

จิรกฤต บุญหมื่น ไวย : การศึกษาเปรียบเทียบเทคนิคการเรียนรู้ของเครื่อง  
สำหรับข้อมูลที่มีปัญหาข้อมูลไม่สมดุล (A COMPARATIVE STUDY OF MACHINE  
LEARNING TECHNIQUES TO DEAL WITH UNBALANCED DATA).  
อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร.เบญจวรรณ โรจนดิษฐ์, 64 หน้า.

การพยากรณ์การขอยกเลิกใช้บริการ/ปัญหาความไม่สมดุล/ต้นไม้การตัดสินใจ/  
ตัวแบบนาอูฟเบย์/ซัพพอร์ตเวกเตอร์แมชชีน

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาปัญหาความไม่สมดุลที่ส่งผลต่อการขอยกเลิกใช้บริการ  
ของลูกค้าธนาคาร โดยใช้เทคนิคการสุ่มตัวอย่างก่อนที่จะนำไปสร้างตัวแบบเพื่อใช้ในการพยากรณ์  
แนวโน้มของลูกค้าที่จะขอยกเลิกใช้บริการ ในการศึกษาครั้งนี้พิจารณาการชักตัวอย่างแบบวิธี  
สังเคราะห์ข้อมูลใหม่ และการชักตัวอย่างลดย่างสุ่ม รวมถึงศึกษาการสร้างตัวแบบด้วย วิธีต้นไม้  
ตัดสินใจ วิธีจำแนกประเภทแบบนาอูฟเบย์ และวิธีซัพพอร์ตเวกเตอร์แมชชีน โปรแกรมหลักที่ใช้  
ในการดำเนินการวิจัยได้แก่ RapidMiner Studio รุ่น 9.6

จากผลการศึกษาพบว่าการใช้การสุ่มตัวอย่างแบบวิธีสังเคราะห์ข้อมูลใหม่ร่วมกับการสร้าง  
ตัวแบบด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน ให้ประสิทธิภาพในการทำนายสูงที่สุด โดยมีค่าความ  
แม่นยำร้อยละ 92.99 ค่าวัดประสิทธิภาพร้อยละ 91.37 ค่า AUC ร้อยละ 96.4 และ อัตราลบเท็จร้อยละ  
ละ 7.01

สาขาวิชาคณิตศาสตร์

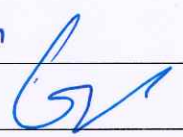
ปีการศึกษา 2562

ลายมือชื่อนักศึกษา

ลายมือชื่ออาจารย์ที่ปรึกษา

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

จิรกฤต



J. Tanthana

JIRAKIT BOONMUNEWAI : A COMPARATIVE STUDY OF  
MACHINE LEARNING TECHNIQUES TO DEAL WITH  
UNBALANCED DATA. THESIS ADVISOR : ASST. PROF.  
BENJAWAN RODJANADID, Ph.D. 64 PP.

CHURN PREDICTION/IMBALANCED PROBLEM/  
DECISION TREE/NAÏVE BAYES/SUPPORT VECTOR MACHINE

The purpose of this research was to study an imbalance problem which affects to the churn prediction for bank customers. In this study, we considered two sampling techniques, synthetic minority over-sampling technique (SMOTE) and random under-sampling, and three prediction modelings, i.e. decision tree classifier, Naïve Bayes classifier and support vector machine classifier. All calculations in the thesis were done by Rapid Miner Studio software version 9.6.

The study showed that the support vector machine classification model with SMOTE sampling technique performed most efficiently, with recall 92.99%, F-score 91.37%, area under curve (AUC) 96.4% and false negative rate 7.01%.

School of Mathematics

Academic Year 2019

Student's Signature Jirakit Boonmuneai

Advisor's Signature Benjawan Rodjanadid

Co-Advisor's Signature J.Tanthand