NCCIT 2011

THE 7TH NATIONAL CONFERENCE ON COMPUTING AND INFORMATION TECHNOLOGY

FACULTY OF TECHNICAL EDUCATION
<TECHED-07>

ผศ.อัครวุฒิ ปรมะปุญญา

PROCEEDINGS OF NCCIT 2011
THE 7TH NATIONAL CONFERENCE ON COMPUTING AND INFORMATION TECHNOLOGY
11-12 MAY 2011
WWW.NCCIT.NET

FACULTY OF INFORMATION TECHNOLOGY
KING MONGKUT'S UNIVERSITY OF TECHNOLOGY NORTH BANGKOK

VOLUME 1

IEEE THAILAND SECTION

บทความวิจัย

การประชุมทางวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 7

11-12 พฤษภาคม 2554

คณะเทคโนโลยีสารสนเทศ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

VOLUME 1

THE 7TH NATIONAL CONFERENCE ON COMPUTING AND INFORMATION TECHNOLOGY

PROCEEDINGS OF NCCIT 2011

THE 7TH NATIONAL CONFERENCE ON COMPUTING AND INFORMATION TECHNOLOGY

11-12 MAY 2011

FACULTY OF INFORMATION TECHNOLOGY
KING MONGKUT'S UNIVERSITY OF TECHNOLOGY NORTH BANGKOK (KMUTNB)
BANGKOK, THAILAND

WWW. NCCIT.NET

IEEE THAILAND SECTION

บทความวิจัย

การประชุมทางวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 7

11-12 พฤษภาคม 2554

คณะเทคโนโลยีสารสนเทศ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

WWW.IT.KMUTNB.AC.TH

Table of Contents

Message from KMUTNB President	ii
Message from General Chair	iii
Conference Organizers	i
Conference Organization Committee.......................................	v
Technical Program Committee...	vi
Keynote Speakers	viii
Technical Program Contents..	xiii
Invited Papers..	xxxiv
Regular Papers...	1
Author Index	498

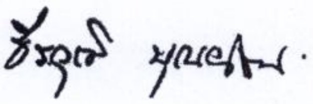
คำปรารภ

ศาสตราจารย์ ดร.ธีรวุฒิ บุณยโสภณ
อธิการบดี มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ปัจจุบันนี้เราไม่สามารถแยกเทคโนโลยีสารสนเทศและการสื่อสาร ออกจากความต้องการในชีวิตประจำวันได้ เนื่องจากมันมีความสำคัญอย่างมากต่อทุกอาชีพ เนื่องจากช่วยอำนวยความสะดวกทางธุรกิจและการพัฒนา ซึ่งไม่เหมือนกับเทคโนโลยีส่วนใหญ่ เนื่องจากเทคโนโลยีสารสนเทศและการสื่อสารเป็นเทคโนโลยีที่ใช้งานได้ทั่วไป ที่จะนำพาไปสู่พัฒนาการปรับปรุงและประสิทธิภาพทางด้านเศรษฐกิจ ดังนั้น การใช้เทคโนโลยีสารสนเทศและการสื่อสารสำหรับธุรกิจและการค้า ในการพัฒนาประเทศจึงต้องการความเข้าใจอย่างถ่องแท้

ด้วยความสำคัญและผลที่ได้รับจากเทคโนโลยีสารสนเทศข้างต้น คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือจึงได้ร่วมมือทางวิชาการจากมหาวิทยาลัยพันธมิตร ได้แก่ มหาวิทยาลัยแห่งรัฐโอกลาโฮมา ประเทศสหรัฐอเมริกา มหาวิทยาลัยเฟิร์นแห่งเมืองฮาเก้น สหพันธ์สาธารณรัฐเยอรมนี มหาวิทยาลัยเคมนิทซ์ สหพันธ์สาธารณรัฐเยอรมนี มหาวิทยาลัยอีดิธโคแวน ประเทศออสเตรเลีย มหาวิทยาลัยราชภัฏมหาสารคาม มหาวิทยาลัยมหาสารคาม มหาวิทยาลัยโยนก และมหาวิทยาลัยนอร์ทกรุงเทพ เพื่อร่วมกันจัดการประชุมทางวิชาการระดับชาติ ด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 7 (NCCIT 2011) ขึ้นระหว่างวันที่ 11-12 พฤษภาคม 2554 ณ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ โดยมีวัตถุประสงค์เพื่อเผยแพร่และแลกเปลี่ยนข่าวสาร วิทยาการ ความรู้ ทักษะและประสบการณ์ จากการค้นคว้าวิจัยและพัฒนาวิชาการด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศของคณาจารย์ นักศึกษา และประชาชนทั่วไป เพื่อนำเทคโนโลยีด้านคอมพิวเตอร์และสารสนเทศ ไปใช้ประโยชน์อย่างจริงจัง

ขอขอบคุณทุกหน่วยงานและคณะบุคคล ที่มีส่วนสนับสนุนในการจัดการประชุมทางวิชาการในครั้งนี้ ขออวยพรให้การประชุมทางวิชาการ NCCIT 2011 ครั้งนี้ ประสบความสำเร็จบรรลุวัตถุประสงค์ทุกประการ



Conference Organizers

King Mongkut's University of Technology
North Bangkok, Thailand
www.kmutnb.ac.th

Oklahoma State University, USA
www.okstate.edu

FernUniversitaet in Hagen, Germany
www.fernuni-hagen.de

Chemnitz University, Germany
www.tu-chemnitz.de

Edith Cowan University, Australia
www.ecu.edu.au

Rajabhat Mahasarakham University
www.rmu.ac.th

Mahasarakham University
www.msu.ac.th

North Bangkok University
www.northbkk.ac.th

Yonok University
www.yonok.ac.th

Institute of Electrical and Electronics Engineers
Thailand Section
www.ieeethailand.org

Conference Organization Committee

Information Contact :

General Chair :

Assoc.Prof.Dr.Monchai Tiantong, KMUTNB

Technical Program Chair :

Prof.Dr.Herwig Unger, FernUni in Hagen, Germany

Conference Treasurer :

Asst.Prof.Dr.Supot Nitsuwat, KMUTNB

Publication Chair :

Asst.Prof.Dr.Anirach Mingkhwan, KMUTNB, IEEE member

Secretary :

Asst.Prof.Dr.Phayung Meesad, KMUTNB

Technical Program Committee

Prof.Dr.Herwig Unger, FernUni, Germany
Prof.Dr.Mark Weiser, OSU, USA
Prof.Dr.Craig Valli, ECU, Australia
Prof.Dr.Wolfram Hardt, CUT, Germany
Assoc.Prof.Dr.Wanchai Rivepiboon, CU
Assoc.Prof.Dr.Viboon Chunkag, KMUTNB
Assoc.Prof.Dr.Boonwat Attachoo, KMITL
Assoc.Prof.Dr.Arunee Intarapirot, RMUTT
Assoc.Prof.Dr.Athakorn Kengpol, KMUTNB
Assoc.Prof.Dr.Sanoe Khunprasert, KMUTNB
Assoc.Prof.Dr.Nitisuk Sawasdijirakij, NBU
Assoc.Prof.Dr.Surapan Yimman, KMUTNB
Assoc.Prof.Dr.Suthep Butdee, KMUTNB
Assoc.Prof.Dr.Prayoot Akkaraekthalin, KMUTNB
Assoc.Prof.Udom Jeenpradub, KMUTNB
Assoc.Prof.Chuleerat Jaruskulchai, KU
Assoc.Prof.Prasong Praneetpograng, SPU
Assoc.Prof.Wirat Pongsiri, MSU
Assoc.Prof.Dr.Somchai Prakan-charoen, KMUTNB
Assoc.Prof.Dr.Somsak Akatimagul, KMUTNB
Assoc.Prof.Dr.Somjit Arch-int, KKU
Asst.Prof.Dr.Phayung Meesad, KMUTNB
Asst.Prof.Dr.Supot Nitsuwat, KMUTNB
Asst.Prof.Dr.Sanit teemuangsai, RMU
Asst.Prof.Dr.Poonsak Kosriyaporn, KMUTNB
Asst.Prof.Dr.Kuntpong Woraratpanya, KMUTNB
Asst.Prof.Dr.Charun Sanrach, KMUTNB
Asst.Prof.Dr.Chaiyapon Thongchaisuratkeul, KMUTNB
Asst.Prof.Dr.Passakorn Ruengrong, NU
Asst.Prof.Dr.Kittinuch Chulikawit, MJU
Asst.Prof.Dr.Winai Bodhisuwan, KU
Asst.Prof.Dr.Manoon Sriwirat, UBU
Asst.Prof.Dr.Panpetch Chinintron, RMUT
Asst.Prof.Dr.Anirat Mingkhwan, KMUTNB
Asst.Prof.Dr.Prachyanun Nilsook, KMUTNB
Asst.Prof.Dr.Parinya Sa-nguansat, RSU
Asst.Prof.Dr.Wittaya Ar-reerach, RMU
Asst.Prof.Dr.Pitsuttha Ar-reerach, RMU
Asst.Prof.Dr.Uraiwan Intarayam, RMUTT
Asst.Prof.Dr.Siriporn Thipkong, RSU
Asst.Prof.Dr.Wasin Chooprayoon, RSU
Asst.Prof.Sittiporn Gerdsum-ang, KMUTNB
Asst.Prof.Burin Rujanapan, Yonok
Asst.Prof.Narisra Chutinara, SWU
Dr.Haythem Zorkta, Aleppo University
Dr.M Muralidhara Rao, Swarnandhra Swarnadea Engineering Collage, India
Dr.Banchong Harangsee, NECTEC
Dr.Panomkhawn Riyamongkol, NU
Dr.Punnee Sittidech, NU
Dr.Passakorn Pratombutr, NECTEC
Dr.Choochat Haruechaiyasak, NECTEC
Dr.Apirak Preechayasombhun, TOT
Dr.Tong Srikacha, TOT
Dr.Nattavee Utakrit, KMUTNB
Dr.Noppasak Tantisattayanon, RMUT
Dr.Somkid Saelee, KMUTNB
Dr.Jirapan Srisompan, KMUTNB
Dr.Suksang Kukanok
Dr.Padej Phomasakha Na Sakolnakorn
Dr.Woraniti Limpakorn
Dr.Pairoj Likitthanasate, TOT
Dr.Banatus Soiraya, KMUTNB
Dr.Wongkot sriurai, UBU
Dr.Sumita Nulmeesri, SSRU
Dr.Sageemas Nawichain, KMUTNB
Dr.Pudsadee Boonrawd, KMUTNB
Dr.Thippaya Chintakovid, KMUTNB
Dr.Cholatip Yawut, KMUTNB
Dr.Sirapat Boonkrong, KMUTNB
Dr.Nawaporn Wisitpongphan, KMUTNB
Dr.Chaiyod Pirak, KMUTNB
Dr.Soradech krutjohn, KMUTNB
Dr.Sittiporn Prawatrungrueng, NBU
Dr.Montean Rattanasiriwongwut, KMUTNB
Dr.Sakchai Tangwannawit, KMUTNB
Dr.Maleerat Sodanil, KMUTNB
Dr.Kamol Limtanyakool, KMUTNB
Dr.Chatchawin Namman, UBU
Dr.Wilaiporn Ledmahakiet, RMUT
Dr.Alisa Kongton, NECTEC
Dr.Doy Sundarasaradula, TOT
Dr.Terawat Piboongungon, TOT
Dr.Noppasak Tantisattayanon, RMUT
Dr.Khaled Faqih, Al al-Bayt University, Jo
Dr.Mahmoud Pirhadi, IAU, Ir

List of Abbreviation

CU	Chulalongkorn University
CUT	Chemnitz University of Technology, Germany
ECU	Edith Cowan University, Australia
FernUni.	Fern University in Hagen, Germany
KKU	Khon-Kaen University
KMITL	King Mongkut's Institute of Technology Ladkrabang
KMUTNB	King Mongkut's University of Technology North Bangkok
KU	Kasetsart University
MJU	Maejo University
MSU	Mahasarakham University
NBU	North Bangkok University
NECTEC	National Electronics and Computer Technology Center
NU	Naresuan University
OSU	Oklahoma State University
RMU	Rajabhat Mahasarakham University
RMUT	Rajamangala University of Technology
RSU	Rangsit University
SPU	Sripathum University
SWU	Srinakarinwirot University
TOT	Telephone of Thailand
UBU	Ubol Ratchathani University
Yonok	Yonok University

Wednesday May 11, 2011	
8:00-9:00	Registration (Nawamintararachinee Building, KMUTNB)
9:00-9.30	Opening Ceremony by Prof. Dr. Teravuti Boonyasopon, *President of King Mongkut's University of Technology North Bangkok*
9:30-10:30	Invited Keynote Speech by Prof.Dr.Serge Tichkiewitch
	Topic : Life Cycle Knowledge Management
10:30-11:00	*Coffee Break*
11:00-12:00	Invited Keynote Speech by Dr.-Ing. Gerd Heinz
	Topic : Interference Networks as a Generalizing Signal Theory between Integral Transformations, Artificial Neural Nets, Optics and Acoustics
12:00-13:00	*Lunch*
13:00-17:00	Parallel Session Presentations
18:00-22:00	Welcome Dinner

Room 1

Session 1 : Data Mining, Knowledge Discovery, and Knowledge Management

Time/Paper-ID	Title/Author	Page
13.00-13.20 **NCCIT2011-03**	**Data Imputation Estimation Using Product Indicator Approaches of Structural Equation Modeling** *Narong Phothi and Somchai Prakancharoen*	1
13.20-13.40 **NCCIT2011-74**	**Agriculture Bibliographic Data Categorization using Feature Selection Techniques** *Pilapan Phonarin, Supot Nitsuwat, and Choochart Haruechaiyasak*	7
13.40-14.00 **NCCIT2011-96**	**An Information Extraction System for Medical Informatics using Text Mining Techniques** *Umaporn Saisangchan, Ekamorn Meesomsak, Natthapong Sripadung, Panida Songram, Kaveepoj Bunluawong, and Phatthanapong Wanchantueng*	13
14.00-14.20 **NCCIT2011-60**	**Comparative Study of Classification Properties between KNN,** **Naïve Bays, Decision Tree and Rule-Based** *Thanyaporn Boonyoung, Aekkarat Lorphichian and Niroot Porngam*	19
14.20-14.40 **NCCIT2011-100**	**Comparative analysis, data mining techniques to determine the suitability of the technique for prediction the different data sets and different data models** *Sajjaporn Waijanya, Thanarak Limjitsomboon and Chayapol Khoonnaret*	24

Time/Paper-ID	Title/Author	Page
14.40-15.00	*Coffee Break*	
15.00-15.20 **NCCIT2011-65**	**A Comparison of Decision Tree, Rule-based, Naïve Bayes and K-Nearest Neighbor for Data Classification** *Gritya Tongpasuk*	30
15.20-15.40 **NCCIT2011-27**	**The Possibility of Using Machine Learning to Predict the Occurrence of Forest Fires in Thailand: A Case Study Chaiyaphum Province** *Narasak Poophayang, Khamron Sunat, Sirapat Chiewchanwattana, and Ngamnij Arch-int*	36
15.40-16.00 **NCCIT2011-90**	**A Comparative Efficiency of Classification with Data Mining Technique** *Patsakorn Singto, Patamaporn Taophan and Akkrawuthi Paramapuneya*	42
16.00-16.20 **NCCIT2011-150**	**A study of related attributes in Vocational Education Institutes and External Quality Assessment with Association Rule Mining** *Chantima Ekwong, Sageemas Na Wichian, and Choochart Haruechaiyasak*	49
16.20-16.40 **NCCIT2011-78**	**Autistic Learning Tracking System : ALTs** *Nanticha Sondoung, Warunee Rattanaprapa, Pichsinee Puttitaweesri, and Chatchada Kaewpruksapimon*	55
18.00-22.00	*Dinner*	
Room 2		
Session 2 : Decision Support and Recommender Systems		
Time/Paper-ID	**Title/Author**	**Page**
13.00-13.20 **NCCIT2011-223**	**A Development of Information Recommender for Notebook Purchasing by using Collaborative Filtering Technique and Utility Theory** *Tanawat Jariyapoom and Chainarong Yensiri*	61
13.20-13.40 **NCCIT2011-15**	**Decision Support Systems for Locating Automatic Teller Machine Based on Investment Analysis** *Santi Teeragul and Rachada Kongkachandra*	67
13.40-14.00 **NCCIT2011-19**	**Automatic Priority Heading System using Decision Tree for Improving Screen Reader** *Jitlada Jermjit and Rachada Kongkachandra*	73
14.00-14.20 **NCCIT2011-124**	**The Data Warehouse Health care checkup system for BASF employee analyze the risk of cholesterol by ID3** *Asadawut Atising and Sirapat Boonkrong*	79

การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล
A Comparative Efficiency of Classification with Data Mining Techniques

พัสกร สิงห์โต (Patsakorn Singto)[1] *อัครวุฒิ ปรมะปุญญา (Akkrawuthi Paramapuneya)*[2]
และปฐมาภรณ์ เถาว์พัน (Patamaporn Taophan)[3]
[1, 2, 3]*ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ*
psingto@hotmail.com, akkrawuthi@hotmail.com, patamatt@gmail.com

บทคัดย่อ

งานวิจัยนี้เสนอการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยเทคนิคการทำเหมืองข้อมูลด้วย Naive Bayesian, KNN, Rule Base และ Decision Tree โดยเลือกพารามิเตอร์ที่แตกต่างกัน ใช้ชุดข้อมูล Ionosphere, Pima Indians Diabetes, The Insurance Company และ Mushroom จากฐานข้อมูล UCI Machine Learning Repository นำข้อมูลมาวิเคราะห์และทำการเปรียบเทียบประสิทธิภาพโดยเลือกพารามิเตอร์ที่แตกต่างกันในแต่ละเทคนิค

การวัดประสิทธิภาพการจำแนกข้อมูลสามารถวัดได้จากความถูกต้องของการจำแนกในแต่ละชุดข้อมูลด้วยอัลกอริธึมการทำเหมืองข้อมูลที่แตกต่างกัน ผลการทดลองการวัดประสิทธิภาพของแบบจำลองการจำแนกข้อมูลด้วย Unseen Data พบว่า ข้อมูลชนิดนามบัญญัติ (Nominal) ถูกจำแนกได้ประสิทธิภาพดีด้วยอัลกอริธึม Decision Tree, Rule Bas และ K-Nearest Neighbor ตามลำดับ สำหรับข้อมูลชนิดต่อเนื่อง (Continuous Data) มีจำนวนแอทริบิวต์มากจะถูกจำแนกได้ประสิทธิภาพดีด้วยอัลกอริธึม Rule Base และข้อมูลมีจำนวนแอทริบิวต์น้อยจะถูกจำแนกได้ประสิทธิภาพดีด้วยอัลกอริธึม K-Nearest Neighbor

คำสำคัญ: เหมืองข้อมูล การจำแนกข้อมูล ต้นไม้การตัดสินใจ เคเนียเรสต์เนเบอร์ กฎพื้นฐาน นาอีฟเบย์เซียน

Abstract

This research proposes to compare the efficiency classification of data mining techniques by Naive Bayesian, KNN, Rule Base, and Decision Tree. Select parameters different set. Use dataset of Ionosphere, Pima Indians Diabetes, The Insurance Company and the Mushroom from UCI Machine Learning Repository. The data was analyzed and compared the efficiency by choosing the different parameters for each technique.

The efficiency measurement of data classification can be measured form the correctness of the classification of each dataset by algorithms making the difference source. The measurement of the model classification data by unseen data. The result found that the nominal data model was well classified by Decision Tree, Rule Base, and K-Nearest Neighbor, respectively. For continuous data and much attribute was well classified by Rule Base. And K-Nearest Neighbor that was good in efficiency and there is little attribute.

Keyword: Data Mining , Classification Data, Decision Tree, K-Nearest Neighbor, Rule Base, Naïve Bayesian.

1. บทนำ

งานวิจัยนี้เป็นการศึกษาความเหมาะสมของการใช้โมเดลในการทำเหมืองข้อมูลที่มีการจำแนกข้อมูลที่มีมิติสูงและสามารถหาความเหมาะสมของข้อมูลจากการหาเทคนิคต่างๆ โดยใช้ข้อมูลจาก UCI Machine Learning Repository และเลือกข้อมูล 4 ชุด ได้แก่ Ionosphere , Pima Indians Diabetes, The Insurance Company (TIC) 2000 และ Mushroom ใช้เป็นกรณีศึกษา โดยค่าพารามิเตอร์ของชุดข้อมูลที่เลือกใช้นั้น มีความซ้ำซ้อนและหลากหลาย ทำให้เกิดความแตกต่างของแต่ละชุดข้อมูล ได้แก่ จำนวนแอทริบิวต์ จำนวนชุดข้อมูล และลักษณะของข้อมูล ผลที่ได้จะสามารถจำแนกกลุ่มข้อมูล ซึ่งการวิจัยครั้งนี้จะใช้เทคนิคการทำเหมืองข้อมูล โดยใช้อัลกอริธึม ได้แก่ Naïve Bayesian, K-Nearest Neighbour, Rule Based และ Decision Tree เพื่อเปรียบเทียบประสิทธิภาพการจำแนกข้อมูล และมีการปรับ

ค่าพารามิเตอร์ที่แตกต่างกัน ทำให้สามารถทราบถึงความเหมาะสมในการใช้การจำแนกข้อมูลประเภทต่างๆได้

2. วัตถุประสงค์ของโครงงาน

2.1 เพื่อศึกษาความเหมาะสมของการใช้อัลกอริธึมในการทำเหมืองข้อมูลที่มีการจำแนกข้อมูลที่มีมิติสูง ได้แก่ Naïve Bayesian, KNN, Rule Base และ Decision Tree

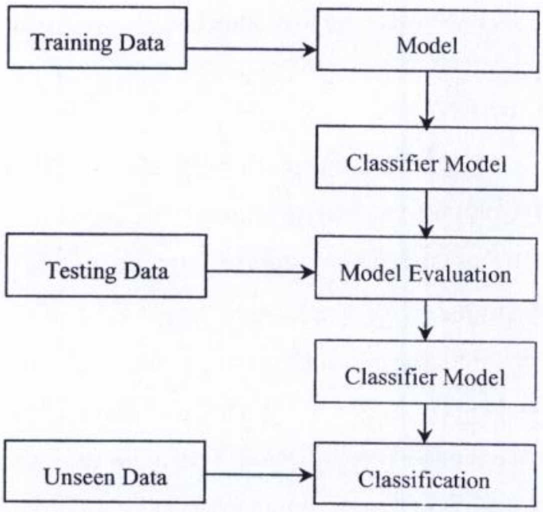
2.2 เพื่อศึกษาความเหมาะสมของอัลกอริธึมต่างๆที่นำมาใช้กับข้อมูลที่มีลักษณะแตกต่างกัน

2.3 เพื่อศึกษาการปรับค่าพารามิเตอร์ที่เหมาะสมกับการทดลอง เพื่อให้ได้ผลลัพธ์ที่ดีขึ้น

3. วิจารณ์งานวิจัยหรือทฤษฎีที่เกี่ยวข้อง

3.1 เหมืองข้อมูล (Data Mining)

เป็นกระบวนการกลั่นกรองสารสนเทศ (Information) ที่ซ่อนอยู่ในฐานข้อมูลขนาดใหญ่ (Knowledge Discovery in Database, KDD) เพื่อทำนายแนวโน้ม และพฤติกรรมโดยอาศัยข้อมูลในอดีต และเพื่อใช้สารสนเทศเหล่านี้ในการสนับสนุนการตัดสินใจ โดยรูปแบบที่ใช้ในการทำเหมืองข้อมูลมีหลายรูปแบบด้วยกัน ในแต่ละรูปแบบจะมีวิธีการที่แตกต่างกันออกไป [1]



ภาพที่ 1: กระบวนการ Classification

3.2 การจำแนกข้อมูล (Classification)

เป็นเทคนิคหนึ่งของการทำเหมืองข้อมูลที่ใช้ในการ Predictive Modeling ซึ่งมีการทำงานแบบ Supervised Learning กล่าวคือ สามารถสร้างแบบจำลองการจัดหมวดหมู่ได้จาก กลุ่มตัวอย่างของข้อมูลที่ได้กำหนดไว้ก่อนล่วงหน้า และสามารถพยากรณ์ตัวอย่างของข้อมูลที่ได้กำหนดไว้ก่อนล่วงหน้า และสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้ด้วย โดยมีกระบวนการและขั้นตอนดังภาพที่ 1 [2]

3.3 นาอีฟเบย์ (Naïve Bayesian)

เป็นอัลกอริธึมการจำแนกข้อมูล [3] ที่ใช้หลักความน่าจะเป็นซึ่งอยู่บนพื้นฐานของ Bayes' Theorem เช่นกำหนดให้การเกิดของเหตุการณ์ต่างๆที่ใช้ในการจำแนกกลุ่มนั้นเป็นอิสระต่อกัน โดยกำหนดให้

P(H) คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ H และ

P(H|E) คือความน่าจะเป็นที่จะเกิดเหตุการณ์ H เมื่อเกิดเหตุการณ์ E

จากตัวแปรที่กำหนดและแนวคิดของ Bayes' Theorem นั้นสามารถทำนายเหตุการณ์ที่พิจารณาได้จากการเกิดของหตุการณ์ต่าง ๆ ได้ดังสมการ

$$P(H|E) = [P(E|H)xP(H)]/P(E) \quad (1)$$

เช่น การทำนายว่าฝนจะตกเมื่อมีเหตุการณ์มีเมฆดำ กำหนดให้

H คือ เหตุการณ์ที่ฝนตก และ

E คือ เหตุการณ์มีเมฆดำ และสามารถทำนายสรุปได้ดังสมการต่อไปนี้

$$P(\text{ฝนตก} \mid \text{เมฆดำ}) = [P(\text{เมฆดำ} \mid \text{ฝนตก}) \times P(\text{ฝนตก})]/P(\text{เมฆดำ})$$

กำหนดให้

1) P(เมฆดำ | ฝนตก) คือ ความน่าจะเป็นที่มีเมฆดำเมื่อฝนตก ซึ่งจะพิจารณาการเกิดเมฆดำเมื่อมีเหตุการณ์ฝนตกเท่านั้น ความน่าจะเป็นนี้สามารถเก็บรวบรวมโดยใช้หลักการณ์ทางสถิติ

2) P (ฝนตก) คือ ความน่าจะเป็นที่ฝนจะตก

3) P (เมฆดำ) คือ ความน่าจะเป็นที่มีเมฆดำ

การคำนวณการจำแนกกลุ่มของเหตุการณ์ที่มีการเกิดของเหตุการณ์ต่างๆ ที่ใช้ในการจำแนกกลุ่มมากกว่า 1 ชนิด ได้ดังสมการต่อไปนี้

$$P(H|E1,E2,...,En) = [P(E1,E2,...,En|H) \times P(H)]/P(E1,E2,....,En) \quad (2)$$

เมื่อกำหนดให้เหตุการณ์ E1, E2,....,En คือ เหตุการณ์ *n* เหตุการณ์ที่ใช้ในการจำแนกกลุ่ม และแต่ละเหตุการณ์ต่างๆที่ใช้ในการจำแนกกลุ่มปีนอิสระต่อกัน [4]

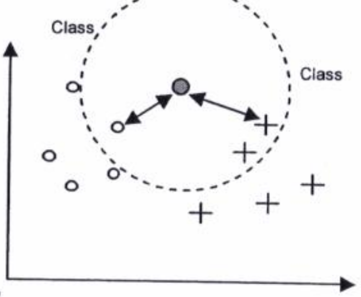
3.4 เคเนียเรสต์เนเบอร์ (K-Nearest Neighbor)

เป็นอัลกอริธึมที่ใช้ในการจัดกลุ่มข้อมูลที่อยู่ใกล้กันเข้าไว้ด้วยกัน โดย k หมายถึงจำนวนข้อมูลที่นำมาพิจารณา เช่น k=3 หมายถึง การพิจารณาข้อมูลจำนวน 3 ชุดที่อยู่ใกล้เคียงจุดที่ต้องการ ประกอบด้วยวิธีการ ดังนี้ 1. กำหนดขนาด k ควรกำหนดเป็นเลขคี่ 2. คำนวณระยะห่างของข้อมูลที่ต้องการพิจารณากับกลุ่มข้อมูลตัวอย่าง 3. จัดเรียงลำดับของระยะห่างและเลือกพิจารณาชุดข้อมูลที่ใกล้จุดที่ต้องการพิจารณาตามจำนวน k ที่กำหนด 4. พิจารณาข้อมูลจำนวน k ชุด และสังเกตว่ากลุ่ม (Class) ไหนที่ใกล้จุดที่พิจารณาเป็นจำนวนมากที่สุด 5. กำหนดกลุ่มให้กับจุดที่พิจารณา ดังภาพที่ 2 กำหนด k = 3 ประกอบด้วย Class (+), Class (○) และจุดพิจารณาคือ ● ดังนั้นข้อมูลที่อยู่ใกล้ ● ที่สุดจำนวน 3 ชุด คือ

+ + เท่ากับ 2

○ เท่ากับ 1

ดังนั้น จึงจัดกลุ่มให้จุดที่พิจารณา ● ให้เป็น Class (+) [5]



ภาพที่ 2: การกระจายตัวของ K-Nearest Neighbor

3.5 กฎพื้นฐาน (Rule based)

กฎพื้นฐาน [6] หรือ rule-based คือ เทคนิคสำหรับการแยกประเภทของระเบียน โดยการใช้รูปแบบกฎ “if…them…” ซึ่งกฎที่นำมาสร้างโมเดลจะถูกนำเสนอ ในรูปแบบการเชื่อมต่อกัน เช่น $R = (r1 \vee r2 ... \vee r_j)$ โดยที่ R คือ กลุ่มของกฎ และ ri คือ กฎที่ใช้แยกหรือกฎแต่ละอัน แต่ละกฎที่ใช้แยกสามารถอธิบายโดยสมการที่ (3)

$$r_i : (\text{เงื่อนไข}) \longrightarrow y_i \quad (3)$$

ด้านซ้ายมือของกฎมักเรียกว่า Rule antecedent หรือ precondition ซึ่งจะมีการเชื่อมต่อของ attribute set: ดังแสดงในสมการที่ (4)

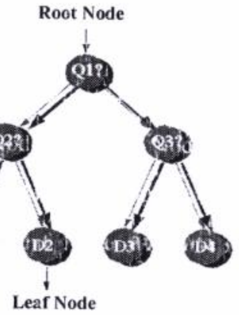
$$Condition_i = \{ (A1\ op\ v1) \wedge (A2\ op\ v2) \wedge (A_k\ op\ v_k) \} \quad (4)$$

ที่ (A_j, v_j) คือ คู่ของ attribute และค่า และ op คือ ตัวปฏิบัติการแบบตรรกะที่ถูกเลือกจาก set $\{=, \neq, <, >, \leq, \geq\}$ ในขณะที่ทางด้านขวามือของกฎเรียกว่า rule consequent ซึ่งจะใช้ทำนาย class y_i [7]

3.6 ต้นไม้การตัดสินใจ (Decision Tree)

ต้นไม้การตัดสินใจเป็น เทคนิคที่ค่อนข้างแพร่หลาย เนื่องจากการตัดสินใจแบบโครงสร้างต้นไม้ เป็นลักษณะที่ทำให้เข้าใจได้ง่าย วิธีนี้คือกระบวนการในการจัดแบ่งข้อมูล (Classification) โดยใช้โครงสร้างของต้นไม้ในการเขียนอธิบายการประมวลผล โดยจำลองกิ่งก้านสาขาของต้นไม้เป็นหลักโดยแตกจากรากทางซ้ายมือ แตกเป็นกิ่งอยู่ทางขวามือแตกกิ่งไปเรื่อย ๆ จนครบเงื่อนไขทั้งหมดและกิ่งสุดท้ายของทุกกิ่งคือกิจกรรมที่ต้องกระทำ เมื่อเป็นตามเงื่อนไขนั้น [8]

อัลกอริธึมพื้นฐานของการสร้างต้นไม้ตัดสินใจ คือ Greedy Algorithm โดยจะสร้างต้นไม้ จากบนลงล่างแบบวนซ้ำ (Recursive) ด้วยวิธีการแบ่งปัญหาใหญ่เป็นปัญหาย่อย (Divide-and-Conquer) ซึ่งรูปแบบของต้นไม้ จะประกอบด้วยโหนดแรกสุดที่เรียกว่า Root Node จาก Root Node ก็จะแตกออกเป็นโหนดลูก และที่โหนดลูกก็จะมีลูกของตัวเอง ซึ่งโหนดในระดับ สุดท้ายจะเรียกว่า Leaf Node ดังภาพที่ 3



ภาพที่ 3: รูปแบบต้นไม้ตัดสินใจ

จากภาพที่ 3 จะเห็นว่า จาก Root Node จนถึง Leaf Node จะมีเพียงเส้นทางเดียวเท่านั้น ซึ่งเส้นทางนี้จะอธิบายถึงกฎที่ใช้

สำหรับการจัดหมวดหมู่ของแต่ละกลุ่ม ซึ่งในแต่ละ Leaf Node นั้นอาจเป็นกลุ่มเดียวกัน ซึ่งเกิดจากเหตุผล ที่แตกต่างกันได้

แต่ละอัลกอริธึมก็ได้นิยามค่าความดีของคุณสมบัติแตกต่างกันไปอัลกอริธึม CART นิยามความดีของคุณสมบัติ โดยใช้ค่าสัมประสิทธิ์จีนี (Gini) แต่ที่แพร่หลายที่สุดคือการ ใช้ค่ามาตรฐานเกน (Gain criterion) ของอัลกอริธึม ID3 [9] และ C4.5 [10] ซึ่งได้จากการคำนวณโดยอาศัยทฤษฎีสารสนเทศ (Information Theory) และค่าเอนโทรปี (Entropy)

4. ขั้นตอนการดำเนินงาน

4.1 การเตรียมข้อมูล

ในการวิจัยครั้งนี้ ผู้วิจัยได้เลือกใช้ข้อมูลจาก UCI Machine Learning Repository จำนวน 4 ชุดข้อมูลเป็นกรณีศึกษา ได้แก่ Ionosphere ข้อมูลการวัดสัญญาณชั้นบรรยากาศนอกโลก เป็นข้อมูลชนิดต่อเนื่อง (Continuous Data) ประกอบด้วยแบบข้อมูลชั้นบรรยากาศ 351 รายการ แต่ละรายการมี 34 แอทริบิวท์ เพื่อตรวจสอบว่าสัญญาณดีหรือไม่ดี Pima Indians Diabetes ข้อมูลโรคเบาหวานของชาวอินเดีย เป็นข้อมูลชนิดต่อเนื่อง ประกอบด้วยแบบของโรคเบาหวาน 768 รายการ แต่ละรายการมี 8 แอททริบิวท์ เพื่อแสดงว่าเป็นหรือไม่เป็นโรคเบาหวาน The Insurance Company (TIC) ข้อมูลลูกค้าที่เป็นเจ้าของกรมธรรม์ เป็นข้อมูลชนิดนามบัญญติ (Nominal Data) ประกอบด้วยแบบข้อมูลลูกค้า 9822 รายการ แต่ละรายการ 85 แอททริบิวท์ เพื่อตรวจสอบว่าได้รับสิทธิการมีหรือไม่มีรถบ้าน และ Mushroom ข้อมูลลักษณะของเห็ดเป็นข้อมูลชนิดนามบัญญติ ประกอบด้วยแบบข้อมูลเห็ด 8124 รายการ 23 แอททริบิวท์ เพื่อแสดงว่าเป็นเห็ดมีพิษหรือไม่มีพิษ และได้นำข้อมูลมาทำการ Cleaning ตัวแปรที่เกิดจากความผิดพลาดในการใส่ข้อมูล และได้ตัดรายการข้อมูลที่ไม่สมบูรณ์ออกไป หลังจากนั้น ทำการลดทอนหรือแบ่งกลุ่มตัวแปรสร้างกลุ่มตัวแปรตามใหม่ตามความเหมาะสม จัดรูปแบบชุดข้อมูลให้อยู่ในรูปแบบ Comma Separate Value File (csv file) แต่ละชุดข้อมูลและจำนวนอินสแทนซ์ มีจำนวนแอทริบิวต์ ดังตารางที่ 1 และแต่ละชุดข้อมูลทำการสุ่มแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลฝึกสอน (Training Set) 66% และ ข้อมูลทดสอบ (Test Set) 34% ดังตารางที่ 2

ตารางที่ 1: ตารางแสดงรายละเอียดแอทริบิวต์

Data set	No. of Instance	No. of Attb.	No. of Nominal Attb.	No. of Numeric Attb.	Missing Values	No. of Class
Ionosphere	351	34	0	34	None	2
Pima Indians Diabetes	768	8	0	8	None	2
The Insurance Company (TIC)	9,822	85	85	0	None	2
Mushroom	8,124	23	23	0	2480	2

ตารางที่ 2: จำนวนรายการข้อมูลชุดฝึกสอน และชุดทดสอบ

Data set	No. of Pattern Training Set for Class			No. of Pattern Test Set for Class			No. of Pattern Total for Class	
	1	2	Tot	1	2	Tot	1	2
Ionosphere	138	93	231	87	33	120	225	126
Pima Indians Diabetes	321	185	506	179	83	262	500	268
The Insurance Company (TIC)	5474	348	5822	3762	238	4000	9236	586
Mushroom	2548	2818	5366	1368	1390	2758	3916	4208

ตารางที่ 3: จำนวนแอทริบิวท์ในแต่ละชุดข้อมูลตามวิธีการประเมินค่าแอทริบิวท์

Attribute Evaluator	Ionosphere	Pima Indians Diabetes	The Insurance Company	Mushroom
Original	34	8	85	23
CfsSubset	13	4	10	5
ChiSquared	20	3	14	12
InfoGain	17	-	13	-
Symmetrical Uncert	18	5	15	-
SVM	-	5	-	-
Consistency Subset	7	-	-	5
OneR	18	3	80	-
ReliefF	19	5	42	12
GainRatio	-	4	16	7
Filtered	21	4	-	12

จากการนำชุดข้อมูลมาประเมินค่าแอทริบิวท์ด้วยวิธีการประเมินค่าแอทริบิวท์ที่เหมาะสม ทำให้ได้ชุดข้อมูลเพิ่มมากขึ้น ดังนี้ Ionosphere มีจำนวน 9 ชุด Pima Indians Diabetes มีจำนวน 9 ชุด The Insurance Company (TIC) มีจำนวน 8 ชุด และ Mushroom มีจำนวน 7 ชุด ดังตารางที่ 3

4.2 การออกแบบโมเดลเพื่อทำการทดลอง

ผู้วิจัยได้ทำการออกแบบโมเดล โดยทำการปรับค่าพารามิเตอร์ให้กับอัลกอริธึม Decision Tree, Rule Base, Naïve Bayesian และ K-Nearest Neighbor ดังตารางที่ 4 – 7

ตารางที่ 4: ค่าพารามิเตอร์ของอัลกอริธึม K-Nearest Neighbor

No.	KNN	distanceFunction	dontNormlize
1	1	EuclideanDistance	FALSE
2	3	EuclideanDistance	FALSE
3	1	EuclideanDistance	TRUE
4	3	EuclideanDistance	TRUE
5	1	ManhattanDistance	FALSE
6	3	ManhattanDistance	FALSE

ตารางที่ 5: แสดงค่าพารามิเตอร์ของอัลกอริธึม Rule Base

No.	Min NumObj	Reduced ErrorPruning	unpruned	useMDL Correction
1	2	FALSE	FALSE	TRUE
2	2	TRUE	FALSE	TRUE
3	1	FALSE	FALSE	TRUE
4	3	FALSE	FALSE	TRUE
5	2	FALSE	FALSE	FALSE
6	3	FALSE	FALSE	FALSE
7	2	TRUE	FALSE	FALSE
8	3	TRUE	FALSE	FALSE
9	2	FALSE	TRUE	FALSE
10	4	FALSE	TRUE	FALSE

ตารางที่ 6: แสดงค่าพารามิเตอร์ของอัลกอริธึม Decision Tree

No.	minNum Object	Reduced ErrorPruning	unpruned	useMDL Correction
1	2	FALSE	FALSE	TRUE
2	2	TRUE	FALSE	TRUE
3	1	FALSE	FALSE	TRUE
4	3	FALSE	FALSE	TRUE
5	3	FALSE	FALSE	FALSE
6	2	TRUE	FALSE	FALSE
7	3	TRUE	FALSE	FALSE
8	2	FALSE	TRUE	FALSE
9	4	FALSE	TRUE	FALSE

ตารางที่ 7: แสดงค่าพารามิเตอร์ของอัลกอริธึม Naive Bayesian

No.	debug	displayModel InOldFormat	useKernel Estimator	useSupervised Discretization
1	FALSE	FALSE	FALSE	FALSE
2	FALSE	FALSE	TRUE	FALSE
3	FALSE	FALSE	FALSE	TRUE
4	TRUE	TRUE	FALSE	FALSE
5	TRUE	TRUE	TRUE	FALSE
6	TRUE	TRUE	FALSE	TRUE

4.3 การดำเนินการฝึกสอนระบบ

การฝึกสอนระบบโดยการนำชุดข้อมูลฝึกสอนที่เตรียมไว้แต่ละชุดเข้าทำฝึกสอนระบบในแต่ละอัลกอริธึมด้วยพารามิเตอร์ที่แตกต่างๆ กัน ดังตารางที่ 4 – 7 โดยให้แต่ละชุดข้อมูลได้เข้าฝึกสอนระบบทุกเทคนิคและทุกพารามิเตอร์

4.4 วิธีการทดสอบระบบ

การทดสอบระบบโดยการนำข้อมูลชุดทดสอบที่เตรียมไว้แต่ละชุดมาดำเนินการทดสอบระบบในแต่ละอัลกอริธึมด้วยพารามิเตอร์ที่แตกต่างๆ กัน และประเมินประสิทธิภาพการจำแนกข้อมูลของระบบด้วยค่าความถูกต้อง (Accuracy)

5. ผลการดำเนินงาน

5.1 ผลการทดสอบ

เป็นการวัดประสิทธิภาพการจำแนกข้อมูลของแต่ละโมเดลด้วยค่าความถูกต้องของการทดสอบระบบด้วยเทคนิคการทำเหมืองข้อมูล 4 เทคนิคที่มีพารามิเตอร์ที่แตกต่างกันกับข้อมูลแต่ละชุด มีผลการทดสอบ ดังตารางที่ 8 - 11

ตารางที่ 8: ค่าความถูกต้องการจำแนกด้วยเทคนิค K-Nearest Neighbour จาก 198 โมเดล

Evaluate Attribute		Num attb.	Parameter Pattern 2	4	5
Ionosophere	SymmetricalUncert	18	93.33	93.33	**97.50**
	Filtered	13	92.50	94.17	94.17
Pima Indians Diabetes	Original	8	71.37	69.47	73.66
	GrainRatioEval	5	**84.73**	66.79	71.76
	CfsSubset	4	72.14	67.56	66.79
The Insurance Company	CfsSubset	10	93.58	**93.78**	92.05
	OneR	80	93.35	93.23	90.13
	InfoGainAttrib	13	93.33	93.58	90.73
Mushroom	Original	23	100	100	100
	ChiSquaredAttrib	12	100	100	100
	Consistency	5	100	100	100

ตารางที่ 9: ค่าความถูกต้องการจำแนกด้วย Naïve Bayesian จาก198 โมเดล

	Evaluate Attribute	Num attb.	Parameter Pattern 1	4	5
Ionosphere	SymmetricalUncert	18	87.50	87.50	91.67
	Filtered	13	92.50	92.50	**97.50**
Pima Indians Diabetes	Original	8	77.10	77.10	73.66
	GrainRatioEval	5	75.95	75.95	76.34
	CfsSubset	4	77.86	**77.86**	73.28
The Insurance Company	CfsSubset	10	**92.53**	92.53	92.40
	OneR	80	77.53	77.53	83.85
	InfoGainAttrib	13	86.30	86.30	87.68
Mushroom	Original	23	95.03	95.03	95.03
	ChiSquaredAttrib	12	94.74	94.74	94.74
	Consistency	5	**98.66**	**98.66**	**98.66**

ตารางที่ 10: ค่าความถูกต้องการจำแนกด้วย Decision Tree จาก 297 โมเดล

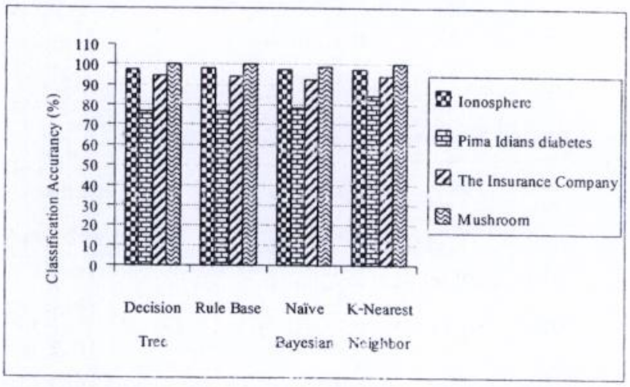
	Evaluate Attribute	Num attb.	Parameter Pattern 2	3	5
Ionosphere	SymmetricalUncert	18	**97.50**	92.50	90.83
	Filtered	13	95.83	92.50	**97.50**
Pima Indians Diabetes	Original	8	72.90	67.56	**77.10**
	GrainRatioEval	5	72.90	73.28	75.57
	CfsSubset	4	72.90	72.52	69.85
The Insurance	CfsSubsetEval	10	94.05	94.05	94.05
	OneR	80	93.93	**94.35**	93.90
	InfoGainAttrib	13	94.05	94.05	94.05
Mushroom	Original	23	100	100	100
	ChiSquaredAttrib	12	100	100	100
	Consistency	5	100	100	100

ตารางที่ 11: ค่าความถูกต้องการจำแนกด้วย Rule Base จาก 330 โมเดล

	Evaluate Attribute	Num attb.	Parameter Pattern 1	2	3
Ionosphere	SymmetricalUncert	18	92.50	90.83	91.67
	Filtered	13	**98.33**	95.83	**98.33**
Pima Indians Diabetes	Original	8	71.76	72.52	70.61
	GrainRatioEval	5	72.52	71.76	72.52
	CfsSubset	4	**76.72**	75.95	75.95
The Insurance Company	CfsSubset	10	93.93	**94.05**	93.78
	OneR	80	91.50	93.48	90.40
	InfoGainAttrib	13	93.75	93.05	**94.05**
Mushroom	Oiginal	23	100	100	100
	ChiSquaredAttrib	12	100	100	100
	Consistency	5	100	100	100

5.2 วิเคราะห์วิจารณ์ เปรียบเทียบ อภิปรายผลการทดลอง

จากตารางที่ 8-11 ค่าความถูกต้องการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล 4 เทคนิคในแต่ละชุดข้อมูล เทคนิคที่เหมาะสมกับข้อมูลแต่ละชุดตามผลของค่าความถูกต้องที่สูงที่สุด คือ เทคนิค Rule Base กับข้อมูล Ionosphere คัดเลือกแอทริบิวท์แบบ Filtered ปรับค่าพารามิเตอร์แบบที่ 1 และ 3 เท่ากับ 98.33% เทคนิค K-Nearest Neighbour ข้อมูล Pima Indians Diabetes คัดเลือกแอทริบิวท์แบบ GrainRatioEval ปรับค่าพารามิเตอร์แบบ 2 เท่ากับ 84.73% เทคนิค Decision Tree ข้อมูล The Insurance Company คัดเลือกแอทริบิวท์แบบ OneR ปรับค่าพารามิเตอร์แบบที่ 3 เท่ากับ 94.35% และ เทคนิค Rule Base, Decision Tree และ K-Nearest Neighbour ข้อมูล Mushroom ใช้แอทริบิวท์ที่มีเริ่มต้นจากชุดข้อมูล และใช้พารามิเตอร์เริ่มต้นของเทคนิคเหมืองข้อมูลเท่ากับ 100%



ภาพที่ 5: เปรียบเทียบความถูกต้องของการจำแนกในแต่ละชุดข้อมูล

6. สรุป

จากการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยอัลกอริธึมการทำเหมืองข้อมูล 4 อัลกอริธึม ได้แก่ Decision Tree, Rule Base และ K-Nearest Neighbour และ Naïve Bayesian กับชุดข้อมูล 4 ชุด คือ ข้อมูลชนิดนามบัญญัติ (Nominal) จำนวน 2 ชุด และข้อมูลชนิดต่อเนื่อง (Continuous Data) จำนวน2 ชุด จากฐานข้อมูล UCI Machine Learning Repository ได้สร้างโมเดลการจำแนกข้อมูลด้วยอัลกอริธึมต่างๆ โดยการปรับค่าพารามิเตอร์ที่เหมาะสม และเลือกแอทริบิวต์ที่มีผลต่อการเปลี่ยนแปลงคลาสมากที่สุด โดยวัดประสิทธิภาพการจำแนกข้อมูลด้วยโมเดลที่แตกต่างกัน จากค่าความถูกต้อง แสดงดังภาพที่ 5 พบว่า ข้อมูลชนิดนามบัญญัติ

(Nominal) คือ The Insurance Company และ Mushroom ถูกจำแนกได้ประสิทธิภาพดีด้วยอัลกอริธึม Decision Tree พารามิเตอร์ที่ 3, Rule Base พารามิเตอร์ที่ 2 K-Nearest Neighbour พารามิเตอร์ที่ 4 และ Naïve Bayesian พารามิเตอร์ที่ 1-4 โดยทุกชุดได้คัดเลือกแอทริบิวท์ด้วยวิธี cfsSubset สำหรับข้อมูลชนิดต่อเนื่อง (Continuous Data) อัลกอริธึม Rule Base พารามิเตอร์ที่ 3 จำแนกข้อมูลได้ประสิทธิภาพดี ด้วยจำนวนแอทริบิวท์มากคัดเลือกแอทริบิวท์ด้วยวิธี Filtered และอัลกอริธึม K-Nearest Neighbour พารามิเตอร์ 2 จำแนกข้อมูลได้ประสิทธิภาพดีกับจำนวนแอทริบิวต์ไม่มากและคัดเลือกแอทริบิวท์ด้วยวิธี GrainRatioEval

จากการวิเคราะห์ผล สามารถสรุปผลได้ว่า อัลกอริธึม K-Nearest Neighbour เหมาะกับข้อมูลมีแอทริบิวท์ไม่มาก และจะดีที่สุดกับข้อมูลที่มีแอทริบิวท์ที่เหมาะสมมากที่สุดด้วยวิธีการประเมินค่าแอทริบิวท์ GrainRatioEval อัลกอริธึม Rule Base เหมาะกับข้อมูลมีแอทริบิวท์มาก และจะดีกับข้อมูลที่เหมาะสมด้วยวิธีการประเมินค่าทริบิวท์ Filtered อัลกอริธึม Decision Tree เหมาะกับข้อมูลชนิดนามบัญญัติ และจะดีที่สุดกับข้อมูลที่มีแอทริบิวท์ที่เหมาะสมมากที่สุดด้วยวิธีการประเมินค่าแอทริบิวท์ CfsSubset อัลกอริธึม Naive Bayesian จากผลการทดลองไม่ปรากฏว่าเหมาะสมกับลักษณะข้อมูลชนิดใดชัดเจน การศึกษาในครั้งต่อไปผู้วิจัยมุ่งศึกษาเทคนิคการทำเหมืองข้อมูลกับข้อมูลที่รวบรวมภายในประเทศเพื่อสามารถค้นหาคำตอบที่ต้องการทราบต่อไป

เอกสารอ้างอิง

[1] อุกฤษ ปัจฉิม. **การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในการทำนายระดับน้ำสูงสุด**. วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมทรัพยากรน้ำ คณะวิศวกรรมศาสตรบัณฑิตวิทยาลัยมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี, 2546.

[2] พยูน พาณิชย์กุล. **การพัฒนาระบบดาต้าไมน์นิ่งโดยใช้ Decision Tree**. โครงการพัฒนาระบบงาน ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ แขนงวิชาวิทยาการสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าคุณทหารลาดกระบัง, 2548.

[3] ภรัณยา อำมฤครัตน์ เดช ธรรมศิริ วาทินี นุ้ยเพียน ณรงค์ โพธิภัทราวุฒิ แสงศิริ ผศ.ดร.พยุง มีสัจ. “การเปรียบเทียบประสิทธิภาพรูปการคัดเลือกและจำแนกข้อมูลด้วยวิธีการทางเครือข่ายประสาทเทียม”. *in The 5th National Conference on Computing and Information Technology.* (5)2009:112.

[4] Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, Second Edition, Morgan Kaufmann publishers: Canada, 2005.

[5] Sun Lee Band, Hyung Jeong Yang and Jae Dong Yang, “Automatic Document Categorization Based on k-NN and Object-Based Thesauri”. *Springer Lecture Notes in Computer Science, Volume 3246/2004*, pp. 101-112

[6] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar.*Introduction to Data Mining*. Addison Wesley. 2005:146,207-208.]

[7] กาญจนา แสงทองพัฒนา. “*การจำแนกประเภทเว็บเพจโดยใช้ลักษณะเด่นบนตัวจำแนกแบบเบย์อย่างง่ายที่ทำงานร่วมกัน*”, วิทยานิพนธ์, บัณฑิตวิทยาลัยมหาวิทยาลัยเกษตรศาสตร์. 2543 :1-2.

[8] Barry de Ville, Decision Trees for Business Intelligence and Data Mining: Using SAS Enterprise Mine, SASPublishing, 2006.

[9] Quinlan, J. R. “Induction of decision tree”, *Machine Learning*,Vol. 1, pp.81-106, 1986.

[10] Quinlan, J. R, “C4.5: Programs for Machine Learning”, San Mateo, CA: Morgan Kaufmann, 1993.

[4] Rajeev, K. and S.C. Soham, "Precise static type analysis for object oriented programs". *SIGPLAN* Not., 2007. 42(2): p. 17-26.

[5] Logozzo, F., *Cibai:"* An Abstract Interpretation-Based Static Analyzer for Modular Analysis and Verification of Java Classes in Verification, Model Checking, and Abstract Interpretation". 2007, *Springer Berlin / Heidelberg*. p. 283-298.

[6] Tal, L., et al., "Putting static analysis to work for verification: A case study", in *Proceedings of the* 2000 *ACM SIGSOFT international symposium on Software testing and analysis*. 2000, ACM: Portland, Oregon, United States.

[7] Logozzo, F., "Separate Compositional Analysis of Class-Based Object-Oriented Languages in Algebraic Methodology and Software Technology". 2004, *Springer Berlin / Heidelberg*. p. 334-348.

[8] Payet, É. and F. Spoto, "Magic-Sets Transformation for the Analysis of Java Bytecode in Static Analysis". 2007, *Springer Berlin / Heidelberg*. p. 452-467.

[9] Fausto, S., "Nullness Analysis in Boolean Form", in *Proceedings of the* 2008 *Sixth IEEE International Conference on Software Engineering and Formal Methods*. 2008, IEEE Computer Society.

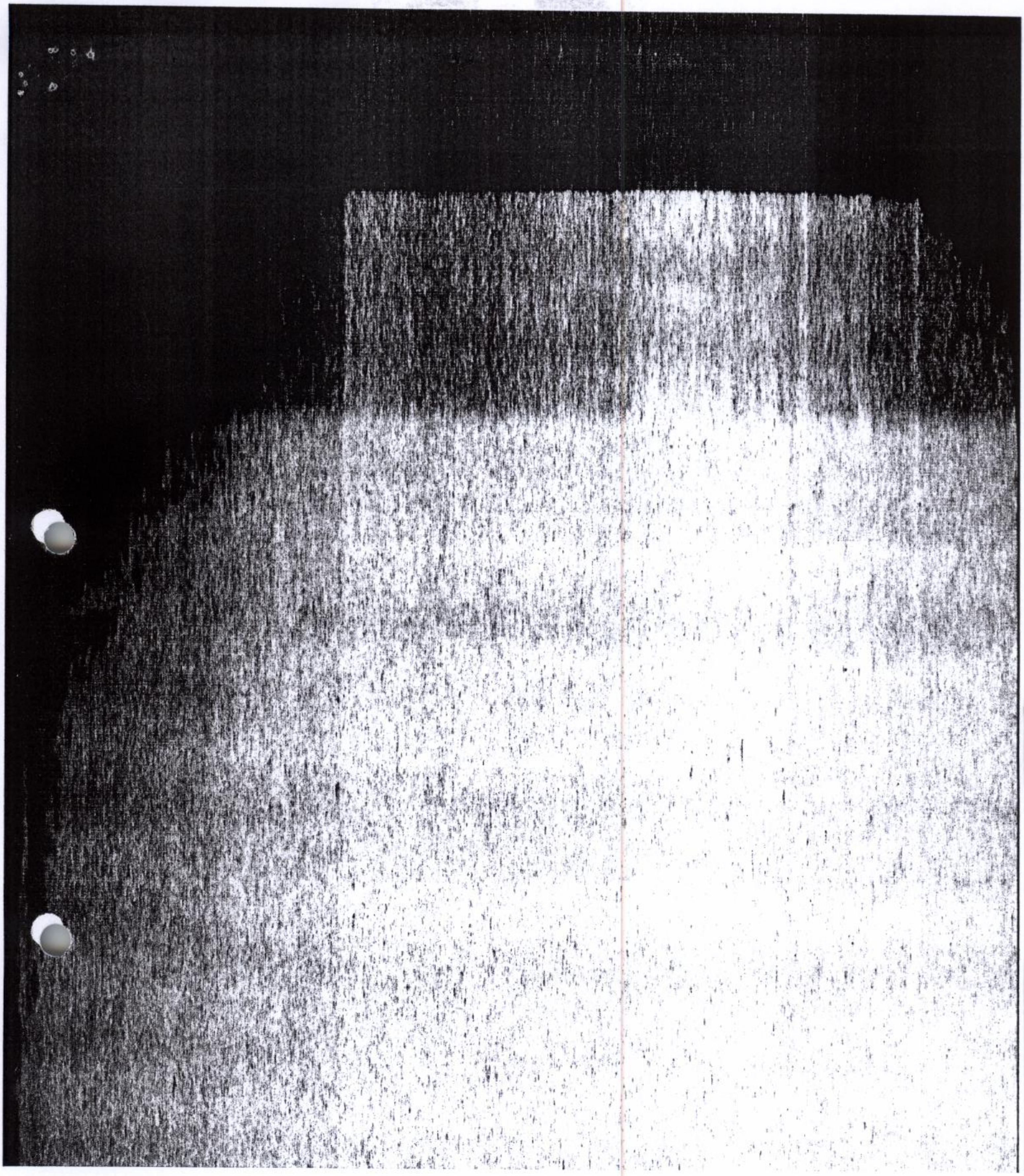
[10] Tsutomu, S. and M. Munehiro, "BDD representation for incompletely specifiedvmultiple-output logic functions and its applications to functional decomposition", in *Proceedings of the* 42*nd annual conference on Design automation*. 2005, *ACM: Anaheim, California, USA*.

[11] Cabodi, G. and M. Murciano, "BDD-Based Hardware Verification Formal Methods for Hardware Verification". Vol. 3965/2006. 2006: *Springer Berlin / Heidelberg*. 30.

[12] Eijk, C.A.J.v., "A BDD-based verification method for large synthesized circuits". *Integr. VLSI J.*, 1997. 23(2): p. 131-149.

[13] Liu, Y., et al., "Binary decision diagram with minimum expected path length", in *Proceedings of the conference on Design, automation and test in Europe*. 2001, IEEE Press: Munich, Germany.

[14] Rudiger, E., G. Wolfgang, and D. Rolf, "Combination of Lower Bounds in Exact BDD Minimization", in *Proceedings of the conference on Design, Automation and Test in Europe - Volume* 1. 2003, IEEE Computer Society.



- KING MONGKUT'S UNIVERSITY OF T[illegible]
- OKLAHOMA STATE UNIVERSITY, [illegible]
- FERN UNIVERSITY IN HAGEN, [illegible]
- CHEMNITZ UNIVERSITY OF T[illegible]
- EDITH COWAN UNIVERSITY, AUSTRALIA