

**PAYMENT ANALYSIS OF PERSONAL INCOME TAX BY USING
DECISION TREE CLASSIFICATION TECHNIQUE**



**A RESEARCH PROJECT SUBMITTED IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(COMPUTER SCIENCE)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2008**

COPYRIGHT OF MAHIDOL UNIVERSITY

การวิเคราะห์การเสียภาษีเงินได้บุคคลธรรมดา โดยใช้เทคนิคการจัดแบ่งประเภทข้อมูล
(PAYMENT ANALYSIS OF PERSONAL INCOME TAX BY USING DECISION TREE
CLASSIFICATION TECHNIQUE)

กชรัตน์ นิลภมร 4837925 SCCS/M

วท.ม. (วิทยาการคอมพิวเตอร์)

คณะกรรมการควบคุมสารนิพนธ์ : คำรัส วงศ์สว่าง , Ph.D, ทรงศรี ตั้งศรีไพโรจน์, Ph.D

บทคัดย่อ

สารนิพนธ์นี้ เราได้พัฒนาระบบที่เรียกว่า PAPIT หรือระบบการวิเคราะห์การเสียภาษีเงินได้บุคคลธรรมดา โดยใช้เทคนิคการจัดแบ่งประเภทข้อมูล ระบบ PAPIT เป็นระบบสารสนเทศที่เป็นเว็บไซต์ และประยุกต์ใช้เทคนิคด้านการการจัดแบ่งประเภทข้อมูล PAPIT มีการสกัดข้อมูลมาจากข้อมูลที่มีอยู่เดิมของกรมสรรพากร ซึ่งข้อมูลของระบบ PAPIT นำมาจากภาษีเงินได้บุคคลธรรมดา เป็นข้อมูลรายละเอียดภาษีของผู้เสียภาษีทุกคน การตรวจสอบข้อมูลการเสียภาษีเป็นงานที่สำคัญมากงานหนึ่งของกรมสรรพากร เนื่องจากเป็นงานที่ต้องใช้เวลา แต่อย่างไรก็ตาม ข้อมูลการเสียภาษีทุกชนิดจะไม่จำเป็นต้องนำมาตรวจสอบ ในทางปฏิบัติเจ้าหน้าที่จะทำการสุ่มตรวจข้อมูลตามที่เจ้าหน้าที่จะเห็นเหมาะสมจึงเป็นเหตุให้การเลือกข้อมูลที่จะนำมาตรวจไม่ถูกต้องเหมาะสม

ระบบจะทำการวิเคราะห์ข้อมูลภาษีเงินได้บุคคลธรรมดาเพื่อเป็นแนวทางหรือแนะนำข้อมูลที่น่าสนใจเพื่อช่วยเจ้าหน้าที่ในการเลือกข้อมูลการเสียภาษีที่จะนำมาตรวจสอบ และทำการทำนายว่าผู้เสียภาษีมีแนวโน้มที่จะหลีกเลี่ยงการเสียภาษีหรือไม่ ซึ่งเรามีการนำเทคนิคการจัดแบ่งประเภท ซึ่งเป็นเทคนิคหนึ่งในการขุดค้นเหมืองข้อมูล มาช่วยในการวิเคราะห์ข้อมูลดังกล่าวข้างต้น เทคนิคการจัดแบ่งประเภทที่เราเลือกใช้ คือ C5.00 ซึ่งเหมาะสมกับข้อมูลที่มีค่าแบบต่อเนื่อง และสามารถสร้างต้นไม้ในการตัดสินใจได้เป็นอย่างดี ระบบที่เราพัฒนาขึ้นได้กำหนดเงื่อนไขต่าง ๆ เพื่อให้เจ้าหน้าที่ของกรมสรรพากรนำไปใช้ในการเลือกข้อมูลการเสียภาษีเพื่อช่วยในการตัดสินใจว่าจะตรวจสอบหรือไม่ ซึ่งการทดลองกฎที่ได้จากระบบที่นำมาให้เจ้าหน้าที่ใช้เพื่อเลือกที่จะตรวจหรือไม่ตรวจ ให้ความถูกต้องถึง 87.18 เปอร์เซ็นต์ และคุณภาพของกฎที่ได้เจ้าหน้าที่ยอมรับว่าถูกต้องอยู่ในระดับดี ซึ่งระบบจะสามารถที่จะช่วยในการทำงานของเจ้าหน้าที่ และสามารถช่วยผู้บริหารในการตัดสินใจในการวางแผนยุทธศาสตร์ในการจัดเก็บภาษี

PAYMENT ANALYSIS OF PERSONAL INCOME TAX BY USING DECISION TREE CLASSIFICATION TECHNIQUE**KOTCHARAT NILPAMORN 4837925 SCCS/M****M.Sc. (COMPUTER SCIENCE)****RESEARCH PROJECT ADVISORS: DAMRAS WONGSAWANG, Ph.D,
SONGSRI TANGSRIPAHOJ, Ph.D.****ABSTRACT**

In this research project, the payment analysis of personal income tax by using decision tree classification technique (PAPIT) is proposed. PAPIT is an information system which provides web-based, data mining applications. It is a system that sort the information from the Revenue Department (RD) collection. In PAPIT, the information is extracted from the Personal Income Tax (PIT) system which provides detailed information of taxation data for each taxpayer. Auditing tax data is one of the most important tasks of the RD. This is a time-consuming task. However, not all tax data needs to be audited. In practice, tax officers randomly select tax data for auditing based on their own decisions. This causes the improper selection of tax data.

PAPIT performs the analysis on PIT information to provide guidelines or to suggest information to help tax officers in the selection of tax data to be audited. For the development of PAPIT, the decision tree classification technique was applied to such information. Particularly, the C5.00 algorithm was the main focus since it supports data analysis of continuous values and produces a good classification tree. Having passed the experiments, PAPIT suggested rules that tax officers could use in order to determine who should be audited and who should not. PAPIT performed both functions with 87.18 % accuracy. This was confirmed by tax officers with a high level of skill in the analysis of taxes. This suggests that PAPIT can help tax officers regarding tax data selection and it can also aid executives in making decisions on taxation policies.

**KEY WORDS : PAYMENT ANALYSIS / DATA MINING / CLASSIFICATION /
C5.00 ALGORITHM / PERSONAL INCOME TAX**

111 pp.