ชื่อเรื่อง : การศึกษาปัจจัยที่มีผลต่อรายได้ของประชากรในประเทศไทย โดยใช้เทคนิคการทำเหมืองข้อมูล

ผู้วิจัย : สิริธร เจริญรัตน์ สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี
ชฎารัตน์ พิพัฒนนันท์ สาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร คณะวิทยาศาสตร์และเทคโนโลยี
ศศิธร มงคลศรีพัฒนา ศูนย์วิจัยมหาวิทยาลัยชิคาโก-มหาวิทยาลัยหอการค้าไทย

ปีที่แล้วเสร็จ : 2554 จำนวน 55 หน้า

บทคัดย่อ

248145

งานวิจัยนี้ศึกษาและวิเคราะห์ข้อมูลการสำรวจด้านการทำงานของประชากร ของสำนักงานสถิติแห่งชาติ โดยนำความรู้ทางด้านการทำเหมืองข้อมูล มาใช้เพื่อสร้างแบบจำลองในการจำแนกประเภทข้อมูล เพื่อศึกษาปัจจัยที่มีผลต่อรายได้ของประชากรในประเทศไทย โดยใช้อัลกอริทึมในต้นไม้การตัดสินใจ C4.5 มาใช้เพื่อสร้างแบบจำลองตามประเภทอุตสาหกรรม ซึ่งแบ่งออกเป็น 4 กลุ่ม คือ (1) การเกษตรและประมง (2) การผลิต (3) การค้าและบริการ และ (4) อื่น ๆ จากนั้นทำการสร้างแบบจำลองจากข้อมูลฝึก โดยทดลองกับค่าความเชื่อมั่น 4 ค่า คือ 0.01, 0.1, 0.25, และ 0.5 ตามลำดับ เพื่อหาค่า Confidence Factor ที่เหมาะสม ผลลัพธ์ที่ได้คือแบบจำลองต้นไม้การตัดสินใจ และทำการทดสอบประสิทธิภาพของแบบจำลองโดยใช้ข้อมูลทดสอบ ซึ่งจะเลือกชุดทดสอบโดยพิจารณาจากค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และขนาดของต้นไม้ เป็นต้น ผลการวิจัยพบว่าปัจจัยที่มีผลต่อรายได้ของประชากรคือ ประเภทการจ้างงาน อาชีพ ระดับการศึกษา อายุ จำนวนชั่วโมงการทำงาน การมีรายได้อื่นๆ นอกเหนือจากรายได้ประจำ ผลที่ได้จากการวิเคราะห์พบว่าอัลกอริทึมในการสร้างต้นไม้การตัดสินใจ C4.5 สามารถนำมาใช้ในการสร้างจำลองทางสังคมศาสตร์ได้อย่างดี

คำสำคัญ : ต้นไม้ตัดสินใจ การทำเหมืองข้อมูล การจำแนกประเภทข้อมูล

Title : The study of Factors Affecting the Income of the Population in Thailand by Using Data Mining Techniques

Researcher :1) Sirithorn Jalernrat
Department of Computer Science , School of Science and Technology
2) Chadarat Phipathananunth
Department of Information and Communication Technology , School of Science and Technology
3) Sasithorn Mongkolsripattana
The University of Chicago University of the Thai Chamber of Commerce Research Center

Year of Accomplishment: 2554

No. of Pages : 55 Pages

Abstract

248145

This research studies and analyses the data from the Labor Forces Survey 2009 of the Office of National Statistics, Thailand. We use a data mining technique to create models for data classification to study the factors affecting income of the Thai population. The decision tree C4.5 algorithm is used to build models in WEKA. We create the models using training data of population in four industries - agriculture and fisheries, manufacturing, trade and services, and other industries. The models are experimented with four confidence factors (CF) i.e. 0.01, 0.1, 0.25, and 0.5 respectively to determine the appropriate CF. We test the performance of the models using test data selected based on the correction, precision, recall and the size of the tree, etc. Results indicate that factors affecting the income of the Thai population are (1) types of employment, (2) occupations, (3) education levels, (4) ages, (5) numbers of hours worked and (6) extra income. Result also shows that the decision tree C4.5 algorithm can be used to construct decision trees for problems in socio-economics effectively.

Keywords: decision trees, data mining, classification data