

ผลกระทบของการทำข้อมูลให้เป็นปกติต่อประสิทธิภาพขั้นตอนวิธีการเรียนรู้ของตัวจำแนกประเภทข้อมูล

The Impact of Data Normalization to the Learning Algorithm Efficiency of Data Classifiers

ชนาธิป หมั่นเพียรสุข<sup>1</sup> สุพจน์ เสงพระพรหม<sup>2</sup> ไกรุ่ง เสงพระพรหม<sup>3</sup> และ ดช ธรรมศิริ<sup>4</sup>

<sup>1,2,3</sup>สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม

85 ถ.มาลัยแมน ต.นครปฐม อ.เมืองนครปฐม จ.นครปฐม 73000

E-mail: <sup>1</sup>kan\_kan.32@hotmail.com, <sup>2</sup>supojn@yahoo.com, <sup>3</sup>kairung2011.heng@gmail.com

<sup>4</sup>สาขาวิชาคอมพิวเตอร์ธุรกิจ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏนครปฐม

85 ถ.มาลัยแมน ต.นครปฐม อ.เมืองนครปฐม จ.นครปฐม 73000

E-mail: <sup>4</sup>dechit@msn.com

<sup>1,2,3,4</sup>หน่วยวิจัยอรรถิยภาพแห่งเครื่องจักร สถาบันวิจัยและพัฒนา มหาวิทยาลัยราชภัฏนครปฐม

85 ถ.มาลัยแมน ต.นครปฐม อ.เมืองนครปฐม จ.นครปฐม 73000 โทรศัพท์: 034-109300 ต่อ 3156

#### บทคัดย่อ

งานวิจัยนี้ได้ทำการศึกษาผลกระทบของการทำข้อมูลให้เป็นปกติที่มีต่อประสิทธิภาพของการจำแนกประเภทข้อมูล โดยการศึกษาได้ทำการทดลองกับวิธีการทำข้อมูลให้เป็นปกติ 3 วิธีพื้นฐาน ได้แก่ วิธีน้อย-มาก วิธีคะแนนเฉลี่ย และวิธีปรับทศนิยม เปรียบประสิทธิภาพกับขั้นตอนวิธีการจำแนกประเภท 4 วิธีหลัก ได้แก่ ต้นไม้ตัดสินใจ เครื่องข่ายประสาทเทียม แบบอย่างง่าย และ เพื่อนบ้านใกล้ที่สุดเค โดยทำการทดสอบกับข้อมูลเกณฑ์มาตรฐานจำนวน 4 ชุดข้อมูล ผลการทดลองพบว่า การทำข้อมูลให้เป็นปกติ สามารถเพิ่มประสิทธิภาพการเรียนรู้ของตัวจำแนกประเภทข้อมูลได้ โดยเฉพาะ วิธีน้อย-มาก สามารถเพิ่มประสิทธิภาพให้กับวิธีการเรียนรู้แบบง่าย คือ การเรียนรู้แบบแบบอย่างง่าย และเพื่อนบ้านใกล้ที่สุดเค ซึ่งให้ผลความถูกต้องของการจำแนกประเภทที่ดีขึ้น

คำสำคัญ: การทำข้อมูลให้เป็นปกติ, การจำแนกประเภทข้อมูล, เหมืองข้อมูล, ขั้นตอนวิธีการเรียนรู้

#### Abstract

This paper presents the impact of data normalization to the learning algorithm efficiency of data classifiers. Three basic methods of data normalization including min-max, z-score and decimal scaling are used to test the efficiency of classification algorithms. The classification techniques used in this research such as decision tree, artificial neural network, naïve bayes and k-nearest neighbor are tested with four benchmark datasets. The results show that data normalization especially min-max method can improve the efficiency of the lazy learner like naïve bayes and KNN to achieve better result in term of the classification accuracy.

Keywords: Normalization, Classification, Data Mining, Learning Algorithm

