

ห้องสมุดงานวิจัย สำนักงานคณะกรรมการการวิจัยแห่งชาติ



236012



การตรวจับธุรกรรมทุจริตทางบัตรเครดิต จากรายการชำระสินค้า
พาณิชย์อิเล็กทรอนิกส์ โดยการทำเหมืองข้อมูล
กรณีศึกษานาการพาณิชย์แห่งหนึ่ง

โดย
ปิยะ ปานทอง

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิชาการบริหารเทคโนโลยี
วิทยาลัยนวัตกรรม มหาวิทยาลัยธรรมศาสตร์
พ.ศ. 2553



การตรวจจับธุรกรรมทุจริตทางบัตรเครดิต จากรายการชำระสินค้า
พาณิชย์อิเล็กทรอนิกส์ โดยการทำเหมืองข้อมูล
กรณีศึกษานาครพาณิชย์แห่งหนึ่ง



โดย

ปิยะ ปานทอง

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิชาการบริหารเทคโนโลยี
วิทยาลัยนวัตกรรมการบริหารเทคโนโลยี
มหาวิทยาลัยนวัตกรรม มหาวิทยาลัยธรรมศาสตร์
พ.ศ. 2553

การตรวจจับธุรกรรมทุจริตทางบัตรเครดิต จากรายการชำระสินค้า
พาณิชย์อิเล็กทรอนิกส์ โดยการทำเหมืองข้อมูล
กรณีศึกษานาการพาณิชย์แห่งหนึ่ง

โดย

นายปิยะ ปานทอง

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิชาการบริหารเทคโนโลยี
วิทยาลัยนวัตกรรม มหาวิทยาลัยธรรมศาสตร์
พ.ศ. 2553

Credit Card Fraud Detection for E-commerce by Using Data Mining;
A Case Study in Commercial Bank

By

Mr.Piya Panthong

An Independent Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science

Technology Management

College of Innovation

Thammasat University

2010

มหาวิทยาลัยธรรมศาสตร์
วิทยาลัยนวัตกรรม

การค้นคว้าอิสระ

ของ

นายปิยะ ปานทอง

เรื่อง

การตรวจจับธุรกรรมทุจริตทางบัตรเครดิต จากรายการชำระสินค้าพาณิชย์อิเล็กทรอนิกส์
โดยการทำเหมืองข้อมูล กรณีศึกษานาคราพาณิชย์แห่งหนึ่ง

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต

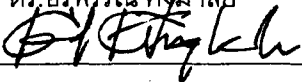
เมื่อ วันที่ 20 เมษายน พ.ศ.2553

ประธานกรรมการสอบการค้นคว้าอิสระ



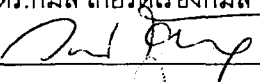
(ดร.อรพรรณ คงมาลัย)

กรรมการและอาจารย์ที่ปรึกษาการค้นคว้าอิสระ



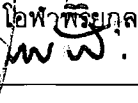
(ดร.กมล เกียรติเรืองกมลลา)

กรรมการสอบการค้นคว้าอิสระ



(ดร.กมลพรพรรณ โอฟ้าทรัพย์กุล)

คณบดี



(รองศาสตราจารย์ ม.ร.ว.พงษ์สวัสดิ์ สวัสดิวัตน์)

บทคัดย่อ

การวิจัยเรื่อง “การตรวจจับธุรกรรมทุจริตทางบัตรเครดิต จากรายการชำระสินค้าพาณิชย์อิเล็กทรอนิกส์ โดยการทำเหมืองข้อมูล กรณีศึกษาธนาคารพาณิชย์แห่งหนึ่ง” มีวัตถุประสงค์เพื่อทราบกลุ่มธุรกรรมพาณิชย์อิเล็กทรอนิกส์ที่ผิดปกติและหาความสัมพันธ์ของกลุ่มธุรกรรมผิดปกติ ด้วยวิธีการทำเหมืองข้อมูล และสร้างแบบจำลองพฤติกรรมของกลุ่มธุรกรรมผิดปกติ ใช้เป็นเกณฑ์ในการคัดกรองธุรกรรมพาณิชย์อิเล็กทรอนิกส์เพื่อป้องกันความเสียหายที่จะเกิดขึ้นจากธุรกรรมทุจริต

การวิจัยนี้ศึกษาจากข้อมูลที่ต้องคัดกรองเก็บไว้ด้วยการทำเหมืองข้อมูล ในขั้นตอนแรกนำข้อมูล ธุรกรรม พาณิชย์อิเล็กทรอนิกส์ มาหาากลุ่มธุรกรรมที่ผิดปกติ โดยการแบ่งกลุ่ม (Clustering Algorithm) ด้วยพารามิเตอร์ Scalable EM เมื่อทราบกลุ่มพฤติกรรมที่ผิดปกติ จึงนำกลุ่มพฤติกรรมที่ผิดปกติเหล่านั้นมาสร้างแบบจำลองพฤติกรรมด้วยกฎความสัมพันธ์ (Association Algorithm) ด้วยพารามิเตอร์ Apriori โดยยึดดำเนินการตามตัวแบบ CRISP-DM

ผลที่ได้จากการทำเหมืองข้อมูล คือ ลักษณะพฤติกรรมจำเพาะของกลุ่มบุคคลที่มีแนวโน้มเกิดธุรกรรมทุจริต เมื่อดูความสัมพันธ์พบว่าพฤติกรรมที่เกิดขึ้นบ่อยมีดังนี้ คือเป็นผู้ถือบัตรที่พักอาศัยอยู่ในเขตคลองเตย ยานนาวา วัฒนา หรือในจังหวัดนนทบุรี นครศรีธรรมราช นราธิวาส มีค่าความเป็นไปได้ทั้งหญิงและชายเท่ากัน มีอายุระหว่าง 34 – 42 ปี หรือ มากกว่า 53 ปีขึ้นไป ทำธุรกรรมพาณิชย์อิเล็กทรอนิกส์ด้วยยอดเงินประมาณ 12,500 – 31,500 บาทต่อครั้ง โดยทำธุรกรรมในช่วงเวลาประมาณ 09.00 – 19:00 น. กลุ่มธุรกรรมพาณิชย์อิเล็กทรอนิกส์ที่มีแนวโน้มเกิดธุรกรรมทุจริต ได้แก่ ประเภท Mail / Phone ORDER, Lodging, Other PR ทำธุรกรรมพาณิชย์อิเล็กทรอนิกส์ประเภทนี้ตั้งแต่ 18 - 80 ครั้งต่อเดือน อย่างไรก็ตามงานวิจัยนี้ยังเป็นเพียงการสร้างแบบจำลองพฤติกรรมกลุ่มธุรกรรมทุจริต ดังนั้นในการใช้งานจริงควรมีการตรวจสอบความถูกต้องและประเมินแบบจำลองก่อนการนำใช้จริง

Abstract

'Credit Card Fraud Detection for E-commerce by Using Data Mining; A Case Study in Commercial Bank' was studied for the purpose of Outlier Transaction and find relationship of behavior by using Data Mining and Make a Model behavior of Fraud Transaction as criteria for detect transactions to prevent damage caused by the fraudulent transactions.

The research had studied organization's collected data. An E-Commerce transaction was used to find Outlier transaction by Clustering Algorithm with Scalable EM parameter in the first step. After abnormal behavior was known, these were constructed for a model of behaviors by Association Algorithm with Apriori parameter. This process was done within a model of CRISP-DM.

The outcome of data mining is perception of the group behavior that is likely transaction fraud by data mining. If the relationship is focused, frequent behaviors that will be found are; the cardholders who have been living in Klongtoey, Yannawa, Wattana, Nonthaburi, Nachonsrithammarat and Narathiwat. It can be happened with male and female. In the age of 34 - 42 years old or more than 53 years old. 12,500 – 31,500 baht is spent per an e-commerce transaction. 09.00 am – 9.00 pm is a period when transactions are usually made. The Mail / Phone ORDER, Lodging and Other PR categories fraud transactions are found about 18 – 80 times per month. However this research is focus on modeling behavior of fraudulent transaction, so in real situation, the model should be checked and evaluated before using.

กิตติกรรมประกาศ

งานวิจัยเฉพาะกรณีฉบับนี้สำเร็จได้ด้วยความช่วยเหลือและร่วมมืออย่างดียิ่งจากหลายบุคคลโดยเฉพาะ ดร.กมล เกียรติเรืองกมลลา ซึ่งเป็นอาจารย์ที่ปรึกษา และให้ข้อเสนอแนะและคำแนะนำด้วยดีมาตลอด รวมถึง ดร.อรพรรณ คงมาลัย และดร.กมลพรรณ โอฬารพิริยกุล ให้เกียรติเป็นกรรมการในการสอบที่ช่วยตรวจสอบความถูกต้อง พร้อมทั้งข้อเสนอแนะที่เป็นประโยชน์ต่องานวิจัย

ขอขอบคุณพนักงานบริษัทกรณีศึกษา ที่เอื้อเฟื้อข้อมูลอันเป็นประโยชน์ต่องานวิจัย และอำนวยความสะดวกในงานวิจัยชิ้นนี้

ขอขอบพระคุณคุณเก๋ ที่ได้ให้คำปรึกษาในเรื่องโปรแกรม ที่ใช้ในการทำเหมืองข้อมูล สุดท้ายนี้ขอขอบพระคุณแม่ และครอบครัวที่อบอุ่น ที่เป็นกำลังใจในการทำงานวิจัยชิ้นนี้

นายปิยะ ปานทอง
มหาวิทยาลัยธรรมศาสตร์
พ.ศ. 2553

สารบัญ

	หน้า
บทคัดย่อ	(1)
กิตติกรรมประกาศ	(3)
สารบัญตาราง	(7)
สารบัญภาพประกอบ	(8)
บทที่	
1. บทนำ.....	1
ความสำคัญของปัญหา.....	1
วัตถุประสงค์และขอบเขตของการวิจัย	2
ระเบียบวิธีวิจัย	2
ขอบเขตงานวิจัย (Scope).....	3
ประโยชน์ที่คาดว่าจะได้รับ	3
ระยะเวลาในการวิจัย (schedule)	4
นิยามศัพท์ นิยามตัวแปร (Definition of terms & variables).....	7
2. กรอบแนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	9
กรอบแนวคิด และ ทฤษฎีที่เกี่ยวข้อง	9
ปัญญาประดิษฐ์ (Artificial Intelligence)	9
การทำเหมืองข้อมูล (Data mining).....	12
กระบวนการเก็บข้อมูลแบบ CRISP-DM	15

หลักการและความรู้ทั่วไปเกี่ยวกับธุรกิจบัตรเครดิต	20
พฤติกรรมผู้บริโภคออนไลน์ (Online Customer Behavior)	23
เอกสารและงานวิจัยที่เกี่ยวข้อง.....	26
งานวิจัยเรื่อง "Credit card fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning"	26
งานวิจัยเรื่อง "Real-time credit card fraud detection using computational intelligence"	26
งานวิจัยเรื่อง "Association rules applied to credit card fraud detection"	26
งานวิจัยเรื่อง "Neural Data Mining for Credit Card Fraud Detection"	26
3. ระเบียบวิธีวิจัย.....	28
รูปแบบงานวิจัย.....	28
ประชากรที่ศึกษาและกลุ่มตัวอย่าง.....	28
เครื่องมือที่ใช้ในการวิจัย.....	29
การทดสอบเครื่องมือที่ใช้ในการวิจัย.....	30
ระยะเวลาในการวิจัย	30
การรวบรวมข้อมูล.....	30
4. ผลการวิจัย	37
ขั้นตอนที่ 1 การเข้าใจในจุดประสงค์ของธุรกิจ.....	38
ขั้นตอนที่ 2 การทำความเข้าใจกับแหล่งที่มาและการเก็บข้อมูล	38
ขั้นตอนที่ 3 การเตรียมข้อมูล	45
ขั้นตอนที่ 4 การสร้างแบบจำลองแล้วเลือกเทคนิคที่เหมาะสม.....	48
ขั้นตอนที่ 5.การตรวจสอบและประเมินผล	73
ขั้นตอนที่ 6 ขั้นตอน การนำไปใช้งาน.....	74

5. สรุปผลการศึกษาและข้อเสนอแนะ.....	78
สรุปผลการศึกษา	79
ข้อจำกัดและอุปสรรคในการทำวิจัย.....	82
ข้อเสนอแนะในการทำวิจัย	82
ข้อเสนอแนะสำหรับงานวิจัยต่อเนื่อง	83
บรรณานุกรม	84
ประวัติการศึกษา.....	86

สารบัญตาราง

ตารางที่		หน้า
1.1	ระยะเวลาในการทำวิจัย	4
2.1	แสดงอัลกอริทึมต่างที่ใช้ในการทำเหมืองข้อมูลในแต่ละแขนง	15
3.1	แสดงข้อมูลขั้นตอนจากแหล่งข้อมูลที่ 1	33
3.2	แสดงข้อมูลขั้นตอนจากแหล่งข้อมูลที่ 2	34
4.1	ลักษณะของข้อมูลที่ใช้ในการทำเหมืองข้อมูล	40
4.2	ลักษณะข้อมูลของผู้ถือบัตรที่ใช้ในการทำเหมืองข้อมูล	41
4.3	ลักษณะคำอธิบายธุรกรรมล้มเหลว	41
4.4	ลักษณะข้อมูลสถานะความปลอดภัยของธุรกรรม	43
4.5	ลักษณะข้อมูลคำอธิบายธุรกรรมล้มเหลว	44
4.6	ลักษณะข้อมูลธุรกรรมที่ใช้ในการวิเคราะห์	46
4.7	ลักษณะข้อมูลของผู้ถือบัตรที่นำมาวิเคราะห์	47
4.8	ลักษณะข้อมูลไม่นำมาวิเคราะห์	48
4.9	เขตข้อมูลของตารางข้อมูลลูกค้า สำหรับวิธีการแบ่งกลุ่ม	50
4.10	จำนวนกลุ่มข้อมูล ค่าเฉลี่ย และ ความถี่ในการใช้งาน	52
4.11	แสดงการกระจายตัวของข้อมูลจำนวนเงินที่ทำธุรกรรม	53
4.12	แสดงการกระจายตัวของข้อมูลความถี่ในการใช้งาน	53
4.13	แสดงการกระจายตัวของข้อมูลรหัสประเภทร้านค้า	54
4.14	แสดงการกระจายตัวของข้อมูลเวลาที่ทำธุรกรรม	55
4.15	ค่าสถิติของข้อมูลกลุ่มที่ 5	55
4.16	ตัวอย่างข้อมูลที่นำมาวิเคราะห์หาความสัมพันธ์ของข้อมูล	57
4.17	แสดงกลุ่มความสัมพันธ์ของข้อมูลจากกลุ่มธุรกรรมทุจริต	57
4.18	ลักษณะข้อมูลกฎความสัมพันธ์	60
4.19	แสดงระยะห่างของแต่ละกลุ่ม	74

สารบัญภาพประกอบ

ภาพที่	หน้า
2.1 แสดงวิวัฒนาการของการทำเหมืองข้อมูลจากอดีตจนถึงปัจจุบัน.....	13
2.2 แสดงความสัมพันธ์แบบจำลองและกระบวนการเก็บข้อมูล ตามตัวแบบ CRISP-D	16
2.3 แสดงขั้นตอนการเก็บข้อมูลตามกระบวนการตามตัวแบบ CRISP-DM	17
2.4 แสดงรายละเอียดในแต่ละขั้นตอนการเก็บข้อมูลตามตัวแบบ CRISP-DM.....	19
2.5 แสดงโครงสร้างองค์กรของธนาคารพาณิชย์แห่งหนึ่ง	20
3.1 แสดงขั้นตอนการเก็บข้อมูลตามกระบวนการ CRISP-DM	32
4.1 ร้อยละของกลุ่มตัวอย่างจำแนกตามประเภทธุรกรรม.....	42
4.2 ร้อยละของกลุ่มตัวอย่างจำแนกตามความผิดปกติของธุรกรรม	43
4.3 แสดงประเภทธุรกรรมที่เกิดความผิดพลาด.....	45
4.4 แสดงความสัมพันธ์ของแต่ละกลุ่ม.....	51
4.5 แสดงจำนวนข้อมูลในแต่ละกลุ่ม	52
4.6 แสดงข้อมูลร้อยละของกลุ่มธุรกรรมผิดปกติ (Outlier)	56
4.7 แสดงความสัมพันธ์ของข้อมูลกลุ่ม Outlier.....	59
4.8 การประเมินแบบจำลอง	60
4.9 ลักษณะพฤติกรรมที่ 1 ที่เข้าข่ายทุจริต.....	71
4.10 ลักษณะพฤติกรรมที่ 2 ที่เข้าข่ายทุจริต.....	73
4.11 ลักษณะพฤติกรรมที่ 3 ที่เข้าข่ายทุจริต.....	73
4.12 ลักษณะพฤติกรรมที่ 4 ที่เข้าข่ายทุจริต.....	73
4.13 ลักษณะพฤติกรรมที่ 5 ที่เข้าข่ายทุจริต.....	73
4.14 ลักษณะพฤติกรรมที่ 6 ที่เข้าข่ายทุจริต.....	74
4.15 ผังงานแสดงกระบวนการตรวจจับธุรกรรมทุจริตของธนาคารพาณิชย์แห่งหนึ่ง.	78
5.1 แสดงลักษณะพฤติกรรมทุจริต	82