



ตัวแบบการจำแนกการประกอบอาชีพนอกระบบของแรงงานในประเทศไทย

นายภาณุ ชุช่วง

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติประยุกต์และวิทยาการวิเคราะห์ข้อมูล

บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ปีการศึกษา 2567

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ตัวแบบการจำแนกการประกอบอาชีพนอกระบบของแรงงานในประเทศไทย



นายภาณุ ชุช่วง

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาหลักสูตร

วิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติประยุกต์และวิทยาการวิเคราะห์ข้อมูล

บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ปีการศึกษา 2567

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ



ใบรับรองโครงร่างการค้นคว้าอิสระ

บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

เรื่อง ตัวแบบการจำแนกการประกอบอาชีพนอกระบบของแรงงานในประเทศไทย

โดย นายภาณุ ชูช่วง

ได้รับอนุมัติให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติประยุกต์และวิทยาการวิเคราะห์ข้อมูล

คณบดีบัณฑิตวิทยาลัย / หัวหน้า

..... ภาควิชา

(ผู้ช่วยศาสตราจารย์ ดร.สุพจน์ จันทรวิวัฒน์)

คณะกรรมการสอบการค้นคว้าอิสระ

..... ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.สุदारัตน์ นิจสุนกิจ)

..... อาจารย์ที่ปรึกษา

(รองศาสตราจารย์ ดร.วิกานดา ผาพันธุ์)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.วิลาสินี ปิระจิตร)

ชื่อ : นายภาณุ ชูช่วง
 ชื่อการค้นคว้าอิสระ : ตัวแบบการจำแนกการประกอบอาชีพนอกระบบของ
 แรงงานในประเทศไทย
 สาขาวิชา : สถิติประยุกต์และวิทยาการวิเคราะห์ข้อมูล
 มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
 อาจารย์ที่ปรึกษาการค้นคว้าอิสระหลัก : รองศาสตราจารย์ ดร.วิกานดา ผาพันธุ์
 ปีการศึกษา : 2567

บทคัดย่อ

วัตถุประสงค์ของการศึกษาค้นคว้าครั้งนี้ เพื่อสร้างตัวแบบที่สามารถจำแนกสถานภาพการทำงาน และเพื่อระบุปัจจัยที่ส่งผลต่อการประกอบอาชีพในระบบหรือนอกระบบของแรงงานในประเทศไทย โดยใช้ข้อมูลการสำรวจแรงงานนอกระบบ ปี พ.ศ. 2567 ของสำนักงานสถิติแห่งชาติ ด้วยการสร้างตัวแบบต้นไม้ป่าสุ่ม และตัวแบบถดถอยโลจิสติก สำหรับตัวแบบถดถอยโลจิสติก ทำการคัดเลือกตัวแปรอิสระเพื่อเข้าสู่ตัวแบบทั้งหมด 5 รูปแบบ ได้แก่ การนำตัวแปรเข้าทั้งหมด คัดเลือกตัวแปรอิสระเพื่อเข้าสู่ตัวแบบทั้งหมด 5 รูปแบบ ได้แก่ การนำตัวแปรเข้าทั้งหมด คัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปร คัดเลือกตัวแปรด้วยวิธีลดตัวแปร คัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปรอิสระแบบขั้นตอน และคัดเลือกตัวแปรโดยพิจารณาจากตัวแปรที่มีความสำคัญของตัวแบบต้นไม้ป่าสุ่ม รวมทั้งรวมทั้งหมด 6 ตัวแบบ ผลการศึกษพบว่าตัวแบบ เทคนิคต้นไม้ป่าสุ่มมีประสิทธิภาพในการจำแนกสถานภาพแรงงานระหว่างการเป็นแรงงานในระบบหรือนอกระบบ ได้ดีที่สุดในที่สุด และจากการพิจารณาความสำคัญของตัวแปรอิสระจากตัวแบบ พบว่า ตัวแปรประเภทของค่าจ้างที่ได้รับ มีบทบาทสำคัญในการจำแนกประเภทสถานภาพแรงงานมากที่สุด รองลงมา ได้แก่ อาชีพ ค่าจ้างที่ได้รับต่อเดือน อายุ ชั่วโมงการทำงานต่อสัปดาห์ จำนวนปีที่ศึกษา ภาค จำนวนสมาชิกในครัวเรือน สถานภาพการสมรส เขตการปกครอง และเพศ ตามลำดับ อีกทั้ง ยังพบว่าตัวแบบ ตัวแบบถดถอยโลจิสติก ทั้ง 5 ตัวแบบ ให้ผลที่ใกล้เคียงกัน โดยทุกตัวแปรอิสระที่พิจารณามีความสำคัญในการจำแนกสถานภาพแรงงานอย่างมีนัยสำคัญทางสถิติ ยกเว้นตัวแปรจำนวนปีที่ศึกษา ซึ่งให้ผลแตกต่างจากตัวแบบต้นไม้ป่าสุ่มที่ตัวแปรจำนวนปีที่ศึกษาค่อนข้างมีความสำคัญในระดับกลางกับการจำแนกสถานภาพแรงงาน

(มีจำนวนทั้งสิ้น 74 หน้า)

คำสำคัญ : แรงงานนอกระบบ, การวิเคราะห์ถดถอยโลจิสติก, ต้นไม้ป่าสุ่ม

อาจารย์ที่ปรึกษาการค้นคว้าอิสระหลัก

Name : Mr. Panu Chuchuang
Independent Study Title : Classification model for informal employment in
Thailand
Major Field : Applied Statistics and Data Analytics
King Mongkut's University of Technology North
Bangkok
Independent Study Advisor : Associate Professor Wikanda Phaphan, Ph.D.
Academic Year : 2024

ABSTRACT

The objective of this study is to develop a model for classifying employment status and identifying factors influencing whether workers in Thailand are engaged in formal or informal employment. The study utilizes data from the 2024 informal employment survey conducted by the National Statistical Office of Thailand. Two classification models, random forest and logistic regression, were developed. For logistic regression, independent variable selection was performed using five methods such as enter regression (including all variables), forward selection, backward elimination, stepwise selection, and feature importance derived from the random forest model, resulting in a total of six models. The results indicate that the Random Forest model provides the highest classification accuracy in distinguishing between formal and informal employment. Analyzing the importance of independent variables, the study finds that the most influential variable in classification is the type of wage received, followed by occupation, monthly wage, age, weekly working hours, years of education, region, household size, marital status, administrative area, and gender, respectively. Furthermore, all five logistic regression models yield similar results, with all independent variables significantly contributing to employment classification, except for years of education. This finding differs from the random forest model, where years of education play a moderately important role in classifying employment status.

(Total 74 Pages)

Keywords: Formal Employment, Logistic Regression, Random Forest

Advisor

กิตติกรรมประกาศ

สารนิพนธ์เรื่องตัวแบบการจำแนกการประกอบอาชีพนอกระบบของแรงงานในประเทศไทยสำเร็จลุล่วงไปได้ด้วยดี ด้วยความอนุเคราะห์และความกรุณาอย่างยิ่งจาก รองศาสตราจารย์ ดร.วิภาดา ฝาพันธ์ อาจารย์ที่ปรึกษาสารนิพนธ์ รองศาสตราจารย์ ดร.พัทธ์ชนก ศรีสุระเดชชัย ผู้ช่วยศาสตราจารย์ ดร.สุดาร์ตน์ นิจสุนกิจ และผู้ช่วยศาสตราจารย์ ดร.วิลาสินี ปิระจิตร ที่กรุณาแนะแนวทางให้คำปรึกษาและข้อแนะนำที่เป็นประโยชน์

ขอขอบคุณสำนักงานสถิติแห่งชาติเป็นอย่างยิ่ง ที่ให้การสนับสนุนข้อมูลระดับย่อย การสำรวจแรงงานนอกระบบ ปี พ.ศ. 2567 ซึ่งเป็นองค์ประกอบสำคัญในการดำเนินการวิจัยครั้งนี้ และขอขอบคุณเจ้าหน้าที่ทุกท่านที่ให้คำแนะนำและอำนวยความสะดวกในการเข้าถึงข้อมูลดังกล่าว

นอกจากนี้ ขอกราบขอบพระคุณคณาจารย์ประจำสาขาวิชาสถิติประยุกต์ทุกท่าน ที่ให้โอกาสทางการศึกษาและประสิทธิ์ประสาทความรู้ให้แก่ผู้วิจัยกระทั่งสำเร็จการศึกษา และขอขอบคุณเจ้าหน้าที่ประจำสาขาวิชาสถิติประยุกต์ที่อำนวยความสะดวกในการทำงาน

สุดท้ายนี้ ข้าพเจ้าขอขอบคุณทุกท่านที่มีส่วนร่วมในการทำให้การศึกษาค้นคว้าครั้งนี้สำเร็จลุล่วง

ภาณุ ชูช่วง

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญ	ช
สารบัญตาราง	ฌ
สารบัญรูปภาพ	ญ
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญ	1
1.2 วัตถุประสงค์	2
1.3 นิยามศัพท์	2
1.4 ขอบเขตการวิจัย	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 สถานการณ์แรงงานนอกระบบของประเทศไทย	4
2.2 เทคนิคต้นไม้ป่าสุ่ม	6
2.3 การวิเคราะห์ถดถอยโลจิสติก	9
2.4 งานวิจัยที่เกี่ยวข้อง	13
บทที่ 3 วิธีการดำเนินงานวิจัย	17
3.1 ข้อมูลที่ใช้	17
3.3 ตัวแปรที่ใช้ในการศึกษา	19
3.4 การสร้างตัวแบบ	22
3.5 การประเมินประสิทธิภาพของตัวแบบ	24
บทที่ 4 ผลการวิเคราะห์	26
4.1 ข้อมูลทั่วไปของหน่วยตัวอย่างที่ใช้ในการวิเคราะห์	26
4.2 ตัวแบบการจำแนกการประกอบอาชีพนอกระบบของแรงงานในประเทศไทย	30
4.2.1 Random Forest	31
4.2.2 Logistic Regression	33
4.3 เปรียบเทียบประสิทธิภาพของตัวแบบ	43

สารบัญ (ต่อ)

	หน้า
บทที่ 5 สรุปและอภิปราย	44
5.1 สรุปผลการวิเคราะห์	44
5.2 ข้อเสนอแนะ	45
บรรณานุกรม	46
ภาคผนวก	47
ชุดคำสั่งโปรแกรม R	48
แบบสอบถามการสำรวจภาวะการทำงานของประชากร ไตรมาส 3 พ.ศ. 2567	63
แบบสอบถามการสำรวจแรงงานนอกระบบ พ.ศ. 2567	71
ประวัติผู้เขียน	74



สารบัญตาราง

ตารางที่	หน้า
3-1 ตัวแปรที่ใช้ในการศึกษา	19
3-2 Confusion Metrix	24
4-1 ข้อมูลหน่วยตัวอย่างจำแนกตามภาค	26
4-2 ข้อมูลหน่วยตัวอย่างจำแนกตามเขตการปกครอง	27
4-3 ข้อมูลหน่วยตัวอย่างจำแนกตามเพศ	27
4-4 ข้อมูลหน่วยตัวอย่างจำแนกตามสถานภาพการสมรส	27
4-5 ข้อมูลหน่วยตัวอย่างจำแนกตามอาชีพ	28
4-6 ข้อมูลหน่วยตัวอย่างจำแนกตามประเภทค่าจ้างที่ได้รับ	29
4-7 ข้อมูลจำนวนสมาชิกในครัวเรือนของหน่วยตัวอย่าง	29
4-8 ข้อมูลอายุของหน่วยตัวอย่าง	29
4-9 ข้อมูลจำนวนปีที่ศึกษาของหน่วยตัวอย่าง	29
4-10 ข้อมูลจำนวนชั่วโมงการทำงานต่อสัปดาห์ของหน่วยตัวอย่าง	30
4-11 ข้อมูลค่าจ้างที่ได้รับต่อเดือนของหน่วยตัวอย่าง	30
4-12 Confusion Matrix ของตัวแบบ Random Forest	32
4-13 Mean Decrease in Gini ของตัวแปรอิสระในตัวแบบการจำแนก Random Forest	32
4-14 Logistic Regression: Enter Regression	33
4-15 Confusion Matrix ของตัวแบบ Logistic Regression: Enter Regression	35
4-16 Logistic Regression: Forward Selection	35
4-17 Confusion Matrix ของตัวแบบ Logistic Regression: Forward Selection	37
4-18 Logistic Regression: Backward Elimination	37
4-19 Confusion Matrix ของตัวแบบ Logistic Regression: Backward Elimination	39
4-20 Logistic Regression: Stepwise Selection	39
4-21 Confusion Matrix ของตัวแบบ Logistic Regression: Stepwise Selection	41
4-22 Logistic Regression: Feature Importance Selection	41
4-23 Confusion Matrix ของตัวแบบ Logistic Regression: Feature Importance	43
4-24 เปรียบเทียบประสิทธิภาพของตัวแบบ	43

สารบัญรูปภาพ

ภาพที่	หน้า
2-1 แนวโน้มของแรงงานในระบบและนอกระบบ พ.ศ. 2563-2567	4
2-2 ผลการสำรวจแรงงานนอกระบบ พ.ศ. 2567	5
2-3 กิจกรรมทางเศรษฐกิจของแรงงานในระบบและนอกระบบ	5
2-4 ส่วนประกอบของต้นไม้ตัดสินใจ	7
2-5 การทำงานของอัลกอริทึม Random Forest	8
4-1 ร้อยละของหน่วยตัวอย่างจำแนกตามสถานภาพแรงงานในระบบและนอกระบบ	26
4-2 ความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ	30
4-3 อันดับความสำคัญของตัวแปรอิสระในตัวอย่างการจำแนก Random Forest	33
5-1 เปรียบเทียบประสิทธิภาพของตัวแบบ	44

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญ

ในยุคปัจจุบัน แรงงานถือเป็นทรัพยากรมนุษย์ที่มีความสำคัญอย่างยิ่งในการขับเคลื่อนเศรษฐกิจและสังคมของประเทศต่างๆ โดยเฉพาะในประเทศที่มีการเติบโตทางเศรษฐกิจและการพัฒนาสังคมอย่างต่อเนื่อง แรงงานเป็นปัจจัยสำคัญในการผลิตสินค้าและบริการที่ตอบสนองต่อความต้องการของตลาด ทั้งในภาคการผลิตและบริการ โดยในประเทศไทย แรงงานมีบทบาทสำคัญทั้งในภาคที่เป็นทางการและภาคที่ไม่เป็นทางการ ซึ่งสามารถแบ่งออกเป็นแรงงานในระบบ (Formal sector) และแรงงานนอกระบบ (Informal sector) การพัฒนาศักยภาพของแรงงานในทั้งสองกลุ่มจึงเป็นปัจจัยสำคัญในการผลักดันเศรษฐกิจของประเทศไปข้างหน้า ทั้งนี้ การพัฒนาแรงงานในระบบมักได้รับการสนับสนุนจากกฎหมายและสวัสดิการต่างๆ เช่น ประกันสังคม และการคุ้มครองจากความเสี่ยงต่างๆ ขณะที่แรงงานนอกระบบยังคงต้องเผชิญกับความเสี่ยงสูงในหลายด้าน

ข้อมูลจากสำนักงานสถิติแห่งชาติ ปี พ.ศ. 2563 แรงงานนอกระบบ มีแนวโน้มลดลงจากร้อยละ 53.8 จนถึงปี พ.ศ. 2565 ร้อยละ 51.0 และค่อยๆ เพิ่มขึ้น โดยปี 2567 ประเทศไทยมีแรงงานนอกระบบประมาณ 21.1 ล้านคน ซึ่งคิดเป็นร้อยละ 52.7 ของผู้มีงานทำ ซึ่งถือเป็นกำลังแรงงานที่เป็นประชากรกลุ่มใหญ่ที่สุดของประเทศ โดยกิจกรรมทางเศรษฐกิจของแรงงานนอกระบบ มากกว่าครึ่งทำงานอยู่ในภาคเกษตรกรรม มีจำนวน 11.4 ล้านคน หรือร้อยละ 54.2 รองลงมาทำงานในภาคการบริการ และการค้าร้อยละ 36.2 และภาคการผลิตร้อยละ 9.6 ขณะที่แรงงานในระบบทำงานอยู่ในภาคการบริการและการค้ามากที่สุด รองลงมาเป็นภาคการผลิตและภาคเกษตรกรรม

แรงงานนอกระบบส่วนใหญ่จะเป็นแรงงานที่ประกอบอาชีพอิสระในกิจการขนาดเล็ก และมีลักษณะเป็นธุรกิจชั่วคราวถึงแม้แรงงานนอกระบบจะมีทั้งในส่วนที่เป็นเจ้าของกิจการที่มีธุรกิจเป็นของตนเอง แต่ส่วนใหญ่ก็ยังประกอบธุรกิจขนาดเล็กในครัวเรือน และมีรายได้น้อยต้องทำงานหนัก รวมทั้งต้องเผชิญกับสภาพปัญหาต่างๆ ในการทำงาน โดยมีสภาพปัญหาหลักๆ ในด้านที่สำคัญ 3 ด้าน ได้แก่ ปัญหาจากการทำงาน ปัญหาจากความไม่ปลอดภัยในการทำงาน และปัญหาจากสภาพแวดล้อมในการทำงาน และอีกปัญหาหนึ่งที่มีความสำคัญ คือการไม่ได้รับสิทธิประโยชน์ตามกฎหมายประกันสังคม ความคุ้มครองในด้านสิทธิประโยชน์ที่เพียงพอจากรัฐ เช่น กรณีการได้รับสิทธิประโยชน์ตามกฎหมายประกันสังคม (กระทรวงแรงงาน, 2566)

การศึกษาตัวแบบการจำแนกการประกอบอาชีพนอกระบบของแรงงานในประเทศไทย จึงมีความสำคัญเพื่อให้สามารถเข้าใจถึงสาเหตุและปัจจัยที่ทำให้แรงงานเลือกทำงานในแต่ละภาคส่วน การศึกษานี้จะช่วยให้สามารถระบุปัจจัยที่มีผลต่อการตัดสินใจในการทำงานทั้งในด้านเศรษฐกิจ สังคม

และสภาพแวดล้อมการทำงาน ซึ่งจะเป็นข้อมูลสำคัญในการพัฒนานโยบายที่สนับสนุนให้แรงงานสามารถเข้าสู่ภาคการจ้างงานที่มีระบบสวัสดิการคุ้มครองและสิทธิประโยชน์จากกฎหมาย ลดช่องว่างความเหลื่อมล้ำทางสังคม และเพิ่มโอกาสในการเข้าถึงสิทธิประโยชน์ที่เป็นธรรม ยกระดับคุณภาพชีวิตของแรงงานในระยะยาว

1.2 วัตถุประสงค์

1.2.1 เพื่อสร้างตัวแบบการจำแนกการประกอบอาชีพนอกระบบของแรงงานในประเทศไทย

1.2.2 เพื่อระบุปัจจัยที่ส่งผลต่อการประกอบอาชีพในระบบหรือนอกระบบของแรงงานในประเทศไทย

1.3 นิยามศัพท์

1.3.1 แรงงานในระบบ หมายถึง ผู้มีงานทำที่ได้รับความคุ้มครองหรือหลักประกันทางสังคมจากการทำงาน ได้แก่

- 1) ข้าราชการ ลูกจ้างประจำ ของราชการส่วนกลาง ราชการส่วนภูมิภาค และราชการส่วนท้องถิ่น
- 2) ลูกจ้างรัฐวิสาหกิจ
- 3) ลูกจ้างกองทุนสงเคราะห์ตาม พรบ. โรงเรียนเอกชน พ.ศ. 2550
- 4) ลูกจ้างของรัฐบาลต่างประเทศหรือองค์การระหว่างประเทศ
- 5) ลูกจ้างที่ได้รับความคุ้มครองตามกฎหมายแรงงาน
- 6) ผู้มีงานทำที่ประกันตนตาม พรบ. ประกันสังคม มาตรา 33, 39 และ 40

1.3.2 แรงงานนอกระบบ หมายถึงผู้มีงานทำที่ไม่ได้รับความคุ้มครอง หรือไม่มีหลักประกันทางสังคมจากการทำงาน เช่นเดียวกับแรงงานในระบบ

1.3.3 ผู้มีงานทำ หมายถึง บุคคลที่มีอายุ 15 ปีขึ้นไป ซึ่งเป็นผู้ที่ทำงานโดยได้รับค่าจ้าง เงินเดือนหรือค่าตอบแทน ที่เป็นตัวเงินหรือสิ่งของ รวมทั้งผู้ที่ไม่ได้รับค่าจ้าง เงินเดือนหรือค่าตอบแทน เนื่องจากทำงานในธุรกิจหรือไร่นาของครัวเรือน

1.4 ขอบเขตการวิจัย

1.4.1 การวิจัยนี้ศึกษาตัวแบบการจำแนกการประกอบอาชีพนอกระบบของแรงงานในประเทศไทย และเพื่อระบุปัจจัยที่ส่งผลต่อการประกอบอาชีพในระบบหรือนอกระบบ โดยพิจารณาตัวแปรจากแบบสอบถามการสำรวจแรงงานนอกระบบ พ.ศ. 2567 ของสำนักงานสถิติแห่งชาติ

1.4.2 การสำรวจแรงงานนอกระบบ ใช้แบบแผนการเลือกตัวอย่างแบบมีชั้นภูมิ 2 ชั้น (Stratified Two – Stage Sampling) และดำเนินการสำรวจพร้อมกันทั่วประเทศในระหว่างวันที่ 1-

12 ของเดือนกรกฎาคม สิงหาคม และ กันยายน พ.ศ. 2567 โดยผนวกข้อมูลกับโครงการสำรวจภาวะการทำงานของประชากร พ.ศ. 2567 ใช้วิธีการสัมภาษณ์หัวหน้าครัวเรือนหรือสมาชิกในครัวเรือนตัวอย่าง

1.4.3 **คัมรวมของการสำรวจแรงงานนอกระบบ** คือ ประชาชนที่อาศัยอยู่ในครัวเรือนส่วนบุคคลและครัวเรือนกลุ่มบุคคลประเภท ครัวเรือนคนงานทั้งในเขตเทศบาลและนอกเขตเทศบาลทุกจังหวัดทั่วประเทศ ยกเว้น ครัวเรือนชาวต่างประเทศ ที่มีเอกสิทธิ์ทางการทูต สำหรับครัวเรือนที่เข้าข่ายการเจนนับของการสำรวจแรงงานนอกระบบ ได้แก่ ครัวเรือนส่วนบุคคลและครัวเรือนกลุ่มบุคคลประเภทครัวเรือนคนงาน ซึ่งจะมีรายชื่อปรากฏในบัญชีรายชื่อครัวเรือนตัวอย่าง บุคคลที่เข้าข่ายการเจนนับ ได้แก่ ผู้ที่เป็นสมาชิกในครัวเรือนส่วนบุคคลตัวอย่าง และผู้ที่ถูกเลือกเป็นตัวอย่าง ในครัวเรือนกลุ่มบุคคลประเภทครัวเรือนคนงาน โดยในการศึกษาครั้งนี้จะพิจารณาเฉพาะผู้ที่อาศัยอยู่ในครัวเรือนส่วนบุคคล

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้ตัวแบบการจำแนกการประกอบอาชีพนอกระบบของแรงงานในประเทศไทย

1.5.2 ทราบถึงปัจจัยที่ส่งผลต่อการประกอบอาชีพในระบบหรือนอกระบบของแรงงานในประเทศไทย

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

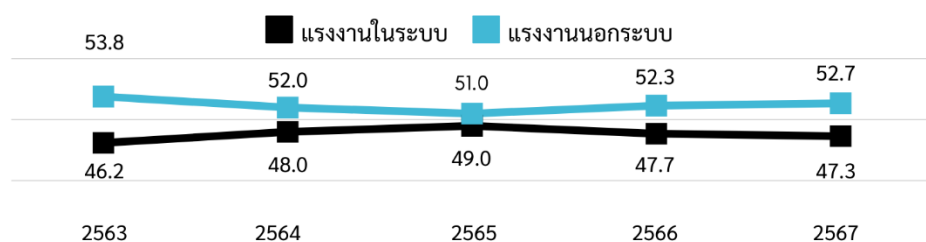
ทฤษฎีและงานวิจัยที่เกี่ยวข้องในการศึกษางานวิจัยนี้ ผู้วิจัยแบ่งออกเป็น 3 หัวข้อหลัก ได้แก่

- (1) สถานการณ์แรงงานนอกระบบในประเทศไทย
- (2) การวิเคราะห์ถดถอยโลจิสติก
- (3) งานวิจัยที่เกี่ยวข้อง

2.1 สถานการณ์แรงงานนอกระบบของประเทศไทย

แรงงานนอกระบบเป็นกลุ่มแรงงานที่มีบทบาทสำคัญต่อการขับเคลื่อนเศรษฐกิจของประเทศ แต่ในขณะเดียวกันก็เป็นกลุ่มที่ไม่ได้รับความคุ้มครองหรือไม่มีหลักประกันทางสังคมจากการทำงาน เช่นเดียวกับแรงงานในระบบ ความคุ้มครองนี้ครอบคลุมถึงสิทธิประโยชน์ด้านประกันสังคม การคุ้มครองแรงงาน และการดูแลด้านสวัสดิการต่างๆ สภาพดังกล่าวทำให้แรงงานนอกระบบมีความเสี่ยงสูงต่อความไม่มั่นคงทางเศรษฐกิจและสังคม หากเกิดปัญหาทางการเงินหรือเจ็บป่วยที่ส่งผลกระทบต่อความสามารถในการทำงาน

สำนักงานสถิติแห่งชาติได้ดำเนินการสำรวจแรงงานนอกระบบในประเทศไทยตั้งแต่ปี พ.ศ. 2548 โดยผนวกข้อมูลกับโครงการสำรวจภาวะการทำงานของประชากรในไตรมาสที่ 3 ของทุกปี (กรกฎาคม - กันยายน) ผลสำรวจให้ข้อมูลเชิงลึกเกี่ยวกับจำนวนแรงงาน ลักษณะทางประชากร ลักษณะการทำงาน และปัญหาต่างๆ ที่เกี่ยวข้องกับแรงงานนอกระบบ โดยนิยามว่าแรงงานนอกระบบคือผู้มีงานทำที่ไม่ได้รับความคุ้มครองหรือไม่มีหลักประกันทางสังคมจากการทำงาน



ภาพที่ 2-1 แนวโน้มของแรงงานในระบบและนอกระบบ พ.ศ. 2563-2567

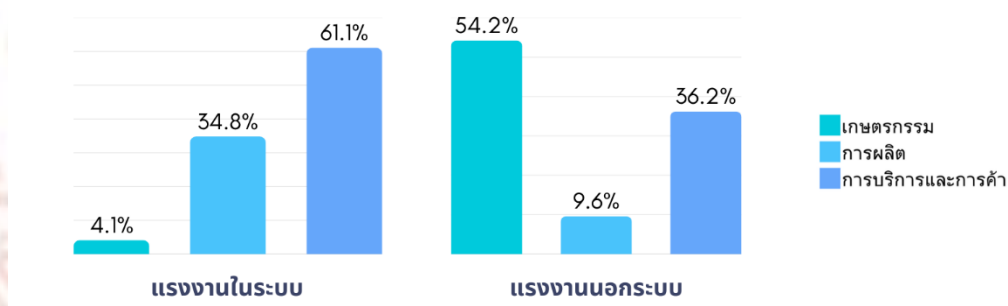
จากภาพที่ 2-1 ข้อมูลในช่วงปี 2563 - 2567 แสดงให้เห็นถึงแนวโน้มที่น่าสนใจเกี่ยวกับจำนวนแรงงานนอกระบบในประเทศไทย ในช่วงสามปีแรก (2563 - 2565) จำนวนแรงงานนอกระบบลดลงอย่างต่อเนื่อง อย่างไรก็ตาม ตั้งแต่ปี 2566 เป็นต้นมา จำนวนแรงงานนอกระบบกลับเพิ่มขึ้นเป็น 52.3% และเพิ่มขึ้นอีกเล็กน้อยในปี 2567 มาอยู่ที่ 52.7% ของจำนวนแรงงานทั้งหมด

แนวโน้มการเพิ่มขึ้นของแรงงานนอกระบบในช่วงปีหลังสะท้อนให้เห็นถึงความท้าทายด้านโครงสร้างเศรษฐกิจของประเทศ



ภาพที่ 2-2 ผลการสำรวจแรงงานนอกระบบ พ.ศ. 2567

การสำรวจแรงงานนอกระบบในปี 2567 พบว่า จากผู้มีงานทำ 40 ล้านคน เป็นแรงงานนอกระบบจำนวน 21.1 ล้านคน คิดเป็น 52.7% ซึ่งมากกว่าแรงงานในระบบที่มี 18.9 ล้านคน หรือคิดเป็น 47.3% โดยเป็นเพศชายมากกว่าเพศหญิง สำหรับกิจกรรมทางเศรษฐกิจของแรงงานนอกระบบมากกว่าครึ่งทำงานอยู่ในภาคเกษตรกรรม มีจำนวน 11.4 ล้านคน หรือ 54.2% รองลงมาทำงานในภาคการบริการและการค้า 36.2% และภาคการผลิต 9.6% ขณะที่แรงงานในระบบทำงานอยู่ในภาคการบริการและการค้ามากที่สุด รองลงมาเป็นภาคการผลิตและภาคเกษตรกรรม



ภาพที่ 2-3 กิจกรรมทางเศรษฐกิจของแรงงานในระบบและนอกระบบ

กระทรวงแรงงาน (2567) กล่าวว่า แรงงานนอกระบบถือเป็นกำลังแรงงานที่เป็นประชากรกลุ่มใหญ่ที่สุดของประเทศ โดยเป็นกลุ่มแรงงานที่มีจำนวนมากที่สุดในประเทศ ซึ่งทำงานอยู่ในเศรษฐกิจนอกระบบที่ไม่ได้รับความคุ้มครองตามกฎหมายคุ้มครองแรงงานหรือกฎหมายประกันสังคม แรงงานนอกระบบส่วนใหญ่จะเป็นแรงงานที่ประกอบอาชีพอิสระ เป็นกิจการขนาดเล็ก และมีลักษณะเป็นธุรกิจชั่วคราวถึงแม้แรงงานนอกระบบจะมีทั้งในส่วนที่เป็นเจ้าของกิจการมีธุรกิจเป็นของตนเอง แต่ส่วนใหญ่มักมีลักษณะเป็นการประกอบธุรกิจขนาดเล็กในครัวเรือน มีรายได้แค่พอเลี้ยงชีพ และส่วนใหญ่มิรายได้น้อย ต้องทำงานหนัก รวมทั้งต้องเผชิญกับสภาพปัญหาต่างๆ ในการทำงาน

โดยมีสภาพปัญหาหลักๆ ในด้านที่สำคัญ 3 ด้าน ได้แก่ ปัญหาจากการทำงาน ปัญหาจากความไม่ปลอดภัยในการทำงาน และปัญหาจากสภาพแวดล้อมในการทำงาน

อีกปัญหาหนึ่งที่มีความสำคัญไม่ยิ่งหย่อนไปกว่ากัน คือ การไม่ได้รับสิทธิประโยชน์ตามกฎหมายประกันสังคม ความคุ้มครองในด้านสิทธิประโยชน์ที่เพียงพอจากรัฐ เช่น กรณีการได้รับสิทธิประโยชน์ตามกฎหมายประกันสังคม แนวคิดเกี่ยวกับการสร้างหลักประกันสังคม คือ การสร้างหลักประกันในการดำรงชีวิตสำหรับสมาชิกที่เข้าร่วมโครงการ เพื่อรับผิดชอบในการเฉลี่ยความเสี่ยงที่อาจเกิดขึ้นจากการเจ็บป่วย คลอดบุตร ทูพพลภาพ เสียชีวิต สงเคราะห์บุตร ชราภาพ และว่างงาน

ดังนั้น กฎหมายจึงกำหนดสิทธิประโยชน์ของแรงงานนอกระบบตามกฎหมายประกันสังคมไว้ในมาตรา 40 แห่งพระราชบัญญัติประกันสังคม พ.ศ. 2533 และที่แก้ไขเพิ่มเติม ที่กำหนดให้บุคคลใดซึ่งมิใช่ลูกจ้างตามมาตรา 33 และไม่ได้เป็นผู้ประกันตนตามมาตรา 39 สามารถสมัครเข้าเป็นผู้ประกันตนโดยสมัครใจได้ โดยแสดงความจำนงต่อสำนักงานประกันสังคม

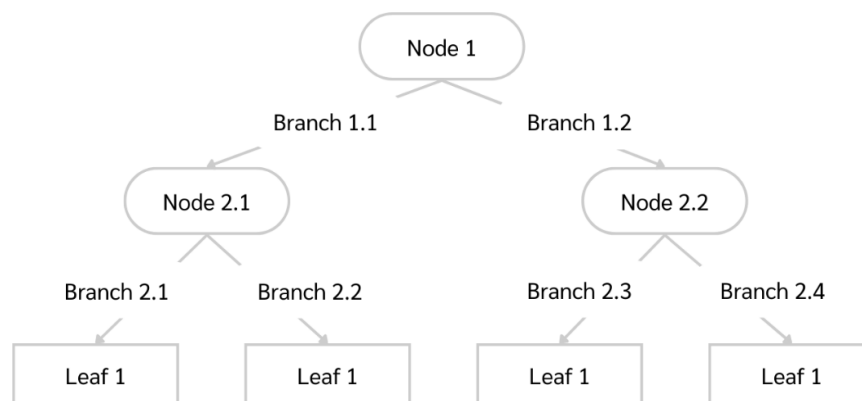
ถึงแม้กฎหมายประกันสังคมจะเปิดโอกาสให้แรงงานนอกระบบสามารถสมัครเป็นผู้ประกันตนได้โดยสมัครใจ ตามมาตรา 40 แห่งพระราชบัญญัติประกันสังคม พ.ศ. 2533 และยังมี การแก้ไขปรับปรุงสิทธิประโยชน์ของผู้ประกันตนตามมาตรานี้ให้เพิ่มมากขึ้น แต่ก็ยังมีความไม่เท่าเทียมกับแรงงานในระบบประกันสังคมที่มีนายจ้าง อีกทั้งแรงงานนอกระบบส่วนใหญ่ยังไม่เห็นความสำคัญในการเข้าสู่ระบบประกันสังคมเท่าที่ควร สิ่งเหล่านี้ถือเป็นเรื่องสำคัญที่ควรจะต้องมีการศึกษาข้อมูล และแนวทางที่เหมาะสม เพื่อเป็นข้อมูลและเป็นแนวทางในการพัฒนา ส่งเสริม และให้ความคุ้มครองแรงงานนอกระบบตามกฎหมายประกันสังคม ให้ได้รับความคุ้มครองที่มีประสิทธิภาพ มีความเหมาะสม และมีคุณภาพชีวิตที่ดียิ่งขึ้นต่อไป

2.2 เทคนิคต้นไม้ป่าสุ่ม

เทคนิคต้นไม้ป่าสุ่ม (Random forest) เป็นเทคนิคการเรียนรู้แบบมีผู้สอนแบบหนึ่งที่ใช้กันง่าย ถูกพัฒนามาจากเทคนิคต้นไม้ตัดสินใจ (Decision tree) ซึ่งเป็นการนำข้อมูลมาสร้างแบบจำลอง มีลักษณะเหมือนโครงสร้างต้นไม้ เป็นการเรียนรู้แบบมีผู้สอน (Supervised learning) คือ สร้างแบบจำลองขึ้นมาจากข้อมูลที่นำมาใช้เรียนรู้ (Training set) และสามารถทำนายกลุ่มของข้อมูลที่ยังไม่เคยนำมาจัดหมวดหมู่ได้โดยผลการทำนายจะขึ้นอยู่กับตัวแปรต้น

รูปแบบโครงสร้างต้นไม้ตัดสินใจ ประกอบด้วย โหนดราก (Root node) ซึ่งจะแตกออกเป็น โหนดลูก โดยจะมีกิ่งของต้นไม้ (Branch) ในการเชื่อมระหว่างโหนดต่างๆ และโหนดลูกระดับสุดท้ายเรียกว่า โหนดใบ (Leaf node) โดยที่แต่ละโหนดของโหนดรากและโหนดลูกจะแสดงค่าคุณลักษณะ (Attribute) ของโหนดที่แตกออกมาใช้ในการทดสอบข้อมูล โดยจำนวนของกิ่งจะเท่ากับจำนวนค่าของคุณลักษณะในโหนดนั้น ส่วนโหนดใบจะแสดงกลุ่ม (Class) ของข้อมูลที่กำหนด ซึ่งสามารถแสดงส่วนประกอบของต้นไม้ตัดสินใจ ดังภาพที่ 2-4 การทำนายกลุ่มของข้อมูลการทำงานจะเริ่มต้น

จากโหนดราก ซึ่งจะนำค่าคุณลักษณะต่างๆ ของข้อมูลไปเปรียบเทียบกับคุณลักษณะของโหนด จากนั้นทำการ ตัดสินใจในการจำแนกกลุ่มของข้อมูลโดยการเปรียบเทียบคุณลักษณะไปเรื่อยๆ จนกระทั่งถึงโหนดใบ จึงได้กลุ่มของข้อมูลที่ถูกกำหนด



ภาพที่ 2-4 ส่วนประกอบของต้นไม้ตัดสินใจ

หลักการพื้นฐานของการสร้างต้นไม้ตัดสินใจ เป็นการสร้างในลักษณะจากบนลงล่าง (Top-Down) เริ่มจากการแบ่งข้อมูลออกเป็นโครงสร้างต้นไม้ โดยที่แผนภาพนั้นจะเป็นโครงสร้างที่มีกฎต่างๆ เกิดขึ้นตามเป้าหมายของการใช้งาน เมื่อได้โครงสร้างต้นไม้แล้วจะสามารถนำโครงสร้างที่ได้นั้นไปใช้งานกับข้อมูลอื่นๆ ได้โดยนำข้อมูลนั้นผ่านกฎการตัดสินใจ (Decision rules) ขั้นตอนการสร้างต้นไม้ตัดสินใจ มีดังนี้ (Han, Kamber, & Pei, 2012)

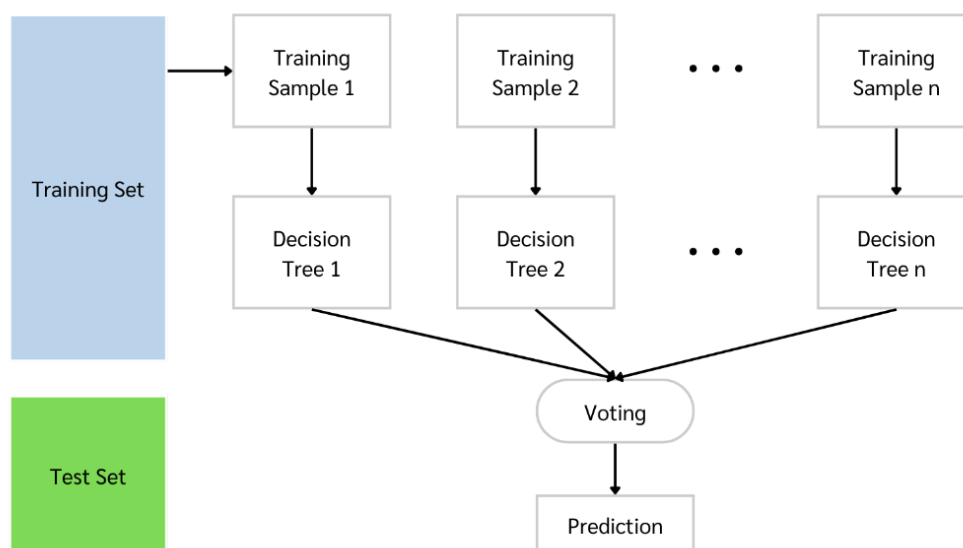
1. ต้นไม้เริ่มต้นโดยมีโหนดเพียงโหนดเดียวแสดงถึงชุดข้อมูลฝึกฝน (Training set)
2. ถ้าข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกันแล้ว ให้โหนดนั้นเป็นใบและตั้งชื่อแยกตามกลุ่มของข้อมูลนั้น
3. ถ้าในโหนดมีข้อมูลหลายกลุ่มปะปนกันอยู่ จะต้องวัดค่าคุณลักษณะ (Attribute measurement) ของแต่ละคุณลักษณะเพื่อใช้เป็นเกณฑ์ในการคัดเลือกคุณลักษณะที่มีความสามารถในการแบ่งแยกข้อมูลออกเป็นกลุ่มต่างๆ ได้ดีที่สุด ซึ่งจะถูกละเลือกให้เป็นตัวทดสอบหรือคุณลักษณะที่ใช้ในการตัดสินใจ
4. กิ่งของต้นไม้จะถูกสร้างขึ้นจากค่าของคุณลักษณะต่างๆ ที่เป็นไปได้ของโหนดทดสอบ และจะแบ่งข้อมูลออกตามกิ่งของต้นไม้ตัดสินใจที่สร้างขึ้น
5. วนซ้ำกระบวนการเดิมเพื่อหาคุณลักษณะที่มีความสามารถในการแบ่งแยกข้อมูลออกเป็นกลุ่มต่างๆ ได้ดีที่สุดสำหรับข้อมูลที่ถูกแบ่งแยกออกมาในแต่ละกิ่ง เพื่อนำ คุณลักษณะนี้มาสร้างเป็นโหนดตัดสินใจต่อไป โดยที่คุณลักษณะที่ถูกเลือกมาเป็นโหนดแล้วจะไม่ถูกเลือกมาอีกสำหรับโหนดในระดับถัดไป

6. แบ่งแยกข้อมูลและแตกกิ่งของต้นไม้ตัดสินใจไปเรื่อยๆ โดยวนซ้ำกระบวนการเดิม ซึ่งจะสิ้นสุดกระบวนการก็ต่อเมื่อเงื่อนไขข้อใดข้อหนึ่งเป็นจริง

สำหรับงานด้านการจำแนกกลุ่มข้อมูล ต้นไม้ป่าสุ่มเกิดจากการรวมกลุ่มกันของโครงสร้างต้นไม้ตัดสินใจ (Hartshorn, 2016) ซึ่งค่าความคลาดเคลื่อนโดยรวมของต้นไม้ป่าสุ่มจะถูกเปลี่ยนให้เป็นค่าลิมิต ทำให้จำนวนของต้นไม้ในป่าเพิ่มขึ้น ค่าความคลาดเคลื่อนโดยรวมจะขึ้นกับความมั่นคงของต้นไม้แต่ละต้น โดยจะใช้วิธีการสุ่มเลือกคุณสมบัติเพื่อการแบ่งแยกโหนดทำให้ค่าความผิดพลาดลดลง

หลักการที่สำคัญของเทคนิคต้นไม้ป่าสุ่ม คือการสร้างต้นไม้ตัดสินใจอย่างง่ายจำนวนมาก ในขั้นตอนการฝึกฝน และใช้วิธีการทำ Majority vote หรือการลงคะแนนเสียงข้างมากช่วยในการตัดสินใจผลการจำแนกกลุ่มของข้อมูล โดยนำผลการจำแนกของต้นไม้แต่ละต้นมารวมกันตัดสินใจ จากนั้นจึงเลือกผลลัพธ์การจำแนกที่ได้รับการโหวตมากที่สุด

ในขั้นตอนการฝึกฝนตัวจำแนกจะใช้เทคนิคที่เรียกว่า Bagging (Bootstrap Aggregation) คือ การสร้างข้อมูลย่อยหลายชุดจากชุดฝึกด้วยวิธีสุ่มแบบใส่คืน (Sampling with Replacement) และสร้างต้นไม้ตัดสินใจขึ้นมาหลายๆ ต้นจากชุดข้อมูลดังกล่าว ซึ่งต้นไม้ตัดสินใจแต่ละต้นจะไม่มี การตัดแต่ง โดยการสุ่มข้อมูลแต่ละครั้งจะมีข้อมูลส่วนหนึ่งที่ไม่ถูกเลือกมาสร้างต้นไม้ตัดสินใจ เรียกว่า Out-of-bag ซึ่งเป็นข้อมูลที่สามารถนำไปใช้ในการตรวจสอบความแม่นยำของต้นไม้ตัดสินใจแต่ละต้น ที่สร้างขึ้นได้ จากการคำนวณค่าความผิดพลาดของข้อมูล Out-of-bag



ภาพที่ 2-5 การทำงานของอัลกอริทึม Random Forest

2.3 การวิเคราะห์ถดถอยโลจิสติก

การวิเคราะห์ถดถอยโลจิสติก (Logistic Regression) เป็นเทคนิคการวิเคราะห์สถิติเชิงคุณภาพ ใช้สำหรับหาความสัมพันธ์ระหว่างตัวแปรอิสระ (มากกว่าหรือเท่ากับ 1 ตัว) กับตัวแปรตาม และพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจ เมื่อแบ่งตามระดับการวัดของตัวแปร สามารถแบ่งออกเป็น 2 ประเภทคือ การวิเคราะห์ถดถอยโลจิสติกทวิ (Binary Logistic Regression) และการวิเคราะห์ถดถอยโลจิสติกแบบพหุกลุ่ม (Multinomial Logistic Regression) ซึ่งการวิเคราะห์ถดถอยโลจิสติกดังกล่าวแตกต่างกันในด้านของตัวแปรตาม โดยการวิเคราะห์การวิเคราะห์ถดถอยโลจิสติกทวิใช้กับตัวแปรตามที่มีค่าได้เพียง 2 ค่า คือ 0 กับ 1 เช่น กลุ่มที่มีเหตุการณ์กับกลุ่มที่ไม่มีเหตุการณ์ หรือใช่และไม่ใช่ เป็นต้น ส่วนการวิเคราะห์ถดถอยโลจิสติกแบบพหุกลุ่มใช้กับตัวแปรตามที่มีมากกว่า 2 ค่า เช่น ตัวแปรตามหมายถึงรูปแบบการขนส่งของแต่ละบุคคล แก่ เครื่องบิน รถยนต์ส่วนบุคคล และรถโดยสารสาธารณะ เป็นต้น (ยุทธ ไกยวรรณ, 2555)

2.3.1 วัตถุประสงค์การวิเคราะห์ถดถอยโลจิสติก

- 1) เพื่อศึกษาความสัมพันธ์ของตัวแปรอิสระที่มีต่อโอกาสที่จะเกิดเหตุการณ์ (ตัวแปรตาม) พร้อมทั้งศึกษาระดับความสัมพันธ์ของตัวแปรอิสระแต่ละตัว
- 2) เพื่อพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจจากสมการที่เหมาะสม หรือใช้สมการโดยการเลือกตัวแปรอิสระที่เหมาะสมเพื่อให้เปอร์เซ็นต์ของความถูกต้องในการพยากรณ์มีค่าสูงสุด

2.3.2 ขั้นตอนของการวิเคราะห์ถดถอยโลจิสติก

สำหรับขั้นตอนการวิเคราะห์ถดถอยโลจิสติก จะคล้ายกับการวิเคราะห์ความถดถอยแบบปกติโดยมีขั้นตอนดังนี้

ขั้นที่ 1 เลือกตัวแปรอิสระที่คาดว่าจะมีความสัมพันธ์กับตัวแปรตาม (โอกาสที่เหตุการณ์จะเกิด) โดยที่ตัวแปรอิสระอาจมีมากกว่า 1 ตัว

ขั้นที่ 2 ตรวจสอบค่าผิดปกติของตัวแปรอิสระแต่ละตัว

ขั้นที่ 3 สร้างสมการ Logistic Response Function แล้วตรวจสอบความถูกต้องเหมาะสมของสมการ

ขั้นที่ 4 ถ้าวัตถุประสงค์เพื่อพยากรณ์ case ใหม่ว่าจะเกิดเหตุการณ์ที่สนใจหรือไม่ จะใช้สมการ $P(\text{เกิดเหตุการณ์}) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}$ ของ case นั้น เมื่อทราบตัวแปรอิสระ x 's ถ้า $P(\text{เกิดเหตุการณ์}) < 0.5$ จะได้ $y=0$ หรือไม่เกิดเหตุการณ์ และ $P(\text{เกิดเหตุการณ์}) \geq 0.5$ จะได้ $y=1$ หรือเกิดเหตุการณ์ (กัลยา, 2559)

2.3.3 เทคนิคการเลือกตัวแปรอิสระเข้าสมการถดถอยโลจิสติก

ในการวิเคราะห์ถดถอยโลจิสติกเป็นการวิเคราะห์เพื่อทำนายโอกาสที่เหตุการณ์ที่สนใจ ($y = 1$) จะเกิดขึ้น และสมการถดถอยโลจิสติกที่ดีจะต้องประกอบด้วยตัวแปรทำนายที่เหมาะสมที่จะทำให้ค่าทำนายโอกาสที่จะเกิดนั้นใกล้เคียงกับความเป็นจริง ในการเลือกตัวแปรทำนายเข้าวิเคราะห์เพื่อให้ได้สมการโลจิสติกที่ดีนั้น มีวิธีเลือก 3 วิธี Enter Method, forward method และ backward method แสดงรายละเอียด ดังนี้ (ยุทธ ไกยวรรณ, 2555)

1) Enter method

เป็นวิธีเลือกตัวแปรทำนายทั้งหมดเข้าสมการถดถอยโลจิสติกพร้อมกันในขั้นตอนเดียว ในการพิจารณาตัวแปรทำนายที่เหมาะสมที่จะนำมาใช้ทำนายในตัวแบบ ผู้วิจัยจะต้องเป็นผู้ตัดสินใจเองว่าตัวแปรทำนายตัวใดบ้างที่มีความสัมพันธ์กับตัวแปรเกณฑ์หรือควรจะอยู่ในสมการถดถอยโลจิสติก โดยพิจารณาค่าสถิติทดสอบ ถ้ามีนัยสำคัญทางสถิติถือว่าตัวแปรทำนายนั้นควรอยู่ในสมการถดถอยโลจิสติก

2) Forward method มีวิธีย่อย 3 วิธีคือ

เป็นการวิเคราะห์ถดถอยโลจิสติกแบบเดินหน้า วิธีการนี้จะคัดเลือกตัวแปรทำนายที่อธิบายความผันแปรของตัวแปรเกณฑ์ได้สูงสุด และมีนัยสำคัญทางสถิติเข้าสมการก่อน จากนั้นจึงเลือกตัวแปรทำนายที่อธิบายความผันแปรของตัวแปรเกณฑ์ได้อันดับรองลงมา และมีนัยสำคัญทางสถิติเข้าสมการตามลำดับ การนำตัวแปรทำนายเข้าสมการจะทำงานเช่นนี้เรื่อย ๆ จนไม่มีตัวแปรทำนายใดที่อธิบายความผันแปรของตัวแปรเกณฑ์อย่างมีนัยสำคัญทางสถิติอีกแล้ว การนำตัวแปรทำนายเข้าสมการแบบ Forward method มีวิธีย่อย ๆ อีก 3 วิธีได้แก่

2.1) วิธี forward stepwise: likelihood ratio

วิธีนี้บางครั้งเรียกว่า Forward LR วิธีนี้เริ่มจากการนำตัวแปรทำนายเข้าสมการทีละ 1 ตัวโดยที่ตัวแปรทำนายที่เลือกเข้าสมการทำให้ค่าทำนายโอกาสที่จะเกิดเหตุการณ์ที่สนใจถูกต้องมากขึ้น เกณฑ์ในการพิจารณาเลือกตัวแปรทำนายเข้าสมการคือ ค่าแสดงความสัมพันธ์ที่มีค่ามากที่สุดก่อน และมีนัยสำคัญทางสถิติเมื่อนำตัวแปรทำนายเข้าสมการ แล้วจะมีการตรวจสอบตัวแปรทำนายนั้นว่าควรจะถูกตัดออกหรือควรจะคงอยู่ในสมการ โดยพิจารณาจากอัตราส่วนความเป็นไปได้ หรือการเปลี่ยนแปลงของ -2LL (-2 likelihood ratio) ถ้าค่า -2LL ลดลงแสดงว่าตัวแปรทำนายควรจะคงอยู่ในสมการ

2.2) วิธี Forward stepwise: Wald

วิธีนี้เหมือนกับวิธี forward LR ทุกประการ เพียงแต่จะพิจารณาจากค่าสถิติของ Wald (Wald statistic) เท่านั้น

2.3) วิธี Forward stepwise: condition

วิธีนี้จะเหมือนกับวิธี Forward LR แตกต่างกันตรงที่ วิธี Forward LR เป็นวิธีที่ไม่มีเงื่อนไข (Unconditional) ส่วนวิธีนี้จะมีเงื่อนไข (Condition) ความแตกต่างของแบบมีเงื่อนไขและไม่มีเงื่อนไขมีดังนี้

1) แบบมีเงื่อนไขใช้กับตัวอย่างขนาดตัวอย่างเล็ก แบบไม่มีเงื่อนไขใช้กับตัวอย่างขนาดใดก็ได้

2) แบบไม่มีเงื่อนไข มีการควบคุมปัจจัยอื่นๆ ที่คาดว่าจะมีอิทธิพลต่อโอกาสที่จะเกิดเหตุการณ์ที่สนใจ เช่น ถ้าผู้วิจัยคาดว่าจะการดื่มแอลกอฮอล์และจำนวนปีที่ดื่มแอลกอฮอล์ทำให้คนเป็นโรคตับแข็ง ตัวแปรเกณฑ์ (y) คือ 1 เป็นโรคตับแข็ง และ 0 คือไม่เป็นโรคตับแข็ง ส่วนตัวแปรทำนาย x_1 คือ ดื่มแอลกอฮอล์และ x_2 คือจำนวนปีที่ดื่มแอลกอฮอล์

3) Backward method

เป็นวิธีที่นำตัวแปรทำนายทั้ง p ตัว เข้าสมการพร้อมกัน จากนั้นจึงพิจารณาตัวแปรทำนายที่อธิบายความผันแปรของตัวแปรเกณฑ์ได้น้อยที่สุดออกจากสมการก่อน ทำเช่นนี้ไปเรื่อยๆ จนกระทั่งเหลือตัวแปรทำนายที่สามารถอธิบายความผันแปรของตัวแปรเกณฑ์ได้อย่างมีนัยสำคัญทางสถิติ การนำตัวแปรทำนายเข้าสมการวิธีนี้มีวิธีย่อย 3 วิธีได้แก่

3.1) วิธี Backward stepwise: Likelihood ratio

วิธีนี้บางทีเรียกว่า Backward LR ซึ่งเป็น วิธีตรงข้ามกับ วิธี Forward stepwise: Likelihood ratio ซึ่งเป็นวิธีที่นำตัวแปรทำนายทั้งหมด p ตัว ($x_1, x_2, x_3, \dots, x_p$) เข้าสมการ แล้วพิจารณาว่าจะนำตัวแปรทำนายตัวใดออกจากสมการ โดยพิจารณานำออกทีละ 1 ตัว โดยพิจารณาจากเกณฑ์การนำตัวแปรทำนายออกจากสมการคือ จะนำตัวแปรทำนายที่ไม่มีผลต่อการทำนายโอกาสที่จะเกิดเหตุการณ์ที่สนใจ ตัวแปรทำนายตัวแรกที่จะนำออกจากสมการจะเป็นตัวแปรที่ไม่มีผลต่อการทำนายโอกาสที่จะเกิดเหตุการณ์ที่สนใจน้อยที่สุด สำหรับเกณฑ์การนำตัวแปรทำนายออกจากสมการจะพิจารณาจากอัตราส่วนความเป็นไปได้หรือจากการเปลี่ยนแปลงของ $-2LL$ เหมือนวิธี Forward stepwise: Likelihood ratio

3.2) วิธี Backward stepwise: Wald

วิธีนี้จะเหมือน backward LR ทุกประการเพียงแต่จะพิจารณาจากค่าสถิติของ Wald (Wald statistic) เท่านั้น

3.3) วิธี Backward stepwise: Condition

วิธีนี้จะเหมือน Backward LR ทุกประการ แตกต่างกันตรงที่วิธี Backward LR เป็นวิธีที่ไม่มีเงื่อนไข (Unconditional) ส่วนวิธีนี้จะมีเงื่อนไข

2.3.4 ตัวแบบการวิเคราะห์ถดถอยโลจิสติก

1) กรณีตัวแปรอิสระ 1 ตัว การวิเคราะห์ถดถอยอย่างง่าย สมการแสดงความสัมพันธ์ระหว่าง x และ y จะ อยู่ในรูปเชิงเส้นดังนี้

$$y = \beta_0 + \beta_1 x + e \quad (2-1)$$

แต่สำหรับการวิเคราะห์ถดถอยโลจิสติก ตัวแปรตามหรือ y มีค่าได้เพียง 2 ค่า คือ ไม่เกิดเหตุการณ์ ($y=0$) และเกิดเหตุการณ์ ($y=1$) ซึ่งความสัมพันธ์ไม่ได้อยู่ในรูปเชิงเส้น เนื่องจากตัวแปรตามมีค่า 0 กับ 1 ความสัมพันธ์จะอยู่ในรูป

$$P(Y) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (2-2)$$

เรียกสมการที่ 2.2 ว่า Logistic response Function โดยที่ $0 \leq P(Y) \leq 1$ เมื่อ (Y) คือ ความน่าจะเป็นของการเกิดเหตุการณ์ Y และ E คือ Exponential Function และความน่าจะเป็นของการไม่เกิดเหตุการณ์ที่สนใจ ($y=0$) มีค่าเท่ากับ $1 - P(Y)$

2) กรณีตัวแปรอิสระมากกว่า 1 ตัว เมื่อตัวแปรอิสระมากกว่า 1 ตัว หรือมีตัวแปรอิสระ n ตัว จะได้สมการความสัมพันธ์ ระหว่าง x และ y ดังนี้

$$P(Y) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}} \quad (2-3)$$

$$P(\text{ไม่เกิดเหตุการณ์}) = 1 - P(Y)$$

จากสมการความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามของการวิเคราะห์ถดถอยโลจิสติกไม่ได้อยู่ในรูปเชิงเส้น โดยปรับให้อยู่ในรูปของ Odds หรือ Odd Ratio ซึ่ง Odds หรือ Odd ratio หมายถึงอัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์ที่สนใจ ($y=1$) กับโอกาสที่จะไม่เกิดเหตุการณ์ดังนั้น Odds ของการเกิดเหตุการณ์แสดงดังสมการที่ 2-4 และแปลงให้เป็นเส้นตรงได้ดังสมการที่ 2-5 และ 2-6

$$\frac{P}{1-P} = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n} \quad (2-4)$$

$$\log \left(\frac{P}{1-P} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n \quad (2-5)$$

$$\text{Log(odds of } P) = \text{logit}(P) = \beta_0 + \beta_1 x + \dots + \beta_n x_n + \epsilon \quad (2-6)$$

โดยที่ P คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์
 β_0 คือ ค่าของ y เมื่อ x เท่ากับ 0
 β_n คือ อัตราการเปลี่ยนแปลงของ y เมื่อ x_n เปลี่ยนไป 1 หน่วย
 โดยที่ตัวแปรอิสระอื่นๆ คงที่
 ε คือ ความคลาดเคลื่อนของการพยากรณ์

2.3.5 เกณฑ์ในการเลือกตัวแบบ

พลากร สีน้อย (2553) จิตรลเวชเกณฑ์การคัดเลือกตัวแบบที่นิยมใช้และพบเห็นได้มากที่สุด คือ สัมประสิทธิ์ตัวกำหนดปรับปรุง (Adjusted Coefficient of Determination) หรือ Adjusted R^2 โดยนักวิจัยมักเข้าใจว่าตัวแบบที่ให้ค่า Adjusted R^2 สูงที่สุด แสดงว่าตัวแบบนั้นเหมาะสมที่สุด แต่ความเป็นจริงตัวแบบที่ให้ค่า Adjusted R^2 สูงที่สุด จะทำให้ตัวแบบที่ได้รับการคัดเลือกมีจำนวนตัวแปรอิสระในตัวแบบมากเกินไป (Over-fit)

ด้วยเหตุนี้ อาไคเคะ จึงได้เสนอเกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's Information Criterion: AIC) โดยตัวแบบที่ให้ค่า AIC ต่ำที่สุด แสดงว่าตัวแบบนั้นมีความเหมาะสมที่สุด ซึ่งเกณฑ์นี้ใช้ได้ดีเมื่อตัวอย่างมีขนาดใหญ่ แต่เมื่อตัวอย่างมีขนาดเล็กเมื่อเทียบกับจำนวนพารามิเตอร์ที่ต้องการประมาณ จะทำให้ตัวแบบที่ได้รับการคัดเลือกมีจำนวนตัวแปรอิสระในตัวแบบมากเกินไป (Over-fit) ซึ่งคำนวณจาก

$$AIC = -2 \ln L + 2k \quad (2-7)$$

โดยที่ L คือ Likelihood ของตัวแบบ
 k คือ จำนวนพารามิเตอร์ในตัวแบบ

2.4 งานวิจัยที่เกี่ยวข้อง

อุบลรัตน์ กอวัฒนสกุล (2563) ได้ศึกษาแรงงานนอกระบบในไทย ในประเด็นความเหลื่อมล้ำค่าแรง และความเสี่ยงในการทำงาน โดยใช้ข้อมูลสำรวจภาวะการทำงานของประชากร ปี พ.ศ. 2563 ของสำนักงานสถิติแห่งชาติ และใช้แบบจำลองการถดถอยโลจิสติก (Logistic Regression) ผลการศึกษาพบว่าลักษณะหรือตัวแปรต่างๆ ของผู้ใช้แรงงาน ได้แก่ เพศ อายุหรือประสบการณ์ทำงาน สถานภาพสมรส พื้นที่อยู่อาศัย ระดับการศึกษา ชั่วโมงการทำงานต่อสัปดาห์ และประเภทของค่าแรง มีความสัมพันธ์ทางสถิติกับโอกาสของการเป็นผู้ใช้แรงงานนอกระบบ เช่น ผู้ใช้แรงงานที่มีอายุหรือประสบการณ์ทำงานน้อย และมีสถานภาพโสดมักจะมีโอกาสเป็นแรงงานนอกระบบสูงกว่า ในทางตรงกันข้าม การเป็นหัวหน้าครัวเรือน การอาศัยอยู่ในเขตเมือง และระดับการศึกษาที่สูง ทำให้โอกาสของผู้ใช้แรงงานในการเป็นแรงงานนอกระบบลดลง นอกจากนี้ ตัวแปรที่เกี่ยวข้องกับงาน เช่น ชั่วโมงการทำงานต่อสัปดาห์และการได้รับค่าแรงรายเดือนมีความสัมพันธ์เชิงลบกับแนวโน้มการเป็นแรงงาน

นอกระบบ กล่าวคือ ผู้ใช้แรงงานที่มีชั่วโมงการทำงานต่อสัปดาห์สูงและได้รับค่าแรงเป็นรายเดือนจะมีโอกาสเป็นแรงงานนอกระบบลดลง

ช่วงชัย ชูปวา และวิภา ชูปวา (2564) ได้ศึกษาปัจจัยที่มีความสัมพันธ์ต่อคุณภาพชีวิตของแรงงานนอกระบบ โดยใช้ข้อมูลการสัมภาษณ์กลุ่มตัวอย่างแรงงานนอกระบบ จังหวัดกาฬสินธุ์ จำนวนทั้งหมด 292 คน และใช้การวิเคราะห์ถดถอยพหุคูณ (Multiple Linear Regression) ผลการศึกษาพบว่า อายุ รายได้เฉลี่ยต่อเดือน และการตรวจร่างกายประจำปี มีความสัมพันธ์ต่อคุณภาพชีวิตของแรงงานนอกระบบ

Marcouiller, Ruiz de Castilla, and Woodruff (1997) ได้ศึกษาช่องว่างค่าจ้างระหว่างแรงงานในระบบและแรงงานนอกระบบในเม็กซิโก เอลซัลวาดอร์ และเปรู โดยใช้ข้อมูลการสำรวจแรงงานในสามประเทศดังกล่าว และใช้แบบจำลองทางเศรษฐมิติเพื่อเปรียบเทียบความแตกต่างของค่าจ้างระหว่างแรงงานในระบบและแรงงานนอกระบบ ผลการศึกษาพบว่าแรงงานนอกระบบได้รับค่าจ้างต่ำกว่าแรงงานในระบบอย่างมีนัยสำคัญ อีกทั้งการศึกษายังพบว่า ช่องว่างค่าจ้างนี้ไม่ได้ขึ้นอยู่กับความสามารถหรือประสบการณ์เพียงอย่างเดียว แต่ยังมีปัจจัยอื่นๆ เช่น สภาพตลาดแรงงานในแต่ละประเทศที่มีอิทธิพลต่อการจ้างงานในแต่ละภาค โดยการศึกษาแสดงให้เห็นถึงความไม่เสมอภาคในตลาดแรงงานที่ส่งผลกระทบต่อคุณภาพชีวิตของแรงงานนอกระบบ

Carneiro และ Henley (2001) ได้ศึกษาการเปรียบเทียบการจ้างงานและรายได้ระหว่างแรงงานในระบบและแรงงานนอกระบบของประเทศบราซิล โดยใช้ข้อมูลการสำรวจทางเศรษฐกิจและสังคมในประเทศบราซิล และใช้แบบจำลองทางเศรษฐมิติ เพื่อวิเคราะห์ความแตกต่างระหว่างการจ้างงานในระบบและแรงงานนอกระบบ ผลการศึกษาพบว่า แรงงานในระบบมีรายได้สูงกว่าแรงงานนอกระบบอย่างมีนัยสำคัญ และแรงงานนอกระบบยังเผชิญกับความไม่มั่นคงทางเศรษฐกิจมากกว่าเนื่องจากขาดสิทธิประโยชน์ทางสังคมและการคุ้มครองทางกฎหมายที่มีในภาคแรงงานในระบบ นอกจากนี้ การศึกษายังพบว่า ตัวแปรต่างๆ เช่น การศึกษา ประสบการณ์การทำงาน และอายุมีอิทธิพลสำคัญต่อการเลือกเข้าทำงานในแต่ละภาค โดยที่แรงงานในระบบมักมีการศึกษาสูงและมีประสบการณ์การทำงานที่มากกว่าแรงงานนอกระบบ

Gunatilaka (2008) ได้ศึกษาลักษณะของแรงงานนอกระบบ ความน่าจะเป็นของการมีงานทำ และปัจจัยกำหนดค่าจ้างในประเทศศรีลังกา โดยใช้ข้อมูลการสำรวจแรงงานของประเทศศรีลังกา และใช้แบบจำลองการถดถอยโลจิสติก (Logistic Regression) ในการวิเคราะห์ความน่าจะเป็นของการทำงานในภาคนอกระบบ ผลการศึกษาพบว่า ตัวแปร เช่น ระดับการศึกษา อายุ และเพศ มีความสัมพันธ์อย่างมีนัยสำคัญกับความน่าจะเป็นของการทำงานในภาคนอกระบบ และค่าจ้างของแรงงานนอกระบบต่ำกว่าแรงงานในระบบ โดยเฉพาะในกลุ่มแรงงานที่มีระดับการศึกษาต่ำและไม่มีทักษะเฉพาะทาง

Gimpelson และ Zudina (2011) ได้ศึกษาแรงงานนอกระบบในรัสเซีย โดยใช้ข้อมูลการสำรวจทางเศรษฐกิจและสังคมของประเทศรัสเซีย เพื่อวิเคราะห์ปัจจัยที่กำหนดการเลือกทำงานในภาคนอกระบบ และใช้แบบจำลองการถดถอยโลจิสติก (Logistic Regression) เพื่อศึกษาอิทธิพลของตัวแปรต่างๆ ที่มีผลต่อการตัดสินใจเลือกทำงานในภาคนอกระบบ ผลการศึกษาพบว่า ตัวแปรที่มีความสัมพันธ์กับการเลือกทำงานในภาคนอกระบบ ได้แก่ การศึกษา โดยกลุ่มคนที่มีระดับการศึกษาต่ำมักจะเลือกทำงานในภาคนอกระบบมากกว่าผู้ที่มีการศึกษาสูง การเข้าถึงสิทธิประโยชน์ทางสังคม เช่น การประกันสุขภาพ หรือสวัสดิการของภาครัฐ หากบุคคลไม่ได้รับการสนับสนุนจากระบบสวัสดิการ จะมีแนวโน้มที่จะเลือกทำงานในภาคนอกระบบ นอกจากนี้ สถานการณ์ทางเศรษฐกิจของประเทศหรือการขาดโอกาสในการเข้าถึงงานในภาคทางการก็เป็นปัจจัยสำคัญที่กระตุ้นให้แรงงานหันไปทำงานในภาคนอกระบบ โดยเฉพาะในช่วงที่เศรษฐกิจตกต่ำและโอกาสในการทำงานในระบบมีจำกัด นอกจากนี้ การขาดการสนับสนุนจากนโยบายภาครัฐหรือการขาดการพัฒนา นโยบายที่เอื้อต่อการรวมแรงงานเข้าสู่ตลาดแรงงานในระบบ ก็เป็นอีกหนึ่งปัจจัยที่ส่งผลให้แรงงานเลือกทำงานในภาคนอกระบบ

De Vreyer และ Roubaud (2013) ได้ศึกษาตลาดแรงงานในเขตเมืองของแอฟริกาใต้สะฮารา โดยใช้ข้อมูลการสำรวจแรงงานในเขตเมืองของหลายประเทศในภูมิภาคแอฟริกาใต้สะฮารา ซึ่งประกอบด้วยข้อมูลเกี่ยวกับรายได้ การศึกษา การเข้าถึงบริการทางสังคม และลักษณะทางประชากรศาสตร์ เช่น อายุ เพศ และสถานภาพครอบครัว การศึกษานี้เน้นการวิเคราะห์ประเภทการจ้างงานที่แบ่งออกเป็นในระบบและนอกระบบเป็นตัวแปรตาม และใช้แบบจำลองการถดถอยโลจิสติก (Logistic Regression) เพื่อทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระต่างๆ ผลการศึกษาพบว่ารายได้ การศึกษา การเข้าถึงบริการทางสังคม สถานการณ์ทำงาน และลักษณะทางประชากรศาสตร์ เช่น อายุ เพศ และสถานภาพครอบครัว แรงงานนอกระบบมักมีรายได้ต่ำกว่าและมีการเข้าถึงบริการทางสังคมน้อยกว่า นอกจากนี้ คนที่มีการศึกษาต่ำและผู้หญิงมักมีแนวโน้มทำงานในภาคนอกระบบมากกว่าผู้ชายหรือผู้ที่มีการศึกษาสูง

García (2017) ได้ศึกษาการจ้างงานนอกระบบในโคลอมเบีย โดยใช้ข้อมูลการสำรวจแรงงานระดับภูมิภาค และใช้แบบจำลองการถดถอยเชิงควอนไทล์ (Quantile Regression) เพื่อวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต่างๆ กับการเลือกประเภทการจ้างงานในกลุ่มคนที่มีระดับรายได้แตกต่างกัน ผลการศึกษาพบว่า การศึกษา รายได้ อายุ สถานะทางครอบครัว และสถานะการทำงาน มีความสัมพันธ์กับการจ้างงานในระบบและนอกระบบ โดยในกลุ่มคนที่มีการศึกษาและรายได้ต่ำมักจะเลือกทำงานในภาคนอกระบบ เนื่องจากขาดโอกาสในการเข้าถึงงานในระบบที่มั่นคงและมีสวัสดิการที่ดีกว่า ส่วนผู้ที่มีการศึกษาหรือรายได้สูงมักมีโอกาสในการเข้าทำงานในภาคที่เป็นทางการหรือในระบบ

มากกว่า นอกจากนี้ การมีสถานะทางครอบครัวหรือสถานะการทำงานที่ไม่มั่นคงก็มีผลต่อการเลือกทำงานในภาคนอกระบบเช่นกัน

Khan และ Hussain (2020) ได้ศึกษาการวัดและปัจจัยกำหนดการจ้างงานนอกระบบในประเทศปากีสถาน โดยใช้ข้อมูลการสำรวจแรงงานของประเทศปากีสถาน และใช้แบบจำลองการถดถอยโลจิสติก (Logistic Regression) ในการวิเคราะห์ปัจจัยที่ส่งผลต่อความน่าจะเป็นของการทำงานในภาคนอกระบบ ผลการศึกษาพบว่าตัวแปร เช่น ระดับการศึกษา อายุ ขนาดครัวเรือน และสถานะการแต่งงาน มีความสัมพันธ์อย่างมีนัยสำคัญกับการจ้างงานในภาคนอกระบบ นอกจากนี้ ยังพบว่าภาคนอกระบบมีความสำคัญในตลาดแรงงานปากีสถาน แต่แรงงานในภาคนี้มักเผชิญปัญหาความไม่มั่นคงทางรายได้และขาดการคุ้มครองทางสังคม

Sibusisiwe Natarsh Mkulisi และ A. Sathiya Susuman (2024) ได้ศึกษาผลกระทบของการเปลี่ยนแปลงทางสังคม เศรษฐกิจ และประชากรศาสตร์ต่อการจ้างงานในภาคนอกระบบของแอฟริกาใต้ในช่วงปี 2017 ถึงปี 2020 โดยใช้ข้อมูลการสำรวจแรงงานของแอฟริกาใต้ และใช้แบบจำลองการถดถอยโลจิสติก (Logistic Regression) ในการวิเคราะห์ความสัมพันธ์ระหว่างปัจจัยต่างๆ เช่น อายุ เพศ เชื้อชาติ สถานภาพการสมรส และการศึกษา ผลการศึกษาพบว่าตัวแปรเพศ อายุ การศึกษา และเชื้อชาติ มีความสัมพันธ์กับการเลือกทำงานในภาคนอกระบบ โดยกลุ่มที่เป็นเยาวชน ผู้หญิง และผู้มีการศึกษาดำเนินแนวโน้มสูงในการจ้างงานในภาคนอกระบบ สะท้อนถึงความจำเป็นในการพัฒนานโยบายที่มุ่งเน้นการสนับสนุนกลุ่มเปราะบางและเพิ่มโอกาสทางเศรษฐกิจในตลาดแรงงานที่กำลังเปลี่ยนแปลงอย่างรวดเร็ว

บทที่ 3

วิธีการดำเนินงานวิจัย

งานวิจัยนี้มีวัตถุประสงค์ศึกษาปัจจัยที่ส่งผลต่อการประกอบอาชีพในระบบหรือนอกระบบของแรงงานในประเทศไทย และเพื่อสร้างตัวแบบที่สามารถจำแนกสถานภาพการทำงานระหว่างการเป็นแรงงานนอกระบบหรือในระบบกับปัจจัยอธิบายต่างๆ ซึ่งมีวิธีดำเนินงานวิจัยดังต่อไปนี้

- 3.1 ข้อมูลที่ใช้
- 3.2 การเตรียมข้อมูล
- 3.3 ตัวแปรที่ใช้ในการศึกษา
- 3.4 การสร้างตัวแบบ
- 3.5 การประเมินประสิทธิภาพของแบบจำลอง

3.1 ข้อมูลที่ใช้

ข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นข้อมูลทุติยภูมิ (Secondary Data) จากการสำรวจแรงงานนอกระบบ ปี พ.ศ. 2567 ของสำนักงานสถิติแห่งชาติ ดำเนินการสำรวจในไตรมาสที่ 3 เดือนกรกฎาคม - กันยายน พ.ศ. 2567 โดยผนวกข้อมูลกับโครงการสำรวจภาวะการทำงานของประชากร พ.ศ. 2567 เพื่อให้ทราบข้อมูลเกี่ยวกับจำนวนและลักษณะทางประชากรที่สำคัญของแรงงานนอกระบบ ประกอบด้วยข้อมูลหลัก 8 ตอน ได้แก่

- ตอนที่ 1 ลักษณะทั่วไปของสมาชิกในครัวเรือน
- ตอนที่ 2 การศึกษา
- ตอนที่ 3 การทำงาน
- ตอนที่ 4 ความต้องการจะทำงานเพิ่ม
- ตอนที่ 5 รายได้ของลูกจ้าง
- ตอนที่ 6 แรงงานนอกระบบ
- ตอนที่ 7 สถานที่ทำงาน/ลักษณะงาน
- ตอนที่ 8 ปัญหาจากการทำงานและการพัฒนาทักษะ

แผนแบบการเลือกตัวอย่างที่ใช้ในการสำรวจแรงงานนอกระบบ เป็นแบบมีชั้นภูมิ 2 ชั้น (Stratified Two – Stage Sampling) โดยมีจังหวัดเป็นสตราตัม และในเขตเทศบาล และนอกเขตเทศบาล เป็นสตราตัมย่อยในการเลือกตัวอย่างขั้นที่หนึ่ง (Primary sampling stage) หน่วยตัวอย่างขั้นที่หนึ่ง คือเขตแ่งนั้บ โดยกรอบตัวอย่างสำหรับการเลือกหน่วยตัวอย่างขั้นที่หนึ่ง ได้มาจากโครงการสำมะโนประชากร พ.ศ. 2553 ที่ได้มีการปรับปรุงให้มีความทันสมัย และจำนวนเขตแ่งนั้บตัวอย่าง ทั้งสิ้น 5,430 เขตแ่งนั้บ สำหรับการเลือกตัวอย่างขั้นที่สอง (Secondary sampling stage)

ครัวเรือนส่วนบุคคลและ สมาชิกในครัวเรือนกลุ่มบุคคลประเภทคนงานเป็นหน่วยตัวอย่างขั้นที่สอง โดยกำหนดจำนวน 16 ครัวเรือนตัวอย่าง ต่อเขตเจนนับตัวอย่างทั้งในเขตเทศบาลและนอกเขตเทศบาล ในกรณีของครัวเรือนกลุ่มบุคคลประเภทคนงานจะเป็น การเลือกสมาชิกตัวอย่าง จากครัวเรือนกลุ่มบุคคลประเภทคนงานทุกครัวเรือน ก่อนดำเนินการเลือกครัวเรือนตัวอย่าง ได้มีการ สร้างกรอบครัวเรือนตัวอย่าง โดยทำการนับและจัด ครัวเรือนและอาคาร/สิ่งปลูกสร้างทุกแห่งในแต่ละ เขตเจนนับ ตัวอย่าง หลังจากนั้นจึงมีการจัดเรียงรายชื่อครัวเรือนส่วนบุคคลใหม่ตามขนาดครัวเรือน ซึ่งวัดด้วยจำนวนสมาชิก ในครัวเรือน แล้วจึงดำเนินการเลือกครัวเรือนตัวอย่างด้วยวิธีการสุ่มแบบ มีระบบ ซึ่งมีกลุ่มตัวอย่างที่ครอบคลุมระดับประเทศ โดยเก็บข้อมูลจากครัวเรือนสุ่มประมาณ 86,880 ครัวเรือน รวมทั้งสิ้นประมาณจำนวน 168,560 หน่วยตัวอย่าง

3.2 การเตรียมข้อมูล

3.2.1 การคัดเลือกข้อมูล

ในการวิเคราะห์ข้อมูลครั้งนี้ จะพิจารณาเฉพาะผู้ที่มีอายุตั้งแต่ 15 ปีขึ้นไป ที่อาศัยอยู่ในครัวเรือนส่วนบุคคล และเป็นผู้ที่มีงานทำ โดยพิจารณาความครบถ้วนของข้อมูลในแต่ละตัวแปร ที่จะนำมาวิเคราะห์ ซึ่งหลังจากดำเนินการตรวจสอบและทำความสะอาดข้อมูลแล้ว มีจำนวนข้อมูลที่สามารถนำมาใช้ในการวิเคราะห์ทั้งหมด 85,479 หน่วยตัวอย่าง

3.2.2 การแปลงรูปแบบข้อมูล

1) ข้อมูลเชิงคุณภาพ

สำหรับข้อมูลเชิงกลุ่ม เช่น เพศ สถานภาพสมรส หรือกลุ่มอาชีพ จะถูกแปลงเป็นตัวเลขโดยอัตโนมัติผ่านกระบวนการเข้ารหัสแบบตัวแปรหุ่น (Dummy Coding) เมื่อทำการวิเคราะห์ในโปรแกรม R วิธีนี้จะสร้างตัวแปรใหม่สำหรับแต่ละกลุ่ม และกำหนดค่าเป็น 1 หากข้อมูลอยู่ในกลุ่มนั้น และ 0 หากไม่ใช่

นอกจากนี้ ตัวแปรระดับการศึกษาที่สำเร็จ ซึ่งเป็นข้อมูลที่จัดเก็บในรูปแบบเชิงกลุ่ม จะถูกแปลงให้อยู่ในรูปของจำนวนปีการศึกษาที่สำเร็จ เช่น ระดับการศึกษาที่สำเร็จของบุคคลนั้น คือ ระดับปริญญาตรี จะถูกแปลงให้เป็นจำนวนปีการศึกษาที่สำเร็จ เท่ากับ 16 ปี

2) ข้อมูลเชิงปริมาณ

ข้อมูลเชิงปริมาณ จะมีการปรับมาตราส่วน (Feature Scaling) โดยใช้ฟังก์ชัน scale() ในโปรแกรม R เพื่อปรับค่าแต่ละตัวแปรให้อยู่ในช่วงที่เหมาะสม โดยใช้ Standardization (Z-score normalization) ซึ่งคำนวณโดย

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma}$$

โดยที่ X คือ ค่าของตัวแปร, μ คือ ค่าเฉลี่ย และ σ คือ ค่าความเบี่ยงเบนมาตรฐาน โดยวิธีนี้ทำให้ค่าของตัวแปรมีค่าเฉลี่ยเป็น 0 และค่าเบี่ยงเบนมาตรฐานเป็น 1 ช่วยให้ข้อมูลมีมาตราส่วนเดียวกัน ลดผลกระทบจากค่าที่มีช่วงของข้อมูลที่แตกต่างกัน

3.3 ตัวแปรที่ใช้ในการศึกษา

ตัวแปรที่ใช้ในการศึกษา ประกอบด้วย ตัวแปรตาม ซึ่งมี 2 ค่า คือ 1 สถานภาพการทำงานนอกระบบ และ 0 สถานภาพการทำงานในระบบ สำหรับตัวแปรอิสระในกรณีที่เป็นข้อมูลเชิงกลุ่มได้มีการแปลงให้เป็นตัวแปรหุ่น (Dummy Variable) ที่มีค่า 0 กับ 1 ซึ่งรายละเอียดของตัวแปรที่ใช้ในการศึกษา แสดงดังตารางที่ 3-1

ตารางที่ 3-1 ตัวแปรที่ใช้ในการศึกษา

ลำดับ	ตัวแปร	สัญลักษณ์	คำอธิบาย
1	ภาค	REG_Central	ตัวแปรหุ่น; บุคคลอาศัยอยู่ภาคกลาง (ไม่รวมกรุงเทพมหานคร) =1, ไม่ใช่ =0
		REG_North	ตัวแปรหุ่น; บุคคลอาศัยอยู่ภาคเหนือ =1, ไม่ใช่ =0
		REG_EastNorth	ตัวแปรหุ่น; บุคคลอาศัยอยู่ภาคตะวันออกเฉียงเหนือ =1, ไม่ใช่ =0
		REG_South	ตัวแปรหุ่น; บุคคลอาศัยอยู่ภาคใต้ =1, ไม่ใช่ =0
		หมายเหตุ:	บุคคลอาศัยอยู่กรุงเทพมหานคร กำหนดเป็นฐาน
2	เขตการปกครอง	AREA	ตัวแปรหุ่น; บุคคลอาศัยอยู่นอกเขตเทศบาล =1, บุคคลอาศัยอยู่ในเขตเทศบาล =0
3	จำนวนสมาชิกในครัวเรือน	NHH	จำนวนสมาชิกในครัวเรือนที่บุคคลอาศัยอยู่
4	เพศ	SEX	ตัวแปรหุ่น; เพศชาย=1, เพศหญิง=0
5	อายุ	AGE	อายุของบุคคล

ลำดับ	ตัวแปร	สัญลักษณ์	คำอธิบาย
6	สถานภาพสมรส	MAR_Single	ตัวแปรหุ่น; บุคคลมีสถานภาพโสด =1, ไม่ใช่ =0
		MAR_MarriedReg	ตัวแปรหุ่น; บุคคลมีสถานภาพสมรส (จดทะเบียน) =1, ไม่ใช่ =0
		MAR_MarriedUnreg	ตัวแปรหุ่น; บุคคลมีสถานภาพสมรส (ไม่จดทะเบียน หรืออยู่กินฉันสามีภรรยา) =1, ไม่ใช่ =0
		MAR_Widowed	ตัวแปรหุ่น; บุคคลมีสถานภาพม่าย =1, ไม่ใช่ =0
		MAR_Separated	ตัวแปรหุ่น; บุคคลมีสถานภาพแยกกันอยู่ =1, ไม่ใช่ =0
		หมายเหตุ:	บุคคลที่มีสถานภาพหย่า กำหนดเป็นฐาน
7	จำนวนปีที่ศึกษา	EDUC	จำนวนปีที่ศึกษา (ตามระดับขั้นที่บุคคลจบการศึกษาสูงสุด)
8	อาชีพ	OCCUP_Manager	ตัวแปรหุ่น; บุคคลประกอบอาชีพผู้จัดการ ทหาร ข้าราชการระดับอาวุโส และผู้บัญญัติกฎหมาย =1, ไม่ใช่ =0
		OCCUP_Professional	ตัวแปรหุ่น; บุคคลประกอบอาชีพผู้ประกอบการวิชาชีพด้านต่างๆ =1, ไม่ใช่ =0
		OCCUP_Technician	ตัวแปรหุ่น; บุคคลประกอบอาชีพเจ้าหน้าที่เทคนิคและผู้ประกอบวิชาชีพที่เกี่ยวข้องกับด้านต่างๆ =1, ไม่ใช่ =0
		OCCUP_Clerk	ตัวแปรหุ่น; บุคคลประกอบอาชีพเสมียน =1, ไม่ใช่ =0

ลำดับ	ตัวแปร	สัญลักษณ์	คำอธิบาย
		OCCUP_ServiceSales	ตัวแปรหุ่น; บุคคลประกอบอาชีพพนักงานบริการและผู้จำหน่ายสินค้า = 1, ไม่ใช่ = 0
		OCCUP_Artisan	ตัวแปรหุ่น; บุคคลประกอบอาชีพช่างฝีมือและผู้ปฏิบัติงานที่เกี่ยวข้อง = 1, ไม่ใช่ = 0
		OCCUP_MachineOperator	ตัวแปรหุ่น; บุคคลประกอบอาชีพผู้ปฏิบัติการเครื่องจักรโรงงานและเครื่องจักร และผู้ปฏิบัติงานด้านการประกอบ = 1, ไม่ใช่ = 0
		OCCUP_Elementary	ตัวแปรหุ่น; บุคคลประกอบอาชีพผู้ประกอบอาชีพงานพื้นฐาน = 1, ไม่ใช่ = 0
		หมายเหตุ:	บุคคลที่ประกอบอาชีพผู้ปฏิบัติงานที่มีฝีมือในด้านการเกษตร ป่าไม้ และประมง กำหนดเป็นฐาน
9	จำนวนชั่วโมงการทำงาน	HR	จำนวนชั่วโมงทำงานต่อสัปดาห์
10	ประเภทของค่าจ้างที่ได้รับ	WAGE_Hourly	ตัวแปรหุ่น; บุคคลได้รับค่าจ้างแบบรายชั่วโมง = 1, ไม่ใช่ = 0
		WAGE_Weekly	ตัวแปรหุ่น; บุคคลได้รับค่าจ้างแบบรายสัปดาห์ = 1, ไม่ใช่ = 0
		WAGE_Monthly	ตัวแปรหุ่น; บุคคลได้รับค่าจ้างแบบรายเดือน = 1, ไม่ใช่ = 0
		WAGE_NonMonetary	ตัวแปรหุ่น; บุคคลได้รับค่าตอบแทนในรูปแบบที่ไม่เป็นตัวเงิน = 1, ไม่ใช่ = 0
		WAGE_Other	ตัวแปรหุ่น; บุคคลได้รับค่าจ้างในรูปแบบอื่นๆ = 1, ไม่ใช่ = 0

ลำดับ	ตัวแปร	สัญลักษณ์	คำอธิบาย
		หมายเหตุ:	บุคคลได้รับค่าจ้างในรูปแบบรายวัน กำหนดเป็นฐาน
11	ค่าจ้างต่อเดือน	APPROX	ค่าจ้างที่บุคคลได้รับต่อเดือน

3.4 การสร้างตัวแบบ

เนื่องจากตัวแปรที่สนใจในการศึกษาครั้งนี้ คือ ตัวแปรสถานภาพการทำงานในระบบและนอกระบบ เป็นตัวแปรแบบสองทางเลือก (Binary Variable) โดยกำหนดให้สถานภาพการทำงานนอกระบบ มีค่าเท่ากับ 1 และสถานภาพการทำงานในระบบ มีค่าเท่ากับ 0 และแบ่งข้อมูลออกเป็น 2 ชุด คือ ชุดฝึกสอน (Training Set) และ ชุดทดสอบ (Test Set) โดยใช้ฟังก์ชัน split() จากแพ็คเกจ caTools ซึ่งทำการสุ่มแบ่งข้อมูลตามตัวแปรตาม โดยกำหนดอัตราส่วนการแบ่งเป็น 75% สำหรับชุดฝึกสอน และ 25% สำหรับชุดทดสอบ

สำหรับตัวแปรอิสระที่พิจารณาในการศึกษานี้ ได้แก่ ภาค เขตการปกครอง จำนวนสมาชิกในครัวเรือน เพศ อายุ สถานภาพสมรส จำนวนปีที่ศึกษา สาขาวิชา อาชีพ ขนาดของกิจการ จำนวนชั่วโมงการทำงาน ประเภทของค่าจ้างที่ได้รับ ค่าจ้างต่อเดือน และสร้างตัวแบบการจำแนกการประกอบอาชีพนอกระบบของแรงงานในประเทศไทย ด้วยตัวแบบ Random Forest และ Logistic Regression ซึ่งสามารถแสดงสมการได้ดังนี้

$$\begin{aligned} \text{logit}(P) = & \beta_0 + \beta_1 \text{REG_Central} + \beta_2 \text{REG_North} + \beta_3 \text{REG_EastNorth} + \beta_4 \text{REG_South} \\ & + \beta_5 \text{AREA} + \beta_6 \text{NHH} + \beta_7 \text{SEX} + \beta_8 \text{AGE} + \beta_9 \text{MAR_Single} \\ & + \beta_{10} \text{MAR_MarriedReg} + \beta_{11} \text{MAR_MarriedUnreg} \\ & + \beta_{12} \text{MAR_Widowed} + \beta_{13} \text{MAR_Separate} + \beta_{14} \text{EDUC} \\ & + \beta_{15} \text{OCCUP_Manager} + \beta_{16} \text{OCCUP_Professional} \\ & + \beta_{17} \text{OCCUP_Technician} + \beta_{18} \text{OCCUP_Clerk} \\ & + \beta_{19} \text{OCCUP_ServiceSales} + \beta_{20} \text{OCCUP_Artisan} \\ & + \beta_{21} \text{OCCUP_MachineOperator} + \beta_{22} \text{OCCUP_Elementary} + \beta_{23} \text{HR} \\ & + \beta_{24} \text{WAGE_Hourly} + \beta_{25} \text{WAGE_Weekly} + \beta_{26} \text{WAGE_Monthly} \\ & + \beta_{27} \text{WAGE_NonMonetary} + \beta_{28} \text{WAGE_Other} + \beta_{29} \text{APPROX} \end{aligned}$$

โดยที่ $\text{logit}(P)$ คือ ฟังก์ชัน logit ซึ่งแสดงโอกาสของบุคคลที่จะมีสถานภาพเป็นแรงงานนอกระบบ

β_0 คือ สัมประสิทธิ์ค่าคงที่

$\beta_1, \beta_2, \dots, \beta_{29}$ คือ ค่าสัมประสิทธิ์การถดถอย

REG คือ ภาค

AREA คือ เขตการปกครอง

NHH คือ จำนวนสมาชิกในครัวเรือน

SEX	คือ เพศ
AGE	คือ อายุ
MAR	คือ สถานภาพสมรส
EDUC	คือ จำนวนปีที่ศึกษา
OCCUP	คือ อาชีพ
HR	คือ จำนวนชั่วโมงการทำงานต่อสัปดาห์
WAGE	คือ ประเภทของค่าจ้างที่ได้รับ
APPROX	คือ ค่าจ้างที่ได้รับต่อเดือน

สำหรับสำหรับการสร้างตัวแบบ Logistic Regression ทำการคัดเลือกตัวแปรอิสระเพื่อเข้าสู่ตัวแบบทั้งหมด 5 รูปแบบ ได้แก่

รูปแบบที่ 1 นำตัวแปรเข้าทั้งหมด (Enter Regression)

เป็นวิธีการเอาตัวแปรอิสระทุกตัวทั้งตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามอย่างมีนัยสำคัญทางสถิติและไม่มีนัยสำคัญทางสถิติเข้าไปวิเคราะห์ในสมการถดถอย

สำหรับรูปแบบที่ 2 ถึงรูปแบบที่ 4 จะใช้ฟังก์ชัน stepAIC() จากแพ็คเกจ MASS ซึ่งใช้สำหรับเลือกตัวแบบที่เหมาะสมที่สุดใน Logistic Regression หรือแบบจำลองเชิงเส้นอื่นๆ โดยใช้ Akaike Information Criterion (AIC) เป็นเกณฑ์ในการเลือกตัวแบบที่ดีที่สุด ซึ่งคำนวณจาก

$$AIC = -2 \ln L + 2k \quad (3-1)$$

โดยที่ L คือ Likelihood ของตัวแบบ

k คือ จำนวนพารามิเตอร์ในตัวแบบ

ค่า AIC ที่ต่ำกว่าหมายถึงตัวแบบที่เหมาะสมกว่า โดยพิจารณาจากความสมดุลระหว่างความแม่นยำ (Goodness of fit) และความซับซ้อน (complexity) ของตัวแบบ ซึ่งมีรายละเอียดของการคัดเลือกตัวแปรดังนี้

รูปแบบที่ 2 คัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปร (Forward Selection)

เป็นวิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการทีละตัวตามลำดับความสัมพันธ์กับ Y โดยตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามมากที่สุดจะถูกคัดเลือกเข้าก่อน กระบวนการนี้จะดำเนินไปเรื่อยๆ โดยเพิ่มตัวแปรที่ช่วยลดค่า AIC ลงมากที่สุดในแต่ละขั้นตอน จนกว่าจะไม่มีตัวแปรใดที่สามารถลดค่า AIC ได้อีก

รูปแบบที่ 3 คัดเลือกตัวแปรด้วยวิธีลดตัวแปร (Backward Elimination)

เป็นวิธีการคัดเลือกตัวแปรอิสระออกจากสมการทีละตัวแปร โดยเริ่มจากการสร้างสมการถดถอยที่นำตัวแปรอิสระทุกตัวเข้าสู่สมการ แล้วจึงคัดเลือกตัวแปรอิสระออกทีละตัว โดยตัวแปรอิสระ

ที่มีความสัมพันธ์กับตัวแปรตามน้อยที่สุดจะถูกคัดออกก่อน กระบวนการนี้จะดำเนินไปเรื่อยๆ จนกว่าจะได้ตัวแบบที่มีค่า AIC ต่ำที่สุด ซึ่งเป็นตัวแบบที่สมดุลระหว่างความแม่นยำและความซับซ้อนของตัวแบบ

รูปแบบที่ 4 คัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปรอิสระแบบขั้นตอน (Stepwise Selection)

เป็นการผสมผสานระหว่าง Forward Selection และ Backward Elimination โดยเริ่มต้นจากตัวแบบที่ไม่มีตัวแปรอิสระ (หรือมีเฉพาะค่าคงที่) แล้วเพิ่มตัวแปรอิสระทีละตัว หากการเพิ่มตัวแปรทำให้ค่า AIC ลดลง ตัวแปรนั้นจะถูกเพิ่มเข้าไปในตัวแบบ หลังจากเพิ่มตัวแปรเข้าไปแล้ว จะมีการตรวจสอบว่ามีตัวแปรใดที่ควรถูกลบออกหรือไม่ หากการลบตัวแปรทำให้ค่า AIC ลดลง ตัวแปรนั้นจะถูกนำออก กระบวนการนี้จะดำเนินต่อไปจนกว่าจะไม่มีเพิ่มหรือลบตัวแปรใดๆ ที่ทำให้ค่า AIC ลดลงได้อีก

รูปแบบที่ 5 คัดเลือกตัวแปรโดยพิจารณาจาก Feature Importance ของ Random Forest Model

เป็นการเลือกตัวแปรเข้าพิจารณาจาก Mean Decrease in Gini (MDG) ซึ่งเป็นวิธีการวัดความสำคัญของตัวแปรอิสระใน Random Forest โดยพิจารณาจากความสามารถของตัวแปรในการลดความไม่บริสุทธิ์ของข้อมูล (Gini Impurity) เมื่อใช้เป็นเกณฑ์แบ่งข้อมูล หากตัวแปรที่มีค่า MDG สูง แสดงว่าตัวแปรนั้นมีบทบาทสำคัญในการจำแนกประเภท ดังนั้น จึงเลือกเฉพาะตัวแปรที่มีค่า MDG สูง และตัดตัวแปรที่มีค่า MDG ต่ำ เมื่อเทียบกับตัวแปรอื่นๆ ออก เพื่อสร้างตัวแบบการถดถอยโลจิสติก

3.5 การประเมินประสิทธิภาพของตัวแบบ

ประเมินประสิทธิภาพของตัวแบบโดยพิจารณาจาก Confusion Matrix โดยการพยากรณ์ของตัวแบบ แสดงดังตารางที่ 3.2

ตารางที่ 3-2 Confusion Matrix

		ค่าจริง	
		Positive	Negative
ค่าพยากรณ์	Positive	TP	FN
	Negative	FP	TN

โดยที่

True Positive (TP) เป็นจำนวนที่ถูก กรณีที่ตัวแบบทำนายว่าเหตุการณ์หรือข้อมูลเป็นจริง (positive) และตรงกับความจริงที่ว่าเป็นจริง ในกรณีนี้ตัวแบบทำนายถูกต้อง

True Negative (TN) เป็นจำนวนที่ถูก กรณีที่ตัวแบบทำนายว่าเหตุการณ์หรือข้อมูลไม่เป็นจริง (negative) และตรงกับความจริงที่ว่าเป็นจริง ในกรณีนี้ตัวแบบทำนายถูกต้อง

False Positive (FP) เป็นจำนวนที่ผิด กรณีที่ตัวแบบทำนายว่าเหตุการณ์หรือข้อมูลเป็นจริง (positive) แต่ความจริงแล้วไม่เป็นจริง ในกรณีนี้ตัวแบบทำนายผิด

False Negative (FN) เป็นจำนวนที่ผิด กรณีที่ตัวแบบทำนายว่าเหตุการณ์หรือข้อมูลไม่เป็นจริง (negative) แต่ความจริงแล้วเป็นจริง ในกรณีนี้ตัวแบบทำนายผิด

หลังจากได้ Confusion Matrix ของแต่ละตัวแบบ จะนำมาคำนวณค่าความถูกต้อง (Accuracy) ความแม่นยำ (Precision) ความครอบคลุม (Recall) คะแนน F1 (F1-Score) และพื้นที่ใต้กราฟ (AUC) โดยค่าที่สูงแสดงว่าตัวแบบนั้นมีความแม่นยำในการพยากรณ์ของตัวแบบทำนาย โดยมีสูตรคำนวณดังต่อไปนี้

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3-2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3-3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3-4)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3-5)$$

$$\text{AUC} = \int_0^1 \text{TPR} \, d(\text{FPR}) \quad (3-6)$$

โดยที่

$$\text{TPR} = \frac{TP}{TP+FN} \quad (\text{True Positive Rate หรือ Recall})$$

$$\text{FPR} = \frac{FP}{FP+TN} \quad (\text{False Positive Rate})$$

AUC มีค่าตั้งแต่ 0 ถึง 1 โดย

AUC = 1 หมายถึงตัวแบบจำแนกได้สมบูรณ์แบบ (ไม่มีข้อผิดพลาด)

AUC = 0.5 หมายถึงตัวแบบไม่มีความสามารถในการจำแนก (สุ่มทาย)

AUC < 0.5 หมายถึงตัวแบบทำงานแย่กว่าการเดาสุ่ม

บทที่ 4

ผลการวิเคราะห์

4.1 ข้อมูลทั่วไปของหน่วยตัวอย่างที่ใช้ในการวิเคราะห์

หน่วยตัวอย่างที่ใช้ในการวิเคราะห์ ประกอบด้วยผู้ที่มีสถานภาพแรงงานในระบบ จำนวน 40,236 หน่วยตัวอย่าง และนอกระบบ 45,243 หน่วยตัวอย่าง หรือคิดเป็นร้อยละ 47.10 และ 52.90 ตามลำดับ แสดงดังภาพที่ 4-1



ภาพที่ 4-1 ร้อยละของหน่วยตัวอย่างจำแนกตามสถานภาพแรงงานในระบบและนอกระบบ

ข้อมูลหน่วยตัวอย่างที่ใช้ในการวิเคราะห์แบ่งออกเป็น 2 กลุ่ม ได้แก่ ตัวแปรอิสระที่เป็นข้อมูลเชิงกลุ่มและตัวแปรอิสระที่เป็นค่าต่อเนื่อง สำหรับตัวแปรอิสระที่เป็นข้อมูลเชิงกลุ่ม ได้แก่ ภาค เขต การปกครอง เพศ สถานภาพการสมรส อาชีพ และประเภทค่าจ้างที่ได้รับ แสดงจำนวนและสัดส่วนของแต่ละกลุ่ม ดังตารางที่ 4-1 ถึง 4-6 สำหรับตัวแปรอิสระที่เป็นค่าต่อเนื่อง ได้แก่ จำนวนสมาชิกในครัวเรือน อายุ จำนวนปีที่ศึกษา จำนวนชั่วโมงทำงานต่อสัปดาห์ และค่าจ้างที่ได้รับต่อเดือน นำเสนอรายละเอียดของค่าสถิติที่สำคัญ ได้แก่ ค่าเฉลี่ย ค่าต่ำสุด ค่าสูงสุด และส่วนเบี่ยงเบนมาตรฐาน ตามที่แสดงในตารางที่ 4-7 ถึง 4-11 โดยรายละเอียดมีดังต่อไปนี้

ตารางที่ 4-1 ข้อมูลหน่วยตัวอย่างจำแนกตามภาค

ภาค	สถานภาพแรงงาน	จำนวน	สัดส่วน
กรุงเทพมหานคร	ในระบบ	3,803	4.45%
	นอกระบบ	1,059	1.24%
ภาคกลาง	ในระบบ	16,143	18.89%
	นอกระบบ	10,617	12.42%
ภาคตะวันออกเฉียงเหนือ	ในระบบ	6,320	7.39%
	นอกระบบ	15,179	17.76%

ภาค	สถานภาพแรงงาน	จำนวน	สัดส่วน
ภาคเหนือ	ในระบบ	6,310	7.38%
	นอกระบบ	10,887	12.74%
ภาคใต้	ในระบบ	7,660	8.96%
	นอกระบบ	7,501	8.78%
รวม		85,479	100.00%

ตารางที่ 4-2 ข้อมูลหน่วยตัวอย่างจำแนกตามเขตการปกครอง

เขตการปกครอง	สถานภาพแรงงาน	จำนวน	สัดส่วน
ในเขตเทศบาล	ในระบบ	24,817	29.03%
	นอกระบบ	21,455	25.10%
นอกเขตเทศบาล	ในระบบ	15,419	18.04%
	นอกระบบ	23,788	27.83%
รวม		85,479	100.00%

ตารางที่ 4-3 ข้อมูลหน่วยตัวอย่างจำแนกตามเพศ

เพศ	สถานภาพแรงงาน	จำนวน	สัดส่วน
ชาย	ในระบบ	20,884	24.43%
	นอกระบบ	27,032	31.62%
หญิง	ในระบบ	19,352	22.64%
	นอกระบบ	18,211	21.30%
รวม		85,479	100.00%

ตารางที่ 4-4 ข้อมูลหน่วยตัวอย่างจำแนกตามสถานภาพการสมรส

สถานภาพการสมรส	สถานภาพแรงงาน	จำนวน	สัดส่วน
โสด	ในระบบ	12,619	14.76%
	นอกระบบ	6,362	7.44%
สมรส (จดทะเบียน)	ในระบบ	12,409	14.52%
	นอกระบบ	20,724	24.24%
สมรส (ไม่จดทะเบียน หรือ อยู่กินฉันสามีภรรยา)	ในระบบ	11,063	12.94%
	นอกระบบ	10,218	11.95%

สถานภาพการสมรส	สถานภาพแรงงาน	จำนวน	สัดส่วน
ม่าย	ในระบบ	1,145	1.34%
	นอกระบบ	3,973	4.65%
หย่า	ในระบบ	1,314	1.54%
	นอกระบบ	1,867	2.18%
แยกกันอยู่	ในระบบ	1,686	1.97%
	นอกระบบ	2,099	2.46%
รวม		85,479	100.00%

ตารางที่ 4-5 ข้อมูลหน่วยตัวอย่างจำแนกตามอาชีพ

อาชีพ	สถานภาพแรงงาน	จำนวน	สัดส่วน
ผู้จัดการ ทหาร ข้าราชการระดับอาวุโสและผู้บัญญัติกฎหมาย	ในระบบ	2,163	2.53%
	นอกระบบ	230	0.27%
ผู้ประกอบการวิชาชีพด้านต่างๆ	ในระบบ	5,036	5.89%
	นอกระบบ	332	0.39%
เจ้าหน้าที่เทคนิคและผู้ประกอบการวิชาชีพที่เกี่ยวข้องกับด้านต่างๆ	ในระบบ	3,219	3.77%
	นอกระบบ	305	0.36%
เสมียน	ในระบบ	3,479	4.07%
	นอกระบบ	157	0.18%
พนักงานบริการและผู้จำหน่ายสินค้า	ในระบบ	7,705	9.01%
	นอกระบบ	10,621	12.43%
ช่างฝีมือและผู้ปฏิบัติงานที่เกี่ยวข้อง	ในระบบ	5,072	5.93%
	นอกระบบ	4,187	4.90%
ผู้ปฏิบัติการเครื่องจักรโรงงานและเครื่องจักร และผู้ปฏิบัติงานด้านการประกอบ	ในระบบ	5,938	6.95%
	นอกระบบ	1,763	2.06%
ผู้ประกอบการอาชีพงานพื้นฐาน	ในระบบ	6,034	7.06%
	นอกระบบ	5,281	6.18%
ผู้ปฏิบัติงานที่มีฝีมือในด้านการเกษตรป่าไม้ และประมง	ในระบบ	1,590	1.86%
	นอกระบบ	22,367	26.17%
รวม		85,479	100.00%

ตารางที่ 4-6 ข้อมูลหน่วยตัวอย่างจำแนกตามประเภทค่าจ้างที่ได้รับ

รูปแบบค่าตอบแทน	สถานภาพแรงงาน	จำนวน	สัดส่วน
รายชั่วโมง	ในระบบ	47	0.05%
	นอกระบบ	75	0.09%
รายวัน	ในระบบ	11,609	13.58%
	นอกระบบ	20,435	23.91%
รายสัปดาห์	ในระบบ	204	0.24%
	นอกระบบ	1,727	2.02%
รายเดือน	ในระบบ	26,998	31.58%
	นอกระบบ	5,188	6.07%
ไม่เป็นตัวเงิน	ในระบบ	1	0.00%
	นอกระบบ	111	0.13%
อื่นๆ	ในระบบ	1,377	1.61%
	นอกระบบ	17,707	20.72%
รวม		85,479	100.00%

ตารางที่ 4-7 ข้อมูลจำนวนสมาชิกในครัวเรือนของหน่วยตัวอย่าง

	ค่าเฉลี่ย	ค่าต่ำสุด	ค่าสูงสุด	ส่วนเบี่ยงเบนมาตรฐาน
ในระบบ	3	1	13	2
นอกระบบ	3	1	15	2
Total	3	1	15	2

ตารางที่ 4-8 ข้อมูลอายุของหน่วยตัวอย่าง

สถานภาพแรงงาน	ค่าเฉลี่ย	ค่าต่ำสุด	ค่าสูงสุด	ส่วนเบี่ยงเบนมาตรฐาน
ในระบบ	41	15	84	12
นอกระบบ	52	15	94	13
รวม	47	15	94	14

ตารางที่ 4-9 ข้อมูลจำนวนปีที่ศึกษาของหน่วยตัวอย่าง

สถานภาพแรงงาน	ค่าเฉลี่ย	ค่าต่ำสุด	ค่าสูงสุด	ส่วนเบี่ยงเบนมาตรฐาน
ในระบบ	11	0	21	4
นอกระบบ	8	0	21	4

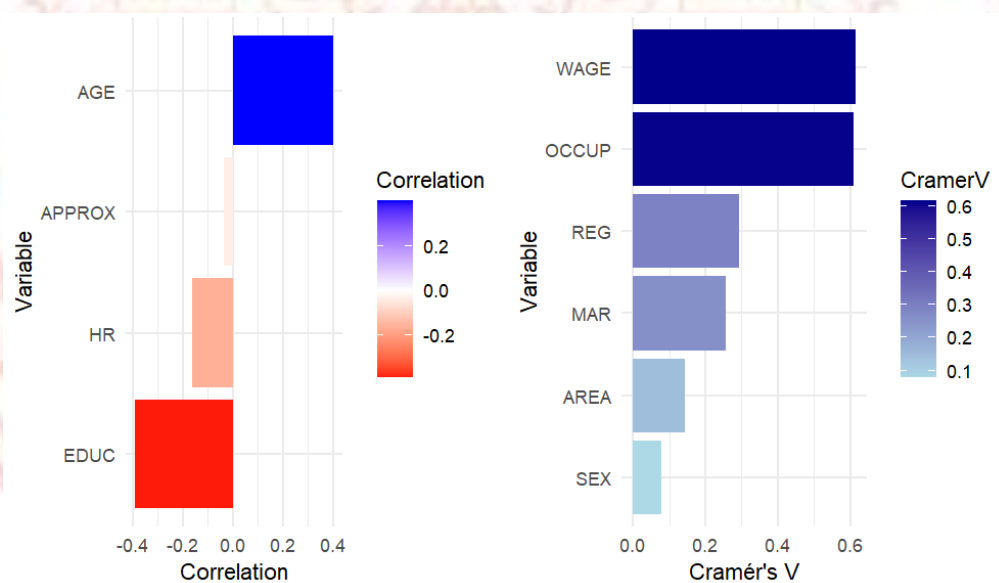
รวม	9	0	21	4
-----	---	---	----	---

ตารางที่ 4-10 ข้อมูลจำนวนชั่วโมงการทำงานต่อสัปดาห์ของหน่วยตัวอย่าง

สถานภาพแรงงาน	ค่าเฉลี่ย	ค่าต่ำสุด	ค่าสูงสุด	ส่วนเบี่ยงเบนมาตรฐาน
ในระบบ	45	0	98	9
นอกระบบ	41	0	98	13
รวม	43	0	98	11

ตารางที่ 4-11 ข้อมูลค่าจ้างที่ได้รับต่อเดือนของหน่วยตัวอย่าง

สถานภาพแรงงาน	ค่าเฉลี่ย	ค่าต่ำสุด	ค่าสูงสุด	ส่วนเบี่ยงเบนมาตรฐาน
ในระบบ	16,069	0	888,888	13,427
นอกระบบ	14,640	0	948,000	23,709
รวม	15,312	0	948,000	19,567



ภาพที่ 4-2 ความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ

จากภาพที่ 4-2 แสดงความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ จากกราฟด้านซ้าย เป็นการพิจารณาค่าความสัมพันธ์แบบไบซีเรียล (Point-Biserial Correlation) ซึ่งใช้สำหรับกรณีที่ตัวแปรตามเป็นตัวแปรสองกลุ่ม (Binary) และตัวแปรอิสระเป็นตัวแปรเชิงปริมาณ พบว่าตัวแปรอายุ (AGE) มีความสัมพันธ์เชิงบวกกับตัวแปรตาม โดยมีค่าสูงที่สุดในกลุ่มตัวแปรเชิงปริมาณ เท่ากับ 0.40 ซึ่งหมายความว่าผู้ที่มีอายุมากขึ้นมีแนวโน้มที่จะอยู่ในกลุ่มแรงงานนอกระบบ ในทางกลับกัน ตัวแปรจำนวนปีที่ศึกษา (EDUC) มีความสัมพันธ์เชิงลบกับตัวแปรตาม โดยมีค่าความสัมพันธ์ เท่ากับ -0.39

ซึ่งบ่งบอกว่าผู้ที่มีจำนวนปีที่ศึกษาสูงขึ้นมีแนวโน้มที่จะไม่อยู่ในกลุ่มแรงงานนอกระบบ สำหรับจำนวน ชั่วโมงการทำงานต่อสัปดาห์ (HR) และค่าจ้างที่ได้รับต่อเดือน (APPROX) มีความสัมพันธ์เชิงลบในระดับต่ำเมื่อเทียบกับตัวแปรอื่นๆ

นอกจากตัวแปรเชิงปริมาณแล้ว ความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรเชิงกลุ่ม โดยใช้ค่าดัชนี Cramér's V ซึ่งเป็นการวัดระดับความสัมพันธ์ระหว่างตัวแปรเชิงกลุ่มกับตัวแปรเชิงกลุ่ม ผลลัพธ์จากกราฟด้านขวาชี้ให้เห็นว่าประเภทค่าจ้างที่ได้รับ (WAGE) และอาชีพ (OCCUP) มีความสัมพันธ์สูงสุด เท่ากับ 0.61 แสดงว่าประเภทค่าจ้างที่ได้รับและอาชีพ มีผลต่อการเป็นแรงงานนอกระบบอย่างมีนัยสำคัญ รองลงมาคือ ตัวแปรภาค (REG) และสถานภาพการสมรส (MAR) มีความสัมพันธ์ระดับปานกลาง เท่ากับ 0.29 และ 0.26 ตามลำดับ บ่งชี้ว่าปัจจัยด้านพื้นที่และสถานภาพการสมรสอาจมีบทบาทต่อการเป็นแรงงานนอกระบบ นอกจากนี้ ตัวแปรเขตการปกครอง (AREA) และเพศ (SEX) แม้จะมีค่าความสัมพันธ์ต่ำกว่า 0.2 แต่ยังคงสะท้อนให้เห็นถึงความแตกต่างบางประการที่อาจมีผลต่อความเป็นไปได้ของการเป็นแรงงานนอกระบบ

อย่างไรก็ตาม ระดับของความสัมพันธ์เพียงอย่างเดียวไม่สามารถชี้ชัดถึงอิทธิพลของตัวแปรเหล่านี้ได้ จำเป็นต้องมีการวิเคราะห์เชิงลึกเพิ่มเติมผ่านแบบตัวแบบในหัวข้อถัดไปเพื่อสามารถระบุปัจจัยที่ส่งผลต่อการประกอบอาชีพนอกระบบของแรงงานในประเทศไทยได้อย่างชัดเจนยิ่งขึ้น

4.2 ตัวแบบการจำแนกการประกอบอาชีพนอกระบบของแรงงานในประเทศไทย

งานวิจัยนี้มุ่งศึกษาเพื่อสร้างตัวแบบการจำแนกการประกอบอาชีพนอกระบบของแรงงานในประเทศไทย โดยทำการหาตัวแบบพยากรณ์ที่ดีที่สุดจาก 2 วิธี ได้แก่ Random Forest และ Logistic Regression สำหรับ Logistic Regression ทำการคัดเลือกตัวแปรอิสระเพื่อเข้าสู่ตัวแบบทั้งหมด 5 รูปแบบ ได้แก่ การนำตัวแปรเข้าทั้งหมด (Enter Regression) คัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปร (Forward Selection) คัดเลือกตัวแปรด้วยวิธีลดตัวแปร (Backward Elimination) คัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปรอิสระแบบขั้นตอน (Stepwise Selection) และคัดเลือกตัวแปรโดยพิจารณาจาก Feature Importance ของ Random Forest Model จากนั้นเปรียบเทียบทั้ง 6 ตัวแบบด้วย Confusion Matrix ซึ่งประกอบด้วย ค่าความถูกต้อง (Accuracy) ความแม่นยำ (Precision) ความครอบคลุม (Recall) คะแนน F1 (F1-Score) และพื้นที่ใต้กราฟ (AUC) โดยผลการวิเคราะห์เป็นดังนี้

4.2.1 Random Forest

จากการสร้างตัวแบบ Random Forest ผลการพยากรณ์สถานภาพแรงงานแสดงดังตารางที่ 4.12 และทำการวัดความสำคัญของตัวแปรอิสระจากตัวแบบ Random Forest โดยพิจารณาจากความสามารถของตัวแปรในการลดความไม่บริสุทธิ์ของข้อมูล (Gini Impurity) เมื่อใช้เป็นเกณฑ์แบ่งข้อมูล หากตัวแปร มีค่า Mean Decrease in Gini (MDG) สูง แสดงว่าตัวแปรนั้นมี

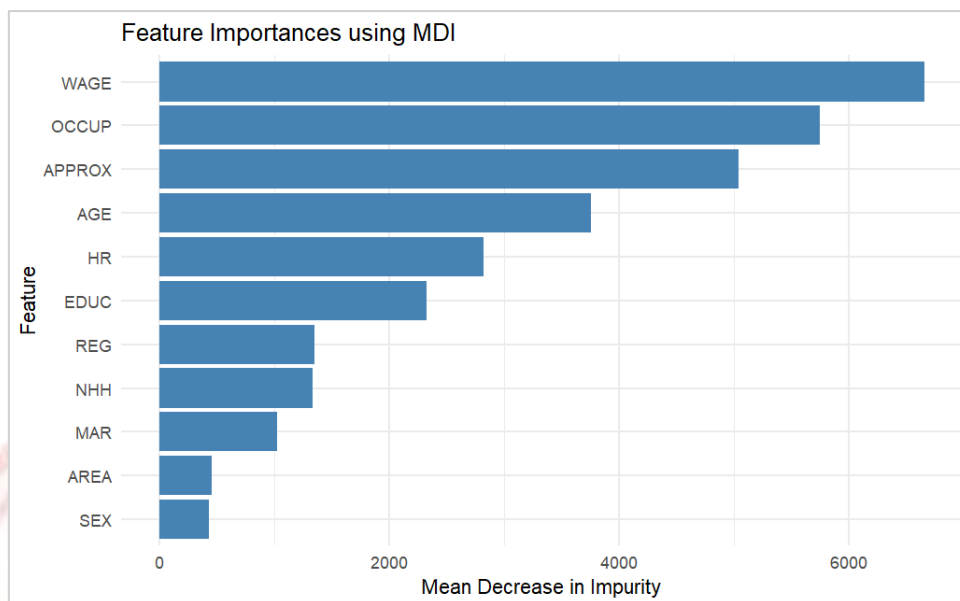
บทบาทสำคัญในการจำแนกประเภท ผลแสดงดังตารางที่ 4-13 และอันดับความสำคัญของตัวแปรอิสระในแบบการจำแนก Random Forest แสดงดังภาพที่ 4-2

ตารางที่ 4-12 Confusion Matrix ของตัวแบบ Random Forest

		ค่าจริง	
		ในระบบ	นอกระบบ
ค่าพยากรณ์	ในระบบ	8,253	1,190
	นอกระบบ	1,536	10,121

ตารางที่ 4-13 Mean Decrease in Gini ของตัวแปรอิสระในแบบการจำแนก Random Forest

ตัวแปรอิสระ	Mean Decrease Gini
REG	1348.333
AREA	453.924
NHH	1333.938
SEX	432.934
AGE	3754.219
MAR	1027.254
HR	2820.116
WAGE	6651.714
APPROX	5036.495
EDUC	2321.366
OCCUP	5744.161



ภาพที่ 4-3 อันดับความสำคัญของตัวแปรอิสระในตัวแบบการจำแนก Random Forest

4.2.2 Logistic Regression

รูปแบบที่ 1 นำตัวแปรเข้าทั้งหมด (Enter Regression)

ตารางที่ 4-14 Logistic Regression: Enter Regression

ตัวแปรอิสระ	Coefficient	Standard Error	P-value	
Intercept	1.974	0.089	$< 2 \times 10^{-16}$	***
REG_Central	0.297	0.056	0.000	***
REG_EastNorth	1.176	0.059	$< 2 \times 10^{-16}$	***
REG_North	1.040	0.059	$< 2 \times 10^{-16}$	***
REG_South	0.221	0.060	0.000	***
AREA_Non	0.157	0.025	0.000	***
NHH	0.060	0.012	0.000	***
SEX_Male	0.068	0.025	0.007	**
AGE	0.575	0.015	$< 2 \times 10^{-16}$	***
MAR_MarriedReg	-0.115	0.061	0.059	.
MAR_MarriedUnreg	-0.116	0.062	0.060	.
MAR_Separated	-0.041	0.076	0.586	
MAR_Single	-0.056	0.063	0.379	
MAR_Widowed	0.123	0.077	0.111	

ระดับนัยสำคัญทางสถิติ: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

ตัวแปรอิสระ	Coefficient	Standard Error	P-value	
HR	-0.222	0.013	$< 2 \times 10^{-16}$	***
WAGE_Hourly	-0.018	0.245	0.942	
WAGE_Monthly	-1.969	0.029	$< 2 \times 10^{-16}$	***
WAGE_NonMonetary	10.567	56.179	0.851	
WAGE_Other	0.502	0.045	$< 2 \times 10^{-16}$	***
WAGE_Weekly	0.317	0.098	0.001	**
APPROX	0.703	0.024	$< 2 \times 10^{-16}$	***
EDUC	0.019	0.016	0.233	
OCCUP_Artisan	-2.177	0.048	$< 2 \times 10^{-16}$	***
OCCUP_Clerk	-3.831	0.109	$< 2 \times 10^{-16}$	***
OCCUP_Elementary	-2.067	0.046	$< 2 \times 10^{-16}$	***
OCCUP_MachineOperat or	-2.640	0.052	$< 2 \times 10^{-16}$	***
OCCUP_Manager	-3.952	0.099	$< 2 \times 10^{-16}$	***
OCCUP_Professional	-4.207	0.087	$< 2 \times 10^{-16}$	***
OCCUP_ServiceSales	-1.474	0.047	$< 2 \times 10^{-16}$	***
OCCUP_Technician	-3.576	0.089	$< 2 \times 10^{-16}$	***
AIC			47,048	

จากตารางที่ 4-14 Logistic Regression รูปแบบที่ 1 เป็นการนำตัวแปรเข้าทั้งหมด (Enter Regression) ประกอบด้วยตัวอิสระที่มีอย่างมีนัยสำคัญทางสถิติและไม่มีนัยสำคัญทางสถิติ พบว่า ภาค เขตการปกครอง จำนวนสมาชิกในครัวเรือน เพศ อายุ สถานภาพการสมรส ประเภทของค่าจ้างที่ได้รับ จำนวนชั่วโมงการทำงานต่อสัปดาห์ ค่าจ้างที่ได้รับต่อเดือน และอาชีพ มีความสัมพันธ์กับสถานภาพแรงงานอย่างมีนัยสำคัญทางสถิติ โดยตัวแบบให้ค่า AIC เท่ากับ 47,048 ซึ่งสามารถเขียนตัวแบบได้ดังนี้

$$\begin{aligned} \text{logit}(P) = & 1.974 + 0.297\text{REG_Central} + 1.040\text{REG_North} + 1.176\text{REG_EastNorth} \\ & + 0.221\text{REG_South} + 0.157\text{AREA} + 0.060\text{NHH} + 0.068\text{SEX} \\ & + 0.575\text{AGE} - 0.056\text{MAR_Single} - 0.115\text{MAR_MarriedReg} \\ & - 0.116\text{MAR_MarriedUnreg} + 0.123\text{MAR_Widowed} \\ & - 0.041\text{MAR_Separate} + 0.019\text{EDUC} - 3.952\text{OCCUP_Manager} \\ & - 4.207\text{OCCUP_Professional} - 3.576\text{OCCUP_Technician} \\ & - 3.831\text{OCCUP_Clerk} - 1.474\text{OCCUP_ServiceSales} \\ & - 2.177\text{OCCUP_Artisan} - 2.640\text{OCCUP_MachineOperator} \\ & - 2.640\text{OCCUP_Elementary} - 0.222\text{HR} - 0.018\text{WAGE_Hourly} \\ & + 0.317\text{WAGE_Weekly} - 1.969\text{WAGE_Monthly} \\ & + 10.567\text{WAGE_NonMonetary} + 0.502\text{WAGE_Other} + 0.703\text{APPROX} \end{aligned}$$

โดยผลการพยากรณ์สถานภาพแรงงานด้วยตัวแบบ Logistic Regression วิธื่อนำตัวแปรเข้าทั้งหมด (Enter Regression) แสดงดังตารางที่ 4-15

ตารางที่ 4-15 Confusion Matrix ของตัวแบบ Logistic Regression: Enter Regression

		ค่าจริง	
		ในระบบ	นอกระบบ
ค่าพยากรณ์	ในระบบ	8,158	1,576
	นอกระบบ	1,901	9,735

รูปแบบที่ 2 คัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปร (Forward Selection)

ตารางที่ 4-16 Logistic Regression: Forward Selection

ตัวแปรอิสระ	Coefficient	Standard Error	P-value	
Intercept	1.966	0.089	$< 2 \times 10^{-16}$	***
OCCUP_Artisan	-2.173	0.047	$< 2 \times 10^{-16}$	***
OCCUP_Clerk	-3.812	0.108	$< 2 \times 10^{-16}$	***
OCCUP_Elementary	-2.068	0.046	$< 2 \times 10^{-16}$	***
OCCUP_MachineOperator	-2.635	0.052	$< 2 \times 10^{-16}$	***
OCCUP_Manager	-3.937	0.098	$< 2 \times 10^{-16}$	***
OCCUP_Professional	-4.185	0.085	$< 2 \times 10^{-16}$	***
OCCUP_ServiceSales	-1.466	0.046	$< 2 \times 10^{-16}$	***
OCCUP_Technician	-3.558	0.088	$< 2 \times 10^{-16}$	***
WAGE_Hourly	-0.023	0.245	0.925	

ระดับนัยสำคัญทางสถิติ: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

ตัวแปรอิสระ	Coefficient	Standard Error	P-value	
WAGE_Monthly	-1.963	0.029	$< 2 \times 10^{-16}$	***
WAGE_NonMonetary	10.566	56.184	0.851	
WAGE_Other	0.504	0.045	$< 2 \times 10^{-16}$	***
WAGE_Weekly	0.318	0.097	0.001	**
AGE	0.569	0.015	$< 2 \times 10^{-16}$	***
REG_Central	0.299	0.056	0.000	***
REG_EastNorth	1.180	0.059	$< 2 \times 10^{-16}$	***
REG_North	1.043	0.059	$< 2 \times 10^{-16}$	***
REG_South	0.223	0.060	0.000	***
APPROX	0.710	0.024	$< 2 \times 10^{-16}$	***
HR	-0.223	0.013	$< 2 \times 10^{-16}$	***
AREA_Non	0.156	0.024	0.000	***
NHH	0.060	0.012	0.000	***
MAR_MarriedReg	-0.115	0.061	0.058	.
MAR_MarriedUnreg	-0.119	0.062	0.053	.
MAR_Separated	-0.042	0.076	0.575	
MAR_Single	-0.055	0.063	0.383	
MAR_Widowed	0.119	0.077	0.123	
SEX_Male	0.067	0.025	0.008	**
AIC		47,047		

จากตารางที่ 4-16 Logistic Regression รูปแบบที่ 2 เป็นการคัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปร (Forward Selection) พบว่า ภาค เขตการปกครอง จำนวนสมาชิกในครัวเรือน เพศ อายุ สถานภาพการสมรส ประเภทของค่าจ้างที่ได้รับ จำนวนชั่วโมงการทำงานต่อสัปดาห์ ค่าจ้างที่ได้รับ ต่อเดือน และอาชีพ มีความสัมพันธ์กับสถานภาพแรงงานอย่างมีนัยสำคัญทางสถิติ โดยตัวแบบให้ค่า AIC เท่ากับ 47,047 ซึ่งสามารถเขียนตัวแบบได้ดังนี้

$$\begin{aligned}
 \text{logit}(P) = & 1.966 + 0.299\text{REG_Central} + 1.043\text{REG_North} + 1.180\text{REG_EastNorth} \\
 & + 0.223\text{REG_South} + 0.156\text{AREA} + 0.060\text{NHH} + 0.067\text{SEX} \\
 & + 0.569\text{AGE} - 0.055\text{MAR_Single} - 0.115\text{MAR_MarriedReg} \\
 & - 0.119\text{MAR_MarriedUnreg} + 0.119\text{MAR_Widowed} \\
 & - 0.042\text{MAR_Separate} - 3.937\text{OCCUP_Manager} \\
 & - 4.185\text{OCCUP_Professional} - 3.558\text{OCCUP_Technician} \\
 & - 3.812\text{OCCUP_Clerk} - 1.466\text{OCCUP_ServiceSales} \\
 & - 2.173\text{OCCUP_Artisan} - 2.635\text{OCCUP_MachineOperator} \\
 & - 2.068\text{OCCUP_Elementary} - 0.223\text{HR} - 0.023\text{WAGE_Hourly} \\
 & + 0.318\text{WAGE_Weekly} - 1.963\text{WAGE_Monthly} \\
 & + 10.566\text{WAGE_NonMonetary} + 0.504\text{WAGE_Other} + 0.710\text{APPROX}
 \end{aligned}$$

ผลการพยากรณ์สถานภาพแรงงานด้วยตัวแบบ Logistic Regression วิธีเพิ่มตัวแปร (Forward Selection) แสดงดังตารางที่ 4-17

ตารางที่ 4-17 Confusion Matrix ของตัวแบบ Logistic Regression: Forward Selection

		ค่าจริง	
		ในระบบ	นอกระบบ
ค่าพยากรณ์	ในระบบ	8,159	1,579
	นอกระบบ	1,900	9,732

รูปแบบที่ 3 คัดเลือกตัวแปรด้วยวิธีลดตัวแปร (Backward Elimination)

ตารางที่ 4-18 Logistic Regression: Backward Elimination

ตัวแปรอิสระ	Coefficient	Standard Error	P-value	
Intercept	1.966	0.089	$< 2 \times 10^{-16}$	***
REG_Central	0.299	0.056	0.000	***
REG_EastNorth	1.180	0.059	$< 2 \times 10^{-16}$	***
REG_North	1.043	0.059	$< 2 \times 10^{-16}$	***
REG_South	0.223	0.060	0.000	***
AREA_Non	0.156	0.024	0.000	***
NHH	0.060	0.012	0.000	***
SEX_Male	0.067	0.025	0.008	**
AGE	0.569	0.015	$< 2 \times 10^{-16}$	***
MAR_MarriedReg	-0.115	0.061	0.058	.
MAR_MarriedUnreg	-0.119	0.062	0.053	.
MAR_Separated	-0.042	0.076	0.575	

ระดับนัยสำคัญทางสถิติ: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

ตัวแปรอิสระ	Coefficient	Standard Error	P-value	
MAR_Single	-0.055	0.063	0.383	
MAR_Widowed	0.119	0.077	0.123	
HR	-0.223	0.013	$< 2 \times 10^{-16}$	***
WAGE_Hourly	-0.023	0.245	0.925	
WAGE_Monthly	-1.963	0.029	$< 2 \times 10^{-16}$	***
WAGE_NonMonetary	10.566	56.184	0.851	
WAGE_Other	0.504	0.045	$< 2 \times 10^{-16}$	***
WAGE_Weekly	0.318	0.097	0.001	**
APPROX	0.710	0.024	$< 2 \times 10^{-16}$	***
OCCUP_Artisan	-2.173	0.047	$< 2 \times 10^{-16}$	***
OCCUP_Clerk	-3.812	0.108	$< 2 \times 10^{-16}$	***
OCCUP_Elementary	-2.068	0.046	$< 2 \times 10^{-16}$	***
OCCUP_MachineOperator	-2.635	0.052	$< 2 \times 10^{-16}$	***
OCCUP_Manager	-3.937	0.098	$< 2 \times 10^{-16}$	***
OCCUP_Professional	-4.185	0.085	$< 2 \times 10^{-16}$	***
OCCUP_ServiceSales	-1.466	0.046	$< 2 \times 10^{-16}$	***
OCCUP_Technician	-3.558	0.088	$< 2 \times 10^{-16}$	***
AIC			47,047	

จากตารางที่ 4-18 Logistic Regression รูปแบบที่ 3 เป็นการคัดเลือกตัวแปรด้วยวิธีลดตัวแปร (Backward Elimination) พบว่า ภาค เขตการปกครอง จำนวนสมาชิกในครัวเรือน เพศ อายุ สถานภาพการสมรส ประเภทของค่าจ้างที่ได้รับ จำนวนชั่วโมงการทำงานต่อสัปดาห์ ค่าจ้างที่ได้รับต่อเดือน และอาชีพ มีความสัมพันธ์กับสถานภาพแรงงานอย่างมีนัยสำคัญทางสถิติ โดยตัวแบบให้ค่า AIC เท่ากับ 47,047 ซึ่งสามารถเขียนตัวแบบได้ดังนี้

$$\begin{aligned}
 \text{logit}(P) = & 1.966 + 0.299\text{REG_Central} + 1.043\text{REG_North} + 1.180\text{REG_EastNorth} \\
 & + 0.223\text{REG_South} + 0.156\text{AREA} + 0.060\text{NHH} + 0.067\text{SEX} \\
 & + 0.569\text{AGE} - 0.055\text{MAR_Single} - 0.115\text{MAR_MarriedReg} \\
 & - 0.119\text{MAR_MarriedUnreg} + 0.119\text{MAR_Widowed} \\
 & - 0.042\text{MAR_Separate} - 3.937\text{OCCUP_Manager} \\
 & - 4.185\text{OCCUP_Professional} - 3.558\text{OCCUP_Technician} \\
 & - 3.812\text{OCCUP_Clerk} - 1.466\text{OCCUP_ServiceSales} \\
 & - 2.173\text{OCCUP_Artisan} - 2.635\text{OCCUP_MachineOperator} \\
 & - 2.068\text{OCCUP_Elementary} - 0.223\text{HR} - 0.023\text{WAGE_Hourly} \\
 & + 0.318\text{WAGE_Weekly} - 1.963\text{WAGE_Monthly} \\
 & + 10.566\text{WAGE_NonMonetary} + 0.504\text{WAGE_Other} + 0.710\text{APPROX}
 \end{aligned}$$

ผลการพยากรณ์สถานภาพแรงงานด้วยตัวแบบ Logistic Regression วิธีลดตัวแปร (Backward Elimination) แสดงดังตารางที่ 4-19

ตารางที่ 4-19 Confusion Matrix ของตัวแบบ Logistic Regression: Backward Elimination

		ค่าจริง	
		ในระบบ	นอกระบบ
ค่าพยากรณ์	ในระบบ	8,159	1,579
	นอกระบบ	1,900	9,732

รูปแบบที่ 4 คัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปรอิสระแบบขั้นตอน (Stepwise Selection)

ตารางที่ 4-20 Logistic Regression: Stepwise Selection

ตัวแปรอิสระ	Coefficient	Standard Error	P-value	
Intercept	1.966	0.089	< 2x10 ⁻¹⁶	***
REG_Central	0.299	0.056	0.000	***
REG_EastNorth	1.180	0.059	< 2x10 ⁻¹⁶	***
REG_North	1.043	0.059	< 2x10 ⁻¹⁶	***
REG_South	0.223	0.060	0.000	***
AREA_Non	0.156	0.024	0.000	***
NHH	0.060	0.012	0.000	***
SEX_Male	0.067	0.025	0.008	**
AGE	0.569	0.015	< 2x10 ⁻¹⁶	***
MAR_MarriedReg	-0.115	0.061	0.058	.
MAR_MarriedUnreg	-0.119	0.062	0.053	.

ตัวแปรอิสระ	Coefficient	Standard Error	P-value	
MAR_Separated	-0.042	0.076	0.575	
MAR_Single	-0.055	0.063	0.383	
MAR_Widowed	0.119	0.077	0.123	
HR	-0.223	0.013	$< 2 \times 10^{-16}$	***
WAGE_Hourly	-0.023	0.245	0.925	
WAGE_Monthly	-1.963	0.029	$< 2 \times 10^{-16}$	***
WAGE_NonMonetary	10.566	56.184	0.851	
WAGE_Other	0.504	0.045	$< 2 \times 10^{-16}$	***
WAGE_Weekly	0.318	0.097	0.001	**
APPROX	0.710	0.024	$< 2 \times 10^{-16}$	***
OCCUP_Artisan	-2.173	0.047	$< 2 \times 10^{-16}$	***
OCCUP_Clerk	-3.812	0.108	$< 2 \times 10^{-16}$	***
OCCUP_Elementary	-2.068	0.046	$< 2 \times 10^{-16}$	***
OCCUP_MachineOperator	-2.635	0.052	$< 2 \times 10^{-16}$	***
OCCUP_Manager	-3.937	0.098	$< 2 \times 10^{-16}$	***
OCCUP_Professional	-4.185	0.085	$< 2 \times 10^{-16}$	***
OCCUP_ServiceSales	-1.466	0.046	$< 2 \times 10^{-16}$	***
OCCUP_Technician	-3.558	0.088	$< 2 \times 10^{-16}$	***
AIC			47,047	

จากตารางที่ 4-20 Logistic Regression รูปแบบที่ 4 เป็นการคัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปรอิสระแบบขั้นตอน (Stepwise Selection) พบว่า ภาค เขตการปกครอง จำนวนสมาชิกในครัวเรือน เพศ อายุ สถานภาพการสมรส ประเภทของค่าจ้างที่ได้รับ จำนวนชั่วโมงการทำงานต่อสัปดาห์ ค่าจ้างที่ได้รับต่อเดือน และอาชีพ มีความสัมพันธ์กับสถานภาพแรงงานอย่างมีนัยสำคัญทางสถิติ โดยตัวแบบให้ค่า AIC เท่ากับ 47,047 ซึ่งสามารถเขียนตัวแบบได้ดังนี้

$$\begin{aligned}
 \text{logit}(P) = & 1.966 + 0.299\text{REG_Central} + 1.043\text{REG_North} + 1.180\text{REG_EastNorth} \\
 & + 0.223\text{REG_South} + 0.156\text{AREA} + 0.060\text{NHH} + 0.067\text{SEX} \\
 & + 0.569\text{AGE} - 0.055\text{MAR_Single} - 0.115\text{MAR_MarriedReg} \\
 & - 0.119\text{MAR_MarriedUnreg} + 0.119\text{MAR_Widowed} \\
 & - 0.042\text{MAR_Separate} - 3.937\text{OCCUP_Manager} \\
 & - 4.185\text{OCCUP_Professional} - 3.558\text{OCCUP_Technician} \\
 & - 3.812\text{OCCUP_Clerk} - 1.466\text{OCCUP_ServiceSales} \\
 & - 2.173\text{OCCUP_Artisan} - 2.635\text{OCCUP_MachineOperator} \\
 & - 2.068\text{OCCUP_Elementary} - 0.223\text{HR} - 0.023\text{WAGE_Hourly} \\
 & + 0.318\text{WAGE_Weekly} - 1.963\text{WAGE_Monthly} \\
 & + 10.566\text{WAGE_NonMonetary} + 0.504\text{WAGE_Other} + 0.710\text{APPROX}
 \end{aligned}$$

ผลการพยากรณ์สถานภาพแรงงานด้วยตัวแบบ Logistic Regression วิธีเพิ่มตัวแปรอิสระแบบขั้นตอน (Stepwise Selection) แสดงดังตารางที่ 4-21

ตารางที่ 4-21 Confusion Matrix ของตัวแบบ Logistic Regression: Stepwise Selection

		ค่าจริง	
		ในระบบ	นอกระบบ
ค่าพยากรณ์	ในระบบ	8,159	1,579
	นอกระบบ	1,900	9,732

รูปแบบที่ 5 คัดเลือกตัวแปรจาก Feature Importance ของตัวแบบ Random Forest

ตารางที่ 4-22 Logistic Regression: Feature Importance Selection

ตัวแปรอิสระ	Coefficient	Standard Error	P-value	
Intercept	2.044	0.087	< 2x10 ⁻¹⁶	***
REG_Central	0.364	0.055	0.000	***
REG_EastNorth	1.234	0.058	< 2x10 ⁻¹⁶	***
REG_North	1.101	0.058	< 2x10 ⁻¹⁶	***
REG_South	0.284	0.059	0.000	***
NHH	0.065	0.012	0.000	***
AGE	0.566	0.015	< 2x10 ⁻¹⁶	***
MAR_MarriedReg	-0.105	0.061	0.083	.
MAR_MarriedUnreg	-0.114	0.062	0.063	.
MAR_Separated	-0.039	0.076	0.605	
MAR_Single	-0.047	0.063	0.456	

ตัวแปรอิสระ	Coefficient	Standard Error	P-value	
MAR_Widowed	0.108	0.077	0.159	
HR	-0.224	0.013	$< 2 \times 10^{-16}$	***
WAGE_Hourly	-0.029	0.245	0.904	
WAGE_Monthly	-1.961	0.029	$< 2 \times 10^{-16}$	***
WAGE_NonMonetary	10.595	56.062	0.850	
WAGE_Other	0.507	0.045	$< 2 \times 10^{-16}$	***
WAGE_Weekly	0.307	0.097	0.002	**
APPROX	0.707	0.024	$< 2 \times 10^{-16}$	***
OCCUP_Artisan	-2.202	0.047	$< 2 \times 10^{-16}$	***
OCCUP_Clerk	-3.886	0.107	$< 2 \times 10^{-16}$	***
OCCUP_Elementary	-2.098	0.046	$< 2 \times 10^{-16}$	***
OCCUP_MachineOperator	-2.661	0.052	$< 2 \times 10^{-16}$	***
OCCUP_Manager	-3.975	0.098	$< 2 \times 10^{-16}$	***
OCCUP_Professional	-4.250	0.084	$< 2 \times 10^{-16}$	***
OCCUP_ServiceSales	-1.529	0.045	$< 2 \times 10^{-16}$	***
OCCUP_Technician	-3.614	0.087	$< 2 \times 10^{-16}$	***
AIC			47090	

จากตารางที่ 4-22 Logistic Regression รูปแบบที่ 5 เป็นการคัดเลือกตัวแปร โดยพิจารณาจาก Feature Importance ของตัวแบบ Random Forest ซึ่งจากภาพที่ 4-3 อันดับความสำคัญของตัวแปรอิสระในตัวแบบการจำแนก Random Forest ตัวแปรเขตการปกครอง (AREA) และตัวแปรเพศ (SEX) มีความสำคัญในระดับต่ำเมื่อเทียบกับตัวแปรอื่นๆ จึงพิจารณานำตัวแปรออกจากตัวแบบ พบว่า ภาค จำนวนสมาชิกในครัวเรือน อายุ สถานภาพการสมรส ประเภทของค่าจ้างที่ได้รับ จำนวนชั่วโมงการทำงานต่อสัปดาห์ ค่าจ้างที่ได้รับต่อเดือน และอาชีพ มีความสัมพันธ์กับสถานภาพแรงงานอย่างมีนัยสำคัญทางสถิติ โดยตัวแบบให้ค่า AIC เท่ากับ 47,090 ซึ่งสามารถเขียนตัวแบบได้ดังนี้

$$\begin{aligned} \text{logit}(P) = & 2.044 + 0.364\text{REG_Central} + 1.101\text{REG_North} + 1.234\text{REG_EastNorth} \\ & + 0.284\text{REG_South} + 0.065\text{NHH} + 0.566\text{AGE} - 0.047\text{MAR_Single} \\ & - 0.105\text{MAR_MarriedReg} - 0.114\text{MAR_MarriedUnreg} \\ & + 0.108\text{MAR_Widowed} - 0.039\text{MAR_Separate} \\ & - 3.975\text{OCCUP_Manager} - 4.250\text{OCCUP_Professional} \\ & - 3.614\text{OCCUP_Technician} - 3.886\text{OCCUP_Clerk} \\ & - 1.529\text{OCCUP_ServiceSales} - 2.202\text{OCCUP_Artisan} \\ & - 2.661\text{OCCUP_MachineOperator} - 2.098\text{OCCUP_Elementary} \\ & - 0.224\text{HR} - 0.029\text{WAGE_Hourly} + 0.307\text{WAGE_Weekly} \\ & - 1.961\text{WAGE_Monthly} + 10.595\text{WAGE_NonMonetary} \\ & + 0.507\text{WAGE_Other} + 0.707\text{APPROX} \end{aligned}$$

ผลการพยากรณ์สถานภาพแรงงานด้วยตัวแบบ Logistic Regression วิธีคัดเลือกตัวแปรโดยพิจารณาจาก Feature Importance ของตัวแบบ Random Forest แสดงดังตารางที่ 4-23

ตารางที่ 4-23 Confusion Matrix ของตัวแบบ Logistic Regression: Feature Importance Selection

		ค่าจริง	
		ในระบบ	นอกระบบ
ค่าพยากรณ์	ในระบบ	8,170	1,587
	นอกระบบ	1,889	9,724

4.3 เปรียบเทียบประสิทธิภาพของตัวแบบ

จากการนำตัวแบบที่ได้ทั้ง 6 ตัวแบบข้างต้น พยากรณ์ข้อมูลชุดทดสอบ 25% หรือจำนวน 21,370 หน่วยตัวอย่าง จะได้ Confusion Matrix ของแต่ละตัวแบบ และคำนวณค่าความถูกต้อง (Accuracy) ความแม่นยำ (Precision) ความครอบคลุม (Recall) คะแนน F1 (F1-Score) และพื้นที่ใต้กราฟ (AUC) โดยทุกค่าจะถูกทำให้อยู่ในรูปของเปอร์เซ็นต์ (100%) ดังตารางที่ 4-24

ตารางที่ 4-24 เปรียบเทียบประสิทธิภาพของตัวแบบ

ตัวแบบ	Accuracy	Precision	Recall	F1-Score	AUC
Random Forest	87.51	84.94	88.10	86.49	87.37
Logistic: Enter	83.73	81.10	83.81	82.43	83.58
Logistic: Forward	83.72	81.11	83.79	82.43	83.58
Logistic: Backward	83.72	81.11	83.79	82.43	83.58
Logistic: Stepwise	83.72	81.11	83.79	82.43	83.58
Logistic: From RF	83.73	81.22	83.73	82.46	83.60

บทที่ 5

สรุปผลและอภิปราย

วัตถุประสงค์ของการศึกษาค้างนี้ เพื่อสร้างตัวแบบที่สามารถจำแนกสถานภาพการทำงาน และเพื่อระบุปัจจัยที่ส่งผลต่อการประกอบอาชีพในระบบหรือนอกระบบของแรงงานในประเทศไทย โดยใช้ข้อมูลการสำรวจแรงงานนอกระบบ ปี พ.ศ. 2567 ของสำนักงานสถิติแห่งชาติ และแบ่งข้อมูล ออกเป็น 2 ชุด คือ ชุดฝึกสอน 75% และ 25% เป็นชุดทดสอบ สำหรับประเมินประสิทธิภาพ ของตัวแบบ 2 วิธี คือ Random Forest และ Logistic Regression สำหรับ Logistic Regression ทำการคัดเลือกตัวแปรอิสระเพื่อเข้าสู่ตัวแบบทั้งหมด 5 รูปแบบ ได้แก่ การนำตัวแปรเข้าทั้งหมด (Enter Regression) คัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปร (Forward Selection) คัดเลือกตัวแปรด้วย วิธีลดตัวแปร (Backward Elimination) คัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปรอิสระแบบขั้นตอน (Stepwise Selection) และคัดเลือกตัวแปรโดยพิจารณาจาก Feature Importance ของตัวแบบ Random Forest รวมทั้งหมด 6 ตัวแบบ จากนั้นเปรียบเทียบประสิทธิภาพของตัวแบบด้วยการสร้าง Confusion Matrix ของแต่ละตัวแบบ และคำนวณค่าความถูกต้อง (Accuracy) ความแม่นยำ (Precision) ความครอบคลุม (Recall) คะแนน F1 (F1-Score) และพื้นที่ใต้กราฟ (AUC)

5.1 สรุปผลการวิเคราะห์



ภาพที่ 5-1 เปรียบเทียบประสิทธิภาพของตัวแบบ

ผลการศึกษพบว่าตัวแบบ Random Forest มีประสิทธิภาพในการจำแนกสถานภาพแรงงาน ระหว่างการเป็นแรงงานในระบบหรือนอกระบบได้ดีที่สุด เนื่องจากให้ค่าความถูกต้อง (Accuracy) ความแม่นยำ (Precision) ความครอบคลุม (Recall) คะแนน F1 (F1-Score) และพื้นที่ใต้กราฟ

(AUC) สูงที่สุด และจากการพิจารณาความสำคัญของตัวแปรอิสระจากตัวแบบ พบว่า ตัวแปรประเภทของค่าจ้างที่ได้รับ มีบทบาทสำคัญในการจำแนกประเภทมากที่สุด รองลงมา ได้แก่ อาชีพ ค่าจ้างที่ได้รับต่อเดือน อายุ ชั่วโมงการทำงานต่อสัปดาห์ จำนวนปีที่ศึกษา ภาค จำนวนสมาชิกในครัวเรือน สถานภาพการสมรส เขตการปกครอง และเพศ ตามลำดับ

อีกทั้ง ยังพบว่าตัวแบบ Logistic Regression ซึ่งทำการคัดเลือกตัวแปรอิสระเข้าสู่ตัวแบบทั้งหมด 5 วิธี ให้ผลค่าความถูกต้อง (Accuracy) ความแม่นยำ (Precision) ความครอบคลุม (Recall) คะแนน F1 (F1-Score) และพื้นที่ใต้กราฟ (AUC) ที่ใกล้เคียงกัน โดยวิธีการคัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปร (Forward Selection) การคัดเลือกตัวแปรด้วยวิธีลดตัวแปร (Backward Elimination) และการคัดเลือกตัวแปรด้วยวิธีเพิ่มตัวแปรอิสระแบบขั้นตอน (Stepwise Selection) ให้ผลการคัดเลือกตัวแปรที่เหมือนกันทั้ง 3 ตัวแบบ โดยมีค่า AIC เท่ากับ 47,047 ซึ่งต่ำกว่าตัวแบบที่ได้จากวิธีนำตัวแปรเข้าทั้งหมด (Enter Regression) และวิธีการคัดเลือกตัวแปรจาก Feature Importance ของตัวแบบ Random Forest ที่มีค่า AIC เท่ากับ 47,048 และ 47,090 ตามลำดับ โดยทุกตัวแปรที่พิจารณามีความสำคัญในการจำแนกสถานภาพแรงงานอย่างมีนัยสำคัญทางสถิติ ยกเว้นตัวแปรจำนวนปีที่ศึกษา ซึ่งให้ผลแตกต่างจากตัวแบบ Random Forest ที่ตัวแปรจำนวนปีที่ศึกษาค่อนข้างมีความสำคัญในระดับกลางกับการจำแนกสถานภาพแรงงาน

5.2 ข้อเสนอแนะ

- 1) ในการศึกษาครั้งถัดไป อาจพิจารณาเทคนิคอื่นๆ เช่น Neural Network หรือ Support Vector Machines (SVM) เพื่อเปรียบเทียบกับ Random Forest และ Logistic Regression
- 2) ในการศึกษาครั้งถัดไป อาจพิจารณาการใช้ข้อมูลจากแหล่งอื่นเพิ่มเติมเพื่อให้ตัวแบบสามารถสะท้อนปัจจัยที่มีผลต่อสถานภาพแรงงานได้ครอบคลุมและแม่นยำยิ่งขึ้น
- 3) หน่วยงานที่เกี่ยวข้องสามารถใช้ข้อมูลเกี่ยวกับปัจจัยสำคัญที่มีผลต่อสถานภาพแรงงาน เพื่อออกแบบนโยบายแรงงานที่เหมาะสม เช่น การกำหนดมาตรการสนับสนุนแรงงานนอกระบบ หรือการพัฒนานโยบายค่าจ้างและสวัสดิการแรงงานให้เหมาะสมกับกลุ่มแรงงานแต่ละประเภท

บรรณานุกรม

- พลากร สีน้อย. (2553). สถิติทดสอบเพื่อคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ.
- อุปลัทธิ์ กอวัฒนสกุล. (2563). แรงงานนอกระบบในไทย: ความเหลื่อมล้ำ ค่าแรง และความเสี่ยงในการทำงาน.
- ช่วงชัย ชูปวา และวิภา ชูปวา. (2564). ปัจจัยที่มีความสัมพันธ์ต่อคุณภาพชีวิตของแรงงานนอกระบบ. กระทรวงแรงงาน. (2566). อนาคตกับการบริหารจัดการแรงงานนอกระบบของไทย.
- สำนักงานสถิติแห่งชาติ. (2567). การสำรวจแรงงานนอกระบบ พ.ศ. 2567.
- Marcouiller, D., Ruiz de Castilla, V., & Woodruff, C. (1997). Formal measures of the informal sector wage gap in Mexico, El Salvador, and Peru. *Economic Development and Cultural Change*, 45 (2), 367– 392.
- Carneiro, F., & Henley, A. (2001). *Modelling formal vs. informal employment and earnings: Microeconometric evidence for Brazil* (Management and Business Working Paper 2001–2015).
- Gunatilaka, R. (2008). *Informal Employment in Sri Lanka: Nature, Probability of Employment, and Determinants of Wages*. International Labour Organization.
- Gimpelson, V.E., & Zudina, A.A. (2011). *Informal Sector in Russia: Dynamics, Structure, Determinants*.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. 3.
- De Vreyer, P., & Roubaud, F. (Eds.). (2013). *Urban labor markets in sub-Saharan Africa*. Washington, DC: World Bank.
- Hartshorn, S. (2016). *Machine Learning With Random Forests And Decision Trees: A Visual Guide For Beginners*.
- García, G.A. (2017). Labor informality: Choice or sign of segmentation? A quantile regression approach at the regional level for Colombia.
- Khan, M. T. A., & Hussain, B. (2020). *Measurement and Determinants of Informal Employment: Evidence from Pakistan*.
- Mkulisi, S. N., & Susuman, A. S. (2024). *Examining the Influence of Socio-Economic and Demographic Transformation on Employment in the Informal Sector from 2017 to 2020*.

ภาคผนวก



ชุดคำสั่งโปรแกรม R

Step 1 Load libraries and Import Dataset

Import dataset

```
# Set the working directory and read the dataset
setwd("D:/KMUTNB/IS/work2")
dataset = read.csv('IES_clean.csv')

str(dataset)
```

```
## 'data.frame': 85479 obs. of 12 variables:
## $ REG : chr "BKK" "BKK" "BKK" "BKK" ...
## $ AREA : chr "Muni" "Muni" "Muni" "Muni" ...
## $ NHH : int 1 1 2 2 1 2 1 4 4 2 ...
## $ SEX : chr "Male" "Female" "Female" "Male" ...
## $ AGE : int 64 81 32 32 19 45 35 35 39 26 ...
## $ MAR : chr "Single" "Widowed" "MarriedUnreg" "MarriedUnreg" ...
## $ HR : int 56 60 48 54 54 48 56 54 48 48 ...
## $ WAGE : chr "Monthly" "Daily" "Daily" "Daily" ...
## $ APPROX: int 9000 30000 10400 13000 8450 10000 15000 10400 13000 9100 ...
## $ RE_SEC: int 0 1 0 0 0 0 0 0 0 0 ...
## $ EDUC : int 12 12 0 6 6 9 9 0 6 6 ...
## $ OCCUP : chr "ServiceSales" "Manager" "Elementary" "Artisan" ...
```

```
head(dataset)
```

```
##   REG AREA NHH   SEX AGE      MAR HR   WAGE APPROX RE_SEC EDUC
## 1 BKK Muni   1  Male  64   Single 56 Monthly  9000     0   12
## 2 BKK Muni   1 Female  81   Widowed 60 Daily 30000     1   12
## 3 BKK Muni   2 Female  32 MarriedUnreg 48 Daily 10400     0    0
## 4 BKK Muni   2 Male  32 MarriedUnreg 54 Daily 13000     0    6
## 5 BKK Muni   1 Female  19   Single 54 Daily  8450     0    6
## 6 BKK Muni   2 Female  45   Divorced 48 Monthly 10000     0    9
##      OCCUP
## 1 ServiceSales
## 2      Manager
## 3 Elementary
## 4      Artisan
## 5 Elementary
## 6 Elementary
```

Biserial correlation

```
continuous_vars <- c("AGE", "HR", "APPROX", "EDUC")
cor_results <- sapply(dataset[, continuous_vars], function(x) -biserial.cor(x, dataset$RE_SEC))
print(cor_results)
```

```
##      AGE      HR      APPROX      EDUC
## 0.39918454 -0.16413690 -0.03644705 -0.38912072
```

Cramer V

```
categorical_vars <- c("REG", "AREA", "SEX", "MAR", "WAGE", "OCCUP")
cramer_v_results <- sapply(categorical_vars, function(var) {
  assocstats(table(dataset$RE_SEC, dataset[[var]]))$cramer})
print(cramer_v_results)
```

```
##      REG      AREA      SEX      MAR      WAGE      OCCUP
## 0.29207132 0.14281280 0.07889312 0.25624478 0.61487014 0.61017889
```


Step 2 Data Preperation

2.1 Encoding the target feature as factor

```
dataset$RE_SEC = factor(dataset$RE_SEC, levels = c(0, 1))
str(dataset$RE_SEC)
```

```
## Factor w/ 2 levels "0","1": 1 2 1 1 1 1 1 1 1 1 ...
```

2.2 Split dataset into Training set and Test set

```
# Splitting the dataset into the Training set and Test set
set.seed(123)
split = sample.split(dataset$RE_SEC, SplitRatio = 0.75)
training_set = subset(dataset, split == TRUE)
test_set = subset(dataset, split == FALSE)
```

```
# Check structure
str(training_set)
```

```
## 'data.frame': 64109 obs. of 12 variables:
## $ REG : chr "BKK" "BKK" "BKK" "BKK" ...
## $ AREA : chr "Muni" "Muni" "Muni" "Muni" ...
## $ NHH : int 1 1 2 1 4 2 2 2 3 1 ...
## $ SEX : chr "Male" "Female" "Male" "Male" ...
## $ AGE : int 64 81 32 35 35 26 22 39 60 51 ...
## $ MAR : chr "Single" "Widowed" "MarriedUnreg" "MarriedUnreg" ...
## $ HR : int 56 60 54 56 54 48 48 48 48 56 ...
## $ WAGE : chr "Monthly" "Daily" "Daily" "Daily" ...
## $ APPROX: int 9000 30000 13000 15000 10400 9100 9100 11000 12000 15000 ...
## $ RE_SEC: Factor w/ 2 levels "0","1": 1 2 1 1 1 1 1 1 1 2 ...
## $ EDUC : int 12 12 6 9 0 6 6 12 7 6 ...
## $ OCCUP : chr "ServiceSales" "Manager" "Artisan" "Artisan" ...
```

```
str(test_set)
```

```
## 'data.frame': 21370 obs. of 12 variables:
## $ REG : chr "BKK" "BKK" "BKK" "BKK" ...
## $ AREA : chr "Muni" "Muni" "Muni" "Muni" ...
## $ NHH : int 2 1 2 4 2 2 5 5 1 4 ...
## $ SEX : chr "Female" "Female" "Female" "Male" ...
## $ AGE : int 32 19 45 39 38 40 39 39 61 62 ...
## $ MAR : chr "MarriedUnreg" "Single" "Divorced" "MarriedUnreg" ...
## $ HR : int 48 54 48 48 48 40 42 42 56 25 ...
## $ WAGE : chr "Daily" "Daily" "Monthly" "Daily" ...
## $ APPROX: int 10400 8450 10000 13000 11000 17000 9100 10000 12000 10000 ...
## $ RE_SEC: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 2 ...
## $ EDUC : int 0 6 9 6 12 9 9 9 0 4 ...
## $ OCCUP : chr "Elementary" "Elementary" "Elementary" "Artisan" ...
```

2.3 Feature Scaling for Continuous Variables

```
# Feature Scaling
training_set["NHH"] = scale(training_set["NHH"])
test_set["NHH"] = scale(test_set["NHH"])

training_set["AGE"] = scale(training_set["AGE"])
test_set["AGE"] = scale(test_set["AGE"])

training_set["HR"] = scale(training_set["HR"])
test_set["HR"] = scale(test_set["HR"])

training_set["EDUC"] = scale(training_set["EDUC"])
test_set["EDUC"] = scale(test_set["EDUC"])

training_set["APPROX"] = scale(training_set["APPROX"])
test_set["APPROX"] = scale(test_set["APPROX"])
```

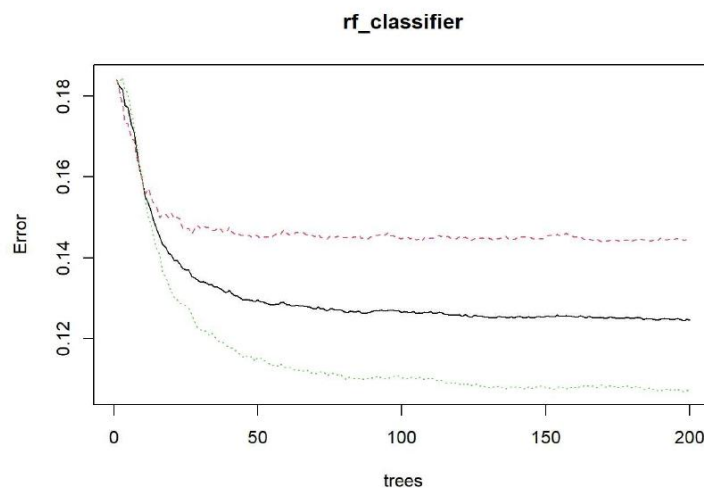
Step 3 Fit Model

3.1 Random Forest Classification

```
set.seed(123)
best_mtry <- floor(sqrt(ncol(training_set) - 1))
rf_classifier <- randomForest(RE_SEC ~ ., data = training_set, ntree = 200, mtry = best_mtry)
rf_classifier
```

```
##
## Call:
## randomForest(formula = RE_SEC ~ ., data = training_set, ntree = 200,      mtry = best_mtry)
##              Type of random forest: classification
##              Number of trees: 200
## No. of variables tried at each split: 3
##
##      OOB estimate of  error rate: 12.46%
## Confusion matrix:
##      0      1 class.error
## 0 25817  4360  0.1444809
## 1  3625 30307  0.1068313
```

```
plot(rf_classifier)
```



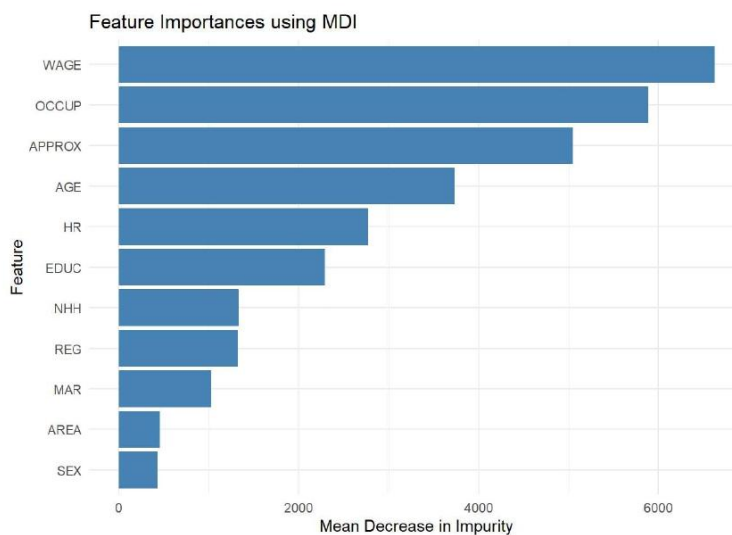
Random Forest Importance Feature

```
importance <- importance(rf_classifier)
print(importance)
```

```
##           MeanDecreaseGini
## REG           1326.3824
## AREA           456.4160
## NHH           1333.9348
## SEX            431.5477
## AGE           3731.7295
## MAR           1026.1782
## HR            2771.1331
## WAGE          6624.6092
## APPROX        5049.3815
## EDUC          2293.1427
## OCCUP         5884.3137
```

```
# Visualize
importance_df <- data.frame(Feature = rownames(importance), Importance = importance(rf_classifier)[, "MeanDecreaseGini"])
```

```
# Plot
ggplot(importance_df, aes(x = reorder(Feature, Importance), y = Importance)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  coord_flip() + labs(title = "Feature Importances using MDI", x = "Feature", y = "Mean Decrease in Impurity") +
  theme_minimal()
```



3.2 Logistic Regression: Enter Method

```
Enter_model <- glm(RE_SEC ~ ., data = training_set, family = binomial)
summary(Enter_model)
```

```
##
## Call:
## glm(formula = RE_SEC ~ ., family = binomial, data = training_set)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.97419    0.08927  22.114 < 2e-16 ***
## REGCentral      0.29686    0.05635   5.268 1.38e-07 ***
## REGEastNorth    1.17634    0.05912  19.898 < 2e-16 ***
## REGNorth        1.04001    0.05920  17.567 < 2e-16 ***
## REGSouth        0.22067    0.05975   3.693 0.000221 ***
## AREANon         0.15678    0.02450   6.400 1.56e-10 ***
## NHH             0.06024    0.01212   4.970 6.68e-07 ***
## SEXMale         0.06772    0.02527   2.680 0.007373 **
## AGE             0.57496    0.01536  37.442 < 2e-16 ***
## MARMarriedReg   -0.11490    0.06084  -1.889 0.058942 .
## MARMarriedUnreg -0.11617    0.06176  -1.881 0.059964 .
## MARSeparated    -0.04122    0.07568  -0.545 0.586033
## MARSingle       -0.05579    0.06345  -0.879 0.379254
## MARWidowed      0.12307    0.07714   1.595 0.110623
## HR              -0.22242    0.01291 -17.233 < 2e-16 ***
## WAGEHourly      -0.01791    0.24538  -0.073 0.941827
## WAGEMonthly     -1.96946    0.02914 -67.577 < 2e-16 ***
## WAGENonMonetary 10.56670    56.17904   0.188 0.850806
## WAGEOther       0.50219    0.04510  11.136 < 2e-16 ***
## WAGEWeekly      0.31712    0.09752   3.252 0.001146 **
## APPROX          0.70343    0.02441  28.813 < 2e-16 ***
## EDUC            0.01890    0.01584   1.193 0.232814
## OCCUPArtisan    -2.17688    0.04752 -45.807 < 2e-16 ***
## OCCUPClerk      -3.83051    0.10888 -35.180 < 2e-16 ***
## OCCUPElementary -2.06652    0.04605 -44.873 < 2e-16 ***
## OCCUPMachineOperator -2.64037    0.05245 -50.341 < 2e-16 ***
## OCCUPManager    -3.95248    0.09856 -40.104 < 2e-16 ***
## OCCUPProfessional -4.20716    0.08720 -48.247 < 2e-16 ***
## OCCUPServiceSales -1.47354    0.04673 -31.532 < 2e-16 ***
## OCCUPTechnician -3.57585    0.08901 -40.175 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 88654  on 64108  degrees of freedom
## Residual deviance: 46988  on 64079  degrees of freedom
## AIC: 47048
##
## Number of Fisher Scoring iterations: 12
```

```
pR2(Enter_model)
```

```
## fitting null model for pseudo-r2
```

```
##              llh      llhNull      G2      McFadden      r2ML
## -2.349405e+04 -4.432694e+04  4.166577e+04  4.699825e-01  4.779129e-01
##              r2CU
## 6.379484e-01
```

```
vif(Enter_model)
```

```
##              GVIF Df GVIF^(1/(2*Df))
## REG      1.207249  4      1.023822
## AREA     1.094032  1      1.045960
## NHH      1.130657  1      1.063324
## SEX      1.169987  1      1.081659
## AGE      1.586453  1      1.259545
## MAR      1.604938  5      1.048445
## HR       1.161605  1      1.077778
## WAGE     1.561186  5      1.045552
## APPROX   1.538215  1      1.240248
## EDUC     1.643166  1      1.281860
## OCCUP    2.755152  8      1.065391
```

3.3 Logistic Regression: Forward Selection

```
initial_model <- glm(RE_SEC ~ 1, data = training_set, family = binomial)
forward_model <- stepAIC(initial_model, direction = "forward", scope = list(lower = initial_model, upper = Enter_model))
```

```
## Start: AIC=88655.88
## RE_SEC ~ 1
##
##           Df Deviance  AIC
## + OCCUP    8   60539 60557
## + WAGE     5   61244 61256
## + AGE      1   77660 77664
## + EDUC     1   78344 78348
## + REG      4   82954 82964
## + MAR      5   84283 84295
## + HR       1   86916 86920
## + AREA     1   87398 87402
## + SEX      1   88247 88251
## + APPROX   1   88567 88571
## <none>      1   88654 88656
## + NIH      1   88654 88658
##
## Step: AIC=60556.69
## RE_SEC ~ OCCUP
##
##           Df Deviance  AIC
## + WAGE     5   52419 52447
## + AGE      1   56574 56594
## + REG      4   58604 58630
## + MAR      5   59120 59148
## + APPROX   1   59364 59384
## + EDUC     1   59672 59692
## + HR       1   60274 60294
## + AREA     1   60497 60517
## + SEX      1   60530 60550
## + NIH      1   60534 60554
## <none>      1   60539 60557
##
## Step: AIC=52447.17
## RE_SEC ~ OCCUP + WAGE
##
##           Df Deviance  AIC
## + AGE      1   49570 49600
## + REG      4   50988 51024
## + APPROX   1   51343 51373
## + MAR      5   51392 51430
## + HR       1   52162 52192
## + EDUC     1   52324 52354
## + AREA     1   52402 52432
## + NIH      1   52412 52442
## <none>      1   52419 52447
## + SEX      1   52417 52447
##
## Step: AIC=49599.78
## RE_SEC ~ OCCUP + WAGE + AGE
##
##           Df Deviance  AIC
## + REG      4   48370 48408
## + APPROX   1   48752 48784
## + HR       1   49396 49428
## + EDUC     1   49476 49508
## + MAR      5   49518 49558
## + NIH      1   49535 49567
## + AREA     1   49537 49569
## + SEX      1   49546 49578
## <none>      1   49570 49600
##
## Step: AIC=48407.93
## RE_SEC ~ OCCUP + WAGE + AGE + REG
##
##           Df Deviance  AIC
## + APPROX   1   47388 47428
## + HR       1   48190 48230
## + EDUC     1   48298 48338
```

```

## + AREA      1  48342 48382
## + NHH       1  48344 48384
## + SEX       1  48348 48388
## <none>      48370 48408
## + MAR       5  48360 48408
##
## Step: AIC=47427.91
## RE_SEC ~ OCCUP + WAGE + AGE + REG + APPROX
##
##           Df Deviance  AIC
## + HR      1  47080 47122
## + AREA    1  47340 47382
## + NHH     1  47362 47404
## + MAR     5  47371 47421
## + SEX     1  47384 47426
## + EDUC    1  47385 47427
## <none>    47388 47428
##
## Step: AIC=47122.34
## RE_SEC ~ OCCUP + WAGE + AGE + REG + APPROX + HR
##
##           Df Deviance  AIC
## + AREA    1  47036 47080
## + NHH     1  47058 47102
## + MAR     5  47066 47118
## + SEX     1  47077 47121
## <none>    47080 47122
## + EDUC    1  47080 47124
##
## Step: AIC=47079.69
## RE_SEC ~ OCCUP + WAGE + AGE + REG + APPROX + HR + AREA
##
##           Df Deviance  AIC
## + NHH     1  47018 47064
## + MAR     5  47021 47075
## + SEX     1  47032 47078
## <none>    47036 47080
## + EDUC    1  47035 47081
##
## Step: AIC=47063.76
## RE_SEC ~ OCCUP + WAGE + AGE + REG + APPROX + HR + AREA + NHH
##
##           Df Deviance  AIC
## + MAR     5  46996 47052
## + SEX     1  47014 47062
## <none>    47018 47064
## + EDUC    1  47017 47065
##
## Step: AIC=47051.55
## RE_SEC ~ OCCUP + WAGE + AGE + REG + APPROX + HR + AREA + NHH +
##           MAR
##
##           Df Deviance  AIC
## + SEX     1  46989 47047
## <none>    46996 47052
## + EDUC    1  46995 47053
##
## Step: AIC=47047.02
## RE_SEC ~ OCCUP + WAGE + AGE + REG + APPROX + HR + AREA + NHH +
##           MAR + SEX
##
##           Df Deviance  AIC
## <none>    46989 47047
## + EDUC    1  46988 47048

```



```
summary(forward_model)
```

```
##
## Call:
## glm(formula = RE_SEC ~ OCCUP + WAGE + AGE + REG + APPROX + HR +
##      AREA + NHH + MAR + SEX, family = binomial, data = training_set)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.96561    0.08897  22.093 < 2e-16 ***
## OCCUPartisan     -2.17268    0.04739 -45.850 < 2e-16 ***
## OCCUPClerk       -3.81219    0.10778 -35.370 < 2e-16 ***
## OCCUPElementary -2.06785    0.04603 -44.922 < 2e-16 ***
## OCCUPMachineOperator -2.63545    0.05227 -50.416 < 2e-16 ***
## OCCUPManager     -3.93713    0.09773 -40.285 < 2e-16 ***
## OCCUPProfessional -4.18494    0.08519 -49.122 < 2e-16 ***
## OCCUPServiceSales -1.46596    0.04629 -31.671 < 2e-16 ***
## OCCUPTechnician  -3.55810    0.08776 -40.541 < 2e-16 ***
## WAGEHourly       -0.02309    0.24528 -0.094 0.924989
## WAGEMonthly      -1.96268    0.02857 -68.687 < 2e-16 ***
## WAGENonMonetary  10.56613    56.18414  0.188 0.850828
## WAGEOther         0.50434    0.04506  11.193 < 2e-16 ***
## WAGEWeekly        0.31831    0.09748  3.266 0.001093 **
## AGE               0.56919    0.01457  39.077 < 2e-16 ***
## REGCentral        0.29864    0.05632  5.303 1.14e-07 ***
## REGEastNorth      1.17964    0.05905  19.977 < 2e-16 ***
## REGNorth          1.04299    0.05915  17.634 < 2e-16 ***
## REGSouth          0.22326    0.05970  3.739 0.000184 ***
## APPROX            0.70987    0.02385  29.769 < 2e-16 ***
## HR               -0.22313    0.01289 -17.306 < 2e-16 ***
## AREANon           0.15565    0.02448  6.359 2.04e-10 ***
## NHH               0.06041    0.01212  4.985 6.20e-07 ***
## MARMarriedReg     -0.11526    0.06083 -1.895 0.058144 .
## MARMarriedUnreg   -0.11922    0.06170 -1.932 0.053324 .
## MARSeparated      -0.04249    0.07568 -0.561 0.574507
## MARSingle         -0.05533    0.06344 -0.872 0.383152
## MARWidowed        0.11891    0.07707  1.543 0.122859
## SEXMale           0.06702    0.02526  2.653 0.007986 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 88654  on 64108  degrees of freedom
## Residual deviance: 46989  on 64080  degrees of freedom
## AIC: 47047
##
## Number of Fisher Scoring iterations: 12
```

```
pR2(forward_model)
```

```
## fitting null model for pseudo-r2
```

```
##          llh          llhNull          G2          McFadden          r2ML
## -2.349451e+04 -4.432694e+04  4.166486e+04  4.699722e-01  4.779055e-01
##          r2CU
## 6.379385e-01
```

```
vif(forward_model)
```

```
##          GVIF Df GVIF^(1/(2*Df))
## OCCUP  2.488342  8      1.058630
## WAGE   1.499076  5      1.041316
## AGE    1.427774  1      1.194895
## REG    1.203558  4      1.023431
## APPROX 1.464284  1      1.210076
## HR     1.159101  1      1.076615
## AREA   1.092361  1      1.045161
## NHH    1.130523  1      1.063260
## MAR    1.586666  5      1.047246
## SEX    1.169247  1      1.081317
```

3.4 Logistic Regression: Backward Elimination

```
backward_model <- stepAIC(Enter_model, direction = "backward")
```

```
## Start: AIC=47048.11
## RE_SEC ~ REG + AREA + NHH + SEX + AGE + MAR + HR + WAGE + APPROX +
## EDUC + OCCUP
##
##           Df Deviance  AIC
## - EDUC      1  46989 47047
## <none>      0  46988 47048
## - SEX        1  46995 47053
## - MAR        5  47013 47063
## - NHH        1  47013 47071
## - AREA       1  47030 47088
## - HR         1  47282 47340
## - APPROX     1  48037 48095
## - REG        4  48311 48363
## - AGE        1  48451 48509
## - WAGE       5  53261 53311
## - OCCUP      8  53844 53888
##
## Step: AIC=47047.02
## RE_SEC ~ REG + AREA + NHH + SEX + AGE + MAR + HR + WAGE + APPROX +
## OCCUP
##
##           Df Deviance  AIC
## <none>      0  46989 47047
## - SEX        1  46996 47052
## - MAR        5  47014 47062
## - NHH        1  47014 47070
## - AREA       1  47030 47086
## - HR         1  47286 47342
## - APPROX     1  48110 48166
## - REG        4  48319 48369
## - AGE        1  48590 48646
## - WAGE       5  53475 53523
## - OCCUP      8  54108 54150
```



```
summary(backward_model)
```

```
##
## Call:
## glm(formula = RE_SEC ~ REG + AREA + NHH + SEX + AGE + MAR + HR +
##       WAGE + APPROX + OCCUP, family = binomial, data = training_set)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.96561      0.08897   22.093 < 2e-16 ***
## REGCentral        0.29864      0.05632    5.303 1.14e-07 ***
## REGEastNorth      1.17964      0.05905   19.977 < 2e-16 ***
## REGNorth          1.04299      0.05915   17.634 < 2e-16 ***
## REGSouth          0.22326      0.05970    3.739 0.000184 ***
## AREANon           0.15565      0.02448   6.359 2.04e-10 ***
## NHH               0.06041      0.01212   4.985 6.20e-07 ***
## SEXMale           0.06702      0.02526    2.653 0.007986 **
## AGE               0.56919      0.01457   39.077 < 2e-16 ***
## MARMarriedReg     -0.11526      0.06083   -1.895 0.058144 .
## MARMarriedUnreg   -0.11922      0.06170   -1.932 0.053324 .
## MARSeparated      -0.04249      0.07568   -0.561 0.574507
## MARSingle         -0.05533      0.06344   -0.872 0.383152
## MARWidowed        0.11891      0.07707    1.543 0.122859
## HR                -0.22313      0.01289  -17.306 < 2e-16 ***
## WAGEHourly        -0.02309      0.24528   -0.094 0.924989
## WAGEMonthly       -1.96268      0.02857  -68.687 < 2e-16 ***
## WAGENonMonetary   10.56613     56.18414    0.188 0.850828
## WAGEOther          0.50434      0.04506   11.193 < 2e-16 ***
## WAGEWeekly        0.31831      0.09748    3.266 0.001093 **
## APPROX            0.70987      0.02385   29.769 < 2e-16 ***
## OCCUPArtisan      -2.17268      0.04739  -45.850 < 2e-16 ***
## OCCUPClerk        -3.81219      0.10778  -35.370 < 2e-16 ***
## OCCUPElementary  -2.06785      0.04603  -44.922 < 2e-16 ***
## OCCUPMachineOperator -2.63545      0.05227  -50.416 < 2e-16 ***
## OCCUPManager      -3.93713      0.09773  -40.285 < 2e-16 ***
## OCCUPProfessional -4.18494      0.08519  -49.122 < 2e-16 ***
## OCCUPServiceSales -1.46596      0.04629  -31.671 < 2e-16 ***
## OCCUPTechnician   -3.55810      0.08776  -40.541 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 88654  on 64108  degrees of freedom
## Residual deviance: 46989  on 64080  degrees of freedom
## AIC: 47047
##
## Number of Fisher Scoring iterations: 12
```

```
pR2(backward_model)
```

```
## fitting null model for pseudo-r2
```

```
##          llh          llhNull          G2          McFadden          r2ML
## -2.349451e+04 -4.432694e+04  4.166486e+04  4.699722e-01  4.779055e-01
##          r2CU
## 6.379385e-01
```

```
vif(backward_model)
```

```
##          GVIF Df GVIF^(1/(2*Df))
## REG      1.203558 4      1.023431
## AREA     1.092361 1      1.045161
## NHH      1.130523 1      1.063260
## SEX      1.169247 1      1.081317
## AGE      1.427774 1      1.194895
## MAR      1.586666 5      1.047246
## HR       1.159101 1      1.076615
## WAGE     1.499076 5      1.041316
## APPROX   1.464284 1      1.210076
## OCCUP    2.488342 8      1.058630
```

3.5 Logistic Regression: Stepwise Selection

```
stepwise_model <- stepAIC(Enter_model, direction = "both")
```

```
## Start: AIC=47048.11
## RE_SEC ~ REG + AREA + NHH + SEX + AGE + MAR + HR + WAGE + APPROX +
## EDUC + OCCUP
##
##      Df Deviance  AIC
## - EDUC    1  46989 47047
## <none>      46988 47048
## - SEX     1  46995 47053
## - MAR     5  47013 47063
## - NHH     1  47013 47071
## - AREA    1  47030 47088
## - HR      1  47282 47340
## - APPROX  1  48037 48095
## - REG     4  48311 48363
## - AGE     1  48451 48509
## - WAGE    5  53261 53311
## - OCCUP   8  53844 53888
##
## Step: AIC=47047.02
## RE_SEC ~ REG + AREA + NHH + SEX + AGE + MAR + HR + WAGE + APPROX +
## OCCUP
##
##      Df Deviance  AIC
## <none>      46989 47047
## + EDUC    1  46988 47048
## - SEX     1  46996 47052
## - MAR     5  47014 47062
## - NHH     1  47014 47070
## - AREA    1  47030 47086
## - HR      1  47286 47342
## - APPROX  1  48110 48166
## - REG     4  48319 48369
## - AGE     1  48590 48646
## - WAGE    5  53475 53523
## - OCCUP   8  54108 54150
```




```
summary(stepwise_model)
```

```
##
## Call:
## glm(formula = RE_SEC ~ REG + AREA + NHH + SEX + AGE + MAR + HR +
##       WAGE + APPROX + OCCUP, family = binomial, data = training_set)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.96561    0.08897   22.093 < 2e-16 ***
## REGCentral        0.29864    0.05632    5.303 1.14e-07 ***
## REGEastNorth      1.17964    0.05905   19.977 < 2e-16 ***
## REGNorth          1.04299    0.05915   17.634 < 2e-16 ***
## REGSouth          0.22326    0.05970    3.739 0.000184 ***
## AREANon           0.15565    0.02448   6.359 2.04e-10 ***
## NHH               0.06041    0.01212   4.985 6.20e-07 ***
## SEXMale           0.06702    0.02526    2.653 0.007986 **
## AGE              0.56919    0.01457   39.077 < 2e-16 ***
## MARMarriedReg     -0.11526    0.06083  -1.895 0.058144 .
## MARMarriedUnreg   -0.11922    0.06170  -1.932 0.053324 .
## MARSeparated      -0.04249    0.07568  -0.561 0.574507
## MARSingle         -0.05533    0.06344  -0.872 0.383152
## MARWidowed        0.11891    0.07707    1.543 0.122859
## HR               -0.22313    0.01289  -17.306 < 2e-16 ***
## WAGEHourly        -0.02309    0.24528  -0.094 0.924989
## WAGEMonthly       -1.96268    0.02857  -68.687 < 2e-16 ***
## WAGENonMonetary   10.56613   56.18414    0.188 0.850828
## WAGEOther         0.50434    0.04506   11.193 < 2e-16 ***
## WAGEWeekly        0.31831    0.09748    3.266 0.001093 **
## APPROX            0.70987    0.02385   29.769 < 2e-16 ***
## OCCUPArtisan      -2.17268    0.04739  -45.850 < 2e-16 ***
## OCCUPClerk        -3.81219    0.10778  -35.370 < 2e-16 ***
## OCCUPElementary  -2.06785    0.04603  -44.922 < 2e-16 ***
## OCCUPMachineOperator -2.63545    0.05227  -50.416 < 2e-16 ***
## OCCUPManager      -3.93713    0.09773  -40.285 < 2e-16 ***
## OCCUPProfessional -4.18494    0.08519  -49.122 < 2e-16 ***
## OCCUPServiceSales -1.46596    0.04629  -31.671 < 2e-16 ***
## OCCUPTechnician  -3.55810    0.08776  -40.541 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 88654  on 64108  degrees of freedom
## Residual deviance: 46989  on 64080  degrees of freedom
## AIC: 47047
##
## Number of Fisher Scoring iterations: 12
```

```
pR2(stepwise_model)
```

```
## fitting null model for pseudo-r2
```

```
##          llh          llhNull          G2          McFadden          r2ML
## -2.349451e+04 -4.432694e+04  4.166486e+04  4.699722e-01  4.779055e-01
##          r2CU
## 6.379385e-01
```

```
vif(stepwise_model)
```

```
##          GVIF Df GVIF^(1/(2*Df))
## REG      1.203558 4      1.023431
## AREA      1.092361 1      1.045161
## NHH       1.130523 1      1.063260
## SEX       1.169247 1      1.081317
## AGE       1.427774 1      1.194895
## MAR       1.586666 5      1.047246
## HR        1.159101 1      1.076615
## WAGE       1.499076 5      1.041316
## APPROX    1.464284 1      1.210076
## OCCUP     2.488342 8      1.058630
```

3.6 Logistic Regression: Variable Selection from RF

```
logistic_rf_model <- glm(RE_SEC ~ . - AREA - SEX - EDUC, data = training_set, family = binomial)
summary(logistic_rf_model)
```

```
##
## Call:
## glm(formula = RE_SEC ~ . - AREA - SEX - EDUC, family = binomial,
##      data = training_set)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    2.04415    0.08739   23.391 < 2e-16 ***
## REGCentral      0.36426    0.05526    6.592 4.35e-11 ***
## REGEastNorth    1.23369    0.05841   21.121 < 2e-16 ***
## REGNorth        1.10108    0.05841   18.852 < 2e-16 ***
## REGSouth        0.28362    0.05891    4.815 1.48e-06 ***
## NHH              0.06526    0.01209    5.399 6.69e-08 ***
## AGE             0.56590    0.01454   38.910 < 2e-16 ***
## MARMarriedReg   -0.10519    0.06070   -1.733 0.08310 .
## MARMarriedUnreg -0.11438    0.06159   -1.857 0.06328 .
## MARSeparated    -0.03916    0.07561   -0.518 0.60455
## MARSingle       -0.04708    0.06317   -0.745 0.45611
## MARWidowed      0.10820    0.07689    1.407 0.15940
## HR              -0.22428    0.01289  -17.396 < 2e-16 ***
## WAGEHourly      -0.02949    0.24493   -0.120 0.90417
## WAGEMonthly     -1.96127    0.02852  -68.762 < 2e-16 ***
## WAGENonMonetary 10.59503    56.06214    0.189 0.85010
## WAGEOther        0.50749    0.04506   11.262 < 2e-16 ***
## WAGEWeekly       0.30662    0.09732    3.151 0.00163 **
## APPROX           0.70697    0.02373   29.790 < 2e-16 ***
## OCCUPArtisan     -2.20242    0.04719  -46.672 < 2e-16 ***
## OCCUPClerk       -3.88599    0.10698  -36.326 < 2e-16 ***
## OCCUPElementary -2.09831    0.04574  -45.876 < 2e-16 ***
## OCCUPMachineOperator -2.66144    0.05212  -51.066 < 2e-16 ***
## OCCUPManager     -3.97541    0.09767  -40.702 < 2e-16 ***
## OCCUPProfessional -4.24997    0.08408  -50.547 < 2e-16 ***
## OCCUPServiceSales -1.52882    0.04488  -34.063 < 2e-16 ***
## OCCUPTechnician  -3.61351    0.08729  -41.394 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 88654  on 64108  degrees of freedom
## Residual deviance: 47036  on 64082  degrees of freedom
## AIC: 47090
##
## Number of Fisher Scoring iterations: 12
```

```
pR2(logistic_rf_model)
```

```
## fitting null model for pseudo-r2
```

```
##          llh          llhNull          G2          McFadden          r2ML
## -2.351813e+04 -4.432694e+04  4.161762e+04  4.694393e-01  4.775206e-01
##          r2CU
## 6.374247e-01
```

```
vif(logistic_rf_model)
```

```
##          GVIF Df GVIF^(1/(2*Df))
## REG    1.157749  4      1.018478
## NHH    1.125672  1      1.060977
## AGE    1.424705  1      1.193610
## MAR    1.537534  5      1.043957
## HR     1.159284  1      1.076701
## WAGE    1.493684  5      1.040940
## APPROX 1.452484  1      1.205190
## OCCUP   2.176170  8      1.049798
```

Step 4 Predict Test set and Confusion Matrix

4.1 Confusion Matrix for Random Forest

```
rf_pred = predict(rf_classifier, newdata = test_set)
rf_cm = table(rf_pred, test_set[, 10])
print(rf_cm)
```

```
##
## rf_pred      0      1
##           0 8543 1193
##           1 1516 10118
```

4.2 Confusion Matrix for Enter Method

```
Enter_pred = predict(Enter_model, newdata = test_set, type = "response")
Enter_pred_class = ifelse(Enter_pred > 0.5, 1, 0)
Enter_cm = table(Enter_pred_class, test_set[, 10])
print(Enter_cm)
```

```
##
## Enter_pred_class  0      1
##                 0 8158 1576
##                 1 1901 9735
```

4.3 Confusion Matrix for forward Selection

```
forward_pred = predict(forward_model, newdata = test_set, type = "response")
forward_pred_class = ifelse(forward_pred > 0.5, 1, 0)
forward_cm = table(forward_pred_class, test_set[, 10])
print(forward_cm)
```

```
##
## forward_pred_class  0      1
##                   0 8159 1579
##                   1 1900 9732
```

4.4 Confusion Matrix for Backward Elimination

```
backward_pred = predict(backward_model, newdata = test_set, type = "response")
backward_pred_class = ifelse(backward_pred > 0.5, 1, 0)
backward_cm = table(backward_pred_class, test_set[, 10])
print(backward_cm)
```

```
##
## backward_pred_class  0      1
##                   0 8159 1579
##                   1 1900 9732
```

4.5 Confusion Matrix for Stepwise Selection

```
stepwise_pred = predict(stepwise_model, newdata = test_set, type = "response")
stepwise_pred_class = ifelse(stepwise_pred > 0.5, 1, 0)
stepwise_cm = table(stepwise_pred_class, test_set[, 10])
print(stepwise_cm)
```

```
##
## stepwise_pred_class  0      1
##                   0 8159 1579
##                   1 1900 9732
```

4.6 Confusion Matrix for Logistic Regression: Variable Selection from RF

```
logistic_rf_pred = predict(logistic_rf_model, newdata = test_set, type = "response")
logistic_rf_pred_class = ifelse(logistic_rf_pred > 0.5, 1, 0)
logistic_rf_cm = table(logistic_rf_pred_class, test_set[, 10])
print(logistic_rf_cm)
```

```
##
## logistic_rf_pred_class  0      1
##                   0 8170 1587
##                   1 1889 9724
```

Step 5 Compare the Performance

```
# Performance Metrics for Random Forest
rf_accuracy = Accuracy(rf_pred, test_set[,10])
rf_precision = Precision(rf_pred, test_set[,10])
rf_recall = Recall(rf_pred, test_set[,10])
rf_f1 = F1_Score(rf_pred, test_set[,10])
rf_auc = AUC(rf_pred, test_set[,10])

# Performance Metrics for Logistic Regression: Enter Model
Enter_accuracy = Accuracy(Enter_pred_class, test_set[,10])
Enter_precision = Precision(Enter_pred_class, test_set[,10])
Enter_recall = Recall(Enter_pred_class, test_set[,10])
Enter_f1 = F1_Score(Enter_pred_class, test_set[,10])
Enter_auc = AUC(Enter_pred_class, test_set[,10])

# Performance Metrics for Logistic Regression: Forward Selection
forward_accuracy = Accuracy(forward_pred_class, test_set[,10])
forward_precision = Precision(forward_pred_class, test_set[,10])
forward_recall = Recall(forward_pred_class, test_set[,10])
forward_f1 = F1_Score(forward_pred_class, test_set[,10])
forward_auc = AUC(forward_pred_class, test_set[,10])

# Performance Metrics for Logistic Regression: Backward Elimination
backward_accuracy = Accuracy(backward_pred_class, test_set[,10])
backward_precision = Precision(backward_pred_class, test_set[,10])
backward_recall = Recall(backward_pred_class, test_set[,10])
backward_f1 = F1_Score(backward_pred_class, test_set[,10])
backward_auc = AUC(backward_pred_class, test_set[,10])

# Performance Metrics for Logistic Regression: Stepwise Selection
stepwise_accuracy = Accuracy(stepwise_pred_class, test_set[,10])
stepwise_precision = Precision(stepwise_pred_class, test_set[,10])
stepwise_recall = Recall(stepwise_pred_class, test_set[,10])
stepwise_f1 = F1_Score(stepwise_pred_class, test_set[,10])
stepwise_auc = AUC(stepwise_pred_class, test_set[,10])

# Performance Metrics for Logistic Regression: Variable Selection from RF
logistic_rf_accuracy = Accuracy(logistic_rf_pred_class, test_set[,10])
logistic_rf_precision = Precision(logistic_rf_pred_class, test_set[,10])
logistic_rf_recall = Recall(logistic_rf_pred_class, test_set[,10])
logistic_rf_f1 = F1_Score(logistic_rf_pred_class, test_set[,10])
logistic_rf_auc = AUC(logistic_rf_pred_class, test_set[,10])

# Create Comparison Table
comparison_table <- data.frame(
  Model = c("Random Forest", "Logistic: Enter", "Logistic: Forward", "Logistic: Backward", "Logistic: Stepwise", "Logistic: Variable Selection from RF"),
  Accuracy = c(rf_accuracy, Enter_accuracy, forward_accuracy, backward_accuracy, stepwise_accuracy, logistic_rf_accuracy),
  Precision = c(rf_precision, Enter_precision, forward_precision, backward_precision, stepwise_precision, logistic_rf_precision),
  Recall = c(rf_recall, Enter_recall, forward_recall, backward_recall, stepwise_recall, logistic_rf_recall),
  F1_Score = c(rf_f1, Enter_f1, forward_f1, backward_f1, stepwise_f1, logistic_rf_f1),
  AUC = c(rf_auc, Enter_auc, forward_auc, backward_auc, stepwise_auc, logistic_rf_auc)
)

# Print Comparison Table
print(comparison_table)
```

```
##              Model Accuracy Precision Recall F1_Score
## 1      Random Forest 0.8732335 0.8492892 0.8774651 0.8631473
## 2      Logistic: Enter 0.8372953 0.8110150 0.8380933 0.8243318
## 3      Logistic: Forward 0.8372017 0.8111144 0.8378517 0.8242663
## 4      Logistic: Backward 0.8372017 0.8111144 0.8378517 0.8242663
## 5      Logistic: Stepwise 0.8372017 0.8111144 0.8378517 0.8242663
## 6 Logistic: Variable Selection from RF 0.8373421 0.8122080 0.8373475 0.8245862
##              AUC
## 1 0.8719083
## 2 0.8358408
## 3 0.8357579
## 4 0.8357579
## 5 0.8357579
## 6 0.8359510
```




การสำรวจภาวะการทำงานของประชากร ไตรมาส 3 พ.ศ. 2567

1. ภาค..... จังหวัด.....

2. อำเภอ/เขต..... ตำบล/แขวง.....

3. บ้านเลขที่..... ถนน..... ตรอก/ซอย.....

4. ในเขตเทศบาล EA..... AREA..... EA.....

นอกเขตเทศบาล EA..... หมู่ที่..... ชื่อหมู่บ้าน..... VIL.....

5. ลำดับที่ EA ตัวอย่าง..... PSU_NO.....

6. ชุด EA ตัวอย่าง..... ชุดครัวเรือนตัวอย่าง..... EA_SET..... SAMSET.....

เดือน..... พ.ศ..... MONTH_YR.....

7. ลำดับที่ครัวเรือนตัวอย่าง.....ประเภท(ส่วนบุคคล.....1 กลุ่มบุคคล(ระบุ)..... 2) HH_NO..... TYPE.....

8. จำนวนสมาชิกในครัวเรือน ชั้นแรงงาน..... คน MEMBERS.....

9. จำนวนสมาชิกในครัวเรือน ชั้นนับจุด..... คน LISTING.....

10. ชื่อผู้ตอบสัมภาษณ์.....

เบอร์โทร E-mail

11. ผลการแรงงานครัวเรือนตัวอย่างนี้ (บันทึกรหัส)

ได้ข้อมูล		รหัส	ไม่ได้ข้อมูล		รหัส
ชั้นนับจุด	ชั้นแรงงาน		ชั้นนับจุด	ชั้นแรงงาน	
1. เป็นครัวเรือนตัวอย่าง			เป็นครัวเรือนตัวอย่าง		
1.1 มีครัวเรือนอาศัยอยู่	แรงงานได้	11	ไปสามครั้งไม่พบ	ผู้ตอบสัมภาษณ์	21
	รื้อถอน ไฟไหม้	12	ไม่ให้ความร่วมมือ		22
	เป็นบ้านว่าง	13	หาบ้านไม่พบ		23
2. ไม่เป็นครัวเรือนตัวอย่าง	แรงงานได้	14	อื่น ๆ (ระบุ)		24
(ครัวเรือนใหม่อยู่แทน					
ครัวเรือนเดิมที่เป็นตัวอย่าง)					

12. วิธีการให้สัมภาษณ์

1. ให้สัมภาษณ์กับพนักงานจ้างนับ
3. นัดหมายให้สัมภาษณ์ในภายหลัง

2. นัดหมายให้สัมภาษณ์ทางโทรศัพท์
4. ตอบแบบสอบถามด้วยตนเอง

--	--	--	--	--	--	--	--

รหัสพนักงานจ้างนับ ID_CODE1 176-182

--	--	--	--	--	--	--	--

รหัสผู้ตรวจสอบ ID_CODE2 183-189

ตอนที่ 1 ลักษณะทั่วไปของสมาชิกในครัวเรือน								
ลำดับที่	ถามทุกคนในครัวเรือน						ถามเฉพาะผู้มีอายุ 15 ปีขึ้นไป	
	ชื่อ-นามสกุล			ความเกี่ยวพันกับหัวหน้าครัวเรือน	เพศ	สัญชาติ	อายุ	สถานภาพสมรส
	คำนำหน้า	ชื่อ	นามสกุล					
	บันทึกชื่อ			บันทึกชื่อ	บันทึกชื่อ	บันทึกชื่อ	(บันทึกอายุเต็มปี ถ้าต่ำกว่า 1 ปี บันทึก "000")	บันทึกชื่อ
	นาย..... 1			หัวหน้าครัวเรือน 01	ชาย.....1	- ถ้ามีสัญชาติ		โสด..... 1
	นาง..... 2			ภรรยาหรือสามี 02	หญิง.....2	ให้บันทึก "สัญชาติ"		สมรส..... 2
	นางสาว..... 3			บุตรที่ยังไม่สมรส 03		เช่น		(จดทะเบียน)
	เด็กชาย..... 4			บุตรที่สมรสแล้ว 04		สัญชาติ ไทย	ถ้าบันทึกอายุ	สมรส..... 3
	เด็กหญิง..... 5			บุตรชาย/บุตรสาว 05		รหัส 7 6 4	● 0 -5 ปี	(ไม่จดทะเบียน หรือ
	ทหารประจำ 6			บุตรของบุตร 06		สัญชาติ จีน	ยุติการสัมภาษณ์	อยู่ในชั้นสามัญกรรยา)
	ทหารปลดประจำ 7			พ่อ/แม่ 07		รหัส 1 5 6	● 6 -14 ปี	หม้าย..... 4
				พ่อ/แม่ของคู่สมรส 08		สัญชาติ หนา	ข้ามไปถามตอนที่ 2	หย่า..... 5
				ปู่/ย่า/ตา/ยาย 09		รหัส 1 0 4	● 15 ปีขึ้นไป	แยกกันอยู่..... 6
				ลูกพี่/น้อง 10		- ถ้าไม่มีสัญชาติ	ถามต่อไป	
				พี่น้อง 11		ให้บันทึก "ไม่มีสัญชาติ"		
				ญาติอื่นๆ 12		รหัส 0 0 0		
				ผู้อาศัย 13		(ถามต่อไป)		
				คนรับใช้/ลูกจ้างในครัวเรือน..... 14				
				ลูกจ้างในธุรกิจของครัวเรือน..... 15				
				สมาชิกครัวเรือนนอกกลุ่มบุคคล..... 00				
F1	F2 1	F2 2	F2 3	F3	F4	F4 1	F5	F6
36	190	191	291	37	39	40	43	46

ตอนที่ 2 การศึกษา เฉพาะผู้มีอายุตั้งแต่ 6 ปีขึ้นไป			ตอนที่ 3 การทำงาน เฉพาะผู้มีอายุตั้งแต่ 15 ปีขึ้นไป		
"... (ชื่อ) ... ขณะนี้ กำลังเรียนชั้นใด"	"... (ชื่อ) ... จบการศึกษาสูงสุดระดับใด"	"... (ชื่อ) ... จบการศึกษาสูงสุด สาขาวิชาใด"	"ระหว่าง 7 วัน ก่อนวันสัมภาษณ์ ... (ชื่อ) ... ได้ทำงานที่ ก่อให้เกิดรายได้/ เงินเดือน/ค่าจ้าง/ ผลกำไร/ผลตอบแทน หรือได้ทำงาน ในฐานะช่วยธุรกิจ ของครัวเรือน โดยไม่ได้รับค่าจ้าง หรือไม่ "	ผู้บันทึก รหัส 2 ในสมุด F10	ผู้บันทึก รหัส 2 ในสมุด F11
บันทึก ระดับ ประเภท และ ชั้น ปี ที่กำลังเรียน	บันทึก ระดับ ประเภท และ ชั้น ปี ที่สำเร็จหลักสูตร	บันทึก สาขาวิชา คณะ หรือ แผนก ที่เรียนสำเร็จ ตามประกาศนียบัตร หรือปริญญาบัตร	บันทึก รหัส	บันทึก รหัส	บันทึก รหัส
ถ้าไม่เรียน บันทึก "ไม่เรียน"	ถ้าจบ ปวช./ประกาศนียบัตร ระดับสูงกว่ามัธยมศึกษาตอนปลาย แต่ต่ำกว่าอนุปริญญา หรืออนุปริญญา ปวส. และปริญญาตรีขึ้นไป (ตามต่อไป)		บันทึก รหัส	บันทึก รหัส	บันทึก รหัส
	ถ้าไม่เคยเรียน บันทึก "ไม่มีการศึกษา"		ทำงาน.....1 (ข้ามไปถามอาชีพ สดม F23)	ได้รับ.....1 (ข้ามไปถามอาชีพ สดม F23)	มี.....1 (ตามต่อไป)
			ไม่ทำงาน.....2 (ตามต่อไป)	ไม่ได้รับ.....2 (ตามต่อไป)	ไม่มี.....2 (ข้ามไปถาม สดม F14)
F7	F8	F9	F10	F11	F12

ตอนที่ 3 การทำงาน ตามเฉพาะผู้มีอายุตั้งแต่ 15 ปีขึ้นไป (ต่อ)					
ผู้บันทึก รหัส 1 ในสมุด F12	ผู้บันทึก รหัส 2 ในสมุด F12	ผู้ทำงานระหว่าง 7 วัน ก่อนวันสัมภาษณ์ รหัส 1 ในสมุด F14	ผู้บันทึก รหัส 2 หรือ 3 ในสมุด F14	ผู้ไม่พร้อมทำงาน รหัส 2 ในสมุด F16	ผู้พร้อมทำงาน รหัส 1 ในสมุด F16
"... ชื่อ ... หยุดทำงาน มานานเท่าใด"	"ก่อนวันสัมภาษณ์ ... (ชื่อ) ... ได้ทำงานทำ/ สมัครงาน บ้างหรือไม่ "	"... (ชื่อ) ...ได้ทำงานทำ/ สมัครงาน อย่างไร " (บันทึกวิธีการที่ใช้มากที่สุด) บันทึกรหัส อ่านหนังสือพิมพ์/นิตยสาร..... 1 วิทยุ/โทรทัศน์..... 2 เพื่อน/ญาติหางานให้..... 3 หาสมัครงานที่หน่วยงาน..... 4 บันทึกรหัส หาใน 7 วัน.....1 (ถามต่อไป) หาใน 8-30 วัน.....2 ไม่ทำงานทำ.....3 (บันทึกรหัส 2,3 ข้ามไปถาม สมุด F16)	" ระหว่าง 7 วัน ก่อนวันสัมภาษณ์ ... (ชื่อ) ...พร้อม ทำงานหรือไม่"	" เหตุผลที่... (ชื่อ) ... ไม่พร้อมทำงาน" บันทึกรหัส ทำงานบ้าน..... 1 เรียนหนังสือ..... 2 รอฤดูกาล..... 3 ยังเด็กหรือชรา..... 4 ป่วย พักการ ฯลฯ..... 5 จนไม่สามารถทำงานได้..... 5 พักผ่อน..... 6 เกษียณการทำงาน..... 7 ดูแลเด็ก/ผู้สูงอายุ/ผู้ป่วย/พิการ 8 อื่นๆ (ระบุ)..... 9 (ยุติการสัมภาษณ์)	"ทำไม... (ชื่อ) ... จึงไม่ทำงานทำ" บันทึกรหัส เชื่อว่าทำงานไม่ได้..... 1 หาแล้วแต่หาไม่ได้..... 2 ไม่ทราบว่าจะไปทำงานอย่างไร..... 3 ไม่สามารถหางานที่เหมาะสมได้..... 4 คืนฟ้าอากาศไม่เอื้ออำนวย..... 5 รอเริ่มงาน/เริ่มธุรกิจใหม่/รอพึงผล/ รอเรียกทำงาน..... 6 กำลังเรียน/อบรม เพื่อเพิ่มทักษะการทำงาน..... 7 อื่นๆ (ระบุ)..... 8 (ถามต่อไป)
F13	F14	F15	F16	F17	F18
☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐

58

59

60

61

62

63

ตอนที่ 3 การทำงาน ตามเฉพาะผู้มีอายุตั้งแต่ 15 ปีขึ้นไป (ต่อ)					
ผู้หางานทำ/พร้อมทำงาน ระหว่าง 7 วันก่อนวันสัมภาษณ์ (ผู้ว่างงาน) รหัส 1 ในสคมก F14 หรือ F16		ผู้เคยทำงานมาก่อน รหัส 1 ในสคมก F20		ผู้มีงานทำหรือเคยทำงานมาก่อน (รหัส 1 ในสคมก F10, F11, F12, F20)	
อาชีพ	อุตสาหกรรม				
"... (ชื่อ) ... ทำงาน/พร้อมทำงาน มานานเท่าใด " บันทึกรหัส ต่ำกว่า 1 เดือน..... 1 1 - 2.9 เดือน..... 2 3 - 5.9 เดือน..... 3 6 - 8.9 เดือน..... 4 9 - 11.9 เดือน..... 5 1 - 1.9 ปี..... 6 2 - 2.9 ปี..... 7 ตั้งแต่ 3 ปีขึ้นไป..... 8 (ถามต่อไป)	"... (ชื่อ) ... เคยทำงาน มาก่อนหรือไม่" บันทึกรหัส เคย.....1 (ถามต่อไป) ไม่เคย.....2 (ยุติการสัมภาษณ์)	"สาเหตุที่ ... (ชื่อ) ... ออกจากงาน/หยุดทำงาน ครั้งสุดท้าย " บันทึกรหัส นายจ้างเลิก/หยุด/ปิดกิจการ 1 หมดสัญญาจ้างแรงงาน... 2 ถูกให้ออก/ไล่ออก/ปดฯ 3 ลาออก..... 4 เลิก/หยุด/ปิดกิจการ..... 5 หมดสัญญาจ้างงาน..... 6 อื่นๆ ระบุ..... 7 (ถามต่อไป)	"... (ชื่อ) ... ออกจากงาน/ หยุดทำงาน ครั้งสุดท้าย มานานเท่าใด" บันทึกรหัส ต่ำกว่า 1 เดือน.....1 1 2.9 เดือน.....2 3 5.9 เดือน.....3 6 8.9 เดือน.....4 9 11.9 เดือน.....5 1 1.9 ปี.....6 2 2.9 ปี.....7 ตั้งแต่ 3 ปีขึ้นไป.....8 (ถามต่อไป)	"... (ชื่อ) ... ทำงานอะไร" ผู้บันทึกรหัส 1 ในสคมก F10 ให้บันทึกอาชีพและตำแหน่งหน้าที่ การงานที่มีชั่วโมงทำงานสูงสุดใน 7 วันก่อนวันสัมภาษณ์ ผู้บันทึกรหัส 1 ในสคมก F11 หรือ F12 ให้บันทึกอาชีพปกติ ผู้บันทึกรหัส 1 ในสคมก F20 ให้บันทึกอาชีพที่ทำครั้งสุดท้าย (ถามต่อไป)	"กิจกรรมหลักที่... (ชื่อ) ...ทำอยู่ เป็นกิจกรรมหรือผลิตอะไร " (ให้บันทึกประเภทของกิจการ ลักษณะงานให้ชัดเจน)
F19	F20	F21	F22	F23	F24

64

65

66

67

68

72

ตอนที่ 3 การทำงาน ตามเฉพาะผู้มีอายุตั้งแต่ 15 ปีขึ้นไป (ต่อ)				
ผู้มีงานทำหรือเคยทำงานมาก่อน (รหัส 1 ในสดมภ์ F10, F11, F12, F20)	ผู้มีงานทำ/ผู้เคยทำงาน บันทึกรหัส 1, 6	ผู้มีงานทำ รหัส 1 ในสดมภ์ F10, F11, F12		
สถานภาพการทำงาน "... (ชื่อ) ...ทำงานในฐานะอะไร " บันทึกรหัส นายจ้าง..... 1 ผู้ทำงานส่วนตัวโดยไม่มีลูกจ้าง..... 2 ผู้ช่วยธุรกิจครัวเรือนโดยไม่ได้รับค่าจ้าง..... 3 ลูกจ้างรัฐบาล..... 4 ลูกจ้างรัฐวิสาหกิจ..... 5 ลูกจ้างเอกชน..... 6 สมาชิกของการรวมกลุ่ม..... 7 (ถ้าบันทึกรหัส 1 หรือ 6 ตามต่อไป บันทึกรหัส 2, 3, 4, 5, 7 ข้ามไปตาม สดมภ์ F27 ยกเว้นผู้บันทึก รหัส 1 ในสดมภ์ F20 ให้ยุติการสัมภาษณ์)	"กิจการ ... (ชื่อ) ... ทำอยู่หรือทำครั้งสุดท้าย มีคนทำงานกี่คน" บันทึกรหัส 1 - 4 คน..... 1 5 - 9 คน..... 2 10 - 19 คน..... 3 20 - 29 คน..... 4 30 - 49 คน..... 5 50 - 100 คน..... 6 101 - 199 คน..... 7 200 คนขึ้นไป..... 8 (ตามต่อไป) (ยกเว้นผู้บันทึกรหัส 1 ในสดมภ์ F20 ให้ยุติการสัมภาษณ์)	"จำนวนชั่วโมงทำงานในระหว่าง 7 วันก่อนวันสัมภาษณ์" (ผู้บันทึกรหัส 1 ในสดมภ์ F10 ให้บันทึกจำนวนชั่วโมงทำงานจริง ในระหว่าง 7 วันก่อนวันสัมภาษณ์ โดยรวมชั่วโมงทำงานช่วงเวลา คอฟฟี่เบรก ไม่รวมเวลาพักอาหารกลางวัน ผู้บันทึกรหัส 1 ในสดมภ์ F11 หรือ รหัส 1 ในสดมภ์ F12 ให้บันทึก "0" ทั้ง 3 สดมภ์)		
		จำนวนชั่วโมงทำงานจริง ของอาชีพ ที่บันทึกใน F23	จำนวนชั่วโมงทำงานจริง ของอาชีพอื่นทุกอาชีพ ที่ทำในระหว่าง 7 วัน ก่อนวันสัมภาษณ์	จำนวนชั่วโมง ทำงานทั้งสิ้น (F27+ F28)
F25	F26	F27	F28	F29

ตอนที่ 4 ความต้องการทำงานเพิ่ม ในระหว่าง 7 วัน ก่อนวันสัมภาษณ์			ตอนที่ 5 รายได้ของผู้มีงานทำ (ยกเว้นนายจ้างและผู้ช่วยธุรกิจครัวเรือนโดยไม่ได้รับค่าจ้าง)		
ผู้มีงานทำ รหัส 1 ในสศมก F10, F11, F12			ค่าจ้าง/รายได้ ผู้มีงานทำ (ยกเว้นนายจ้างและผู้ช่วยธุรกิจครัวเรือนโดยไม่ได้รับค่าจ้าง) (ผู้ที่บันทึกรหัส 1 ในสศมก F10, F11, F12 และมีสถานภาพการทำงานเป็นรหัส 2 หรือ 4 หรือ 5 หรือ 6 หรือ 7 ในสศมก F25)		
"ในระหว่าง 7 วันก่อนวันสัมภาษณ์ ... (ชื่อ) ... ยังมีเวลาเหลือพอที่จะทำงานเพิ่มและ ต้องการจะทำงานเพิ่มหรือไม่" บันทึกรหัส มีเวลาและต้องการทำงานเพิ่ม.....1 (ถามต่อไป) มี/ไม่มีและไม่ต้องการทำงานเพิ่ม.....2 (ผู้มีสถานภาพการทำงานเป็นรหัส 2, 4, 5, 6, 7 ในสศมก F25 ขึ้นไปถามสศมก F33 รหัส 1 หรือ 3 ขึ้นไปถามสศมก F40)	ผู้มีงานทำที่มีเวลา และต้องการทำงาน รหัส 1 ในสศมก				
	"... (ชื่อ) ... ต้องการทำงาน เพิ่มขึ้นสัปดาห์ละ กี่ชั่วโมง" บันทึกจำนวน เป็นชั่วโมง (ถามต่อไป)	"... (ชื่อ) ... ได้ทำ งานทำ/ ติดต่อ สมัครงาน เพิ่มบ้าง หรือไม่" บันทึกรหัส หา.....1 ไม่หา...2 (ถามต่อไป)	"งานที่ ... (ชื่อ) ... ทำอยู่ได้รับค่าจ้าง/รายได้ ในประเภทใด" บันทึกรหัส เป็นตัวเงิน :- รายชั่วโมง.....1 รายวัน.....2 รายสัปดาห์.....3 รายเดือน.....4 อื่นๆ (ระบุ).....5 ไม่ทราบ.....6 ไม่เป็นตัวเงิน.....7 (ตอบรหัส 1-3 ถามต่อไป รหัส 4-7 ขึ้นไปถาม สศมก F35)	สำหรับผู้ที่บันทึกรหัส 1, 2, 3 ในสศมก F33 ค่าจ้าง/รายได้ที่ได้รับนับได้กับบาท ต่อชั่วโมงหรือต่อวัน หรือต่อสัปดาห์" (บันทึกจำนวนขีดขวา) (ถามต่อไป)	สำหรับผู้ที่บันทึกรหัส 1-7 ในสศมก F33 ค่าจ้าง/รายได้ที่ได้รับประมาณ เดือนละเท่าไร" (บันทึกจำนวนขีดขวา) (ถามต่อไป)
F30	F31	F32	F33	F34	F35
☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐

85

86

88

89

90

95

ตอนที่ 5 รายได้ของผู้มีงานทำ (ต่อ) (ยกเว้นนายจ้างและผู้ช่วยธุรกิจครัวเรือนโดยไม่ได้รับค่าจ้าง)				"เป็นสมาชิก คนสุดท้าย ของครัวเรือน หรือไม่"	"ใครเป็นผู้ตอบ แบบสัมภาษณ์"
ผลประโยชน์ตอบแทนอื่นๆ ผู้มีงานทำ (ยกเว้นนายจ้าง) (ผู้ที่บันทึกรหัส 1 ในสแตมป์ F10, F11, F12 และมีสถานภาพการทำงานเป็นรหัส 2 หรือ 4 หรือ 5 หรือ 6 หรือ 7 ในสแตมป์ F25)					
"ระหว่าง 30 วันก่อนวันสัมภาษณ์ ... (ชื่อ) ... นอกจากค่าจ้าง ที่ได้รับแล้วยังได้รับรายได้หรือผลประโยชน์อื่นหรือไม่" (ถ้าได้รับบันทึก จำนวนเงิน ถ้าไม่ได้รับบันทึก "0")					
โบนัส	ค่าล่วงเวลาเดือนละ	เงินอื่น ๆ เดือนละ (เช่น เงินประจำตำแหน่ง เบี้ยประชุม เงินค่าอาหาร เสื้อผ้า หรือเครื่องแต่งกายอื่นๆ ค่าทิป ค่าที่อยู่อาศัย ค่าพาหนะ เป็นต้น)	ผลประโยชน์ตอบแทน ที่ไม่เป็นตัวเงิน (เช่น อาหาร เสื้อผ้าหรือ เครื่องแต่งกายอื่นๆ ที่อยู่อาศัย รถรับ-ส่ง เป็นต้น)	บันทึกรหัส ไม่เป็น.....0 เป็น.....1	บันทึกรหัส ตนเอง.....1 สมาชิกในครัวเรือนตอบแทน.....2 ไม่เป็นสมาชิกในครัวเรือนตอบแทน.....3
F36	F37	F38	F39	F40	F41

101

107

112

117

122

145

ตอนที่ 6 แรงงานนอกระบบ				
ถามเฉพาะผู้มีงานทำ (รหัส 1 ในสดมภ์ F10, F11, F12) ที่มีอายุ 15 ปีขึ้นไป				
ผู้บันทึก รหัส 4 ในสดมภ์ F25	ผู้บันทึกรหัส 1 หรือรหัส 6 ในสดมภ์ F25	ผู้บันทึกรหัส 4 ในสดมภ์ F47 หรือรหัส 6 ในสดมภ์ F25	ผู้บันทึกรหัส 4 ในสดมภ์ F47 หรือ รหัส 1,2,3,6,7 ในสดมภ์ F25	ผู้บันทึกรหัส 5 ในสดมภ์ F50 และรหัส 6 ในสดมภ์ F25
"... (ชื่อ) ... ทำงานในฐานะ" บันทึกให้ ข้าราชการ/ลูกจ้างประจำ..... 1 พนักงานราชการ..... 2 ลูกจ้างของรัฐบาลต่างประเทศ/..... 3 องค์การระหว่างประเทศ..... 3 (ข้ามไปถามตอนที่ 7) ลูกจ้างชั่วคราว/ผู้รับจ้างเหมา..... 4 (ข้ามไปถามสดมภ์ F49)	"... (ชื่อ) ... ได้รับความคุ้มครอง ตามกฎหมายต่อไปหรือไม่" บันทึกให้ ได้รับ..... พรบ. โรงเรียนเอกชน พ.ศ. 2550 (เป็นสมาชิกกองทุนสหเคราะห์)..... 1 (ข้ามไปถามตอนที่ 7) พรบ. คุ้มครองผู้รับงานไปทำที่บ้าน พ.ศ. 2553..... 2 กฎกระทรวงฉบับที่ 14 (พ.ศ.2555) ออกตามความใน พรบ. คุ้มครองแรงงาน พ.ศ. 2541..... 3 กฎกระทรวงคุ้มครองแรงงานใน งานเกษตรกรรม พ.ศ. 2557..... 4 ไม่ได้รับ..... 5 (ถามต่อไป)	"... (ชื่อ) ... ได้รับ สิทธิการลาป่วย/ลาพักผ่อน ประจำปี โดยไม่หักค่าจ้างหรือไม่" บันทึกให้ ได้รับ..... 1 ไม่ได้รับ..... 2 (ถามต่อไป)	"... (ชื่อ) ... เป็น ผู้ประกันตนในระบบประกัน สังคมของสำนักงาน ประกันสังคมหรือไม่" บันทึกให้ เป็น..... มาตรา 33..... 1 มาตรา 39..... 2 มาตรา 40..... 3 ไม่ทราบมาตรา..... 4 (ข้ามไปถามตอนที่ 7) ไม่เป็น..... 5 (ถามต่อไป)	"ในระยะเวลา 3 เดือน ก่อนวันสัมภาษณ์ ... (ชื่อ) ... ทำงานอย่างต่อเนื่องกับ นายจ้างคนเดิมโดยได้รับ เงินเดือน/ค่าจ้าง ตามที่ได้ ตกลงกันไว้อย่างสม่ำเสมอ หรือไม่" บันทึกให้ ใช่..... 1 ไม่ใช่..... 2 (ถามต่อไป)
F47	F48	F49	F50	F51
☐	☐	☐	☐	☐
☐	☐	☐	☐	☐
☐	☐	☐	☐	☐
☐	☐	☐	☐	☐
☐	☐	☐	☐	☐
☐	☐	☐	☐	☐

ตอนที่ 7 สถานที่ทำงาน/ลักษณะงาน				
ถามเฉพาะผู้มีงานทำ (รหัส 1 ในสดมภ์ F10, F11, F12) ที่มีอายุ 15 ปีขึ้นไป				
"ประเภทของสถานที่ทำงานที่ ...(ชื่อ)...ทำงานอยู่"	ผู้บันทึกรหัส 1, 2, 3, 6, 7 ในสดมภ์ F25			
	"สถานประกอบการที่ ...(ชื่อ)...ทำงาน จัดเป็นรูปแบบใด"	"สถานประกอบการ ที่...(ชื่อ)...ทำงาน ได้จดทะเบียนกับ หน่วยงานราชการ หรือไม่"	"สถานประกอบการ ที่...(ชื่อ)...ทำงาน ได้จัดทำบัญชี หรือไม่"	
บันทึกรหัส	บันทึกรหัส	บันทึกรหัส	บันทึกรหัส	บันทึกรหัส
สถานประกอบการ/สำนักงาน/ร้านค้าของตนเอง แต่ไม่ใช่ที่อยู่อาศัย..... 1				
สถานประกอบการ/สำนักงาน/ร้านค้าของนายจ้าง แต่ไม่ใช่ที่อยู่อาศัยของนายจ้าง..... 2	บริษัท/หจก./หสน..... 1	จด..... 1	ทำ..... 1	
แผงลอยข้างถนน/ในตลาด..... 3	ส่วนบุคคล..... 2	ไม่จด..... 2	ไม่ทำ..... 2	
สถานที่ก่อสร้าง..... 4	กลุ่มแม่บ้าน..... 3	ไม่ทราบ..... 3	ไม่ทราบ..... 3	
ที่อยู่อาศัยหรือบริเวณที่อยู่อาศัยของตนเอง..... 5	อื่นๆ (ระบุ)..... 4			
ที่อยู่อาศัยของนายจ้าง..... 6	(ถามต่อไป)	(ถามต่อไป)	(ถามต่อไป)	
ที่ทำงานไม่เป็นหลักแหล่ง..... 7				
พื้นที่เพาะปลูก..... 8				
อื่นๆ (ระบุ)..... 9				
(ถามต่อไป)				
F52	F53	F54	F55	
☐	☐	☐	☐	
☐	☐	☐	☐	
☐	☐	☐	☐	
☐	☐	☐	☐	
☐	☐	☐	☐	
☐	☐	☐	☐	
☐	☐	☐	☐	

167

168

169

170

ตอนที่ 8 ปัญหาจากการทำงานและการพัฒนาทักษะ			
ถามเฉพาะผู้มีงานทำ (รหัส 1 ในสมุด F10, F11, F12) ที่มีอายุ 15 ปีขึ้นไป			
... (ชื่อ) ... มีปัญหาในเรื่องเหล่านี้บ้างหรือไม่			...(ชื่อ)...ต้องการอบรมหรือพัฒนาทักษะในด้านใดมากที่สุด"
"จากการทำงาน"	"จากสภาพแวดล้อมในการทำงาน"	"จากความปลอดภัยในการทำงาน"	
บันทึกรหัส	บันทึกรหัส	บันทึกรหัส	บันทึกรหัส
ไม่มีปัญหา..... 0	ไม่มีปัญหา..... 0	ไม่มีปัญหา..... 0	ไม่ต้องการ..... 0
มีปัญหา:-	มีปัญหา:-	มีปัญหา:-	ต้องการ:-
คำตอบแทน..... 1	สถานที่ทำงาน	สารเคมี..... 1	อุตสาหกรรม..... 1
งานหนัก..... 2	คับแคบ..... 1	เครื่องจักร/เครื่องมือที่เป็นอันตราย.. 2	คหกรรม..... 2
ทำงานไม่ตรงเวลาปกติ..... 3	สกปรก..... 2	อันตรายต่อหู /ตา..... 3	ศิลปหัตถกรรม..... 3
งานขาดความต่อเนื่อง..... 4	อากาศไม่ถ่ายเท..... 3	ทำงานในที่สูง/ใต้น้ำ/ใต้ดิน..... 4	การเกษตรกรรม..... 4
ชั่วโมงทำงานมากเกินไป..... 5	อิริยาบถในการทำงาน..... 4	ความไม่สงบ / การก่อการร้าย..... 5	ภาษาต่างประเทศ..... 5
ไม่มีวันหยุด..... 6	ฝุ่นละออง คิว้น กลิ่น..... 5	อื่นๆ (ระบุ)..... 6	คอมพิวเตอร์และเทคโนโลยี..... 6
ไม่สามารถลาหยุดหรือ	เสียงดัง..... 6	(ถามต่อไป)	การบริหารจัดการ และบริการ..... 7
ลาพักผ่อนได้..... 7	แสงสว่าง..... 7		การท่องเที่ยว การโรงแรม และภัตตาคาร..... 8
ไม่มีสวัสดิการ..... 8	อื่นๆ (ระบุ)..... 8		อื่นๆ (ระบุ)..... 9
(ถามต่อไป)	(ถามต่อไป)		(จบการสัมภาษณ์)
F56	F57	F58	F59
☐	☐	☐	☐
☐	☐	☐	☐
☐	☐	☐	☐
☐	☐	☐	☐
☐	☐	☐	☐
☐	☐	☐	☐

171

172

173

174

ประวัติผู้เขียน

ชื่อ	นายภาณุ ชูช่วง
ชื่อการค้นคว้าอิสระ	ตัวแบบการจำแนกการประกอบอาชีพนอกระบบของแรงงานในประเทศไทย
สาขาวิชา	สถิติประยุกต์และวิทยาการวิเคราะห์ข้อมูล คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ประวัติ	เกิดเมื่อวันที่ 15 สิงหาคม พ.ศ. 2539 สำเร็จการศึกษาระดับปริญญาตรี เศรษฐศาสตรบัณฑิต มหาวิทยาลัยสงขลานครินทร์ ประวัติการทำงาน พ.ศ. 2565-2566 นักประเมินราคาทรัพย์สิน กรมธนารักษ์ พ.ศ. 2566–ปัจจุบัน นักวิชาการสถิติ สำนักงานสถิติแห่งชาติ