

การพัฒนาระบบแนะนำภาพยนตร์ด้วยโครงสร้างข้อมูลกราฟ

The Development of Movie Recommendation System with Graph Data Structure

ศุจินธร ทรงสิทธิเดช^{1,*} และณกร อินทร์พยุง¹

Suchinthorn Songsittidet^{1,*} and Nakorn Indra-Payoong¹

¹ คณะโลจิสติกส์ มหาวิทยาลัยบูรพา ประเทศไทย; Faculty of Logistics, Burapha University, ChonBuri, Thailand

* Corresponding author email: suchinthorn.s@psru.ac.th

บทคัดย่อ

วัตถุประสงค์: เพื่อออกแบบและประเมินประสิทธิภาพวิธีการแนะนำภาพยนตร์โดยการประยุกต์ใช้แนวคิดจากโครงสร้างกราฟข้อมูล

วิธีการศึกษา: การศึกษาครั้งนี้ใช้ตัวอย่างชุดข้อมูล MovieLens ประกอบด้วย จำนวนข้อมูลคะแนนความชอบ 100,000 เรคคอร์ด ภาพยนตร์ 1,682 เรื่อง โดยมาจากผู้ชมภาพยนตร์ 943 คน ซึ่งมีการวิเคราะห์ 2 ส่วน ได้แก่ 1) การแนะนำด้วยคะแนนความชอบภาพยนตร์ (Rating) แบ่งกลุ่มผู้ชมภาพยนตร์อาศัยวิธี K - mean และ 2) การแนะนำด้วยวิธีต้นไม้แบบทอดข้ามของค่าน้ำหนักมากที่สุดในโครงสร้างข้อมูลแบบกราฟ แบ่งกลุ่มตามคุณลักษณะผู้ชมภาพยนตร์

ข้อค้นพบ: การแนะนำภาพยนตร์ Top – 10 ด้วยคะแนนความชอบจากการแบ่งกลุ่มผู้ชมภาพยนตร์ 5 กลุ่ม ด้วย K – mean Clustering ได้ค่าเฉลี่ยความถูกต้องในการแนะนำ ร้อยละ 28.16 ในขณะที่การแนะนำด้วยโครงสร้างข้อมูลแบบกราฟในการแบ่งกลุ่มผู้ชมภาพยนตร์จากคุณลักษณะของชุดข้อมูล ได้แก่ เพศ ช่วงอายุ และอาชีพ จำนวน 111 กลุ่ม ได้ค่าเฉลี่ยความถูกต้องในการแนะนำ ร้อยละ 87.45

การประยุกต์ใช้จากการศึกษานี้: การประยุกต์ใช้วิธีค่าน้ำหนักมากที่สุดในแบบทอดข้ามในโครงสร้างข้อมูลแบบกราฟเป็นการแนะนำอาศัยความสัมพันธ์ของชุดข้อมูลการชมภาพยนตร์จึงทำให้การแนะนำมีประสิทธิภาพเพิ่มขึ้น

คำสำคัญ: การแนะนำ โครงสร้างข้อมูลกราฟ ค่าน้ำหนักมากที่สุดในแบบทอดข้าม

Abstract

Purpose: The objective of this research is to design and evaluate the efficiency a movie recommendation process with graph data structure.

Methodology: The MovieLens dataset contains 100,000 records on 1,682 movies from 943 users. There are two parts of the study 1) the recommendation based on movie preference ratings by K - mean clustering method and 2) the recommendation based on a spanning tree of maximum weights in graph data structure by user's attributions.

Findings: The recommendations for top – 10 movies based on movie preference ratings from 5 user groups by K – mean Clustering. The result has shown that the average recommendation accuracy is 28.16%. In addition to the recommendation for top-10 movies based on graph data structure from 111 user groups by user's attributions, such as sex, age range, and occupation found that the average recommendation accuracy is 87.45%.

Applications of this study: The results indicated that the proposed maximum weight spanning tree in graph data structure can recommend movies to watching more efficiently.

Keywords: Recommendation system, Graph data structure, Maximum spanning tree

1. บทนำ

เทคโนโลยีที่เปลี่ยนแปลงไปอย่างรวดเร็วมีผลต่อการดำรงชีวิตที่เปลี่ยนแปลงไปตามยุคสมัย ทำให้ผู้ใช้บริการต้องการความสะดวกสบาย ผู้ให้บริการจึงให้ความสำคัญต่อความสามารถในการตอบสนองความต้องการต่อลูกค้าให้ได้มากขึ้น โดยการประยุกต์ใช้โปรแกรมในลักษณะหลากหลาย ระบบให้คำแนะนำ (Recommender System) เป็นเครื่องมือหนึ่งสำหรับช่วยแนะนำรายการสินค้าและบริการ เพื่อแก้ปัญหาการตัดสินใจของผู้ใช้บริการจากข้อมูลที่มีเป็นจำนวนมาก ระบบให้คำแนะนำช่วยให้ผู้ใช้บริการเลือกสินค้าได้ตรงความต้องการอย่างมีประสิทธิภาพ อีกทั้งยังเป็นระบบการบริการข้อมูลสนับสนุนการตัดสินใจของผู้ใช้บริการและให้คำแนะนำอย่างเหมาะสมทั้งผลิตภัณฑ์หรือสินค้า ข้อมูลข่าวสารและการบริการ ในพื้นที่ร้านค้าอิเล็กทรอนิกส์ รวมถึงการให้บริการมัลติมีเดีย การให้บริการหาเส้นทางอินเทอร์เน็ตและพื้นที่อื่น ๆ อีกมากมาย (Falk, 2019) ซึ่งเป็นส่วนหนึ่งของเหตุผลที่บริษัทต่าง ๆ ประยุกต์ใช้เพื่อพัฒนาติดตั้งระบบในการวิเคราะห์พฤติกรรมของผู้ใช้พิจารณาการแนะนำผลิตภัณฑ์ บริการและข้อมูลที่เหมาะสมที่สุด (Walek & Fojtik, 2020)ระบบการแนะนำเป็นระบบที่ถูกสร้างมาเพื่อพยายามแนะนำเสนอข้อมูล ผลิตภัณฑ์ ผู้คน สินค้าหรือบริการที่คาดว่าผู้ใช้น่าจะสนใจหรืออาจเป็นข้อมูลที่ต้องการสำหรับการดำเนินธุรกิจ ความท้าทายของการศึกษาระบบแนะนำนอกจากการปรับปรุงอัลกอริทึมให้เหมาะสมกับแต่ละปัญหาแล้วอีกส่วนหนึ่งที่สำคัญ

คือ การปรับปรุงคุณภาพสำหรับผู้ใช้ที่ต้องการแนะนำให้นำเชื่อถือที่จะช่วยค้นหาสินค้าที่ชอบได้อย่างถูกต้องและสม่ำเสมอ (Sarwar, et al., 2001)

ชุดข้อมูลการซื้อของลูกค้าที่มีมากมายสามารถนำไปใช้ประโยชน์ด้านการแนะนำรายการสินค้าให้กับลูกค้าในครั้งต่อไปด้วยการใช้ข้อมูลส่วนบุคคลของลูกค้า ข้อมูลพื้นฐานเกี่ยวกับพฤติกรรมการซื้อสินค้า ความชอบและความสนใจของลูกค้าใช้เป็นเครื่องมือสำคัญด้านการตลาด ซึ่งข้อมูลการซื้อของลูกค้านี้เป็นกลไกที่เป็นผลจากกระบวนการตัดสินใจของมนุษย์และช่วยคาดการณ์ความสนใจของลูกค้าด้วย การแนะนำผลิตภัณฑ์ที่แตกต่างและหลากหลายจะสามารถปรับให้เข้ากับรสนิยมของลูกค้าได้ (Kuzelewska, 2014) ซึ่งเป็นส่วนหนึ่งในการเพิ่มยอดขายได้จริง การพิจารณาความซับซ้อนเพื่อแนะนำรายการสินค้าเพื่อให้ได้ผลลัพธ์ที่ใกล้เคียงกัน จากการคาดการณ์ความชอบโดยดูจากพฤติกรรมการซื้อในอดีตของลูกค้ารวมถึงลูกค้ารายอื่นด้วย (Ivarsson & Lindgren, 2016) การแนะนำรายการสินค้าที่น่าสนใจให้แก่ลูกค้าสามารถทำได้อย่างแม่นยำมากขึ้น โดยการพิจารณาจากความชอบหรือการให้คะแนนจากลูกค้าโดยอ้างอิงการให้คะแนนจากลูกค้าที่มีคุณสมบัติคล้ายคลึงกันให้แก่ลูกค้าเป้าหมายจึงเป็นเทคนิคการแนะนำที่ได้รับความนิยม

ตัวอย่างชุดข้อมูลที่ใช้ในการทดสอบครั้งนี้ใช้ชุดข้อมูล MovieLens ดำเนินการโดย Grouplens ซึ่งชุดข้อมูล MovieLens ได้เผยแพร่ครั้งแรกในปี ค.ศ.1998 ใช้อธิบายการแสดงความชอบดูภาพยนตร์ของผู้ชมได้ (Harper & Konstan, 2015) ข้อมูลความชอบดูภาพยนตร์นำเสนอในรูปแบบตารางข้อมูลของผู้ใช้ระบบในแต่ละรายการภาพยนตร์พร้อมด้วยรหัสโปรเจกต์ที่สามารถบ่งบอกพื้นที่ที่อยู่อาศัยได้ โดยผลคะแนนความชอบที่แสดงด้วยดาวตั้งแต่ 0 ถึง 5 ดาว ซึ่งเป็นที่นิยมในงานวิจัยเกี่ยวกับการแนะนำ ชุดข้อมูล MovieLens เป็นผลที่ได้มาจากผู้ใช้งานเป็นระยะเวลามากกว่า 1 ปีและเป็นระบบออนไลน์แบบพลวัต ซึ่งมีเพียงผู้ใช้งานเท่านั้นที่สามารถให้คะแนนภาพยนตร์ที่เคยชมได้ และคะแนนของผู้ใช้ระบบจะมีอิทธิพลต่อการคาดการณ์คะแนนในภาพยนตร์แต่ละเรื่องเหล่านั้นด้วย (Cosley, et al., 2003) ดังนั้นชุดข้อมูล MovieLens จึงถูกใช้เป็นชุดข้อมูลสำหรับการวิจัยอย่างแพร่หลาย

การศึกษานี้เป็นการนำเสนอวิธีการแนะนำของชุดข้อมูลภาพยนตร์โดยเปรียบเทียบการแนะนำด้วยการแบ่งกลุ่ม (Clustering) ของผู้ใช้งานคือ ลูกค้าที่ชมภาพยนตร์ เพื่อศึกษาการแนะนำภาพยนตร์จากการให้ค่าคะแนนความชอบ (Rating) และการแนะนำด้วยความสัมพันธ์โครงสร้างข้อมูลกลุ่มผู้ชมภาพยนตร์ (Graph Data Structure) นำมาทดสอบความถูกต้องในการแนะนำเพื่อประเมินประสิทธิภาพที่จะนำมาใช้แนะนำภาพยนตร์ต่อผู้ใช้ได้ (Adomavicius & Kwon, 2007)

2. วัตถุประสงค์

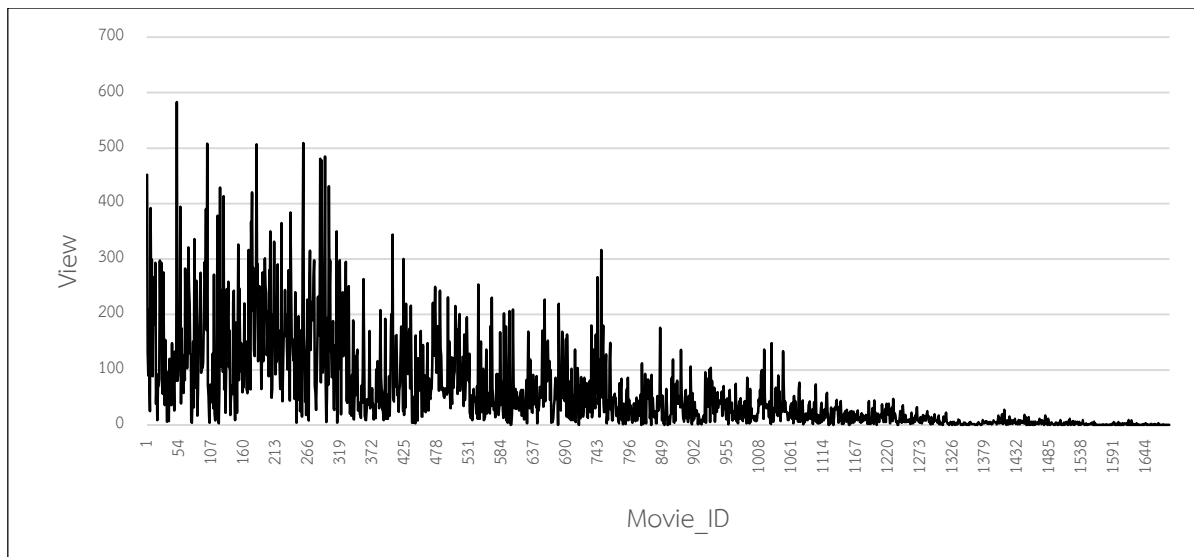
1) เพื่อออกแบบวิธีการแนะนำภาพยนตร์ด้วยคะแนนความชอบและด้วยโครงสร้างข้อมูลกราฟจากชุดข้อมูลการชมภาพยนตร์

2) เพื่อประเมินประสิทธิภาพวิธีการแนะนำภาพยนตร์จากชุดข้อมูลการชมภาพยนตร์

3. วิธีการศึกษา

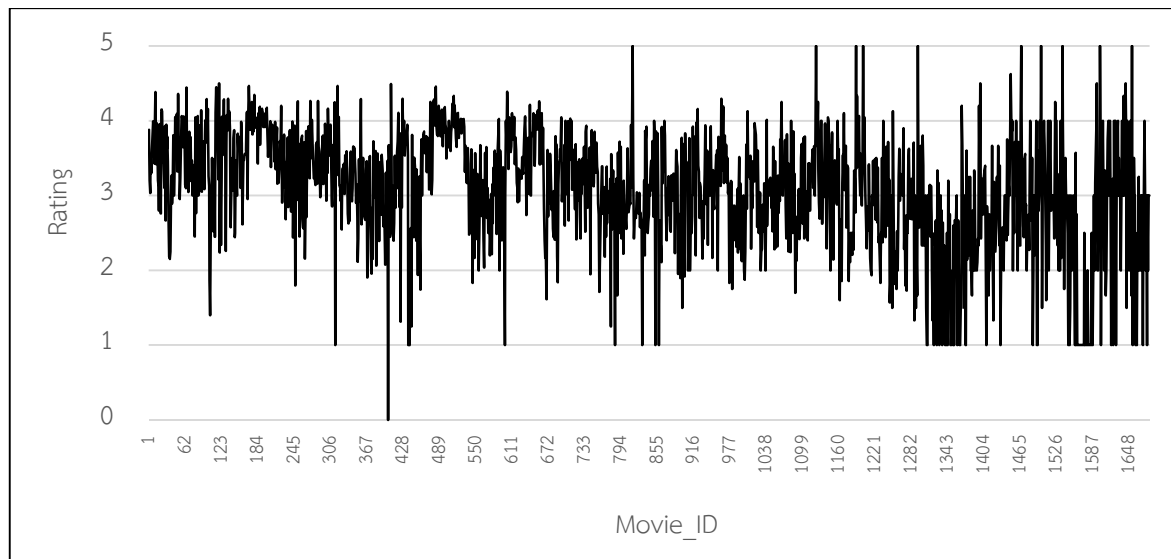
3.1 วิธีการและการออกแบบ

การศึกษานี้ใช้ชุดข้อมูลของผู้ใช้งานในระบบที่อาศัยการให้คะแนนความชอบที่มีต่อรายการสินค้า โดยชุดข้อมูลที่นำมาทดสอบนี้ผู้วิจัยได้เลือกชุดข้อมูลจาก MovieLens นำมาหาความสัมพันธ์ระหว่างผู้ใช้และรายการภาพยนตร์เพื่อให้ตรงตามวัตถุประสงค์ของการทำวิจัย ซึ่งการนำชุดข้อมูลจาก MovieLens เป็นข้อมูลที่เปิดเผยจากเว็บไซต์ Grouplens.org ใช้ในการทดสอบเป็นกลุ่มตัวอย่าง โดยการรวมชุดข้อมูลความชื่นชอบของผู้ใช้ระบบที่มีต่อรายการสินค้าในที่นี้คือภาพยนตร์ของผู้ใช้ระบบที่เคยชมภาพยนตร์ไปแล้ว ชุดข้อมูลที่ให้ทดสอบครั้งนี้ ใช้ชุดข้อมูล MovieLens_100K ซึ่งเก็บชุดข้อมูลภาพยนตร์ (Movie) ตั้งแต่ปี ค.ศ. 1526 – ค.ศ. 2000 ประเภทหนัง เช่น การ์ตูน ตลกเบาสมอง ผจญภัย สารคดี ต่อสู้ เป็นต้น ทั้งหมด 1,682 เรื่อง ชุดข้อมูลผู้ใช้ระบบ ประกอบด้วย เพศ อายุ และอาชีพ ทั้งหมด 943 คน และชุดข้อมูลการให้คะแนนความชอบ 1 – 5 (Rating Score) 100,000 ระเบียบข้อมูล (Record) โดยแสดงในภาพที่ 1



ภาพที่ 1 ปริมาณการเข้าชมภาพยนตร์ในระบบ

จากภาพที่ 1 ข้อมูลปริมาณการเข้าชมภาพยนตร์ระหว่างภาพยนตร์แสดงด้วย รหัส ID และความถี่ในการรับชมภาพยนตร์ในแต่ละเรื่องจากผู้ชมภาพยนตร์ทั้งหมด 1,682 เรื่อง ตัวอย่างเช่น Movie_ID1 เรื่อง Toy Story ซึ่งเข้าฉายในปี ค.ศ.1995 เป็นภาพยนตร์ประเภท ผจญภัย เบาสมอง การ์ตูน จินตนาการสำหรับเด็ก มีจำนวนผู้ชม 450 คนจาก 943 คน เป็นต้น โดยการให้คะแนนความชอบของภาพยนตร์ (Rating Score) แต่ละเรื่องแสดงดังภาพที่ 2



ภาพที่ 2 การให้คะแนนความชอบภาพยนตร์ในระบบ

จากภาพที่ 2 แสดงข้อมูลระหว่างค่าเฉลี่ยของคะแนนความชอบของภาพยนตร์และภาพยนตร์ที่แสดงด้วยรหัส ID ของภาพยนตร์ในแต่ละเรื่องจากผู้ชมภาพยนตร์ทั้งหมด 1,682 เรื่อง ตัวอย่างเช่น Movie_ID1 เรื่อง Toy Story ที่มีผู้เคยชมภาพยนตร์ได้ให้คะแนนความชอบไว้ทั้งหมด 450 คน มีคะแนนความชอบเฉลี่ย 3.86 คะแนน เป็นต้น ซึ่ง Adomavicius, Manouselis & Kwon. (2011) ได้กล่าวถึงการแนะนำโดยการให้คะแนนเป็นวิธีที่ตรงไปตรงมาด้วยคะแนนที่จะคาดการณ์ให้ผู้ใช้ที่มีปัจจัยพื้นฐานเหมือนกันซึ่งเป็นระบบการแนะนำแบบดั้งเดิม แม้ว่าจะเป็นหลักการที่ถูกใช้เป็นจำนวนมากแต่การตัดสินใจควรพิจารณาหลายมุมมองเพื่อใช้ในการตัดสินใจทางเลือกที่ดีที่สุด

คุณลักษณะของข้อมูลผู้ใช้งานระบบ ประกอบด้วย เพศ อายุ และอาชีพ ผู้วิจัยจึงทำการแบ่งชุดข้อมูลออกเป็นข้อมูลการเรียนรู้ 900 ชุดข้อมูล (95%) และข้อมูลทดสอบ 43 ชุดข้อมูล (5%) จากข้อมูลทั้งหมดของผู้ใช้ระบบเพื่อทดสอบความถูกต้องในการแนะนำภาพยนตร์ เนื่องจากชุดข้อมูลผู้ชมภาพยนตร์ในระบบมีจำนวนไม่มาก อีกทั้งการแนะนำภาพยนตร์ด้วยการจัดกลุ่มจึงต้องการสัดส่วนของชุดข้อมูลการเรียนรู้จำนวนมากเพื่อทำการแบ่งกลุ่มผู้ชมภาพยนตร์ โดยจัดกลุ่มคุณสมบัติของข้อมูลผู้ใช้งานระบบทั้งหมดดังตารางที่ 1

ตารางที่ 1 ข้อมูลของชุดการเรียนรู้และทดสอบ

ข้อมูล	การบัญญัติ	ข้อมูลการเรียนรู้	ข้อมูลทดสอบ
เพศ	1: ชาย	643	27
	2: หญิง	257	16
ช่วงอายุ	1: < 14 ปี	11	0
	2: 15-23 ปี	181	9
	3: 24-41 ปี	461	19

ข้อมูล	การบัญชี	ข้อมูลการเรียนรู้	ข้อมูลทดสอบ
ช่วงอายุ	4: 42-56 ปี	204	12
	5: 57-75 ปี	43	3
อาชีพ	1: administrator	77	2
	2: artist	26	2
	3: doctor	6	1
	4: educator	90	1
	5: engineer	65	0
	6: entertainment	16	0
	7: executive	31	3
	8: healthcare	15	0
	9: homemaker	7	1
	10: lawyer	12	1
	11: librarian	48	1
	12: marketing	26	0
	13: programmer	26	1
	14: retired	44	3
	15: salesman	65	10
	16: scientist	14	0
	17: student	11	7
	18: technician	28	2
	19: writer	186	2
	20: none	9	1
	21: other	98	1

3.2 ความคล้ายคลึง (Similarity)

การใช้เทคนิคในการหาความคล้ายคลึงระหว่างผู้ใช้ ซึ่งเป็นเทคนิคที่ใช้กันบ่อยที่สุดเพื่อหาความเชื่อมโยงความสัมพันธ์ด้วยข้อมูลความถี่ในการดูภาพยนตร์ของผู้ใช้ระบบในโครงสร้างข้อมูล (Adomavicius & Kwon, 2007) จากชุดข้อมูลผู้ใช้ระบบที่เข้าชมภาพยนตร์ทั้งหมด 943 คน ของผู้ใช้ระบบที่เคยชมภาพยนตร์ทั้งหมด 1,682 เรื่อง การคำนวณหาค่าความคล้ายคลึงในลักษณะคู่ข้อมูล I และ J ด้วยวิธี Cosine Similarity ดังแสดงในสมการ ดังนี้

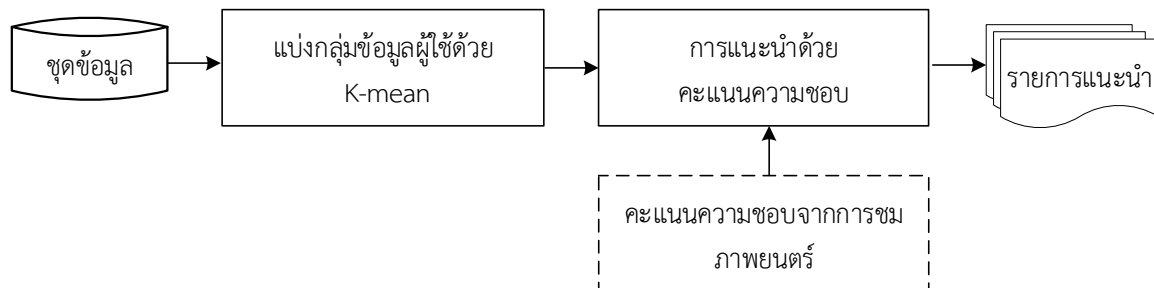
$$\text{Cosine Similarity (I, J)} = \frac{\sum_{i=1}^N I_i J_i}{\sqrt{\sum_{i=1}^N (I_i)^2} \times \sqrt{\sum_{i=1}^N (J_i)^2}}$$

3.3 การแบ่งกลุ่ม

การวิเคราะห์ข้อมูลด้วยชุดข้อมูลเป็นกลุ่มที่ข้อมูลมีความคล้ายกันจัดไว้ในกลุ่มเดียวกัน เทคนิคที่ใช้ส่วนมากมี 2 เทคนิค คือ Hierarchical cluster analysis และ K - mean cluster analysis ซึ่งมีวัตถุประสงค์ที่ต่างกัน โดยการข้อมูลแบบ K - mean เป็นวิธีที่ถูกนำมาใช้บ่อยมากที่สุด เนื่องจากไม่ยุ่งยากซับซ้อนและเข้าใจง่าย โดยอาศัยศูนย์กลางของข้อมูลแต่ละกลุ่มและคำนวณระยะทางกับข้อมูลในชุดข้อมูลด้วยวิธียูคลิดี้น จากนั้นข้อมูลจะถูกจัดให้อยู่ในกลุ่มศูนย์กลางที่มีระยะใกล้ที่สุด (Kesorn, 2021)

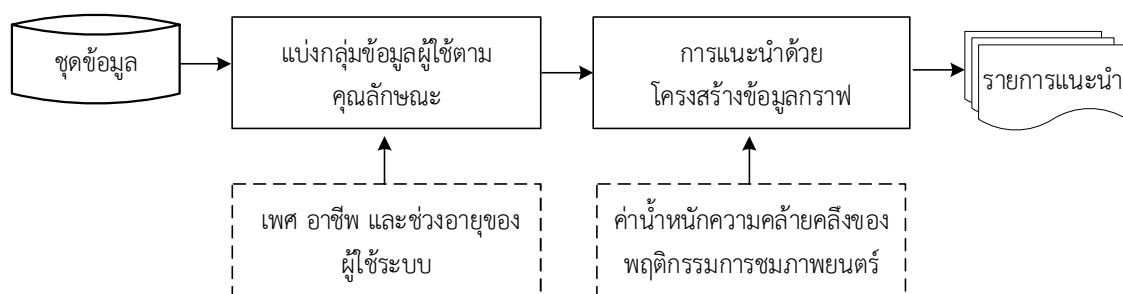
ดังนั้น การศึกษานี้เป็นการนำเสนอวิธีการใช้ชุดข้อมูลตัวอย่าง ซึ่งออกแบบและประเมินผลการแนะนำโดยจำแนกชุดข้อมูลการเรียนรู้และชุดข้อมูลทดสอบจากผู้ชมภาพยนตร์ เพื่อสนับสนุนการพัฒนาระบบโครงสร้างข้อมูลกราฟสำหรับระบบการแนะนำสินค้าหรือบริการ โดยออกแบบการวิเคราะห์ข้อมูล ดังนี้

1) การวิเคราะห์การแนะนำด้วยคะแนนความชอบ (Rating) โดยแบ่งกลุ่มผู้ชมภาพยนตร์ด้วยวิธี K-mean เพื่อแนะนำผู้ชมภาพยนตร์ตามกลุ่ม



ภาพที่ 3 กรอบแนวคิดของการแนะนำด้วยคะแนนความชอบ

2) การวิเคราะห์การแนะนำด้วยโครงสร้างข้อมูลกราฟและใช้วิธี Maximum spanning tree (MST) ในการแนะนำภาพยนตร์ที่มีความคล้ายคลึงกันมากที่สุด โดยแบ่งกลุ่มผู้ชมภาพยนตร์ด้วยคุณลักษณะของผู้ชมภาพยนตร์ ประกอบด้วย เพศ อายุและอาชีพ เพื่อแนะนำผู้ชมภาพยนตร์ตามกลุ่ม



ภาพที่ 4 กรอบแนวคิดของการแนะนำด้วยโครงสร้างข้อมูลกราฟ

4. ผลการศึกษา

4.1 การวิเคราะห์การแนะนำด้วยคะแนนความชอบ (Rating)

การให้ความสนใจวิธีการแนะนำที่พิจารณาถึงความชอบโดยรวมของผู้ใช้ จากความชอบที่มีต่อลักษณะต่าง ๆ ของวัตถุมากกว่าหนึ่งลักษณะ โดยข้อมูลการชมภาพยนตร์ 1,682 เรื่อง ของผู้ชมภาพยนตร์ทั้งหมด 900 คน สามารถออกแบบการแนะนำภาพยนตร์ จากความสนใจของผู้ชมภาพยนตร์โดยการแบ่งกลุ่มผู้ชมภาพยนตร์ด้วยข้อมูลส่วนตัว ได้แก่ เพศ อายุ และอาชีพ ด้วยวิธีการแบ่งกลุ่มข้อมูล K – mean ซึ่งเป็นวิธีที่เหมาะสมกับจำนวนข้อมูลและได้รับความนิยมเพื่อแบ่งกลุ่มข้อมูล สามารถแบ่งได้ 5 กลุ่ม ดังตารางที่ 2

ตารางที่ 2 จำนวนกลุ่มผู้ชมภาพยนตร์ชุดการเรียนรู้

Number of cases in each cluster		
Cluster	1	117
	2	174
	3	242
	4	218
	5	149
Valid		900

จากตารางที่ 2 พบว่า จำนวนสมาชิกกลุ่มด้วยวิธี K-mean จากข้อมูลชุดการเรียนรู้ด้วยข้อมูลส่วนตัวผู้ชมภาพยนตร์ได้ 5 กลุ่มดังนี้ กลุ่มที่ 1 จำนวน 117 คน กลุ่มที่ 2 จำนวน 174 คน กลุ่มที่ 3 จำนวน 242 คน กลุ่มที่ 4 มีจำนวน 218 คนและกลุ่มที่ 5 มีจำนวน 149 คน สามารถแสดงข้อมูลจากการให้คะแนนความชอบ (Rating score) โดยเฉลี่ยแต่ละกลุ่มได้ดังตารางที่ 3

ตารางที่ 3 การให้คะแนนความชอบในกลุ่มผู้ชมภาพยนตร์

_Id	M1	M2	M3	M4	M5	----->	M1678	M1679	M1680	M1681	M1682
G1	1.70	0.26	0.25	0.67	0.21	----->	0.00	0.00	0.00	0.00	0.00
G2	1.64	0.40	0.22	0.86	0.34	----->	0.00	0.00	0.00	0.00	0.00
G3	1.94	0.57	0.41	0.90	0.33	----->	0.00	0.01	0.01	0.00	0.00
G4	2.00	0.41	0.28	0.81	0.25	----->	0.00	0.00	0.00	0.00	0.00
G5	1.74	0.54	0.19	0.65	0.34	----->	0.00	0.00	0.00	0.02	0.00

การแบ่งจำนวนของผู้ชมภาพยนตร์ตามกลุ่มด้วยเพศ ช่วยอายุ และอาชีพ ได้ 5 กลุ่ม นำไปคำนวณค่าเฉลี่ยของคะแนนความชอบจากคะแนนที่มีผู้ชมภาพยนตร์ 1,682 เรื่องได้ให้ไว้ตั้งแต่ 1 – 5 คะแนน เพื่อแนะนำภาพยนตร์ให้แก่ผู้ชมภาพยนตร์ที่อยู่กลุ่มเดียวกันได้ เมื่อทราบจำนวนผู้ชมภาพยนตร์แล้วสามารถคำนวณหาการแนะนำจากการให้คะแนนความชอบจากการประวัติชมภาพยนตร์ของผู้ชมภาพยนตร์ในแต่ละ

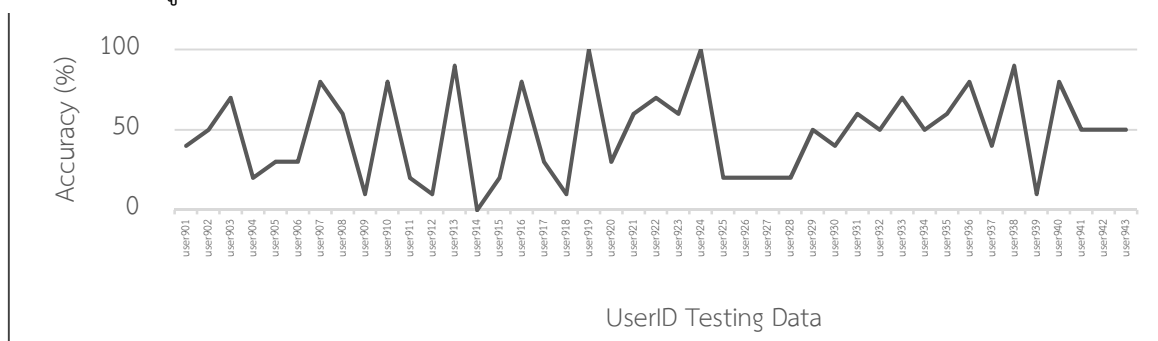
กลุ่มได้โดยจะเลือกแนะนำภาพยนตร์ที่เหมาะสมที่สุดจากการให้ Rating ของผู้ชม โดยภาพยนตร์ที่ได้แนะนำ 10 อันดับแรก (Top – 10) ของแต่ละกลุ่ม ตัวอย่างภาพยนตร์ที่ถูกแนะนำให้ผู้ชมในระบบที่ User_ID: 943 ดังภาพที่ 5

USER_ID: 943

Rank	Movie_Name (Entry)
1	Star Wars (1977)
2	Return of the Jedi (1983)
3	Contact (1997)
4	Scream (1996)
5	Pulp Fiction (1994)
6	Empire Strikes Back (1980)
7	Toy Story (1995)
8	Fargo (1996)
9	Liar Liar (1997)
10	Twelve Monkeys (1995)

ภาพที่ 5 ตัวอย่างภาพยนตร์ที่แนะนำด้วยคะแนนความชอบ

จากภาพที่ 5 ตัวอย่างภาพยนตร์ที่แนะนำด้วยคะแนนความชอบให้แก่ผู้ใช้ในระบบของ User_ID: 943 ซึ่งเป็นสมาชิกกลุ่มรหัส G3 จากการแบ่งกลุ่มข้อมูลด้วยวิธี K – mean พบว่า ภาพยนตร์ที่ถูกชมประกอบด้วย Star Wars, Return of the Jedi, Toy Story และ Raider of the Lost Ark ทั้งหมด 5 เรื่อง โดยการวิจัยครั้งนี้ผู้วิจัยทำการทดสอบความถูกต้อง (Accuracy) เพื่อพิจารณาการประเมินผลลัพธ์การแนะนำภาพยนตร์ให้แก่ผู้ชมทั้งหมด 43 คน แสดงดังภาพที่ 6



ภาพที่ 6 ความถูกต้องในการแนะนำด้วยคะแนนความชอบ (Rating)

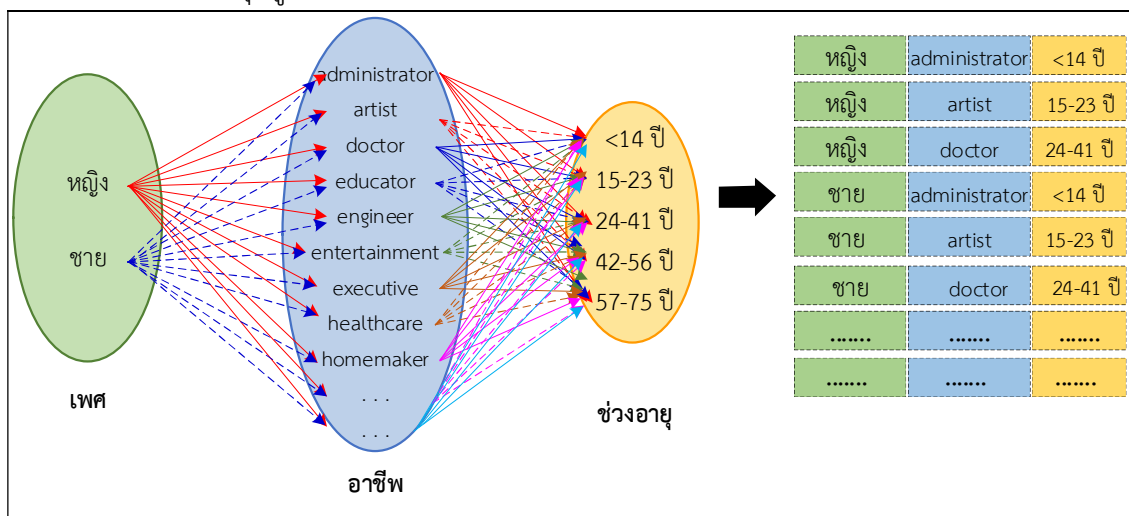
จากภาพที่ 6 แสดงความสัมพันธ์ระหว่างเปอร์เซ็นต์ความถูกต้อง (Accuracy) และผู้ชมภาพยนตร์ที่เป็นชุดข้อมูลทดสอบจำนวน 43 คน พบว่าเปอร์เซ็นต์ความถูกต้องมีค่าเฉลี่ยเท่ากับ 28.16% และในข้อมูลทดสอบที่ค่าความถูกต้องมีค่า 0% จากผู้ชมภาพยนตร์ 1 คน เนื่องจากเป็นผู้ชมภาพยนตร์ที่ไม่มีคะแนนความชอบภาพยนตร์ตามประวัติการชมภาพยนตร์ในกลุ่มผู้ชมเลย ดังนั้นความสนใจรายบุคคลในการ

ชมหรือขึ้นชอบภาพยนตร์ที่บ่งบอกโดยการให้คะแนนความชอบภาพยนตร์จึงมีผลต่อการเลือกชมภาพยนตร์จากการแนะนำ การแบ่งกลุ่มผู้ชมภาพยนตร์จำเป็นต้องคำนึงถึงคุณลักษณะของข้อมูลเฉพาะของผู้ชมภาพยนตร์ด้วย การแนะนำจากคะแนนความชอบเพียงอย่างเดียวไม่สามารถแสดงถึงความสัมพันธ์หรือความชอบที่คล้ายคลึงกันของผู้ชมภาพยนตร์ได้ จึงนำมาสู่การประยุกต์ใช้แนวคิดจากโครงสร้างกราฟข้อมูลเพื่อแนะนำภาพยนตร์เพื่อให้มีความถูกต้องในการแนะนำภาพยนตร์ให้มากยิ่งขึ้น

4.2 การประยุกต์ใช้โครงสร้างกราฟข้อมูลเพื่อแนะนำภาพยนตร์

การเชื่อมโยงโหนด (Connected) ภายใต้งราฟโดยมีน้ำหนักกราฟข้อมูลบนเส้นเชื่อมโยงระหว่างโหนดการประยุกต์ใช้โครงสร้างกราฟข้อมูลด้วยหลักการต้นไม้แบบทอดข้าม (Spanning Tree) จากผลรวมของค่าน้ำหนักที่มากที่สุดของเส้นเชื่อมโยงด้วยความคล้ายคลึง (Similarity) พฤติกรรมความถี่การชมภาพยนตร์ในโครงสร้างกราฟข้อมูลกลุ่มผู้ชมภาพยนตร์ตามคุณลักษณะส่วนบุคคลและทดสอบเปอร์เซ็นต์ความถูกต้อง รายละเอียดมีดังนี้

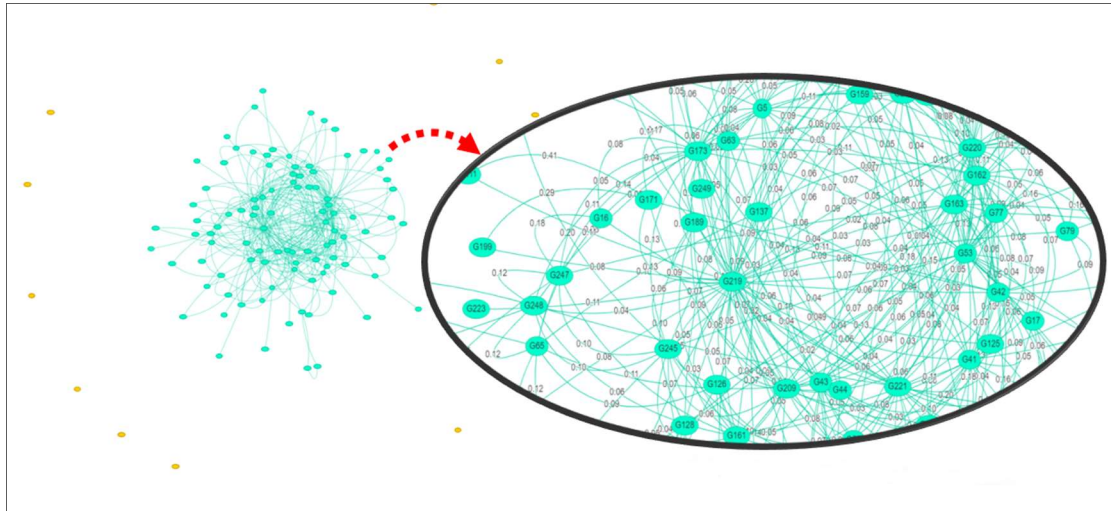
4.2.1 โครงสร้างกราฟของชุดข้อมูลผู้ใช้ในระบบ จากจำนวนผู้ชมภาพยนตร์ทั้งหมด 900 คน ประกอบด้วยคุณลักษณะ เพศ ช่วงอายุ และอาชีพ ได้ 252 กลุ่มโดย พบว่า จำนวน 111 กลุ่มที่มีจำนวนสมาชิกผู้ชมภาพยนตร์ จากข้อมูลผู้ชมภาพยนตร์ทั้งหมด โดยคุณสมบัติของลูกค้ำที่เข้ามาใช้บริการสามารถจัดความเชื่อมโยงเพื่อแบ่งกลุ่มผู้ชมภาพยนตร์ ดังแสดงในภาพที่ 7



ภาพที่ 7 การแบ่งกลุ่มผู้ชมภาพยนตร์

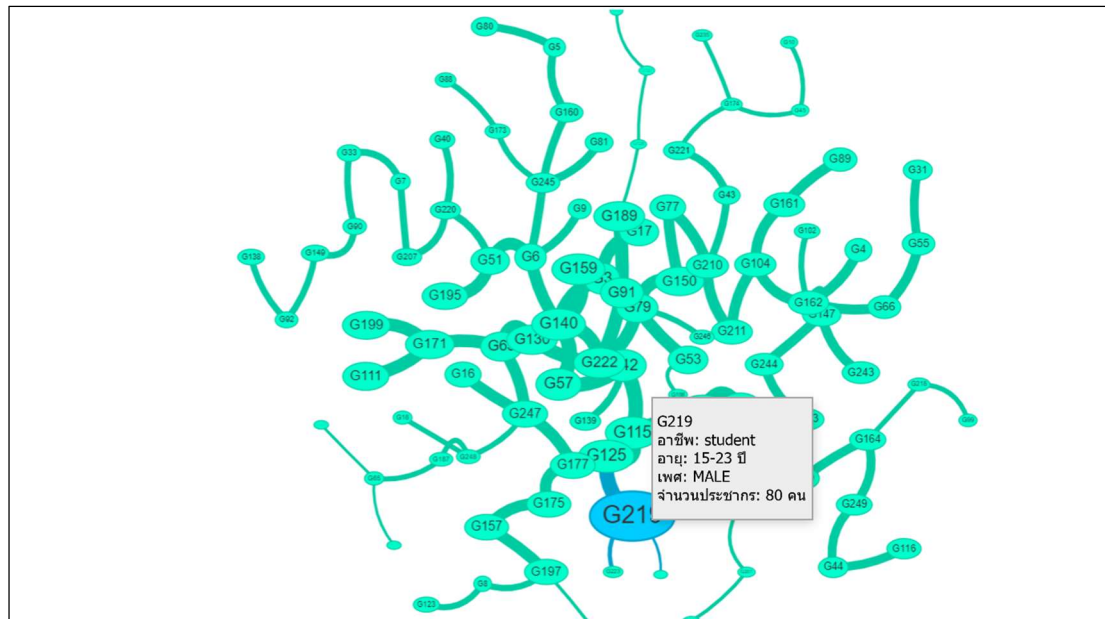
จากภาพที่ 7 คุณลักษณะของผู้ชมภาพยนตร์ประกอบด้วย เพศ (2 เพศ) อาชีพ (21 อาชีพ) และช่วงอายุ (5 ช่วงอายุ) ได้จำนวนกลุ่มจากคุณลักษณะ 252 กลุ่ม ซึ่งชุดข้อมูลเรียนรู้ 900 ชุดข้อมูลพบกลุ่มผู้ชมภาพยนตร์เพียง 111 กลุ่ม เพื่อนำมาเชื่อมโยงในโครงสร้างชุดข้อมูลด้วยค่าความคล้ายคลึงแบบ Cosine Similarity จากชุดข้อมูลจำนวนภาพยนตร์ที่ได้รับการชมของกลุ่มผู้ใช้ในระบบ (S) ให้เป็นค่าน้ำหนักในเส้นความเชื่อมโยง (Edge) จากการชมภาพยนตร์สำหรับใช้ในการแนะนำภาพยนตร์แสดงด้วยโหนดสีเขียวและ

โหนดสีเหลืองแสดงถึงกลุ่มของผู้ชมภาพยนตร์ที่ไม่มีความคล้ายคลึงกับกลุ่มผู้ชมอื่น ๆ ซึ่งความเชื่อมโยงดังภาพที่ 8



ภาพที่ 8 โครงสร้างกราฟของผู้ชมภาพยนตร์

จากภาพที่ 8 จากกลุ่มผู้ชมภาพยนตร์ที่แบ่งตามคุณลักษณะได้ 111 กลุ่มนั้น พบว่า โหนดที่มีค่าความคล้ายคลึงกันระหว่างกลุ่มผู้ชมภาพยนตร์มีทั้งหมด 99 Node (โหนดสีเขียว) และ 98 Edge เพื่อประยุกต์ใช้หลักการ Spanning Tree จากค่าน้ำหนักที่สูงสุด ซึ่งค่าน้ำหนักที่แสดงใน Edge วิเคราะห์พฤติกรรมชมด้วยความถี่ในการชมภาพยนตร์ภายในกลุ่มที่มีลักษณะคุณสมบัติตรงกัน เพื่อแนะนำภาพยนตร์ Top – 10 ให้กับผู้ชมภาพยนตร์ที่เป็นชุดทดสอบจำนวน 43 คนตามกลุ่มที่ตรงกับคุณลักษณะของผู้ชมภาพยนตร์ ตัวอย่างโครงสร้างข้อมูลของกลุ่มผู้ชมภาพยนตร์ ซึ่งผู้ชมภาพยนตร์ที่เป็นชุดทดสอบสำหรับรหัส User_ID: 943 ดังภาพที่ 9



ภาพที่ 9 โครงสร้างข้อมูลของกลุ่มผู้ชมภาพยนตร์กลุ่มรหัส G219

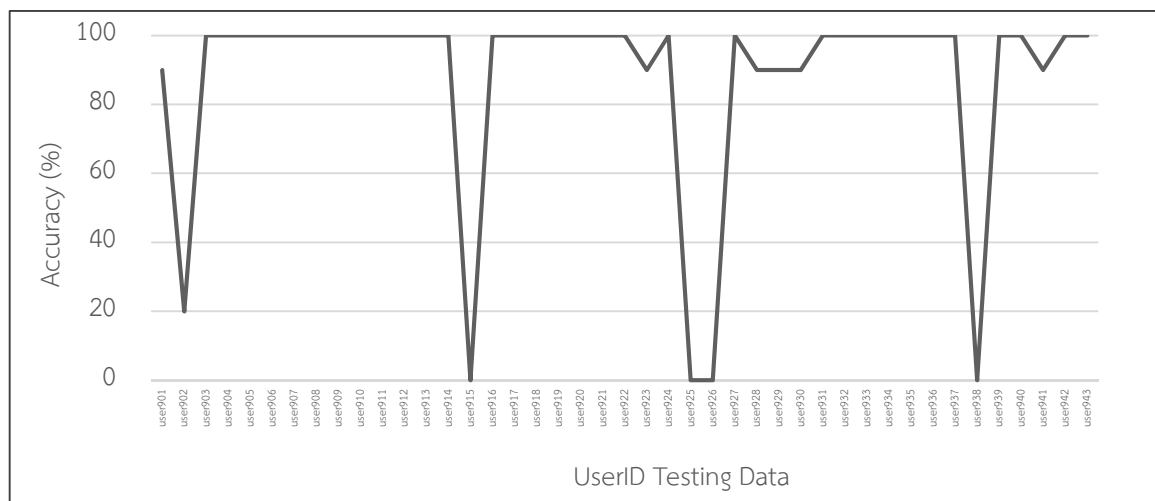
จากภาพที่ 9 ตัวอย่างโครงสร้างข้อมูลของกลุ่มผู้ชมภาพยนตร์ที่โหนดเริ่มต้นเป็นกลุ่มผู้ชมภาพยนตร์รหัส G219 ซึ่งผู้ชมภาพยนตร์ที่ใช้รหัส User_ID: 943 มีคุณลักษณะตรงกัน ได้แก่ อาชีพนักเรียน อายุอยู่ในช่วง 15 – 23 ปีและเป็นเพศชาย สามารถประยุกต์ใช้หลักการ Spanning Tree จากค่าน้ำหนักที่สูงสุด โดยภาพยนตร์ที่ได้แนะนำ 10 อันดับแรก (Top – 10) ของแต่ละกลุ่ม ตัวอย่างภาพยนตร์ที่ถูกแนะนำให้ผู้ชมในระบบรหัส User_ID: 943 ดังภาพที่ 10

USER_ID: 943

Rank	Movie_Name (Entry)
1	Star Wars (1977)
2	Contact (1997)
3	Toy Story (1995)
4	Fargo (1996)
5	Willy Wonka and the Chocolate Factory (1971)
6	Shining, The (1980)
7	Chasing Amy (1997)
8	Full Monty, The (1997)
9	Air Force One (1997)
10	Babe (1995)

ภาพที่ 10 ตัวอย่างภาพยนตร์ที่แนะนำด้วยคะแนนความชอบด้วยหลักการ Spanning Tree

จากภาพที่ 10 ตัวอย่างการทดสอบความถูกต้องในการแนะนำภาพยนตร์ให้แก่ผู้ใช้ในระบบรหัส User_ID: 943 สมาชิกกลุ่มรหัส G219 จากการแบ่งกลุ่มข้อมูลตามคุณลักษณะของผู้ใช้ระบบ พบว่า ภาพยนตร์ที่ถูกชม ประกอบไปด้วย Star Wars, Contact, Toy Story, Fargo, Willy Wonka and the Chocolate Factory, The Shining, Chasing Amy, The Full Monty, Air Force One และ Babe ทั้งหมด 10 เรื่อง ซึ่งเป็นภาพยนตร์ที่ได้รับการแนะนำ Top - 10 โดยการวิจัยครั้งนี้ผู้วิจัยทำการทดสอบความถูกต้อง (Accuracy) เพื่อพิจารณาการประเมินผลลัพธ์การแนะนำภาพยนตร์ให้แก่ผู้ชมทั้งหมด 43 คน แสดงดังภาพที่ 11



ภาพที่ 11 ความถูกต้องในการแนะนำด้วยหลักการ Spanning Tree

จากภาพที่ 11 แสดงความสัมพันธ์ระหว่างเปอร์เซ็นต์ความถูกต้อง (Accuracy) และผู้ชมภาพยนตร์ที่เป็นชุดข้อมูลทดสอบจำนวน 43 คน พบว่าเปอร์เซ็นต์ความถูกต้องมีค่าเฉลี่ย 87.45% จะเห็นว่าผู้ชมภาพยนตร์จำนวน 4 คน มีเปอร์เซ็นต์ความถูกต้องเท่ากับ 0% เนื่องจากผู้ชมภาพยนตร์ดังกล่าวไม่มีคุณสมบัติตรงกับกลุ่มของผู้ชมภาพยนตร์ที่อยู่ในกลุ่มชุดข้อมูลเรียนรู้และไม่มีความคล้ายคลึงกับกลุ่มผู้ชมอื่น ๆ จึงไม่พบความสัมพันธ์ของที่เชื่อมโยงกันในโครงสร้างข้อมูลเป็นผลให้ไม่สามารถแนะนำภาพยนตร์ให้แก่ผู้ชมภาพยนตร์นั้นได้

5. สรุปผลการศึกษา

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อออกแบบและประเมินประสิทธิภาพวิธีการแนะนำภาพยนตร์โดยการประยุกต์ใช้แนวคิดจากโครงสร้างกราฟข้อมูลและคะแนนความชอบด้วยการนำเสนอรูปแบบการแนะนำภาพยนตร์ Top-10 ที่มีความนิยมสูงสุดจากข้อมูลภาพยนตร์ 1,682 เรื่องที่มีผู้ชมภาพยนตร์ทั้งหมด 943 คน และแบ่งกลุ่มเพื่อให้สามารถลดจำนวนชุดข้อมูลลง ใช้เวลาการคำนวณความเชื่อมโยงได้เร็วและง่ายมากขึ้น จากชุดข้อมูลผู้ชมภาพยนตร์ทั้งหมดสามารถแบ่งเป็นชุดการเรียนรู้ 900 คน และชุดทดสอบ 43 คนเพื่อนำมา

ทำนายการแนะนำภาพยนตร์ แบ่งออกเป็น 2 ส่วน ได้แก่ 1) การคำนวณหาการแนะนำด้วยคะแนนความชอบ โดยการแบ่งกลุ่มด้วยเทคนิค K – mean จากการแนะนำภาพยนตร์ให้กลุ่มชุดทดสอบ พบว่า เปอร์เซนต์ความถูกต้อง มีค่าเฉลี่ยเท่ากับ 28.16% และในข้อมูลทดสอบที่ค่าความถูกต้องมีค่า 0% จากผู้ชมภาพยนตร์ 1 คน และ 2) การแนะนำด้วยการประยุกต์ใช้หลักการ Spanning Tree จากการอาศัยแบ่งกลุ่มด้วยคุณลักษณะรายบุคคลของผู้ชมภาพยนตร์ ประกอบด้วย เพศ อาชีพและช่วงอายุ คำนวณน้ำหนักความคล้ายคลึงของข้อมูลพฤติกรรมกรรมการชมภาพยนตร์ที่มากที่สุดเพื่อทดสอบการแนะนำ พบว่า เปอร์เซนต์ความถูกต้อง มีค่าเฉลี่ยเท่ากับ 87.45% และในข้อมูลทดสอบที่ค่าความถูกต้องมีค่า 0% จากผู้ชมภาพยนตร์ 4 คน ทำให้พบว่าการประยุกต์ใช้หลักการ Spanning Tree ในโครงสร้างข้อมูลกราฟให้ความถูกต้องในการแนะนำสูง จากการพิจารณาความสัมพันธ์ของคุณลักษณะของผู้ชมภาพยนตร์หรือลูกค้าและพฤติกรรมกรรมการชมภาพยนตร์หรือการเลือกซื้อสินค้า ซึ่งการแบ่งกลุ่มผู้ชมภาพยนตร์ที่อาศัยคุณลักษณะรายบุคคลของผู้ชมและค่าความคล้ายคลึงของข้อมูลพฤติกรรมกรรมการชมภาพยนตร์ในโครงสร้างข้อมูลกราฟ สามารถสร้างโครงข่ายของชุดข้อมูลลูกค้าและสินค้าทั้งหมดเพื่อการแนะนำได้

6. อภิปรายผล

การศึกษาการออกแบบและทดสอบวิธีการที่ใช้ในการแนะนำสินค้าโดยอาศัยชุดข้อมูลการชมภาพยนตร์จากฐานข้อมูล MovieLens เป็นชุดข้อมูลที่สามารถนำมาใช้ทดสอบเทคนิคการแนะนำสินค้าและพัฒนาเทคนิคการแนะนำเพื่อสอดคล้องกับคุณลักษณะของกลุ่มชุดข้อมูลที่น่ามาศึกษาได้เป็นอย่างดี ผู้วิจัยยังพบงานวิจัยที่ใช้ชุดข้อมูล MovieLens 100K ด้วย เทคนิคที่แตกต่างกันไป Leng, et al. (2020) ใช้วิธี Multi-Graph Attention Network จากชุดข้อมูล MovieLens พบค่า RMSE 0.890 โดยได้ศึกษาชุดข้อมูลของ Douban, Flixster และ YahooMusic ด้วย Zhang & Chen. (2020) ใช้วิธี matrix completion model แยกตัวแปรของ rating เช่น อายุ ประเภทหนัง แล้วนำมาคาดการณ์เปรียบเทียบกับ graph neural network พบค่า RMSE 0.905 และ Han, et al. (2021) ประยุกต์ใช้ Global Kernel Model ตาม rating Matrix พบค่า RMSE 0.889 ซึ่งประสิทธิภาพของ algorithm ที่มีความแตกต่างกัน เช่น ความเร็ว ขนาดชุดข้อมูล เป็นต้น ดังนั้นการแนะนำด้วยการแบ่งกลุ่มตามคุณลักษณะของผู้ใช้ในระบบและประยุกต์ใช้หลักการ Spanning Tree จากค่าน้ำหนักที่สูงสุดในโครงสร้างข้อมูลกราฟสามารถแนะนำสินค้าที่มีลักษณะตรงตามความต้องการได้มากยิ่งขึ้น โดยเฉพาะชุดข้อมูลที่มีความหลากหลายและซับซ้อน การจัดกลุ่มชุดข้อมูลที่เหมาะสมยังเป็นข้อจำกัดที่อาจส่งผลต่อการทำนายการเลือกบริโภคสินค้าและประสิทธิภาพในการแนะนำสินค้า โดยที่การแบ่งกลุ่มลูกค้าที่มีคุณลักษณะหลากหลายนั้นเหมาะกับชุดข้อมูลที่มีจำนวนมากเพื่อให้ครอบคลุมต่อการแนะนำลูกค้าทั้งหมด ดังนั้นการศึกษาโครงข่ายของข้อมูลลูกค้าและสินค้าที่พิจารณาถึงความสัมพันธ์ของข้อมูลมากกว่ารายการสินค้าที่มีปริมาณการจำหน่ายสูงที่สุดหรือรายการสินค้าที่ได้รับความนิยมสูงที่สุด การ

วิเคราะห์ด้านความสัมพันธ์ของชุดข้อมูลของลูกค้า เช่น เพศ อายุ อาชีพ ภูมิภานา พฤติกรรมการบริโภค เป็นต้น หรือความสัมพันธ์ของชุดข้อมูลของสินค้า เช่น รสชาติ กลิ่น สี ช่วงเวลาการขาย เป็นต้น จะสนับสนุนการตัดสินใจของลูกค้าได้แม่นยำขึ้น และสามารถนำไปประยุกต์ใช้กับข้อมูลการซื้อขายของธุรกิจต่างๆ เช่น ร้านค้า ร้านอาหารหรือโรงแรมได้อย่างมีประสิทธิภาพด้วย

เอกสารอ้างอิง

- Adomavicius, G., & Kwon, Y. (2007). New recommendation techniques for multicriteria rating systems. *IEEE Intelligent Systems*, 22(3), 48-55.
- Adomavicius, G., Manouselis, N., & Kwon, Y. (2011). **Multi-criteria recommender systems**. In **Recommender systems handbook**. Springer, Boston, MA.
- Cosley, D., Lam, S. K., Albert, I., Konstan, J. A., & Riedl, J. (2003). Is seeing believing? How recommender system interfaces affect users' opinions. In **Proceedings of the SIGCHI conference on Human factors in computing systems**, (585-592). New York: Association for Computing Machinery.
- Falk, K. (2019). **Practical recommender systems**. New York: Manning Publications.
- Han, S. C., Lim, T., Long, S., Burgstaller, B., & Poon, J. (2021). GLocal-K: Global and Local Kernels for Recommender Systems. In **Proceedings of the 30th ACM International Conference on Information & Knowledge Management**. (3063-3067). New York: Association for Computing Machinery.
- Harper, F. M., & Konstan, J. A. (2015). The MovieLens Datasets: History and Context. *ACM Transactions on Interactive Intelligent Systems*, 5(4), 1-19.
- Ivarsson, J., & Lindgren, M. (2016). **Movie recommendations using matrix factorization**. Retrieved from <https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-186400>
- Kesorn, K. (2021). **Data Science System**. Phitsanulok: Faculty of Science, Naresuan University.
- Kuzelewska, U. (2014). Clustering algorithms in hybrid recommender system on movielens data. *Studies in logic, grammar and rhetoric*, 37(1), 125-139.
- Leng, Y., Ruiz, R., Dong, X., & Pentland, A. S. (2020). Interpretable recommender system with heterogeneous information: A geometric deep learning perspective. *International Journal of Mechanical and Production Engineering Research and Development (IJMPERD)*, 10(3), 2411-2430.
- Sarwar, B., Karypis, G., Konstan, J. & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithm. **Proceeding of the 10th International World Wide Web Conference**, 285-295.
- Walek, B., & Fojtik, V. (2020). A hybrid recommender system for recommending relevant movies using an expert system. *Expert Systems with Applications*, 158, 1-18.
- Zhang, M., & Chen, Y. (2020). Inductive matrix completion based on graph neural networks. *The International Conference on Learning Representations (ICLR)*, 1-14.