วิจัยเชิงสำรวจเพื่อศึกษาความแตกต่างของเทคนิคเหมืองข้อมูลและรูปแบบการเรียนรู้
A Survey of Difference Data Mining Techniques and Learning Styles

ธิดารัตน์ เหลืองรุ่งเรือง (Tidarat Luangrungruang)[1] และอุรฉัตร โคแก้ว (Urachart Kokaew)[2]
[1]สาขาวิชาเทคโนโลยีสารสนเทศ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น
tidarat.l@kkumail.com, urachart@kku.ac.th

บทคัดย่อ

ประชากรในประเทศต้องได้รับการศึกษาที่เท่าเทียมกัน ไม่ว่าบุคคลนั้นจะมีความผิดปกติทางด้านร่างกาย อารมณ์ และสติปัญญา การศึกษาจะเป็นเครื่องมือที่ช่วยพัฒนาตนส่งผลให้มีคุณภาพชีวิตที่ดีขึ้น การจัดการศึกษาให้สอดคล้องกับการเรียนรู้แต่ละบุคคล ซึ่งมีลักษณะการจัดการศึกษาที่เฉพาะ งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอการทำเหมืองข้อมูลทางการศึกษา เพื่อค้นหารูปแบบการเรียนรู้ ซึ่งอาศัยเทคนิคต่าง ๆ ของเหมืองข้อมูล เช่น ต้นไม้ตัดสินใจ (Decision Tree), กฎของเบย์ (Bayesian Learning), แบบเนอีฟเบย์ (Naïve Bay)

คำสำคัญ: รูปแบบการเรียนรู้, เหมืองข้อมูล

Abstract

The population in the country must be equally educated. Regardless of the person's physical, emotional, and intellectual disabilities, education is a tool that helps develop oneself, resulting in better quality of life. Management of education in accordance with individual learning. It has a unique educational management style. This research aims to present educational data mining to find learning pattern based on data mining techniques such as Decision Tree, Bayesian Learning, Naïve Bay.

Keyword: Learning style, Data mining

1. บทนำ

บทการพัฒนาประเทศให้มั่งคงและยั่งยืนนั้น ต้องมาจากพื้นฐานที่แข็งแรง นั่นคือประชากรในประเทศต้องได้รับการศึกษาที่เท่าเทียมกันทุกคน สำหรับนักเรียนผู้พิการทางการได้ยินมีรูปแบบการเรียนรู้ที่ต่างกัน จึงต้องจัดการศึกษาให้สอดคล้องกับการเรียนรู้แต่ละบุคคล การทำความเข้าใจในสไตล์การเรียนรู้ของนักเรียน จึงมีความจำเป็นอย่างยิ่งในการนำข้อมูลสารสนเทศนั้นมาออกแบบการจัดการเรียนรู้ให้สอดคล้องกับสไตล์การเรียนรู้

ดังนั้น จึงมีแนวคิดในการนำเทคโนโลยีการทำเหมืองข้อมูลมาใช้ในการพยากรณ์ด้านการศึกษาโดยใช้เทคนิคต่าง ๆ เพื่อพิจารณาเทคนิคที่เหมาะสม ในการนำมาวิเคราะห์ข้อมูลปัจจัยที่ส่งผลต่อการเรียนรู้ เพื่อให้เกิดประโยชน์ต่อการวางแผนการเรียนการสอน

2. ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

2.1 เทคนิคเหมืองข้อมูล เป็นกระบวนการกระทำกับข้อมูลที่มีจำนวนมากแบบอัตโนมัติ เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลนั้น โดยที่รูปแบบที่ค้นพบนี้จะต้องมีความหมายและนำไปใช้ประโยชน์ในด้านต่าง ๆ ได้ [1] [2] [3] อัลกอริทึมที่ใช้ในการจำแนก ประเภทข้อมูลมีหลากหลายชนิด [4] เช่น ต้นไม้ตัดสินใจ (Decision Tree), กฎของเบย์ (Bayesian Learning), แบบเนอีฟเบย์ (Naïve Bay)

2.2 ต้นไม้ตัดสินใจ (Decision Tree) การเรียนรู้ของต้นไม้ตัดสินใจ (Decision Tree) เป็นการเรียนรู้โดยการจำแนกประเภท (Classification) ข้อมูลออกเป็นกลุ่ม (class) ต่าง ๆ โดยใช้คุณลักษณะ (attribute) [5] ข้อมูลในการจำแนกประเภทต้นไม้ตัดสินใจที่ได้จากการเรียนรู้ทำให้ทราบว่า คุณลักษณะใดเป็นตัวกำหนดการจำแนกประเภท และคุณลักษณะแต่ละตัวมีความสำคัญมากน้อยต่างกันอย่างไร เพราะฉะนั้นการจำแนก

ประเภทมีประโยชน์ช่วยให้ผู้ใช้สามารถวิเคราะห์ข้อมูลและตัดสินใจได้อย่างถูกต้องยิ่งขึ้น

ส่วนประกอบของผลลัพธ์ของการเรียนรู้ต้นไม้ตัดสินใจ

- โหนดภายใน (Internal node) คือ คุณลักษณะต่าง ๆ ของข้อมูล ซึ่งเมื่อข้อมูลใด ๆ ตกลงมาที่โหมด จะใช้คุณสมบัตินี้เป็นตัวตัดสินใจว่าข้อมูลจะไปในทิศทางใดโดยหนวดภายในที่เป็นจุดเริ่มต้นของต้นไม้ เรียกว่า โหนดราก - กิ่ง (Branch, Link) เป็นค่าของคุณลักษณะในโหนดภายในที่แตกกิ่งออก ซึ่งโหนดภายในจะแตกกิ่งเป็นจำนวนเท่ากับจำนวนค่าของคุณลักษณะในโหนดภายในนั้น

- โหนดใบ (Leaf node) คือกลุ่มต่าง ๆ ซึ่งเป็นผลลัพธ์ในการจำแนกประเภทข้อมูล

2.3 กฎของเบย์ (Bayesian Learning) [6]

การจำแนกประเภทโดยใช้กฎของเบย์เป็นการจำแนกประเภทแบบมีผู้สอนที่ต้องมีฉลากบอกประเภทข้อมูลอยู่เช่นเดียวกับ การเรียนรู้แบบต้นไม้ตัดสินใจการเรียนรู้เชิงอินสแตนซ์ แต่ความแตกต่างของกฎของเบย์จะใช้ความรู้ก่อนหน้ามาหาความน่าจะเป็นเพื่อทำนายการจำแนกประเภทข้อมูลในอนาคต

การจำแนกประเภทโดยใช้กฎของเบย์

เป็นการจำแนกประเภทโดยใช้หลักสถิติในการพยากรณ์ความน่าจะเป็นของสมาชิก

- ทฤษฎีของเบย์ (Bayesian theorem)

- เป็นการเรียนรู้เพิ่มเติม: ตัวอย่างใหม่ที่ได้จากการมาถูกนำมาปรับเปลี่ยนการแจกแจงซึ่งมีผลต่อการเพิ่ม/ลดความน่าจะเป็น ทำให้มีการเรียนรู้ที่เปลี่ยนไปวิธีการนี้ตัวแบบจะถูกปรับเปลี่ยนไปตามตัวอย่างใหม่ที่ได้โดยผนวกกับความรู้เดิมที่มี

- การทำนายค่าคลาสเป้าหมายของตัวอย่างใช้ความน่าจะเป็นมากที่สุดของทุกสมมติฐาน

2.4 แบบเนอีฟเบย์ (Naïve Bay)

เป็นวิธีที่ใช้กันอย่างแพร่หลาย [7] และใช้งานง่าย เป็นหนึ่งในอัลกอริทึมที่มีการเรียนรู้เร็วและสามารถจัดการกับคุณลักษณะหรือคลาสได้หลากหลาย [8]

เป็นการใช้หลักการความน่าจะเป็นทางสถิติเข้ามาช่วยในการทำนายความเป็นสมาชิก ซึ่งอยู่บนพื้นฐานของ Bayes' Theorem วิธีของเนอีฟเบย์เป็นวิธีการจำแนก ที่สามารถเรียนรู้เพิ่มเติมได้โดยการเพิ่มชุดข้อมูลที่ใช้ในการฝากสอน ข้อดีของวิธีการนี้ คือ มีความง่าย ต่อการนำไปใช้งาน เนื่องจากรูปแบบการคำนวณนั้นง่ายและผลลัพธ์ที่ได้สามารถประยุกต์ใช้ได้ดีเหมาะกับกรณีเซตตัวอย่างมีจำนวนมาก แต่มีข้อเสีย คือ จะให้ผลลัพธ์ที่ดีก็ต่อเมื่อคุณลักษณะของข้อมูลที่นำมาทดสอบเป็นอิสระต่อกัน

2.5 รูปแบบการเรียนรู้ VARK แนวคิดของ Neil Fleming [9] [10] [11]

วิธีการเรียนรู้ของตนเองว่าสามารถเรียนรู้ได้เร็วผ่านช่องทางใด กำหนดไว้ 4 ช่องทางดังนี้

1) ผู้เรียนที่เรียนรู้ผ่านทางสายตา (Visual)

2) ผู้เรียนที่เรียนรู้ด้วยการฟัง (Aural)

3) ผู้เรียนที่เรียนรู้ด้วยการอ่านหรือการเขียน (Read/Write)

4) ผู้เรียนที่เรียนรู้ด้วยการเคลื่อนไหวทางร่างกาย (Kinesthetic)

2.6 รูปแบบการเรียนรู้ของเดวิด เอ คอล์บ (David A. Kolb) [12] [13] เป็นการจำแนกรูปแบบการเรียนรู้ตามลักษณะการรับรู้ และประมวลผลสารสนเทศซึ่งแบ่งผู้เรียนออกเป็น 4 ประเภท คือ แบบอเนกนัย แบบดูดซึม แบบเอกนัย และแบบปรับปรุง

2.7 รูปแบบแบบการเรียนของ Felder and Silverman

รูปแบบนี้มาจากการศึกษาการเรียนรู้ของนักศึกษาวิศวกรรมศาสตร์ (Felder and Silverman, 1988) [14] และต่อมาได้ขยายแนวคิดไปสู่การศึกษาทั่วไปในสาขาวิทยาศาสตร์ (Felder, 1993) [15] โดยพิจารณาแบบการเรียนออกเป็น 5 มิติ แต่ละมิติมี 2 ขั้วที่ต่อเนื่องกัน ได้แก่

- การรับรู้ (ผ่านประสาท สัมผัส – การหยั่งรู้)
- การนำเข้า (การมอง – การฟัง)
- การจัดระบบ (อุปนัย – นิรนัย)
- การประมวลผล (การกระทำ – การคิดไตร่ตรอง)
- การทำ ความเข้าใจ (เป็นขั้นตอน – แบบองค์รวม)

งานวิจัยที่เกี่ยวข้อง

W. Paireekreng and T. Prexawanprasut [16] มีจุดมุ่งหมายเพื่อแก้ปัญหาในการระบุรูปแบบการเรียนรู้สำหรับนักเรียนใหม่ รูปแบบการเรียนรู้รูปแบบการเรียนรู้ได้รับการเสนอและเทคนิคการจำแนกชุดได้ดำเนินการ ผลการวิจัยแสดงให้เห็นว่าเทคนิคการจำแนกชุดที่นำเสนอทำได้ดีกว่าเทคนิคการจำแนกประเภทอื่น ๆ ที่ใช้ในการทดลอง โดยใช้รูปแบบการเรียนรู้แบบ Felder และ Silverman ทำให้ทราบว่าอัลกอริทึม Artificial Neural Networks (ANN) ให้ค่า accuracy rate มากที่สุดที่ 88.30

Phanthipha Petchboonmee, Duangkamol Phonak, and Monchai Tiantong [17] มีวัตถุประสงค์เพื่อการจัดรูปแบบการเรียนรู้และเปรียบเทียบประสิทธิภาพการจัดรูปแบบการเรียนรู้ของ David Kolb ของนักศึกษาภาควิชาคอมพิวเตอร์สารสนเทศ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา

Y. H. Hung, R. I. Chang, and C. F. Lin [18] การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อศึกษาความสัมพันธ์ระหว่างรูปแบบการเรียนกับผลการเรียนรู้ เพื่อให้คำแนะนำในการปรับตัวเพื่อเพิ่มประสิทธิภาพในการแก้ปัญหาได้มีการพัฒนากลไกการระบุลักษณะการเรียนรู้แบบไฮบริด (HLSI) ขึ้นอยู่กับอัลกอริทึมการจัดกลุ่มแบบ k-means

Oranuch Pantho and M. Tiantong [19] มีวัตถุประสงค์เพื่อจำแนกรูปแบบการเรียนรู้ VARK ของผู้เรียน โดยใช้ขั้นตอนวิธีการตัดสินใจ Tree C4.5 ข้อมูลเกี่ยวกับรูปแบบการเรียนรู้ของผู้เรียนเก็บรวบรวมโดยใช้แบบสอบถามที่ตอบโดยนักศึกษา 1,205 ราย ผลการศึกษาพบว่าการจำแนกประเภทการเรียนรู้ VARK ตามขั้นตอนของ Tree Decision C4.5 มีความถูกต้อง 83.40% และได้รับกฎ 108 ข้อ สรุปได้ว่าสามารถใช้อัลกอริทึมการตัดสินใจ Tree C4.5 เพื่อจำแนกรูปแบบการเรียนรู้ VARK ของผู้เรียนได้อย่างมีประสิทธิภาพ

Y.-C. Chang, W.-Y. Kao, C.-P. Chu, and C.-H. Chiu [20] บทความนี้เสนอกลไกการจำแนกประเภทการเรียนรู้เพื่อจำแนกรูปแบบการเรียนรู้ของนักเรียน กลไกที่เสนอช่วยปรับปรุงการจำแนก k-neighbor neighbor (k-NN) และรวมกับอัลกอริทึมทางพันธุกรรม (GA) เพื่อแสดงให้เห็นถึงความสามารถในการทำงานของกลไกที่เสนอกลไกการนำเสนอถูกนำมาใช้ในระบบการจัดการเรียนรู้แบบเปิด พฤติกรรมการเรียนรู้ของนักเรียนชั้นประถมศึกษาปีที่ 117 จำแนกตามกลไกที่เสนอ ผลการทดลองแสดงให้เห็นว่ากลไกการจำแนกประเภทที่เสนอสามารถจำแนกรูปแบบการเรียนรู้ของนักเรียนได้อย่างมีประสิทธิภาพ

T. R. Patil and S. Sherekar [21] บทความนี้ให้ข้อมูลเกี่ยวกับการประเมินประสิทธิภาพตามกรณีที่ถูกต้องและไม่ถูกต้องของการจำแนกข้อมูลโดยใช้Naïve Bayes และอัลกอริทึมการจำแนก J48 ขั้นตอนวิธี Naive Bayes ขึ้นอยู่กับความน่าจะเป็นและขั้นตอนวิธี j48 ขึ้นอยู่กับโครงสร้างการตัดสินใจ เอกสารฉบับนี้นำเสนอการประเมินผลเชิงเปรียบเทียบของผู้จัดจำแนก NAIVE BAYES และ J48 ในบริบทของชุดข้อมูลธนาคารเพื่อเพิ่มอัตราการเป็นบวกที่แท้จริงและลดอัตราการผิดนัดผิดพลาดของ defaulters แทนที่จะใช้ความแม่นยำในการจัดหมวดหมู่ที่สูงขึ้นโดยใช้เครื่องมือของ WEKA

N. B. H. Ahmad and S. M. Shamsuddin [22] เปรียบเทียบประสิทธิภาพของตัวจำแนกประเภทต่าง ๆ เช่น Bayes โครงสร้างการตัดสินใจและกฎการจัดหมวดหมู่ในการจำแนกรูปแบบการเรียนรู้ของนักเรียน การตั้งค่าและพฤติกรรมของนักเรียนในขณะที่ใช้ระบบ e - learning ได้รับการสังเกตและวิเคราะห์และได้เลือกแอตทริบิวต์ 20 รูปแบบ ในรูปแบบการเรียนรู้ของ Felder Silverman

3. วิธีดำเนินงาน

งานวิจัยนี้เป็นการค้นหางานวิจัยที่เกี่ยวข้องกับการรูปแบบการเรียนรู้และนำมาวิเคราะห์

4. ผลการดำเนินงาน

การเปรียบเทียบรูปแบบการเรียนรู้และเทคนิคที่ใช้

งานวิจัย	สิ่งที่ศึกษา	ผลการศึกษา
An Integrated Model for Learning Style Classification in University Students Using Data Mining Techniques [16]	รูปแบบการเรียนรู้ Felder Silverman และ Classification	Artificial Neural Networks มีผลลัพธ์ที่ดีที่สุด 88.30%
A Comparative Data Mining Technique for David Kolb's Experiential Learning Style Classification [17]	รูปแบบการเรียนรู้ David Kolb's และ Classification	J48 มีค่าผลลัพธ์ที่สูง 85.65%
Hybrid learning style identification and developing adaptive problem - solving learning activities [18]	รูปแบบการเรียนรู้ Felder Silverman และ Clustering	สามารถระบุรูปแบบการได้และรูปแบบการเรียนรู้แบบผสมผสานมีความสัมพันธ์กับประสิทธิภาพการเรียนรู้มาก
Using Decision Tree C4.5 Algorithm to Predict VARK Learning Styles [19]	รูปแบบการเรียนรู้ VARK ของ Felder Silverman	83.40%
A learning style classification mechanism for e - learning [20]	รูปแบบการเรียนรู้ VARK ของ Felder Silverman	สามารถระบุรูปแบบการเรียนรู้ได้อย่างมีประสิทธิภาพ
Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification [21]	การเปรียบเทียบอัลกอริทึม Naive Bayes และอัลกอริทึม J48	J48 มีประสิทธิภาพมากกว่าตัวจำแนก Naive Bayes
A Comparative Analysis of Mining Techniques for Automatic Detection of Student's Learning style [22]	รูปแบบการเรียนรู้ Felder Silverman	Tree ให้ความแม่นยำสูงสุด โดยมีความถูกต้อง 92.5%

5. สรุป

งานวิจัยนี้เป็นการศึกษาเทคนิคการทำเหมืองข้อมูลและรูปแบบการเรียนรู้ โดยผู้วิจัยได้ทำการศึกษางานวิจัยที่เกี่ยวข้องเพื่อค้นหาเทคนิคที่เหมาะสม และจากการศึกษาพบว่าเทคนิค

ต้นไม้ตัดสินใจ (Decision Tree) และโครงข่ายประสาทเทียม (Artificial Neural Networks) มีค่าความถูกต้อง และมีความเหมาะสมในการนำไปใช้สร้างแบบจำลองที่ส่งผลต่อการเรียนรู้

6. เอกสารอ้างอิง

[1] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[2] T. C. Sharma and M. Jain, "WEKA approach for comparative study of classification algorithm," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 2, no. 4, pp. 1925-1931, 2013.

[3] C. F. Lin, Y.-c. Yeh, Y. H. Hung, and R. I. Chang, "Data mining for providing a personalized learning path in creativity: An application of decision trees," *Computers & Education,* vol. 68, pp. 199-210, 2013.

[4] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowledge and information systems,* vol. 14, no. 1, pp. 1-37, 2008.

[5] X. Amatriain and J. M. Pujol, "Data mining methods for recommender systems," in *Recommender Systems Handbook*: Springer, 2015, pp. 227-262.

[6] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine learning,* vol. 29, no. 2-3, pp. 131-163, 1997.

[7] Y. Ko, "How to use negative class information for Naive Bayes classification," *Information Processing & Management,* vol. 53, no. 6, pp. 1255-1268, 2017.

[8] C.-H. Lee, "An Information-Theoretic Filter Approach for Value Weighted Classification Learning in Naive Bayes," *Data & Knowledge Engineering,* 2017.

[9] N. D. Fleming, *Teaching and learning styles: VARK strategies*. IGI Global, 2001.

[10] N. D. Fleming, "I'm different; not dumb. Modes of presentation (VARK) in the tertiary classroom," in *Research and Development in Higher Education, Proceedings of the 1995 Annual Conference of the Higher Education and Research Development Society of Australasia (HERDSA), HERDSA*, 1995, vol. 18, pp. 308-313.

[11] N. Fleming and D. Baume, "Learning Styles Again: VARKing up the right tree!," *Educational developments,* vol. 7, no. 4, p. 4, 2006.

[12] D. A. Kolb, "Disciplinary inquiry norms and student learning styles: Diverse pathways for growth," *The modern american college,* pp. 21-43, 1981.

[13] D. A. Kolb, *Experiential learning: Experience as the source of learning and development*. FT press, 2014.

[14] R. M. Felder and L. K. Silverman, "Learning and teaching styles in engineering education," *Engineering education,* vol. 78, no. 7, pp. 674-681, 1988.

[15] R. M. Felder, "Reaching the Second Tier--Learning and Teaching Styles in College Science Education," *Journal of college science teaching,* vol. 22, no. 5, pp. 286-90, 1993.

[16] W. Paireekreng and T. Prexawanprasut, "An integrated model for learning style classification in university students using data mining techniques," in *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2015 12th International Conference on*, 2015, pp. 1-5: IEEE.

[17] P. Petchboonmee, D. Phonak, and M. Tiantong, "A comparative data mining technique for david kolb's experiential learning style classification," *International Journal of Information and Education Technology,* vol. 5, no. 9, p. 672, 2015.

[18] Y. H. Hung, R. I. Chang, and C. F. Lin, "Hybrid learning style identification and developing adaptive problem-solving learning activities," *Computers in Human Behavior,* vol. 55, pp. 552-561, 2016.

[19] P. Oranuch and T. Monchai, "Using Decision Tree C4. 5 Algorithm to Predict VARK Learning Styles," 2016.

[20] Y.-C. Chang, W.-Y. Kao, C.-P. Chu, and C.-H. Chiu, "A learning style classification mechanism for e-learning," *Computers & Education,* vol. 53, no. 2, pp. 273-285, 2009.

[21] T. R. Patil and S. Sherekar, "Performance analysis of Naive Bayes and J48 classification algorithm for data classification," *International Journal of Computer Science and Applications,* vol. 6, no. 2, pp. 256-261, 2013.

[22] N. B. H. Ahmad and S. M. Shamsuddin, "A comparative analysis of mining techniques for automatic detection of student's learning style," in *Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on*, 2010, pp. 877-882: IEEE.