (Web Application) โดยใช้โปรแกรมภาษา PHP ร่วมกับโปรแกรม Adobe Dreamweaver CS 6.0 ในการพัฒนาเพื่อเพิ่มความสะดวกในการใช้งาน ให้สามารถใช้งานได้จริง เพื่อประกอบการตัดสินใจในการใช้จัดตั้งงบประมาณประจำปี และช่วยในการวิเคราะห์ลดอัตราการเกิดอุบัติเหตุ เพื่อการบริหารการจัดการกองทุนเงินทดแทนที่ดียิ่งขึ้น

2. อภิปรายผล

2.1 การพัฒนาคลังข้อมูล

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาคลังข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน ซึ่งใช้ข้อมูลจากฐานข้อมูลของสำนักงานกองทุนเงินทดแทน สำนักงานประกันสังคม โดยมีการรวบรวมข้อมูล และทำการวิเคราะห์ออกแบบคลังข้อมูลโดยใช้โครงสร้างแบบสตาร์ (Star Schema) และทำการพัฒนาคลังข้อมูลด้วยโปรแกรม Microsoft SQL Server 2008 R2 และใช้เครื่องมือ SQL Business Intelligence Development Studio ในการสร้างกระบวนการอีทีแอล (Extract Transform and Load) แล้วนำมาสร้างคิวบ์ (Cube) ด้วยเครื่องมือ Analysis Service Project เพื่อให้สามารถพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ โดยเป็นการประมวลผลเชิงวิเคราะห์ออนไลน์ (Online Analytical Processing) ผู้ใช้งานสามารถเรียกดู

รายงานผ่านการใช้เทคนิค Pivot Table ของโปรแกรม Microsoft Excel 2007 จากหลายมิติมุมมองที่ได้ออกแบบสำหรับผู้บริหารใช้ดูข้อมูลจำนวนการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน ในแต่ละปีงบประมาณ หัวหน้าส่วนงานกองทุนเงินทดแทนใช้ดูข้อมูลสถานประกอบการที่ประสบอันตราย ประเภทกิจการที่ประสบอันตราย เพื่อวิเคราะห์การเรียกเก็บอัตราเงินสมทบ และเจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทนใช้ดูข้อมูลรายบุคคล เพื่อนำไปวิเคราะห์การจ่ายค่ารักษาพยาบาล ซึ่งสอดคล้องกับงานวิจัยของ (นายอภิชาติ ปัญญา : 2557) ในการสร้างคลังข้อมูลการเข้าใช้เว็บไซต์มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ลำปาง ซึ่งผลการวิจัยพบว่า สามารถนำมาใช้ในการวางแผนพัฒนาปรับปรุงและดูแลเว็บไซต์ อีกทั้งยังทำให้ผู้บริหารได้รับรายงานในรูปแบบหลายมิติที่สามารถนำมาประกอบการตัดสินใจ เพื่อประโยชน์ในการบริหารเว็บได้อย่างมีประสิทธิภาพ

2.2 การทำเหมืองข้อมูล

ในงานวิจัยครั้งนี้ ผู้วิจัยได้นำเสนอการใช้เทคนิคเหมืองข้อมูลในการสร้างตัวแบบพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินจากกองทุนเงินทดแทน โดยใช้ชุดข้อมูล 2 แบบ คือ ข้อมูลแบบไม่จัดกลุ่ม (Original Data) และข้อมูลแบบจัดกลุ่ม (Data Cluster) และนำมาผ่านวิธีการคัดเลือกคุณลักษณะ (Feature Selection) ซึ่งสอดคล้องกับงานวิจัยของ (วงกต ศรีอุไร : 2557) ร่วมกับเทคนิคการทำเหมืองข้อมูล ด้วยวิธีการจำแนกประเภท (Classification) โดยเลือกใช้เทคนิคโครงข่ายประสาทเทียมมัลติเลเยอร์เพอร์เซฟตรอน (Multilayer Perceptron) เทคนิคการจำแนกแบบเบย์ (Naïve Bayes) และต้นไม้ตัดสินใจ (Decision Tree) โดยใช้อัลกอริทึม J48 ในส่วนชุดข้อมูลแบบจัดกลุ่ม (Clustering Data) ด้วยอัลกอริทึม K-means แล้วทำการทดสอบประสิทธิภาพด้วยวิธี K-fold Cross Validation โดยเลือกใช้ K เท่ากับ 10 ซึ่งสอดคล้องกับงานวิจัยของ (เสกสรรค์ วิลัยลักษณ์ วิภา เจริญภัณฑารักษ์ และดวงดาว วิชาดากุล, 2558) เพื่อนำมาใช้ประกอบการตัดสินใจของผู้บริหารในการจัดตั้งงบประมาณประจำปี

นอกจากนี้ สำหรับการพยากรณ์การใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน ผลที่ได้จากการพยากรณ์ของตัวแบบที่สร้างขึ้นนั้น มีร้อยละของความถูกต้องค่อนข้างสูง โดยผลการพยากรณ์ถูกนำเสนอเป็นแนวโน้มของการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทนโดยวัดอัตราความเสี่ยง โดยจะแบ่งเป็น High, Medium หรือ Low เพื่อเป็นข้อมูลประกอบการตัดสินใจในการใช้จัดตั้งงบประมาณประจำปี และใช้ในการวางแผนลดอัตราการเกิดประสบอันตรายจากการทำงาน สำหรับผู้บริหาร หัวหน้าส่วนงานกองทุนเงินทดแทน และเจ้าหน้าที่ปฏิบัติงานกองทุนเงินทดแทน

ซึ่งผลของการพยากรณ์ที่ได้ ไม่ได้หมายความว่าผู้ใช้งานระบบการพยากรณ์จะได้ผลการพยากรณ์ที่ตรงตามความต้องการร้อยเปอร์เซ็นต์ แต่หมายความว่าข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการนั้น มีผลการพยากรณ์ในรูปแบบเดียวกับกลุ่มของอตราความเสี่ยงนั้นๆ ซึ่งเป็นเพียงข้อมูลที่ใช้ประกอบการตัดสินใจในการใช้จัดตั้งงบประมาณประจำปี และวางแผนลดอัตราการเกิดประสบอันตรายจากการทำงาน เป็นต้น

3. ปัญหาและอุปสรรค

3.1 จากนำข้อมูลที่ได้ออกมาจากระบบของหน่วยงาน ซึ่งได้ข้อมูลออกมาในรูปแบบของไฟล์เอ็กเซล (Excel File) และข้อมูลที่ได้มานั้นมีความเสียหายของข้อมูล และข้อมูลส่วนใหญ่เป็นข้อมูลที่ไม่สามารถนำมาออกแบบฐานข้อมูล (Database) ได้ทันที ต้องทำการจัดรูปแบบข้อมูลให้สามารถนำเข้าฐานข้อมูล (Database) ในคลังข้อมูลได้

3.2 การทำกระบวนการอีทีแอล (Extract Transform and Load) ซึ่งในกระบวนการนี้ต้องทำการคัดลอกข้อมูล เคลียร์ข้อมูล และทำการโหลดข้อมูลนั้น เนื่องจากข้อมูลที่ได้มาจากระบบเป็นข้อมูลที่มีปัญหา และข้อมูลที่ทำการสร้างคลังข้อมูลทั้งตารางข้อเท็จจริง (Fact Table) และตารางมิติ (Dimension Table) ไม่ตรงกัน ซึ่งโปรแกรมที่ใช้ในการทำกระบวนการอีทีแอลจึงเกิดข้อผิดพลาด (Error) ดังนั้นผู้วิจัยจึงต้องหาสาเหตุและแก้ไขปัญหา โดยทำการแก้ไขในส่วนของตารางข้อเท็จจริง (Fact Table) และตารางมิติ (Dimension Table) ให้มีข้อมูลที่ตรงกันเพื่อที่จะสามารถรันโปรแกรมในส่วนของกระบวนการอีทีแอลได้ตามปกติ

3.3 การจัดทำวิจัยครั้งนี้ ไม่สามารถนำข้อมูลด้านการเงินของกองทุนเงินทดแทน สำนักงานประกันสังคม มาใช้ในการทำคลังข้อมูลและสร้างตัวแบบพยากรณ์ได้ เนื่องจากข้อมูลด้านการเงิน ซึ่งแบ่งเป็น ค่ารักษา ค่าทดแทน และค่าสูญเสียอวัยวะ เป็นข้อมูลที่มีการเปลี่ยนแปลงตลอดเวลา ซึ่งกองทุนเงินทดแทน สำนักงานประกันสังคม ให้ความคุ้มครองลูกจ้างเป็นระยะเวลายาวนานมากถึง 15 ปี จึงทำให้ข้อมูลด้านการเงินมีการเปลี่ยนแปลงทุกเดือน ทุกปี จนกว่าจะมีการจ่ายค่ารักษาพยาบาล ค่าทดแทน และค่าสูญเสียอวัยวะจนครบจำนวนเงินและครบจำนวนปีที่ต้องจ่ายตามพระราชบัญญัติกองทุนเงินทดแทน พ.ศ.2537

4. ข้อเสนอแนะ

4.1 ข้อเสนอแนะในการนำผลการวิจัยไปใช้

4.1.1 การศึกษาวิจัยครั้งนี้เป็นการประยุกต์ใช้เทคนิคของการทำเหมืองข้อมูลในการวิเคราะห์หาความรู้ใหม่ และนำผลที่ได้มาช่วยให้ผู้บริหารใช้ในการตัดสินใจการบริหารงานกองทุนเงินทดแทน สำนักงานประกันสังคม

4.1.2 การพัฒนาคลังข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการ ที่เบิกเงินกองทุนทดแทน เพื่อช่วยให้ข้อมูลจากฐานข้อมูลมาจัดเก็บเข้าสู่คลังข้อมูล ทำให้ได้ข้อมูลที่ถูกต้องตรงความต้องการของผู้ใช้งาน และสามารถนำไปวิเคราะห์เพื่อออกรายงานช่วยในการตัดสินใจและปฏิบัติงานได้ดีมากยิ่งขึ้น

4.1.3 การพัฒนาระบบสนับสนุนการประมวลผลวิเคราะห์เชิงออนไลน์ เพื่อให้ได้รายงานตามความต้องการของผู้บริหารในรูปแบบหลายมิติ โดยผู้บริหารสามารถใช้รายงานมาช่วยตัดสินใจจัดตั้งงบประมาณ หัวหน้าส่วนงานกองทุนเงินทดแทนใช้ดูรายงานเพื่อวิเคราะห์การกำหนดอัตราการเรียกเก็บเงินสมทบ จากรายงานการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน และเจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทน ใช้วิเคราะห์ในการจ่ายค่ารักษาพยาบาลของลูกจ้างต่อไป จากรายงานในรูปแบบหลายมิติทำให้การปฏิบัติงานและการบริหารงานมีประสิทธิภาพมากยิ่งขึ้น

4.2 ข้อเสนอแนะในการวิจัยครั้งต่อไป

4.2.1 การวิจัยครั้งต่อไปควรนำข้อมูลที่มากกว่า 2 ปี มากทำการสร้างคลังข้อมูลเพื่อใช้ในการวิเคราะห์ข้อมูล และจะสามารถมองเห็นถึงความแตกต่างการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนทดแทน มากกว่าเดิม

4.2.2 การวิจัยครั้งต่อไปควรเพิ่มเทคนิควิธีการจำแนกประเภท (classification) ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) ไม่สามารถให้คำตอบที่เป็นค่าต่อเนื่องได้ ซึ่งจะแสดงผลเป็นค่าระดับ หรือ ค่าช่วง ซึ่งในการวิจัยครั้งหน้าควรเลือกใช้วิธีโครงข่ายประสาทเทียม (Neural Network) ที่ให้ค่าการพยากรณ์แบบต่อเนื่องได้

4.2.3 การวิจัยครั้งต่อไปควรเพิ่มเพิ่มปัจจัยด้านการเงินและด้านอื่นๆ ที่น่าจะมีผลต่อการใช้สิทธิของลูกจ้างที่ประสบอันตรายจากการทำงาน มาพิจารณาเป็นปัจจัยในการพยากรณ์และนำไปพัฒนาเป็นแอปพลิเคชันเพื่อใช้กับสมาร์ทโฟน เพื่อสะดวกในการใช้งานต่อไป

ภาคผนวก

แบบประเมินความพึงพอใจ
ระบบคลังข้อมูลการใช้สิทธิลูกจ้างของสถานประกอบการที่เบิกเงินกองทุนเงินทดแทน

คำชี้แจง

1. แบบสอบถามชุดนี้สำหรับผู้ใช้งาน เป็นแบบประเมินความพึงพอใจการใช้งานคลังข้อมูลผลการเรียนของนักเรียน เพื่อนำไปใช้ในการปรับปรุงต่อไป

2. การแสดงความคิดเห็นของผู้ใช้งานที่มีต่อคลังข้อมูลผลการเรียนของนักเรียนที่ได้พัฒนาขึ้นจะประกอบด้วยส่วนของรายการประเมินที่อยู่ด้านซ้ายมือ และมาตราส่วนประมาณค่าอยู่ด้านขวามือ จำนวน 5 ช่อง โปรดทำเครื่องหมาย √ ลงในช่องทางด้านขวามือที่ตรงกับความคิดเห็นของท่าน โดยกำหนดค่าความหมายดังนี้

5 เท่ากับ ระดับความพึงพอใจมากที่สุด
4 เท่ากับ ระดับความพึงพอใจมาก
3 เท่ากับ ระดับความพึงพอใจปานกลาง
2 เท่ากับ ระดับความพึงพอใจน้อย
1 เท่ากับ ระดับความพึงพอใจน้อยที่สุด

รายละเอียดแบบสอบถาม

แบบสอบถามแบ่งเป็น 2 ตอน ดังนี้

ตอนที่ 1 ข้อมูลทั่วไป

ตอนที่ 2 ความพึงพอใจการใช้งานและข้อเสนอแนะ

ตอนที่ 1 ข้อมูลทั่วไป

เพศ ☐ ชาย ☐ หญิง

อายุ ☐ 15-30 ปี ☐ 31-45 ปี ☐ 46-60 ปี

สถานภาพ ☐ ผู้บริหาร ☐ หัวหน้าส่วนงานกองทุนเงินทดแทน

☐ เจ้าหน้าที่ปฏิบัติงานด้านกองทุนเงินทดแทน

หน่วยงานที่สังกัด……..

ตอนที่ 2 ความพึงพอใจการใช้งาน

รายประเมิน	**ระดับการประเมิน**				
	มากที่สุด (5)	**มาก (4)**	**ปานกลาง (3)**	**น้อย (2)**	**น้อยที่สุด (1)**
1. มีความสะดวกและง่ายต่อการใช้งาน					
2. การออกแบบรายงานเข้าใจได้ง่าย					
3. ความถูกต้องของข้อมูล					
4. ช่วยลดขั้นตอนในการทำงาน					
5. ข้อมูลตรงกับความต้องการผู้ใช้งาน					

ตอนที่ 2 ความคิดเห็นและข้อเสนอแนะ

.....

.....

.....

.....

.....

.....

*** ขอขอบคุณสำหรับการตอบแบบประเมิน ***

บรรณานุกรม

บรรณานุกรม

กิตติพงษ์ กลมกล่อม. (2552). *การออกแบบและพัฒนาคลังข้อมูล (Data warehouse)*. กรุงเทพฯ: เคทีพี คอมพ์แอนด์ คอนซัลท์.

กุลเทพ จิรลักข์ และเอกสิทธิ์ วิวัฒนาประสิทธิ์. (2555). “กรณีศึกษาการทำคลังข้อมูล.” ใน *ประมวลสาระชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ (Data warehouse data mining and business intelligence)* หน่วยที่ 7 หน้า 299 - 383 นนทบุรี มหาวิทยาลัยสุโขทัยธรรมาธิราช บัณฑิตศึกษา สาขาวิชาวิทยาศาสตร์และเทคโนโลยี.

คเชนทร์ อึ่งสกุล. (2552). *การพัฒนาระบบคลังข้อมูลด้านแรงงานของสำนักงานสถิติจังหวัดเชียงใหม่* (ค้นคว้าอิสระวิทยาศาสตรมหาบัณฑิตไม่ได้ตีพิมพ์). มหาวิทยาลัยเชียงใหม่, เชียงใหม่.

ชฎิภกรน์ ทรายหมอ. (2556). *การพยากรณ์ภาวะแทรกซ้อน ของผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทาเหมืองข้อมูล* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยสุโขทัยธรรมาธิราช, นนทบุรี.

ชัยเลิศ พิชิตพรชัย. (2555). “การพัฒนาระบบคลังข้อมูล” ใน *ประมวลสาระชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ (Data warehouse data mining and business intelligence)* หน่วยที่ 1-7 นนทบุรี มหาวิทยาลัยสุโขทัยธรรมาธิราช บัณฑิตศึกษา สาขาวิทยาศาสตร์และ เทคโนโลยี.

ปรีชา ลิ้มตระกูล. (2559). *การพัฒนาตัวแบบการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยสุโขทัยธรรมาธิราช, นนทบุรี.

เลิศ เลิศศิริโสภณ. (2541). ถึงเวลาของดาต้าแวร์เฮ้าส์แล้วหรือยัง *BCM Magazine*, l9, (115), 9 กันยายน 2541.

วงกต ศรีอุไร. (2557). “การจำแนกผู้ป่วยโรคอ้วนลงพุงโดยใช้วิธีการคัดเลือกคุณลักษณะและโครงข่ายประสาทเทียม” *วารสารวิทยาศาสตรมหาวิทยาลัยศรีนครินทรวิโรฒ,* 1 ,91-102.

วรารัตน์ รุ่งวรวุฒิ. (2554). *เอกสารประกอบการบรรยาย วิชาการทำเหมืองข้อมูล (Data Mining)*. ขอนแก่น: มหาวิทยาลัยขอนแก่น, ขอนแก่น.

วิภา เจริญภัณฑารักษ์. (2555). “หลักการพื้นฐานของการทำเหมืองข้อมูล” ใน *ประมวลสาระชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ (Data warehouse data mining and*

business intelligence) หน่วยที่ 8 หน้า 1-62 นนทบุรี มหาวิทยาลัยสุโขทัยธรรมาธิราช บัณฑิตศึกษา สาขาวิชาวิทยาศาสตร์และเทคโนโลยี.

สำนักงานประกันสังคม กระทรวงแรงงาน. (2537). *พระราชบัญญัติเงินทดแทน พ.ศ.2537* นนทบุรี: สำนักงานประกันสังคม กระทรวงแรงงาน. นนทบุรี.

_____. (2533). *พระราชบัญญัติประกันสังคม พ.ศ.2533* นนทบุรี: สำนักงานประกันสังคม กระทรวงแรงงาน. นนทบุรี.

สุวรรณี อัศวกุลชัย. (2555). "หลักการพื้นฐานของคลังข้อมูล" ใน *ประมวลสาระชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ (Data warehouse data mining and business intelligence)* หน่วยที่ 1 หน้า 1-45 นนทบุรี มหาวิทยาลัยสุโขทัยธรรมาธิราช บัณฑิตศึกษา สาขาวิชาวิทยาศาสตร์และเทคโนโลยี.

เสกสรรค์ วิลัยลักษณ์. (2558). *การใช้เทคนิคการทำเหมืองข้อมูลเพื่อการพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยสุโขทัยธรรมาธิราช, นนทบุรี.

อภิชาติ ปัญญา. (2557). *การทำเหมืองข้อมูลสำหรับการพัฒนาเว็บไซต์ กรณีศึกษาเว็บไซต์มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ลำปาง* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยสุโขทัยธรรมาธิราช, นนทบุรี.

Ballard, C. and others. (1998). *Data Modelling Techniques for data warehouse*,. California: IBM Corporation, International Technical Support Organization.

Best, John W. (1963). *Research in Education*. New Jersey: Prentice-Hall.

C.Lakshmi Devasena. (2013). Classification of multivariate data sets without missing values using memory based Classifiers-An Effectiveness evaluation. International Journal of Artificial Intelligence & Applications. VOL. 4, No. 1, (2013): 129-142.

Cui, Z., Damiani, E., & Leida, M. (2007). Benefits of Ontologies in Real Time Data Access. Digital Ecosystems and Technologies Conference, DEST.

Kurniawan and Halim. (2013). Use Data Warehouse and Data Mining to Predict Student Academic Performance in School: A Case Study. IEEE International Conference on Teaching 2013, Kuta, Indonesia.

Han, Jiawei and Kamber, Micheline. (2006). *Data Mining: Concepts and Techniques* (2nd ed). New York: Morgan Kaufmann.

https://th.wikipedia.org/wiki/ธุรกิจอัจฉริยะ Retrieved August 29, 2016.

http://www.cs.waikato.ac.nz/~ml/weka/ Retrieved August 29, 2016.

http://www.scribd.com/doc/23295033/3-Tier-Achtecture Retrieved August 29, 2016.

Hye Rin Kim. (2009). Prediction of Forest Fires using Data Mining Methods (Master's Project). Retrieved from http://www.stats.uwo.ca/faculty/aim/2011/FireProject/R/project.pdf.

Mohammad Rifaie, Erwin J. Bls and Abdel-Rahman M. Muhsen. (2008). Data Warehouse Architecture for GIS Applications. Proceeding of iiWAS2008, Linz, Austria.

Nicholson, S. and Stanton, J. (2003). Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. Hershey PA: Idea Group Publishing.

Roiger and Geatz. (2003). Data mining A tutorial-based primer. Pearson Internation Edition, USA: Addison-Wesley.

Ron Kohavi. (1995). A study of cross-validation and bootstrap for accuracy estimation And Model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. Vol. 2, no. 12, pp. 1137-1143.

Umair Shafique and Haseeb Qaiser. (2014). A Comparative Study of Data Mining Process Model (KDD, CRISP-DM and SEMMA). International Journal of Innovation and Scientific Research. VOL. 2, No. 1, (2014): 217-222.

Stackowiak, R., Rayman, J. & Greenwald, R. (2007). Oracle Data Warehousing and Business Intelligence Solutions, Indianapolis: Wiley Publishing, Inc.

Turban, E. (1967). The Use of Mathematical Models in Plant Maintenance Decision Making. Management Science, 13, 6, pp. 342-359.

Yoshua Bengio, Yves Grandvalet. (2015). Bias in Estimating the Variance of K-Fold Cross-Validation. pp. 75-95.

ประวัติผู้วิจัย

ชื่อ	นายอิทธิ บรรลือศักดิ์
วัน เดือน ปีเกิด	29 พฤษภาคม 2533
สถานที่เกิด	นครสวรรค์
ประวัติการศึกษา	ปริญญาบริหารธุรกิจบัณฑิต มหาวิทยาลัยราชภัฏสวนดุสิต 2555
สถานที่ทำงาน	สำนักงานประกันสังคมกรุงเทพมหานครพื้นที่ 9 10/1273-8 ถนนนวมินทร์ (สุขาภิบาล1) แขวงคลองกุ่ม เขตบึงกุ่ม กรุงเทพมหานคร
ตำแหน่ง	นักวิชาการประกันสังคม