

เทคนิคการจัดกลุ่ม K-Means แบบการคำนวณคอนตัม

นายภาณุวัฒน์ ธนาภรณ์จีนพงษ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2567



929906852

CU iThesis 6470323821 thesis / recv: 28052568 21:50:52 / seq: 12

K-Means Clustering Using Quantum Computing

Mr. Panuwat Tanapornchinpong

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of
Science in Computer Science
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2024



929906852

CU iThesis 6470323821 thesis / recv: 28052568 21:50:52 / seq: 12

หัวข้อวิทยานิพนธ์	เทคนิคการจัดกลุ่ม K-Means แบบการคำนวณควอนตัม
โดย	นายภาณุวัฒน์ ธาราภรณ์จีนพงษ์
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ศาสตราจารย์ ดร.ประภาส จงสิตย์วัฒนา

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.วิทยา วัฒนสุโกประสิทธิ์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ณัฐฉิ หนูไพโรจน์)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ศาสตราจารย์ ดร.ประภาส จงสิตย์วัฒนา)

..... กรรมการภายนอกมหาวิทยาลัย
(ศาสตราจารย์ ดร.บุญเจริญ ศิริเนาวกุล)



929906852

CU-IThesis 6470323821 thesis / rev: 28052568 21:50:52 / seq: 12

ภาณุวัฒน์ ธารภรณ์ชินพงษ์ : เทคนิคการจัดกลุ่ม K-Means แบบการคำนวณควอนตัม.
(K-Means Clustering Using Quantum Computing) อ.ที่ปรึกษาหลัก : ศ. ดร.ประภาส
จงสถิตย์วัฒนา

วิทยานิพนธ์ฉบับนี้ศึกษาอัลกอริทึม K-Means แบบผสมระหว่างควอนตัมและคลาสสิก
สำหรับการจัดกลุ่มข้อมูลผู้ป่วยโรคหัวใจ โดยใช้วงจร swap-test ของควอนตัมในการคำนวณ
ระยะทาง และได้ทำการทดสอบบนควอนตัมคอมพิวเตอร์จำลองใน 2 แนวทาง คือแบบที่มี
สัญญาณรบกวน และแบบอุดมคติ ด้วยชุดข้อมูลจริงที่มีมากกว่า 1,000 รายการ ผลการทดลอง
แสดงให้เห็นว่า วิธีควอนตัมทั้งสองสามารถทำความแม่นยำได้สูงถึง 0.83 และให้ค่า F1-score
ใกล้เคียงกับ K-Means แบบคลาสสิก (0.82–0.83) แม้ในกรณีค่าจากควอนตัมคอมพิวเตอร์ที่มี
สัญญาณรบกวน ผลการศึกษานี้ชี้ให้เห็นถึงศักยภาพในการใช้งานจริงของวิธีจัดกลุ่มที่ได้รับการ
เสริมด้วยควอนตัม

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์
ปีการศึกษา 2567

ลายมือชื่อนิติต
ลายมือชื่อ อ.ที่ปรึกษาหลัก



929906852

CU :Thesis 6470323821 thesis / recv: 28052568 21:50:52 / seq: 12

6470323821 : MAJOR COMPUTER SCIENCE

KEYWORD: Quantum computing, Swap Test Circuit, K-means clustering

Panuwat Tanapornchinpong : K-Means Clustering Using Quantum Computing.

Advisor: Prof. Prabhas Chongstitvatana

This thesis investigates hybrid quantum-classical K-Means algorithms for clustering heart disease patient data. Using quantum swap-test circuits for distance calculation, two approaches were tested on both noisy and ideal quantum simulators. A real-world dataset of over 1,000 records. The results show that both quantum methods achieve accuracy up to 0.83 and F1-scores comparable to the classical K-Means baseline (0.82–0.83), even when executed on quantum simulators with real-device noise models. These findings highlight the practical potential of quantum-enhanced clustering methods.

Field of Study: Computer Science

Academic Year: 2024

Student's Signature

Advisor's Signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จล่วงด้วยดี เนื่องด้วยได้รับความอนุเคราะห์จากอาจารย์ที่ปรึกษา
วิจัย ศาสตราจารย์ ดร.ประภาส จงสอติย์วัฒนา ซึ่งได้ให้คำแนะนำ แนวคิด ตลอดจนแนวทางในการ
แก้ไขปัญหาต่าง ๆ ด้วยความเอาใจใส่อย่างดียิ่งตลอดระยะเวลาการทำวิจัย ขอกราบขอบพระคุณเป็น
อย่างสูงไว้ ณ ที่นี้ด้วย

ขอขอบคุณ ผู้ช่วยศาสตราจารย์ ดร. ณัฐวุฒิ หนูไพโรจน์ ประธานกรรมการสอบ
วิทยานิพนธ์ และ ศาสตราจารย์ ดร. บุญเจริญ ศิริเนาวกุล กรรมการสอบวิทยานิพนธ์ ที่ได้ให้คำแนะนำ
และข้อชี้แนะแนวทางในการทำงานวิจัย

สุดท้ายนี้ ขอขอบคุณครอบครัว และเพื่อน ๆ ที่เป็นกำลังใจสำคัญ และคอยให้การสนับสนุน
มาโดยตลอด ทั้งด้านแรงใจและความช่วยเหลือในทุก ๆ ด้าน ซึ่งล้วนมีส่วนสำคัญยิ่งต่อความสำเร็จ
ของวิทยานิพนธ์ฉบับนี้

ภานุวัฒน์ ธนาภรณ์ชินพงษ์



929906852

CU-1Thesis 6470323821 thesis / recv: 28052568 21:50:52 / seq: 12

สารบัญ

หน้า

.....	ก
บทคัดย่อภาษาไทย	ก
.....	ง
บทคัดย่อภาษาอังกฤษ	ง
กิตติกรรมประกาศ	จ
สารบัญ	ฉ
สารบัญตาราง	ช
สารบัญภาพ	ฌ
บทที่ 1 บทนำ	10
1.1 ความเป็นมาและความสำคัญของปัญหา	10
1.2 วัตถุประสงค์ของการวิจัย	11
1.3 ขอบเขตงานวิจัย	11
1.4 ขั้นตอนและวิธีการดำเนินการวิจัย	11
1.5 คุณค่าทางวิชาการและประโยชน์ที่คาดว่าจะได้รับการวิจัย	12
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	1
2.1 ทฤษฎีที่เกี่ยวข้อง	1
2.2 งานวิจัยที่เกี่ยวข้อง	7
บทที่ 3 วิธีดำเนินการวิจัย	12
3.1 ข้อมูลที่ใช้ศึกษา (Dataset)	12
3.2 การเตรียมข้อมูลควอนตัม (Quantum data preparation)	14
3.3 ควอนตัมอัลกอริทึม (Quantum K-means algorithms)	15



929906852

CU-IThesis 6470323821 thesis / recv: 28052568 21:50:52 / seq: 12

3.4 การทดสอบและประเมินผล (Test and Evaluation)	19
บทที่ 4 วิธีดำเนินการวิจัย.....	21
4.1 ผลการวิจัยการจำแนกโรคหัวใจจาก ข้อมูลชุดโรคหัวใจ	21
บทที่ 5 สรุปผลการวิจัยและเสนอแนะ.....	26
5.1 สรุปผลการวิจัย	26
5.2 ข้อเสนอแนะ	26
บรรณานุกรม	28
ประวัติผู้เขียน	31



929906852

CU iThesis 6470323821 thesis / recv: 28052568 21:50:52 / seq: 12

สารบัญตาราง

หน้า

ตารางที่ 1 Summary of evaluation metrics.....23

ตารางที่ 2 Comparison with other work.....23

สารบัญภาพ

หน้า

รูปที่ 1: รูปของ Bloch's Sphere ก่อนและหลัง ของคิวบิตสถานะ $ 0\rangle$ ที่ผ่าน H-gate.....	5
รูปที่ 2 :ขั้นตอนการทำงานของ K-means Clustering	6
รูปที่ 3 :ตัวแปรของข้อมูลที่เลือกเพื่อประมวลผล	12
รูปที่ 4 :ชุดข้อมูลก่อนและหลังในการลบค่าผิดปกติ	13
รูปที่ 5 :ชุดข้อมูลที่ถูกประมวลผลแล้ว	13
รูปที่ 6 : Swap test full circuit.....	14
รูปที่ 7: Quantum circuit of algorithm 1	16
รูปที่ 8:Quantum circuit of algorithm 2	17
รูปที่ 9: Classical K-Means result and True Label	21
รูปที่ 10: Experiment result on simulator with noise(ibm_brisbane)	22
รูปที่ 11: Experiment result on ideal simulator(qasm_simulator)	22



929906852

CU iThesis 6470323821 thesis / recv: 28052568 21:50:52 / seq: 12

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

คอมพิวเตอร์ควอนตัมนั้นเป็นคอมพิวเตอร์ที่ใช้หลักการกลศาสตร์ควอนตัม (Quantum mechanics) โดยผลการทดลองและปรากฏการณ์ต่างๆ เกิดจากการควบคุมอะตอมที่มีขนาดเล็กเพื่อใช้ประมวลผลตามที่เรากำลังต้องการ ซึ่งผลลัพธ์ต่างๆ ไม่สามารถอธิบายได้ด้วยกลศาสตร์นิวตัน โดยปรากฏการณ์ที่สำคัญของกลศาสตร์ควอนตัมที่นำมาใช้ ได้แก่ หลักการทับซ้อนทางควอนตัม (Quantum superposition) ที่อะตอมตัวหนึ่งสามารถอยู่หลายตำแหน่งได้ในเวลาเดียวกัน และหลักการความพัวพันเชิงควอนตัม (Quantum entanglement) อะตอมสองตัวที่อยู่ห่างกันสามารถสื่อสารกันได้ โดยไม่ต้องมีการส่งสัญญาณหากัน คุณสมบัติดังกล่าวได้ถูกนำมาใช้สร้างเป็นคิวบิต (Qubits) ในคอมพิวเตอร์ควอนตัม ซึ่งเปรียบเสมือนบิต (Bits) ในคอมพิวเตอร์คลาสสิกนอกจากนั้น ยังมีส่วนช่วยให้การคำนวณดียิ่งขึ้น

ในหลายงานวิจัย ได้มีการพิสูจน์ว่าคอมพิวเตอร์ควอนตัมสามารถแก้ปัญหาที่คอมพิวเตอร์คลาสสิกไม่สามารถทำได้ เช่น การใช้ Shor's algorithm ในการถอดรหัสแบบ RSA , Grover's algorithm ที่ใช้ในการทำ Unstructured search algorithm เป็นต้น อย่างไรก็ตาม คอมพิวเตอร์ควอนตัมนั้นยังต้องถูกพิสูจน์ว่ามีประสิทธิภาพมากกว่าคอมพิวเตอร์คลาสสิกในด้านไหนบ้าง ในปัจจุบันได้มีการนำคอมพิวเตอร์ควอนตัมมาใช้ในเรื่องการเรียนรู้ของเครื่อง (Machine learning) ซึ่งมีแนวโน้มไปในทางที่ดี จึงทำให้ได้รับความสนใจในการวิจัยเป็นอย่างมาก โดยแนวคิดดังกล่าวเรียกว่า Quantum machine learning ซึ่งในปัจจุบัน คอมพิวเตอร์ควอนตัมยังอยู่ในยุค Noisy Intermediate-Scale Quantum (NISQ) ทำให้ผลลัพธ์ที่ได้จากการประมวลผลนั้นยังมี Error อยู่บ้าง ดังนั้น เพื่อให้ได้ผลลัพธ์ที่มีประสิทธิภาพสูงสุด อัลกอริทึมที่ใช้จะเป็นการทำงานร่วมกันระหว่างคอมพิวเตอร์ควอนตัมและคอมพิวเตอร์คลาสสิก

ในปัจจุบัน การเรียนรู้ของเครื่อง ได้ถูกนำมาประยุกต์ใช้แก้ปัญหาในด้านต่างๆ อย่างหลากหลาย เช่น การแพทย์ การตลาด การกีฬา และด้านวิศวกรรม เป็นต้น อย่างไรก็ตาม การเติบโตของข้อมูลขนาดใหญ่ (Big Data) เป็นความท้าทายอย่างหนึ่งสำหรับการนำการเรียนรู้ของเครื่องมาประยุกต์ใช้ จึงต้องมีการเพิ่มพลังการคำนวณ (Computing power) และพัฒนาอัลกอริทึม (Algorithm) ใหม่ๆ เพื่อให้สามารถรองรับข้อมูลขนาดใหญ่ โดยประมวลผลได้อย่างมีประสิทธิภาพและใช้เวลาในการประมวลผลน้อย จึงเป็นโจทย์ที่สำคัญ



การจัดกลุ่มข้อมูล(Clustering) เป็นการเรียนรู้ของเครื่องประเภทหนึ่ง ที่เป็นเทคนิคการเรียนรู้โดยไม่มีผู้สอน (Unsupervised learning) โดยวิธีการคือข้อมูลต่างๆจะถูกใส่เข้ามา โดยไม่มีการกำหนดค่าเป้าหมายของแต่ละข้อมูล จากนั้นให้เครื่องจักรวิเคราะห์จากการจำแนกและสร้างแบบแผนจากข้อมูลที่ได้รับมา โดยมีเทคนิคหนึ่งของการจัดกลุ่มข้อมูลที่ถูกใช้อย่างแพร่หลาย นั่นคือ K-means clustering algorithm โดยมีหลักการคือ ข้อมูลภายในกลุ่มจะต้องมีความคล้ายกับจุดกึ่งกลางของกลุ่มตัวเองมากกว่าจุดกึ่งกลางของกลุ่มอื่น ซึ่งตัวแปรที่ใช้วัดความคล้ายนั้นคือระยะทางที่วัดระหว่างข้อมูลกับจุดกึ่งกลางของกลุ่ม โดยในปัจจุบัน การนำอัลกอริทึมของการจัดกลุ่มข้อมูลนั้นสามารถเป็นไปได้ที่ประมวลผลบนคอมพิวเตอร์ควอนตัม

ในงานวิจัยชิ้นนี้จึงมุ่งเน้นการศึกษาทดลองออกแบบวงจรควอนตัมเพื่อใช้ในการสร้างโมเดลที่ใช้จัดกลุ่มข้อมูล(Clustering) ด้วยวิธี Quantum K-Means เพื่อให้สามารถรู้ถึงความสามารถและศักยภาพของคอมพิวเตอร์ควอนตัม ผลการทดลองจะถูกนำมาเทียบกับ โมเดลที่ใช้อัลกอริทึม K-means Clustering ด้วยคอมพิวเตอร์คลาสสิก โดยเราจะใช้การจำลองเครื่องควอนตัมคอมพิวเตอร์ (Quantum simulator) จาก QISKIT ที่พัฒนาโดย IBM ในการทดสอบ

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาขั้นตอนวิธีการสร้างวงจรควอนตัมเพื่อใช้ในการทำโมเดลการเรียนรู้ของเครื่องแบบไม่มีผู้สอน
2. เพื่อศึกษาประสิทธิภาพและข้อจำกัดในการใช้วงจรควอนตัมในการการเรียนรู้ของเครื่องแบบไม่มีผู้สอน

1.3 ขอบเขตงานวิจัย

1. งานวิจัยนี้ศึกษาการออกแบบและใช้งาน Quantum K-means เป็น โมเดลการเรียนรู้เพื่อแก้ปัญหา Clustering
2. ทำการทดสอบประเมินผลการทดสอบโมเดล โดยจะมีการวัดค่า accuracy, silhouette และ v-measure เพื่อใช้เปรียบเทียบระหว่างอัลกอริทึมของคอมพิวเตอร์ควอนตัมกับอัลกอริทึมของคอมพิวเตอร์คลาสสิก
3. งานวิจัยนี้ใช้การจำลองของควอนตัมคอมพิวเตอร์จาก IBM QISKIT Simulator

1.4 ขั้นตอนและวิธีการดำเนินการวิจัย

1. ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง
2. สร้างวงจรควอนตัมเพื่อใช้ในการแปลงข้อมูลแบบคลาสสิกให้เป็นสถานะทางควอนตัม

3. ทำการสร้างวงจรควอนตัมในส่วนของการกำหนดจุด Centroid เริ่มต้นแบบ K-means++ และในส่วนของการคำนวณระยะทางระหว่างจุด centroid โดยใช้ Swap gate บนคอมพิวเตอร์ควอนตัม
4. สร้างโมเดล Quantum K-means โดยการนำวงจรควอนตัมในข้อที่ 1 และ ข้อที่ 2 มาทดแทนขั้นตอนต่างๆในโมเดล K-means ของคอมพิวเตอร์คลาสสิก
5. สร้างโมเดล Classic K-means บนคอมพิวเตอร์คลาสสิก
6. ทำการจัดเตรียมข้อมูล(Data pre-processing) โดยมีรายละเอียดดังนี้
 - 6.1. ปรับแก้ไขความถูกต้องของชุดข้อมูล
 - 6.2. ปรับรูปแบบของข้อมูลให้เหมาะสม
 - 6.3. ทำการกำจัดค่าที่ผิดปกติออกจากชุดข้อมูล (outlier removal)
 - 6.4. ทำการลบข้อมูลซ้ำ (Duplicate Records)
 - 6.5. ทำการลดมิติข้อมูล(Dimensionality Reduction) เพื่อลดเวลาในการสอนโมเดลในคอมพิวเตอร์ควอนตัม
7. นำชุดข้อมูลในรูปแบบต่างๆที่มีการจัดเตรียมเรียบร้อยแล้ว เข้าประมวลผลด้วยโมเดลควอนตัมจากข้อ 4 และ โมเดลคลาสสิกจากข้อ 5
8. ประเมินสมรรถนะ (Evaluation) ของผลการทดลองจากทั้ง 2 โมเดล โดยมีการวัดค่า เช่น accuracy, silhouette และ v-measure เป็นต้น
9. สรุปผลการทดลอง

1.5 คุณค่าทางวิชาการและประโยชน์ที่คาดว่าจะได้รับการวิจัย

ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัยชิ้นนี้คือ เข้าใจวิธีการออกแบบและการใช้งานวงจรควอนตัมเพื่อสร้างโมเดลการเรียนรู้ของเครื่องแบบไม่มีผู้สอน สามารถรับรู้ถึงความสามารถและขีดจำกัดของโมเดลการเรียนรู้ของเครื่องด้วยอัลกอริทึมควอนตัม เมื่อเปรียบเทียบกับโมเดลการเรียนรู้ของเครื่องด้วยอัลกอริทึมแบบคลาสสิก นอกจากนี้ได้คาดหวังว่าสามารถนำข้อดีที่ได้จากการวิจัยไปพัฒนาโมเดลที่มีประสิทธิภาพมากขึ้น



929906852

CU-thesis 6470323821 thesis / recv: 28052568 21:50:52 / seq: 12

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

1. Noisy Intermediate Scale Quantum (NISQ) Computers

ยุคของคอมพิวเตอร์ควอนตัมที่มี คิวบิตประมาณ 50-100 โดยมีความสามารถที่จะแก้ไขปัญหาก็ได้เร็วกว่าคอมพิวเตอร์คลาสสิก[1] แต่การประมวลผลภายในคอมพิวเตอร์ยุคนี้จะมีปัญหาที่มาจาก noise ที่เกิดจาก quantum gate และ เกิดจากความเสถียรของการคงสภาพของคิวบิตในสถานะต่างๆที่ต่ำ เป็นผลให้ขนาดของวงจรควอนตัมถูกจำกัด ดังนั้นการออกแบบวงจรควอนตัมที่มี depth ที่ต่ำ ข้อกำหนดหลักของวงจรควอนตัมในยุค NISQ

2. Quantum Computing

การคำนวณแบบควอนตัมนั้น เป็นการประมวลผลโดยใช้หลักการกลศาสตร์ควอนตัม (Quantum mechanics) ซึ่งหลักการพื้นฐานและทฤษฎีต่าง ๆ นั้นจะอ้างอิงจาก Quantum Computation and Quantum Information [2] โดยผู้เขียน Michael A. Nielsen & Isaac L. Chuang

3. Quantum Bit

Quantum Bit หรือ Qubit เป็นหน่วยที่เล็กที่สุดของข้อมูล ที่ใช้ในคอมพิวเตอร์ควอนตัม โดยอ้างอิงจากหน่วยที่เล็กที่สุดของข้อมูลในคอมพิวเตอร์คลาสสิก เรียกว่า บิต(Bit) สำหรับบิตนั้นจะมีค่าได้เฉพาะ 2 ค่า คือ 1 หรือ 0 ซึ่งในแต่ละช่วงเวลา บิตจะมีค่าได้รูปแบบเดียวเท่านั้น แต่สำหรับ Qubit นั้น เนื่องจากมีคุณสมบัติจากสิ่งที่เรียกว่า การทับซ้อนของควอนตัม (Superposition) ทำให้ค่าของ Qubit นั้นบอกเป็นความน่าจะเป็นระหว่าง ค่า 1 หรือ 0 โดยเราสามารถทราบค่าจริงได้ก็ต่อเมื่อทำการวัดค่า ซึ่งจะทำให้คุณสมบัติทางควอนตัมหายไป สำหรับค่าของ Qubit สามารถเขียนในรูปของสมการ(1)ตามด้านล่าง

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \quad (1)$$

$$|\alpha|^2 + |\beta|^2 = 1 \quad (2)$$

โดย $|0\rangle$, $|1\rangle$ คือ สถานะพื้นฐาน(Basis State) และ α , β คือ ค่า probability amplitude coefficient ของ basis state $|0\rangle$, $|1\rangle$ ตามลำดับ โดยตัวแปรดังกล่าว มีค่าเป็นจำนวนเชิงซ้อน(complex number) และมีลักษณะ normalize ตามดังสมการ(2) ซึ่งตัวแปรดังกล่าว

สามารถสื่อได้ว่า มีโอกาสที่ Qubit มีค่า 0 ด้วยค่าความน่าจะเป็นคือ $|\alpha|^2$ และมีค่า 1 ด้วยค่าความน่าจะเป็นคือ $|\beta|^2$

4. Quantum computation

การเปลี่ยนสถานะของ Qubit จากสถานะหนึ่งไปเป็นสถานะหนึ่งนั้น เป็นวิธีที่ใช้ในการคำนวณแบบควอนตัม ซึ่งวิธีนี้จะคล้ายกับที่ใช้ในคอมพิวเตอร์คลาสสิก คือเปลี่ยนค่าบิตโดยใช้ Logic Gate ที่เป็นส่วนประกอบในวงจรไฟฟ้า สำหรับคอมพิวเตอร์ควอนตัม ใช้ Quantum Gate เปลี่ยนค่าหรือเปลี่ยนสถานะของ Qubit ซึ่งออกแบบให้ใช้งานภายในวงจรควอนตัม (Quantum Circuit) โดยค่าของ quantum gate นั้นสามารถแสดงได้ในรูปแบบเมทริกซ์ ขนาด $2^n \times 2^n$ โดย n คือจำนวน qubit ที่ gate ต้องใช้งาน หรือแสดงในรูปแบบการหมุนภายใน Bloch's sphere ประเภทของ Quantum Gate นั้น มีรายละเอียดดังนี้

4.1. Single qubit gates

ประเภทของ quantum gate ที่ทำงานได้โดยใช้เพียง 1 qubit ในการใช้งาน โดยเกทที่ใช้งานภายในงานวิจัยนี้มีตัวอย่างดังนี้

4.1.1. X-Gate

ควอนตัมเกทที่จะสลับค่าระหว่างค่าจาก $|0\rangle$ ไปเป็น $|1\rangle$ และ $|1\rangle$ ไปเป็น $|0\rangle$ ซึ่งจะคล้ายกับ NOT gate คอมพิวเตอร์คลาสสิก

	$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
---	--

4.1.2. Y-Gate

ควอนตัมเกททำหน้าที่คล้าย X-Gate แต่จะหมุนค่ารอบแกน Y

	$Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$
---	---

4.1.3. Z-Gate

ควอนตัมเกททำหน้าที่คล้าย X-Gate แต่จะหมุนค่ารอบแกน Z โดยมีชื่อเรียกอีกอย่างว่า phase shift gate

q[0] —  —	$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
--	---

4.1.4. Hadamard Gate

Hadamard gate หรือ H-gate เป็นควอนตัมเกตพื้นฐานที่เปลี่ยนคิวบิตจากสถานะปกติเป็นสถานะ Superposition

q[0] —  —	$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$
--	--

4.1.5. R_x Rotation Gate

เกตที่ใช้ในการปรับค่า โดยการใช้นั้นจะกำหนดเป็นองศาในการหมุน โดยเกตนี้อาจใช้ในการหมุนรอบแกน x

q[0] —  —	$R_x(\theta) = \begin{bmatrix} \cos(\theta/2) & -i \sin(\theta/2) \\ -i \sin(\theta/2) & \cos(\theta/2) \end{bmatrix}$
--	--

4.1.6. R_y Rotation Gate

เกตที่ใช้ในการปรับค่า โดยการใช้นั้นจะกำหนดเป็นองศาในการหมุน โดยเกตนี้อาจใช้ในการหมุนรอบแกน y

q[0] —  —	$R_y(\theta) = \begin{bmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{bmatrix}$
--	---

4.1.7. R_z Rotation Gate

เกตที่ใช้ในการปรับค่า โดยการใช้นั้นจะกำหนดเป็นองศาในการหมุน โดยเกตนี้อาจใช้ในการหมุนรอบแกน z

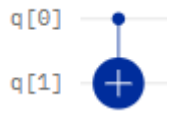
q[0] —  —	$R_z(\theta) = \begin{bmatrix} \exp(-i\theta/2) & 0 \\ 0 & \exp(i\theta/2) \end{bmatrix}$
--	---

4.2. Multiple qubit gates

ประเภทของ quantum gate ที่ทำงานได้โดยใช้มากกว่า 1 qubit ขึ้นไป โดยเกทที่ใช้ทำงานภายใน งานวิจัยนี้ มีตัวอย่างดังนี้

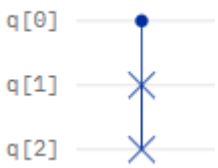
4.2.1. Controlled-X Gate

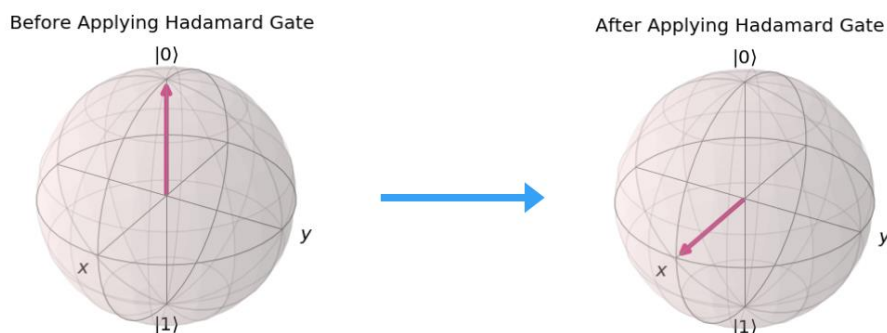
Controlled-X Gate หรือ Controlled Not Gate เป็น Gate ที่ต้องการ 2 Qubits ในการทำงาน โดยใช้ Qubit ที่ 1 ในการควบคุมการเปลี่ยนสถานะของ Qubit ที่ 2 โดยหาก Qubit ที่ 1 มีสถานะ คือ $|1\rangle$ จะทำการเปลี่ยนค่าของ Qubit ที่ 2

	$CNOT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$
---	---

4.2.2. Controlled swap Gate

Controlled swap Gate หรือ Fredkin gate เป็น gate ที่ต้องการ 3 qubits การทำงานต้องใช้ 1 qubit ในการควบคุมการเปลี่ยนสถานะของ 2 qubit ที่เหลือ โดยหาก qubit ที่ควบคุมมีค่าเป็น $|1\rangle$ qubit ที่เหลือจะถูกเปลี่ยนสถานะ

	$CSWAP = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
---	--



รูปที่ 1: รูปของ Bloch's Sphere ก่อนและหลัง ของคิวบิตสถานะ $|0\rangle$ ที่ผ่าน H-gate

5. Quantum Algorithm

Quantum Algorithm เป็นขั้นตอนที่ใช้การแก้ปัญหา โดยใช้ในการทำงานบนควอนตัมคอมพิวเตอร์ ซึ่งใช้ความสามารถต่างๆ ในทฤษฎีควอนตัม ได้แก่ Quantum Superposition และ Quantum entanglement เป็นต้น โดยการใช้งานนั้นจะถูกเขียนในวงจรควอนตัม ซึ่งสิ่งที่ทำให้ Quantum algorithm มีความน่าสนใจกว่า classic algorithm คือ มีความเป็นไปได้ที่จะสามารถแก้ปัญหาที่คอมพิวเตอร์ ไม่สามารถแก้ได้ ในระยะเวลาที่เป็นไปได้ สิ่งที่น่าสนใจของ Quantum algorithm คือมีโอกาสที่จะสามารถแก้ปัญหาที่คอมพิวเตอร์คลาสสิกไม่สามารถแก้ได้ โดยใช้เวลาที่เป็นไปได้หรือ ถึงแม้ จะใช้ supercomputer แล้วก็ตาม ซึ่งสิ่งนี้คือ Quantum supremacy

อัลกอริทึมที่ทำให้เกิด quantum supremacy นั้นคือ Shor's algorithm [3] ที่ใช้สำหรับแก้ปัญหา Integer factorization และ Grover's algorithm[4] ที่ใช้สำหรับการทำ Unstructured search algorithm โดยการใช้หลักการ Amplitude amplification

6. K-means Clustering

K-means clustering [5] เป็นอัลกอริทึมที่ใช้จัดกลุ่มข้อมูลแบบไม่มีผู้สอน (Unsupervised clustering method) อัลกอริทึมจะจัดกลุ่มข้อมูลตามคุณสมบัติที่คล้ายกัน โดยจะมีจุดกึ่งกลางของแต่ละกลุ่มข้อมูล (centroid) ซึ่ง centroid นั้นคือค่าเฉลี่ยของข้อมูลทั้งหมดที่อยู่ในกลุ่มนั้น

การจัดกลุ่มแบบนี้ มีเงื่อนไขคือ ต้องมีการกำหนดจำนวนกลุ่มที่ต้องการจำแนกไว้ก่อน ซึ่งหากเรานำชุดข้อมูลเข้าสู่กระบวนการในการจัดกลุ่มนั้น จะเริ่มจากสุ่มตำแหน่งของ centroid โดยมีจำนวนตามกลุ่มที่กำหนด หลังจากนั้น อัลกอริทึมจะเลือกข้อมูลจัดเข้ากลุ่มที่มีระยะทางห่างจาก centroid ของกลุ่มนั้นน้อยที่สุด โดยส่วนที่เลือกข้อมูลสามารถเขียนเป็นสมการ (3) ได้

ดังนั้น กำหนดให้ K คือจำนวนกลุ่มของข้อมูลที่กำหนด S คือชุดข้อมูลที่อยู่ในกลุ่ม โดยที่ $S = \{s_1, s_2, s_3, s_4, \dots, s_K\}$ C คือจุด centroid ของแต่ละกลุ่ม โดยที่ $C = \{c_1, c_2, c_3, c_4, \dots, c_K\}$

$$\operatorname{argmin} \sum_{i=1}^K \sum_{x \in S_i} \|x - c_i\|^2 \quad (3)$$

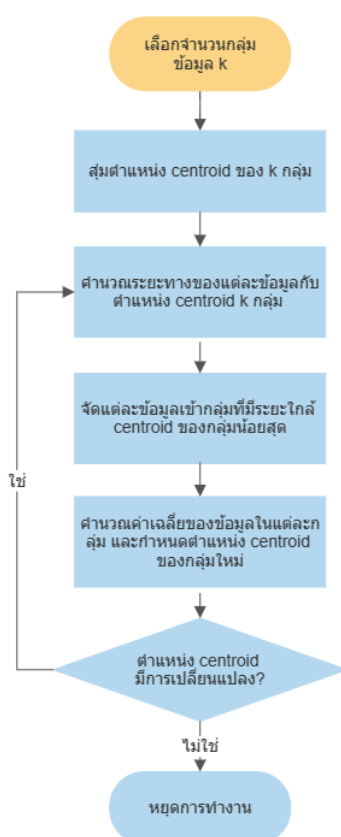
เมื่อจัดข้อมูลเข้ากลุ่มหมดแล้ว อัลกอริทึมจะปรับตำแหน่ง centroid ของกลุ่มใหม่ โดยเฉลี่ยตำแหน่งของข้อมูลที่อยู่ในกลุ่มทั้งหมด ซึ่งสามารถเขียนเป็นสมการ (4) ได้ดังนี้ โดยที่ L_i คือขนาดของชุดข้อมูล S_i

$$c_i = \frac{1}{L_i} \sum_{n=1}^{L_i} (x \in S_i) \quad (4)$$

จากนั้นจะทำซ้ำในการเลือกข้อมูลเข้ากลุ่มและปรับตำแหน่ง centroid จนกว่าตำแหน่งของ centroid ไม่มีการเปลี่ยนแปลง โดย time complexity ของ K-means จะขึ้นอยู่กับ 3 อย่าง ได้แก่ จำนวน feature N ของแต่ละข้อมูล, ขนาดข้อมูลทั้งหมด M และ จำนวน cluster K ซึ่งคือ MNK)

สำหรับขั้นตอนการทำงานของ K-means Clustering สามารถแสดงลำดับขั้นตอนได้ดังรูปที่

2



รูปที่ 2 : ขั้นตอนการทำงานของ K-means Clustering

7. K-means++

K-means++ [6] เป็นอัลกอริทึมที่พัฒนาในส่วนของการสุ่มจุด centroid เพิ่มเติมจาก K-means โดย K-means++ จะช่วยเพิ่มความเร็วและความแม่นยำของอัลกอริทึมในการจัดกลุ่มข้อมูล ซึ่งมีขั้นตอนดังนี้ กำหนดให้ $D(x)$ คือระยะทางที่สั้นที่สุดจากจุดข้อมูลไปยังจุดcentroid ที่ใกล้ที่สุด

- สุ่มจุด centroid จุดแรก ด้วยสุ่มแบบuniform จากชุดข้อมูล X
- กำหนดจุด centroid ที่ 2 โดยเลือกจากจุดข้อมูล x จากชุดข้อมูล X ด้วยความน่าจะเป็น ตามดังสมการ (5)

$$\frac{D(x)^2}{\sum_{x \in X} D(x)^2} \quad (5)$$

- กำหนดจุด centroid ที่เหลือ โดยทำซ้ำตามข้อ b จนครบทุกจุด centroid ที่กำหนด

2.2 งานวิจัยที่เกี่ยวข้อง

1. q-means: A quantum algorithm for unsupervised machine learning

ในปี ค.ศ.2018 Iordanis Kerenidis, Jonas Landman, Alessandro Luongo, และ Anupam Prakash[7] ได้นำเสนออัลกอริทึม q-means ที่ใช้ในคอมพิวเตอร์ควอนตัม มีการนำเสนอขั้นตอน initializationแบบ K-means++ โดยอธิบายในรูปแบบของทฤษฎีในแต่ละขั้นตอนและบทพิสูจน์ของทฤษฎีนั้นๆ สำหรับการSimulation ได้นำเสนอ δ -k-means ที่เป็น q-means ที่ทดสอบในคอมพิวเตอร์คลาสสิก เนื่องจากคอมพิวเตอร์มีขนาดไม่ใหญ่พอในการทดลองด้วยทฤษฎีที่เสนอ ซึ่ง δ -k-means นั้นเป็น k-means ที่เพิ่ม noise เข้าไปในอัลกอริทึม โดยจะใส่ noise ในขั้นตอนการเลือกข้อมูลเข้ากลุ่มcentroid และในขั้นตอนการคำนวณจุด centroid ใหม่ สำหรับการเปรียบเทียบประสิทธิภาพ ได้ทดสอบกับชุดข้อมูล MNIST เปรียบเทียบระหว่าง k-means และ δ -k-means ในระดับ noise ต่างๆ ซึ่งผลลัพธ์คือ δ -k-means มี accuracy ที่ใกล้เคียงกับ k-means โดยที่ δ -k-means มีrunning time ที่ต่ำกว่า สำหรับชุดข้อมูลที่มีขนาดใหญ่ โดยการเปรียบเทียบทางทฤษฎี

2. K-Means Clustering on Noisy Intermediate Scale Quantum Computers

ในปี ค.ศ.2019 Sumsam Ullah Khan , Ahsan Javed Awan ,และ Gemma Vall-Llosera[8] ได้ทำการศึกษาการออกแบบอัลกอริทึม K-means เพื่อใช้ภายในควอนตัมคอมพิวเตอร์ เนื่องจากควอนตัมมี decoherence time ที่ต่ำ ทางผู้วิจัยจึงได้นำเสนอรูปแบบวงจรควอนตัมที่มี depth ที่ต่ำ เพื่อเพิ่ม accuracy ของโมเดลในการจัดกลุ่มข้อมูล ซึ่งมี 3 แนวทาง ได้แก่ quantum interference ,negative rotations และ destructive interference โดยวงจรที่ออกแบบนั้นทางผู้วิจัยได้มีการทดสอบในควอนตัมคอมพิวเตอร์ของ IBM ด้วยชุดข้อมูล ได้แก่ Random dataset, MNIST , IRIS เป็นต้น ซึ่งผลลัพธ์ที่ได้คือวงจรควอนตัมที่มี depth ต่ำทั้ง 3 แบบ ให้ accuracy ที่สูงใกล้เคียงกับ accuracy ของอัลกอริทึม K-mean ในคอมพิวเตอร์คลาสสิก

3. Quantum k-means algorithm based on Trusted server in Quantum Cloud Computing

ในปี ค.ศ.2020 Changqing Gong ,Zhaoyang Dong ,Abdullah , Gani, และ Han Qi [9] ได้ศึกษาอัลกอริทึม Quantum K-mean ที่ใช้งานบน Quantum Cloud computing โดยได้เสนอ Quantum K-mean ซึ่งมี Core subroutine คือ Swap test ที่ใช้ในขั้นตอนการหาความเหมือนระหว่างข้อมูลและกลุ่มข้อมูล และ Grover optimization ที่ใช้ในขั้นตอนคัดเลือกข้อมูลเข้ากลุ่ม โดยใช้ค่าที่ได้จากขั้นตอน Swap test โดยอัลกอริทึมนี้จะแบ่งเป็น 2 ส่วน ในฝั่งของ Client จะเป็นการแปลงข้อมูลเป็นสถานะควอนตัม และทำการเข้ารหัสข้อมูลก่อนส่งให้ server และรับผลลัพธ์ที่ได้พร้อมถอดรหัส เพื่อแสดงผลลัพธ์ ในฝั่งของ server จะทำการประมวลผลและส่งกลับให้ clients ซึ่ง Core subroutine จะอยู่กับฝั่ง Server ซึ่งผู้วิจัยได้ทำการทดลองวงจรควอนตัม ผ่าน คอมพิวเตอร์ควอนตัมของ IBM Qiskit โดยการเข้ารหัสนั้นได้ใช้วิธี Quantum homomorphic encryption scheme(QHE) ซึ่งผลการทดลองแสดงให้เห็นว่าคอมพิวเตอร์ควอนตัมสามารถประมวลผลอัลกอริทึม Quantum K-mean โดยมีการเข้ารหัสด้วยวิธีในทางควอนตัมได้

4. Patient Data Analysis with the Quantum Clustering Method

ในปี ค.ศ.2022 Shradha Deshmukh , Bikash K. Behera , และ Preeti Mulay[10] ได้นำเสนอการอัลกอริทึมแบบ Quantum K-means โดยผสมระหว่างอัลกอริทึมคลาสสิกและควอนตัม ซึ่งภายในอัลกอริทึมนี้ มีการใช้ Swap test เป็นขั้นตอนในการหาระยะทางระหว่างข้อมูลและจุด Centroid และในขั้นตอนจัดข้อมูลเข้ากลุ่มและขั้นตอนการหา centroid ใหม่ นั้นทางผู้วิจัยได้ออกแบบให้ประมวลผลอยู่ในคอมพิวเตอร์ควอนตัมทั้งหมด โดยวงจรที่ออกแบบนั้นได้ทดสอบด้วยคอมพิวเตอร์ควอนตัมแบบ Simulator ด้วยชุดข้อมูลทั้งแบบจำลองขึ้นมา

และ แบบข้อมูลจริง สำหรับข้อมูลจริงนั้น ได้แก่ Mucormycosis และ Cardiovascular ซึ่งจากการทดสอบในชุดข้อมูล Cardiovascular พบว่า Quantum K-means นั้นมีความแม่นยำสูงถึงประมาณ 92 % เมื่อเทียบกับ K-means และ K-means++ มีความแม่นยำอยู่ที่ 82% และ 79% ตามลำดับ

5. Quantum K-means clustering method for detecting heart disease using quantum circuit approach

ในปี ค.ศ.2022 S S Kavitha, และ Narasimha Kaulgud [11] ได้นำเสนอการใช้โมเดล Quantum K-means clustering เพื่อ predict โรคหัวใจ โดยขั้นตอนนั้นจะเริ่มจากใช้คอมพิวเตอร์ในการจัดการข้อมูลด้วยวิธี Dimensionality reduction และ outlier removal ในขั้นตอนของการจัดกลุ่มจะใช้ Quantum K-mean ซึ่งมี Core subroutine คือ Swap test ที่ใช้ในขั้นตอนการหาระยะห่างระหว่างข้อมูลกับจุด centroid ในส่วนการจัดข้อมูลเข้ากลุ่มนั้นใช้หลักการ Nearest-neighbour โดยโมเดลที่นำเสนอทดสอบบนคอมพิวเตอร์ควอนตัมแบบ Simulator ซึ่งผลทดสอบพบว่า โมเดลที่นำเสนอ มีความแม่นยำสูงถึง 96% เมื่อเทียบโมเดล K-means ที่มีความแม่นยำเพียง 94%

6. Intelligent recommender system based on quantum clustering and matrix completion

ในปี ค.ศ.2022 Oumayma Ouedrhiri, Oumayma Banouar, Salah El Hadaj, และ Said Raghay [12] ได้นำเสนอระบบ Recommender โดยใช้ Quantum K-means algorithm ร่วมกับ singular value decomposition (SVT) algorithm และได้ทำการทดลองบนชุดข้อมูล MoviesLens เป็นข้อมูลระหว่างลักษณะของ user กับ การให้คะแนนหนัง โดยเป้าหมายนั้นคือการคาดการณ์คะแนนหนัง สำหรับ Quantum K-mean มี Core subroutine คือ Swap test ซึ่งใช้ในขั้นตอน Initialization แบบ K-means++ และขั้นตอนการหาระยะห่างระหว่างข้อมูลกับจุด centroid ในส่วนของ SVT จะถูกใช้เป็น post processing หลังจากได้ผลลัพธ์จาก Quantum K-means เพื่อ predict คะแนนหนัง โดยผลลัพธ์ที่ได้เมื่อเทียบกับอัลกอริทึม K-means +SVT นั้นมีผลทดสอบที่ดีกว่า โดยมีค่า precision สูงกว่า 5%, ค่า recall สูงกว่า 9% และ ค่า F-measure สูงกว่า 8%

7. Application of Quantum Computing to Accurate Positioning in 6G Indoor Scenarios

ในปี ค.ศ.2022 H. Urgelles, P. Picazo-Martínez, and J. F. Monserrat [13] ได้นำเสนอวิธีการระบุตำแหน่งของอุปกรณ์ภายในอาคารผ่านเครือข่าย 6G โดยประยุกต์ใช้ ควอนตัม



คอมพิวเตอร์ ในการคำนวณระยะทางระหว่างอุปกรณ์กับจุดเชื่อมต่อ (Access Point) ที่อยู่ในอาคาร สำหรับการคำนวณระยะทางนั้น ผู้วิจัยได้ประยุกต์ใช้สูตร Euclidean Distance ในรูปแบบที่สามารถดำเนินการบนคอนตัมคอมพิวเตอร์ได้ โดยการสร้างวงจรคอนตัมที่จำลองการคำนวณดังกล่าว ทั้งนี้ผลลัพธ์ที่ได้นั้น พบว่าระยะทางที่คำนวณได้ด้วยวิธีที่น่าเสนอเปรียบเทียบกับ วิธี Euclidean Distance แบบดั้งเดิมนั้น มีค่าที่ใกล้เคียงกัน และเมื่อมีการจำลองสภาพแวดล้อมภายในอาคารโดยแบ่งพื้นที่เป็น 4 ส่วน ระบบสามารถระบุตำแหน่งของอุปกรณ์ในแต่ละส่วนได้อย่างถูกต้อง นอกจากนี้ งานวิจัยยังได้ทำการศึกษาผลของจำนวนครั้งในการรันวงจรคอนตัม (shots) ที่เหมาะสม พบว่า การใช้ 4,000 shots ให้ค่าความแม่นยำที่เพียงพอ เมื่อเพิ่มจำนวนมากกว่านี้ ค่าความคลาดเคลื่อน (error) ลดลงเพียงเล็กน้อยโดยไม่มีนัยสำคัญ นอกจากนี้ได้ทำการวัดเวลาในการประมวลผลเปรียบเทียบระหว่างคอนตัมคอมพิวเตอร์กับคอมพิวเตอร์แบบคลาสสิก พบว่า เมื่อขนาดของข้อมูลเพิ่มขึ้นและมีความซับซ้อนมากขึ้น คอนตัมคอมพิวเตอร์สามารถประมวลผลได้รวดเร็วกว่าอย่างมีนัยสำคัญ

8. Quantum clustering with k-Means: A hybrid approach

ในปี ค.ศ.2024 A. Poggiali, A. Berti, A. Bernasconi, G. M. Del Corso, and R. Guidotti [14] นำเสนออัลกอริทึม k-Means แบบไฮบริดที่ผสมผสานการประมวลผลคอนตัมเข้ากับการประมวลผลแบบคลาสสิก เพื่อเพิ่มประสิทธิภาพในการจัดกลุ่มข้อมูล โดยมี 3 รูปแบบ ซึ่งเน้นออกแบบวงจรคอนตัม ให้มีความซับซ้อนน้อยลง โดยสามารถ Time Complexity ของขั้นตอนการจัดกลุ่ม จาก $O(nk)$ เหลือ $O(1)$ โดยที่ n คือจำนวนข้อมูล และ k คือจำนวน Centroid. โดยเมื่อทดลองเปรียบเทียบกับ Classical K-Means พบว่า รูปแบบอัลกอริทึมที่น่าเสนอใช้เวลาในการประมวลผลน้อยกว่า Classical K-Means อย่างไรก็ตาม ความแม่นยำในการจัดกลุ่มของอัลกอริทึมที่น่าเสนอ ยังมีความแม่นยำน้อยกว่า

9. Practical quantum k-means clustering: Performance analysis and applications in energy grid classification

S. DiAdamo, C. O'Meara, G. Cortiana, and J. Bernabé- Moreno[15] นำเสนอการวิเคราะห์เชิงลึกเกี่ยวกับการใช้งาน Quantum K-Means Clustering บนฮาร์ดแวร์ควอนตัมจริงผ่านระบบ IBM Quantum Cloud โดยมีจุดมุ่งหมายเพื่อใช้ในการจัดกลุ่มโครงข่ายไฟฟ้าย่อยในประเทศเยอรมนี โดยได้ใช้ swap test ในการคำนวณระยะห่างแบบคอนตัม โดยเปรียบเทียบระหว่าง amplitude embedding และ angle embedding รวมถึงเปรียบเทียบผลลัพธ์จาก การ

จำลอง simulation และ ฮาร์ดแวร์จริง พบว่า noise บนเครื่องจริงทำให้ความแม่นยำลดลงอย่างชัดเจน โดยเฉพาะใน angle embedding และเสนอวิธีใหม่เรียกว่า vector subspace parallelization ซึ่งแบ่งเวกเตอร์ความมิติสูงเป็นคู่ๆ แล้วคำนวณระยะทางแบบขนาน เพื่อเพิ่มความแม่นยำ ผลการทดลอง พบว่า เมื่อใช้ข้อมูลเพียงแค่ 2 มิติ ให้ผลดี แต่เมื่อเพิ่มมิติหรือจำนวนกลุ่ม ความแม่นยำลดลงอย่างชัดเจน สำหรับความแม่นยำเมื่อใช้วิธี parallelized subspace estimation มีค่าaccuracy สูงถึง 84.8% เทียบกับ 17% ของวิธีเดิม

10. Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction

H. Gupta, H. Varshney, T. K. Sharma, N. Pachauri, and O. P. Verma [16] เปรียบเทียบประสิทธิภาพของอัลกอริธึมการเรียนรู้ของเครื่องควอนตัม (Quantum Machine Learning - QML) กับการเรียนรู้เชิงลึก (Deep Learning - DL) สำหรับการทำนายโรคเบาหวาน โดยเน้นการปรับปรุงประสิทธิภาพผ่านเทคนิคการเตรียมข้อมูล ผลการทดลองพบว่า DL แสดงประสิทธิภาพที่เหนือกว่าโมเดล QML ทุกตัวชี้วัดที่ประเมิน เช่น โมเดล DL มีความแม่นยำ 95% ในขณะที่ QML มีความแม่นยำที่ 86%



บทที่ 3

วิธีดำเนินการวิจัย

ในงานวิจัยนี้ จะมีการสร้างโมเดลควอนตัมสำหรับการวิเคราะห์แบบจัดกลุ่ม (Clustering) และทดสอบข้อมูลเปรียบเทียบกับโมเดลการจัดกลุ่มบนคลาสสิกคอมพิวเตอร์ ซึ่งจะแบ่งได้เป็น 2 ส่วนคือ Quantum data preparation และ Quantum K-mean algorithm

3.1 ข้อมูลที่ใช้ศึกษา (Dataset)

1. ข้อมูลโรคหัวใจ

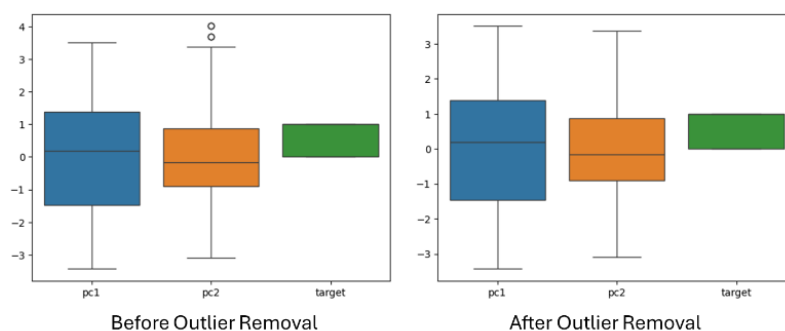
ชุดข้อมูลที่ใช้ในการศึกษานี้ นำมาจากฐานข้อมูลสาธารณะ [17] โดยประกอบด้วยตัวแปรทั้งหมด 13 รายการ ซึ่งรวมถึงตัวแปรเป้าหมาย 1 รายการ มีข้อมูลผู้ป่วยรวมทั้งสิ้น 1,025 ตัว สำหรับการวิเคราะห์แบบจัดกลุ่ม (Clustering) ได้เลือกใช้ตัวแปร 8 รายการที่เป็นตัวแปรประเภทเชิงคุณภาพ (Categorical) และตัดข้อมูลตัวแปรเชิงปริมาณ (Numerical และ Continuous) ออกเพื่อให้สอดคล้องกับเทคนิคการเข้ารหัสที่เหมาะสมกับข้อมูลประเภทเชิงคุณภาพ รูปที่ 3 นั้นได้แสดงรายละเอียดของตัวแปรที่เลือกใช้

Feature	Category – Encoded Value
Sex	Male → 1, Female → 2
Chest Pain Type	Typical angina → 1, Non-anginal pain → 2, Atypical angina → 3, Asymptomatic → 4
Fasting Blood Sugar	Lower than 120 mg/ml → 1, Greater than 120 mg/ml → 2
Rest ECG	ST-T wave abnormality → 1, Normal → 2, Left ventricular hypertrophy → 3
Exercise Induced Angina	No → 1, Yes → 2
Slope of ST Segment	Flat → 0, Downsloping → 1, Upsloping → 2
Vessels Colored by Fluoroscopy	Zero → 1, One → 2, Two → 3, Three → 4, Four → 5
Thalassemia	Fixed Defect → 0, Reversable Defect → 1, Normal → 2, No → 3

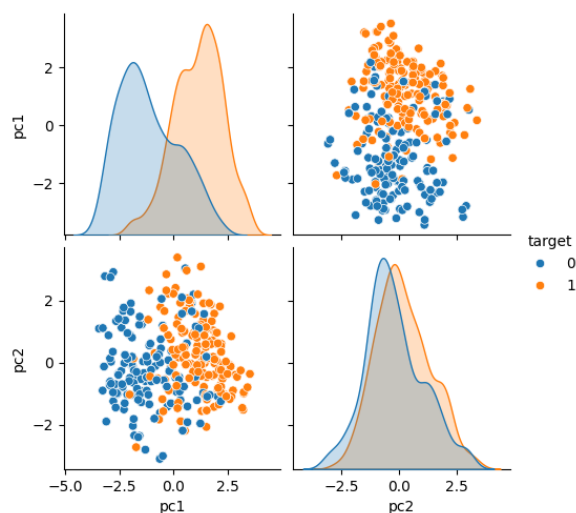
รูปที่ 3 : ตัวแปรของข้อมูล que เลือกเพื่อประมวลผล

ทั้งนี้การจำลองข้อมูลด้วยคอนตัมคอมพิวเตอร์ด้วยคอมพิวเตอร์คลาสสิกนั้นใช้ทรัพยากรที่สูง จึงต้องมีการกรองข้อมูล ให้ประมวลผลเฉพาะข้อมูลที่สำคัญ โดยขั้นตอนการเตรียมข้อมูล ดังนี้ การปรับขอบเขตข้อมูลแบบมาตรฐาน(Standard Scaling) เพื่อให้ตัวแปรมีค่าอยู่ในช่วงมาตรฐานเดียวกัน จากนั้นทำการลบค่าผิดปกติ (Outliers) ออกโดยใช้วิธีช่วงควอไทล์ (Interquartile Range: IQR) เพื่อตัดข้อมูลที่มีค่าเบี่ยงเบนมากไปดังรูปที่ 4 และลบข้อมูลซ้ำ (Duplicate Records) เพื่อลดอคติในการเรียนรู้และลดภาระในการคำนวณ จากนั้นจะถูกลดมิติของข้อมูลโดยใช้เทคนิคการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA) โดยชุดข้อมูลจะถูกทำให้ลดเหลือเพียง 2 มิติ เพื่อความรวดเร็วในการประมวลผลคอนตัมคอมพิวเตอร์ หลังผ่านกระบวนการเตรียมข้อมูลแล้ว ชุดข้อมูลจึงถูกลดลงเหลือเพียง 300 ตัว ซึ่งแสดงผลในรูปแบบ Pair Plot ตามรูปที่

5



รูปที่ 4 :ชุดข้อมูลก่อนและหลังในการลบค่าผิดปกติ



รูปที่ 5 :ชุดข้อมูลที่ถูกประมวลผลแล้ว

3.2 การเตรียมข้อมูลควอนตัม (Quantum data preparation)

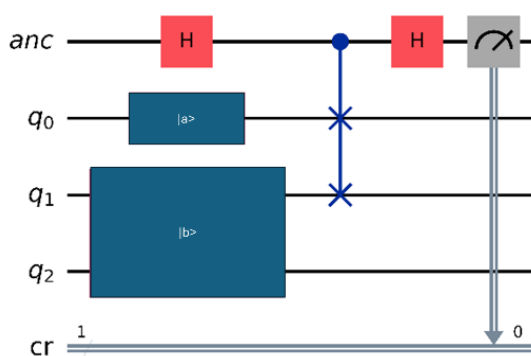
เนื่องจากควอนตัมคอมพิวเตอร์ ไม่สามารถนำชุดข้อมูลจากคลาสสิกคอมพิวเตอร์ใช้ได้ ในทันที จำเป็นต้องมีการแปลงข้อมูลแบบคลาสสิก ให้เป็นสถานะทางควอนตัม ก่อนนำไปใช้ ซึ่งวิธีการแปลงข้อมูลเป็นสถานะควอนตัม มีหลายวิธีที่สามารถเลือกใช้ ซึ่งในโมเดลควอนตัมที่นำเสนอ นั้น จะใช้วิธี Amplitude Encoding ซึ่งเป็นเทคนิคที่แปลงข้อมูลเข้าสู่ Amplitude ของสถานะควอนตัม โดยกำหนดให้ข้อมูลคลาสสิก คือ $x = (x_0, x_1, \dots, x_{N-1})$ ซึ่งจะสามารถอธิบายเป็นสมการ 6 และ 7

$$|\psi\rangle = \sum_{j=0}^{N-1} x_j |j\rangle \quad (6)$$

โดยที่ x_j คือองค์ประกอบของเวกเตอร์คลาสสิก และ $|j\rangle$ แทน binary representation ของ j ในระบบควอนตัมที่มี $\log_2(N)$ คิวบิต ทั้งนี้ เพื่อให้ $|\psi\rangle$ เป็นสถานะควอนตัมที่ต้องการ เวกเตอร์ x ต้องผ่านการทำให้เป็นเวกเตอร์หน่วย (Normalization) โดยต้องเป็นไปตามเงื่อนไขตามสมการ 7:

$$\sum_{j=0}^{N-1} |x_j|^2 = 1 \quad (7)$$

โมเดลควอนตัมที่นำเสนอ นั้น จะใช้ควอนตัมคอมพิวเตอร์ในการคำนวณระยะทางระหว่างจุด 2 จุด คล้ายกับการคำนวณระยะทางระหว่างจุดแบบ Euclidean distance ใน K-means algorithm ซึ่งจะใช้วงจรควอนตัมที่เรียกว่า Swap Test Circuit โดยจะวัดความน่าจะเป็นการซ้อนทับ (overlap) ระหว่างสถานะควอนตัมสองสถานะ ทั้งนี้การ encoding และ การคำนวณระยะทางในควอนตัม ได้ใช้แนวทางที่พัฒนาจาก [13] โดยมีลักษณะวงจรตามรูปที่ 6



รูปที่ 6 : Swap test full circuit

สถานะควอนตัม $|a\rangle$ และ $|b\rangle$ นั้นถูกแปลงมาจาก vector x, y ตามลำดับ ซึ่ง x และ y เปรียบเสมือนจุด 2 จุด ที่จะนำมาคำนวณระยะทางระหว่างกัน

$$|a\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes |x\rangle + |1\rangle \otimes |y\rangle) \quad (8)$$

$$|b\rangle = \frac{1}{\sqrt{2}}(|x\rangle |0\rangle + |x\rangle |1\rangle) \quad (9)$$

โดยที่

$$Z = |x|^2 + |y|^2 \quad (10)$$

Euclidean distance D ระหว่างเวกเตอร์สองตัวคำนวณได้โดย

$$D = \sqrt{2 - Z|\langle a | b \rangle|^2} \quad (11)$$

เทคนิคนี้พบว่า $|a\rangle$ ต้องการ 1 คิวบิต ในขณะที่ $|b\rangle$ ต้องการคิวบิตเป็นจำนวน $O(\log(N))+1$ โดยที่ N คือมิติของเวกเตอร์ข้อมูลคลาสสิก ทำให้เหมาะสำหรับใช้งานควอนตัมคอมพิวเตอร์ในยุคปัจจุบัน (NISQ)

3.3 ควอนตัมอัลกอริทึม (Quantum K-means algorithms)

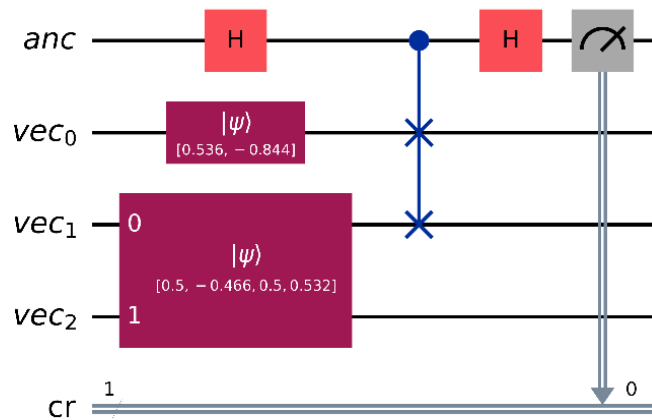
ในงานวิจัยนี้ มีการนำเสนอแนวทาง 2 รูปแบบสำหรับอัลกอริทึม Quantum K-Means สำหรับแนวทางแรกใช้กลยุทธ์แบบ 1 จุด ต่อ 1 เซนทรอยด์ (1-point-1-centroid: 1p1k) ซึ่งทำการเปรียบเทียบแต่ละจุดข้อมูลกับเซนทรอยด์ทั้งหมดแบบรายจุด โดยใช้การประมวลผลเชิงควอนตัมภาพรวมของขั้นตอนการทำงานแสดงไว้ใน Algorithm 1 โครงสร้างของแนวทางนี้มีลักษณะคล้ายคลึงกับอัลกอริทึม K-Means แบบดั้งเดิม โดยการสุ่มค่าเริ่มต้นของเซนทรอยด์ดำเนินการด้วยวิธี K-Means++ (บรรทัดที่ 1) ส่วนการอัปเดตเซนทรอยด์จะดำเนินการแบบคลาสสิก โดยใช้ค่าเฉลี่ยของจุดข้อมูลที่ถูกจัดกลุ่มร่วมกัน (บรรทัดที่ 14) จุดที่แตกต่างอย่างสำคัญอยู่ใน บรรทัดที่ 5-9 ซึ่งใช้การคำนวณเชิงควอนตัมเพื่อประมาณค่าระยะทาง โดยในขั้นตอนนี้ จุดข้อมูลและเซนทรอยด์แต่ละคู่จะถูกเข้ารหัสเป็นสถานะควอนตัม และถูกคำนวณโดย Swap test circuit ซึ่งอัลกอริทึมนี้เป็นโมเดลแบบผสมผสานระหว่างควอนตัมและคลาสสิก (Hybrid Quantum-Classical Model)

Algorithm 1 Quantum K-Means Algorithm**Require:** Dataset - D , Clusters - k , Max iter - T , quantum shots - s **Ensure:** Centroids - C , Assignments - A

```

1:  $C \leftarrow \text{InitCentroids}(D, k)$  ▷ Random init
2:  $\text{Converged} \leftarrow \text{False}$ ,  $t \leftarrow 0$ 
3: while not Converged and  $t < T$  do
4:    $\text{Dist} \leftarrow []$ 
5:   for each  $x \in D$  do
6:      $d_x \leftarrow []$ 
7:     for each  $c \in C$  do
8:        $Q \leftarrow \text{Encode}(x, c)$  ▷ Quantum Encoding
9:        $d \leftarrow \text{QDist}(Q, s)$  ▷ Quantum swap test
10:      Append  $d$  to  $d_x$ 
11:    end for
12:    Append  $d_x$  to  $\text{Dist}$ 
13:  end for
14:   $A \leftarrow \text{Assign}(\text{Dist})$  ▷ Classical Step
15:   $C' \leftarrow \text{Update}(D, A, k)$ 
16:  if  $C' = C$  then ▷ Check if centroids remain the same
17:    Converged  $\leftarrow \text{True}$ 
18:  else
19:     $C \leftarrow C'$ 
20:  end if
21:   $t \leftarrow t + 1$ 
22: end while
23: return  $(C, A)$ 

```



รูปที่ 7: Quantum circuit of algorithm 1

สำหรับแนวทางที่ 2 คือ กลยุทธ์แบบ 1 จุดต่อหลายเซนทรอยด์ (1-point-many-centroids: 1pMK) ซึ่งใช้ Quantum Parallelism ในการคำนวณระยะทางระหว่างจุดข้อมูลเพียงจุดเดียวกับเซนทรอยด์ทั้งหมดในเวลาเดียวกัน วิธีการนี้ช่วยลดจำนวนครั้งที่ต้องเรียกใช้งานวงจรควอนตัมในแต่ละ

ละรอบของอัลกอริธึมได้อย่างมีนัยสำคัญ ใน Algorithm 2 จะใช้วงจรควอนตัมเพียงวงจรเดียวในการเข้ารหัส 1 จุดข้อมูลและเซนทรอยด์ทั้งหมดพร้อมกัน(บรรทัดที่ 6) จากนั้นจะวัดค่ากับระยะทางของแต่ละเซนทรอยด์ได้ในครั้งเดียวแบบขนาน

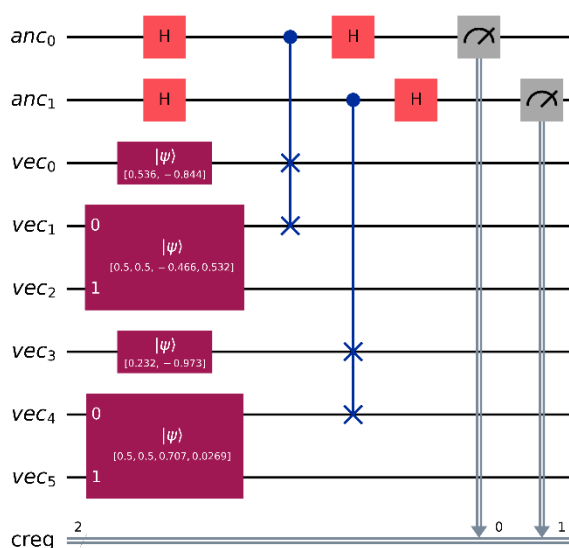
Algorithm 2 Quantum K-Means Algorithm (1P-MK)

Require: Dataset - D , Clusters - k , Max iter - T , Quantum shots - s

Ensure: Centroids - C , Assignments - A

```

1:  $C \leftarrow \text{InitCentroids}(D, k)$  ▷ Random initialization
2:  $\text{Converged} \leftarrow \text{False}$ ,  $t \leftarrow 0$ 
3: while not Converged and  $t < T$  do
4:    $\text{Dist} \leftarrow []$  ▷ Stores all distances
5:   for each  $x \in D$  do ▷ Iterate over each data point
6:      $Q \leftarrow \text{Encode}(x, C)$  ▷ Quantum encoding of all centroids in parallel
7:      $d_x \leftarrow \text{QDist}(Q, s)$  ▷ Quantum swap test for all centroids
8:     Append  $d_x$  to  $\text{Dist}$ 
9:   end for
10:   $A \leftarrow \text{Assign}(\text{Dist})$  ▷ Classical cluster assignment
11:   $C' \leftarrow \text{Update}(D, A, k)$  ▷ Update centroids based on new assignments
12:  if  $C' = C$  then ▷ Check if centroids remain the same
13:    Converged  $\leftarrow \text{True}$ 
14:  else
15:     $C \leftarrow C'$ 
16:  end if
17:   $t \leftarrow t + 1$ 
18: end while
19: return  $(C, A)$ 
  
```



รูปที่ 8: Quantum circuit of algorithm 2

เมื่อวิเคราะห์ความซับซ้อนเชิงเวลา (Time complexity) ของ Algorithm 1 และ Algorithm 2 โดยพิจารณาในส่วนของการหาคะแนน พบว่า Algorithm 1 (1p1k) เปรียบเทียบข้อมูลแต่ละจุดกับ centroid ที่ละตัว จะมี time complexity อยู่ที่ $O(K(\log N+S))$ ต่อจุดข้อมูลหนึ่งจุด โดย K คือจำนวน Centroid, N คือจำนวนมิติข้อมูล และ S คือจำนวนการวัดซ้ำ (shots) ที่ต้องใช้เพื่อประมาณผลลัพธ์จาก ancilla ซึ่งมีความน่าจะเป็น (probabilistic)

สำหรับ Algorithm 2 (1pMK) วงจรจะใช้ ancilla หลายตัว โดยแต่ละตัวควบคุมการวัดระยะทางกับเซนทรอยด์หนึ่งตัว ทำให้สามารถเปรียบเทียบกับหลายเซนทรอยด์ได้พร้อมกันภายในวงจรเดียว โดยมีข้อดีคือ สามารถลดจำนวนครั้งในการเรียกใช้งานวงจรจากเดิม K ครั้ง เหลือเพียง K/b ครั้ง โดยที่ b คือจำนวน ancilla ที่ใช้ต่อรอบ ซึ่งช่วยลด latency และ overhead ที่เกี่ยวข้องกับการจัดการควอนตัมคิวเตอร์ได้ในทางปฏิบัติ อย่างไรก็ตาม Time complexity คือ $O(K(\log N+S))$ ต่อจุดข้อมูลหนึ่งจุด

หากพิจารณาภาพรวมของอัลกอริทึมแบบเต็ม (Full loop) ซึ่งรวมทั้งการคำนวณระยะทางและการอัปเดต centroid จะพบว่า Algorithm 1 มี Time complexity โดยรวมอยู่ที่ $O(TMK(\log N + S))$ โดยที่ T คือจำนวนรอบการทำงานซ้ำ (iterations), M คือจำนวนจุดข้อมูล, K คือจำนวนเซนทรอยด์, N คือจำนวนมิติข้อมูล และ S คือจำนวน shots ในขณะที่ Algorithm 2 (Practical 1pMK) มี Time complexity ในขั้นตอนการหาระยะทางระดับเดียวกับ Algorithm 1 แต่จำนวนครั้งในการเรียกใช้งานวงจรสามารถลดลงจาก K ครั้ง เหลือ K/b ครั้งต่อจุดข้อมูล เมื่อใช้ ancilla พร้อมกัน b ตัว ดังนั้น Time complexity ของ Full loop จะเป็น $O(TM[K/b](\log N+S))$ โดยในกรณีที่ b มีค่าสูงเทียบเท่า K ค่า Time complexity จะลดลงเป็น $O(TM(\log N+S))$

เมื่อเปรียบเทียบการสเกลของทั้งสองวิธี จะเห็นว่า Algorithm 1 มีความซับซ้อนที่แปรผันโดยตรงกับจำนวน centroid K อย่างชัดเจน เนื่องจากต้องเรียกใช้งานวงจรควอนตัมแยกกัน K ครั้ง ต่อจุดข้อมูลหนึ่งจุด ในขณะที่ Algorithm 2 สามารถลดอัตราการเติบโตของความซับซ้อนลงได้ โดยใช้การประมวลผลแบบขนานผ่าน ancilla หลายตัว (ancilla parallelism) ซึ่งช่วยให้สามารถคำนวณระยะทางระหว่างจุดข้อมูลกับ centroid หลายตัวภายในวงจรเดียวได้ ลดภาระด้านเวลาในการประมวลผล โดยเฉพาะในกรณีที่จำนวน Centroid มีขนาดใหญ่ อย่างไรก็ตาม ควรพิจารณาความเหมาะสมของการใช้ควอนตัมคอมพิวเตอร์ เช่น จำนวน qubit ที่มีอยู่ในระบบจริงซึ่งยังคงจำกัดในยุค NISQ

เมื่อเปรียบเทียบกับวิธี Classical K-Means ซึ่งมี Time complexity เท่ากับ $O(TMNK)$ โดยที่ T คือจำนวนรอบการทำซ้ำ, M คือจำนวนข้อมูล, N คือจำนวนมิติ และ K คือจำนวนเซนทรอยด์ จะเห็นว่าวิธีคลาสสิกสามารถประมวลผลได้อย่างมีประสิทธิภาพบนฮาร์ดแวร์ทั่วไป เนื่องจากไม่มี overhead จากการเตรียมสถานะควอนตัมหรือการวัดซ้ำ (shots) อย่างไรก็ตาม เมื่อเทคโนโลยีควอนตัมในอนาคตมีการพัฒนาและสามารถลดจำนวน shots ที่ใช้ในการประมวลผลลงได้ ประสิทธิภาพของ Quantum K-Means โดยเฉพาะ Algorithm 2 อาจเข้าใกล้หรือดีกว่า Classical K-Means ได้ ในกรณีที่ข้อมูลมีมิติจำนวนมากหรือมีจำนวน Centroid มาก

นอกจากนี้เมื่อเปรียบเทียบกับ[11] ได้นำเสนออัลกอริทึม Quantum K-Means โดยใช้แนวทาง circuit-based approach และมี Time complexity เป็น $O(TMK)$ ซึ่งได้วิเคราะห์ว่า ประสิทธิภาพที่ดีกว่า Classical K-Means อย่างไรก็ตาม โครงสร้างของอัลกอริทึมยังคงใช้รูปแบบการเปรียบเทียบระยะทางแบบรายคู่ ระหว่างจุดข้อมูลและ Centroid โดยไม่ใช้เทคนิคคล้ายกับ Algorithm 2 ดังนั้นจึงยังไม่สามารถลดความซับซ้อนเชิงเวลาได้ในเชิงโครงสร้างอย่างชัดเจนเมื่อเทียบกับกัน

3.4 การทดสอบและประเมินผล (Test and Evaluation)

การทดสอบโมเดลควอนตัมที่นำเสนอ ทั้ง 2 แนวทาง ได้แก่ 1p1k และ 1pMk โดยจะทดสอบบนควอนตัมคอมพิวเตอร์รูปแบบจำลอง 2 รูปแบบ คือ แบบจำลองควอนตัมที่มีสัญญาณรบกวน (Noisy Quantum Simulator) ซึ่งจำลองลักษณะการทำงานของควอนตัมคอมพิวเตอร์จริง โดยใช้ รุ่น ibm_brisbane เป็นแบบจำลอง และ แบบจำลองจำลองควอนตัมในอุดมคติ (Ideal Quantum Simulator) ซึ่งเป็นการประเมินสมรรถนะของอัลกอริทึมในสภาพแวดล้อมที่ไม่มีสัญญาณรบกวน โดยการทดลองทั้งหมดใช้ชุดข้อมูลโรคหัวใจประสิทธิภาพของอัลกอริทึมควอนตัมทั้งสองถูกประเมินโดยการเปรียบเทียบซึ่งกันและกัน รวมถึงเปรียบเทียบกับอัลกอริทึม K-Means ในคลาสสิกคอมพิวเตอร์

สำหรับการประเมินผลนั้นจะมีการวัดค่าเพื่อประสิทธิภาพการจัดกลุ่มของโมเดลควอนตัมที่นำเสนอ ได้แก่

- Sum of Squared Errors(SSE)

- Silhouette Score(sil) ซึ่งสะท้อนความเหมาะสมของการจัดกลุ่ม
- ค่าความสอดคล้องกับกลุ่มข้อมูลจริง (V-Measure) สำหรับเปรียบเทียบความแม่นยำในการจำแนกกลุ่ม
- ค่าความแม่นยำ(Accuracy)
- Precision
- Recall
- F1-score



929906852

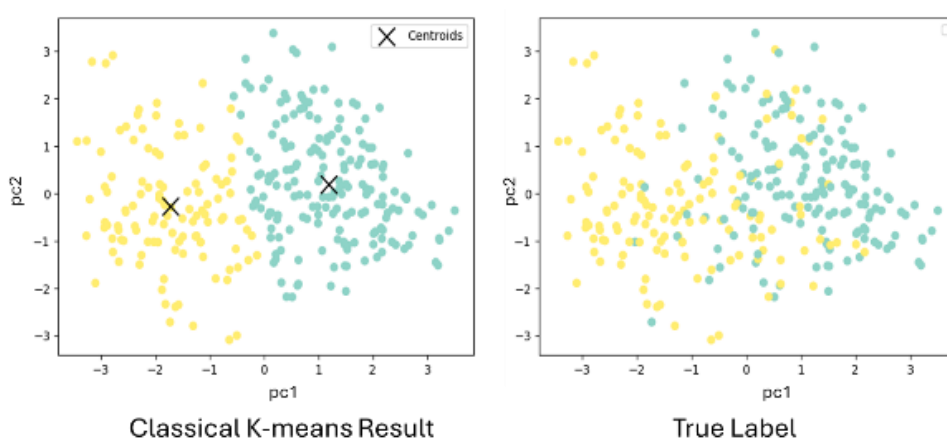
CU iThesis 6470323821 thesis / recv: 28052568 21:50:52 / seq: 12

บทที่ 4

วิธีดำเนินการวิจัย

4.1 ผลการวิจัยการจำแนกโรคหัวใจจาก ข้อมูลชุดโรคหัวใจ

ในรูปที่ 9 คือชุดข้อมูลที่ถูกจัดการข้อมูลแล้ว โดยมีจำนวนข้อมูล 300 อัน แล้วถูกนำมาจำแนกด้วย K-means ในคลาสสิกคอมพิวเตอร์ และ จำแนกตาม Ground Truth label ตามลำดับ เพื่อมาใช้เป็นค่าพื้นฐานสำหรับเปรียบเทียบ

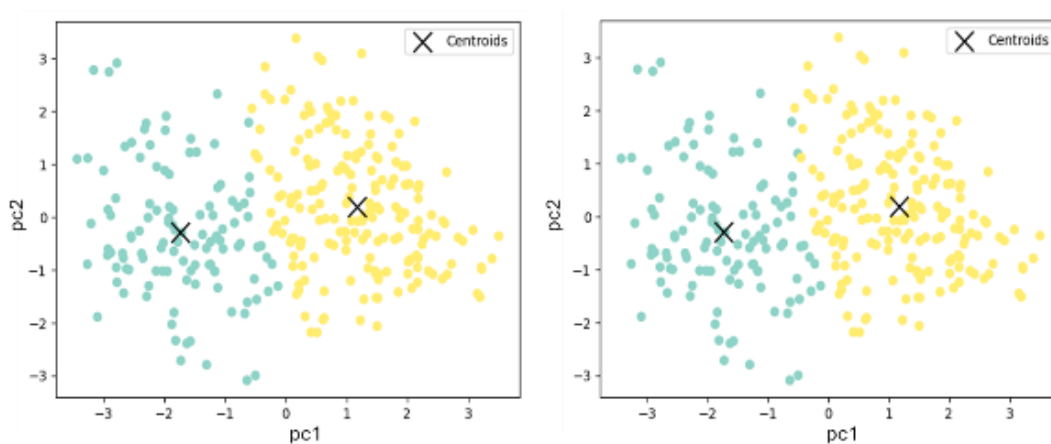


รูปที่ 9: Classical K-Means result and True Label

ในรูปที่ 10 แสดงผลการทดลองที่ได้จากการใช้ Quantum Simulator ที่มี noise จากควอนตัมคอมพิวเตอร์จริง บนอัลกอริทึมทั้งสอง ได้แก่ Algorithm 1 (1p1k) และ Algorithm 2 (1pMK) ให้ผลลัพธ์ในการจัดกลุ่มที่มีความใกล้เคียงกันมาก และมีความคล้ายกับผลลัพธ์ของ K-Means แบบคลาสสิกอย่างชัดเจน แม้ว่าผลการจัดกลุ่มของโมเดลควอนตัมจะมีความคล้ายคลึงกับคลาสสิก แต่ยังพบ error ในบางส่วน เมื่อเปรียบเทียบกับ True label อย่างไรก็ดีตาม error ดังกล่าวไม่ได้เกิดจากความคลาดเคลื่อนจากการคำนวณ แต่คาดว่าเกิดจากตำแหน่งของจุดข้อมูลที่อยู่ใกล้กับขอบเขตระหว่างกลุ่ม (Cluster Boundaries) ซึ่งส่งผลให้เกิดความกำกวมในการจัดกลุ่ม

สำหรับการทดลองที่ได้จากการใช้ Quantum Simulator ที่เป็น ideal noise ดังรูปที่ 11 พบว่าผลลัพธ์นั้นยังคงคล้ายเดิม โดยผลการจัดกลุ่มของโมเดลควอนตัมจะมีความคล้ายคลึงกับคลาสสิก

Experiment result on simulator with noise(ibm_brisbane)

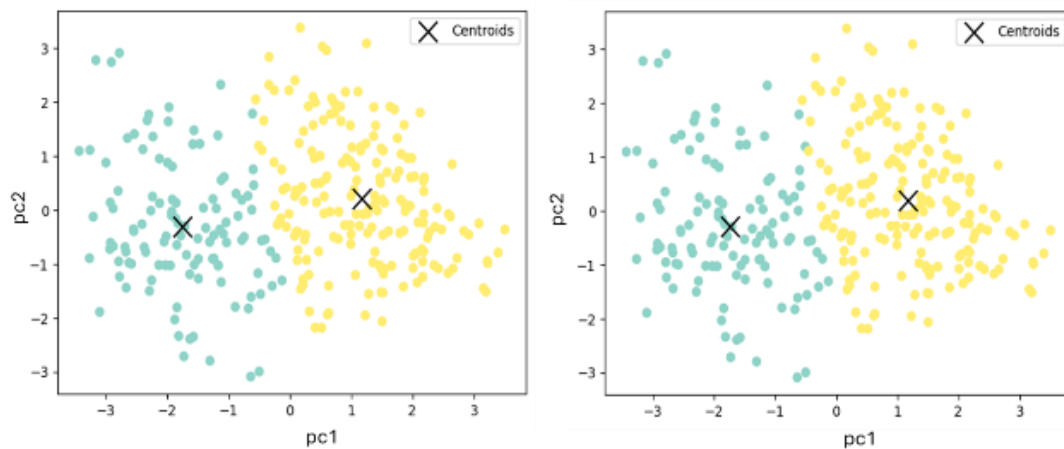


Algorithm 1 result

Algorithm 2 result

รูปที่ 10: Experiment result on simulator with noise(ibm_brisbane)

Experiment result on ideal simulator (qasm_simulator)



Algorithm 1 result

Algorithm 2 result

รูปที่ 11: Experiment result on ideal simulator(qasm_simulator)

ตารางที่ 1 Summary of evaluation metrics

Algorithm	Accuracy	Precision	Recall	F1	SSE	Sil	V-meas
Classical K-mean	0.82	0.83	0.82	0.82	665.06	0.433	0.334
1p1k (ibm_brisbane)	0.82	0.83	0.82	0.83	665.17	0.433	0.328
1pMk (ibm_brisbane)	0.82	0.83	0.82	0.83	665.17	0.434	0.353
1p1k(qasm_simulator)	0.83	0.83	0.82	0.83	665.17	0.435	0.346
1pMk(qasm_simulator)	0.82	0.83	0.82	0.83	665.17	0.433	0.328

ตารางที่ 2 Comparison with other work

Algorithm	Accuracy	Precision	Recall	F1
XGboost [18]	0.97	0.95	0.97	0.93
Correlation Based Feature Selection [19]	0.85	0.80	0.85	0.79
Quantum Circuit [11]	0.96	0.97	0.94	0.96
Classical K-mean	0.82	0.83	0.82	0.82
1p1k(ibm_brisbane)	0.82	0.83	0.82	0.83
1pmk(ibm_brisbane)	0.82	0.83	0.82	0.83
1p1k(qasm)	0.83	0.83	0.82	0.83
1pmk(qasm)	0.82	0.83	0.82	0.83



ผลการทดลองที่สรุปในตารางที่ 1 แสดงให้เห็นว่าอัลกอริธึม Quantum K-Means ที่นำเสนอทั้งสองแนวทาง ได้แก่ 1p1k และ 1pMK สามารถให้ประสิทธิภาพในการจัดกลุ่มที่ใกล้เคียงกับอัลกอริธึม K-Means แบบคลาสสิก ได้ในทุกเกณฑ์การประเมิน โดยเฉพาะค่าความแม่นยำ (Accuracy) ที่อยู่ในช่วงระหว่าง 0.82–0.83 รวมถึงค่า Precision, Recall, และ F1-score ที่มีความแตกต่างเพียงเล็กน้อย

นอกจากนี้ ตัวชี้วัดผลอื่นๆ เช่น SSE, Silhouette Score, และ V-Measure ก็ยังคงมีความสอดคล้องกันระหว่างอัลกอริธึมแบบคลาสสิกและควอนตัม แม้ในกรณีที่ทำการทดลองบนควอนตัมคอมพิวเตอร์แบบใช้ noise จริงของ ibm_brisbane ผลลัพธ์ดังกล่าวแสดงให้เห็นว่าอัลกอริธึม Quantum K-Means มีความเสถียร ภายใต้ข้อจำกัดของฮาร์ดแวร์จริง

ในตารางที่ 2 ได้แสดงการเปรียบเทียบผลลัพธ์กับงานวิจัยอื่นๆ ที่ใช้ชุดข้อมูลโรคหัวใจชุดเดียวกัน โดยพบว่าโมเดลขั้นสูง เช่น XGBoost [18] สามารถทำค่า Accuracy ได้สูงถึงประมาณ 0.97 เนื่องจากการเรียนรู้แบบมีผู้สอน (supervised learning) ที่สามารถใช้ข้อมูล label เพื่อปรับขอบเขตการตัดสินใจได้อย่างแม่นยำ อย่างไรก็ตาม หากพิจารณาเฉพาะแนวทางแบบไม่มีผู้สอน (unsupervised clustering) อัลกอริธึม Quantum K-Means ที่นำเสนอในงานวิจัยนี้ยังคงแสดงให้เห็นถึงประสิทธิภาพที่สามารถแข่งขันได้ เมื่อเปรียบเทียบกับวิธี Correlation-Based Feature Selection [19] ซึ่งทำค่า Accuracy ได้ที่ 0.85 โมเดลควอนตัมที่มีผลลัพธ์ที่ใกล้เคียงกัน แม้จะอยู่ภายใต้ข้อจำกัดของการเรียนรู้แบบไม่มีผู้สอน เมื่อเทียบกับงานวิจัย Quantum Circuit Clustering [11] ซึ่งรายงานค่า Accuracy ที่ 0.96 ผลลัพธ์ของวิธีที่นำเสนอมีผลที่ต่ำกว่า

อย่างไรก็ตาม ปัจจัยสำคัญที่ส่งผลต่อค่า Accuracy ของการจัดกลุ่มที่ได้อยู่ประมาณ ~0.82–0.83 คือ การทับซ้อนระหว่าง True label 2 กลุ่ม ตามที่แสดงไว้ใน ภาพที่ 8 ซึ่งสะท้อนว่าทั้งสองกลุ่มไม่สามารถแยกกันได้อย่างชัดเจน ดังนั้น ข้อจำกัดของประสิทธิภาพ จึงไม่ได้มาจากโมเดลควอนตัม

เมื่อพิจารณาโมเดลควอนตัม ถูกใช้ในปัญหาที่มีความซับซ้อนมากขึ้น เช่น การจัดกลุ่มในชุดข้อมูลที่มีจำนวนตัวแปรมากกว่า 2 มิติ Quantum K-Means ทั้งสองแบบจะเผชิญกับ ข้อจำกัดที่แตกต่างกัน

- Algorithm 1 (1p1k) ทำการเปรียบเทียบข้อมูลแต่ละจุดกับเซนทรอยด์แต่ละตัวแบบรายคู่ ส่งผลให้สามารถปรับใช้กับชุดข้อมูลขนาดใหญ่ได้ง่ายกว่า เนื่องจากต้องใช้คิวบิตน้อยและสามารถประมวลผลวนรอบได้ แต่ต้องรันวงจรหลายครั้งกว่า ส่งผลให้เวลานาน
- ในทางตรงกันข้าม Algorithm 2 (1pMK) จะแปลงข้อมูลเซนทรอยด์ทั้งหมดไว้ในวงจรควอนตัมเพียงวงจรเดียว และเปรียบเทียบกับข้อมูล ใช้ประโยชน์จาก Quantum Parallelism ช่วยลดจำนวนรอบการรันวงจร แต่มีข้อเสียคือ ต้องการจำนวนคิวบิตสูงขึ้น ส่งผลให้มีความเสี่ยงต่อข้อจำกัดของฮาร์ดแวร์และความผิดพลาดจาก noise มากขึ้นเมื่อจำนวนเซนทรอยด์หรือมิติเพิ่มขึ้น

ดังนั้นการเลือกใช้อัลกอริทึม Quantum K-Means ที่เหมาะสมสำหรับชุดข้อมูลที่มีความซับซ้อนและขนาดใหญ่ จึงควรพิจารณาระหว่าง quantum parallel กับข้อจำกัดของทรัพยากรควอนตัมคอมพิวเตอร์



929906852

บทที่ 5

สรุปผลการวิจัยและเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้ได้ศึกษาการประยุกต์ใช้อัลกอริทึม K-Means แบบควอนตัม-คลาสสิกผสมผสาน (Hybrid Quantum-Classical) สำหรับการจัดกลุ่มข้อมูลผู้ป่วยที่เกี่ยวข้องกับโรคหัวใจ โดยผสานการคำนวณระยะทางเชิงควอนตัมผ่านวงจร Swap Test เข้ากับกระบวนการจัดกลุ่มแบบคลาสสิก งานวิจัยนี้แสดงให้เห็นถึง ศักยภาพของการประมวลผลเชิงควอนตัมในการเพิ่มประสิทธิภาพการคำนวณและการขนาน (Parallelism) ภายในกระบวนการจัดกลุ่มแบบคลาสสิก มีการนำเสนออัลกอริทึมควอนตัม 2 แนวทาง ได้แก่ Algorithm 1: 1 จุดข้อมูลต่อ 1 เซนทรอยด์ (1p1k) และ Algorithm 2: 1 จุดข้อมูลต่อหลายเซนทรอยด์ (1pMK) ทั้งสองอัลกอริทึมได้รับการทดสอบโดยใช้ชุดข้อมูลจากฐานข้อมูลสาธารณะ ที่มีข้อมูลผู้ป่วยมากกว่า 1,000 ข้อมูล โดยผลการทดลองจากตัวจำลองควอนตัมทั้งแบบมีสัญญาณรบกวน (noisy simulator) และแบบอุดมคติ (ideal simulator) พบว่า ผลลัพธ์ของการจัดกลุ่มมีความใกล้เคียงกับอัลกอริทึม K-Means แบบคลาสสิก และยังสามารถเปรียบเทียบประสิทธิภาพจากตัวชี้วัดต่างๆ เช่น Accuracy, F1-score, Silhouette และ SSE เป็นต้น อย่างไรก็ตาม อัลกอริทึมการจัดกลุ่มแบบควอนตัมยังคงเผชิญกับข้อจำกัดจากฮาร์ดแวร์ควอนตัมในปัจจุบัน ได้แก่ จำนวนคิวบิตที่จำกัด, ความลึกของวงจร (Circuit Depth), สัญญาณรบกวน (Noise) ที่เกิดขึ้นระหว่างการคำนวณ เป็นต้น โดยเฉพาะ Algorithm 2 จะพบปัญหา หากใช้กับข้อมูลที่มีมิติสูงหรือจำนวนกลุ่มมาก

ถึงแม้ว่าจะมีข้อจำกัดดังกล่าว ผลการทดลองยังคงชี้ให้เห็นถึง ศักยภาพของการจัดกลุ่มด้วยควอนตัมในงานวิเคราะห์ข้อมูลทางการแพทย์ โดยเฉพาะอย่างยิ่งในการสนับสนุนการวินิจฉัยโรคในระยะแรกและการประเมินความเสี่ยงของผู้ป่วยในอนาคต

5.2 ข้อเสนอแนะ

จากผลการศึกษาในครั้งนี้ สามารถสรุปข้อเสนอแนะสำหรับการวิจัยในอนาคตได้ดังนี้:

1. การออกแบบอัลกอริทึมที่เป็นควอนตัมทั้งหมด (Full Quantum Algorithm) เพื่อให้สามารถใช้ประโยชน์จากคุณสมบัติควอนตัม ได้มากยิ่งขึ้น และลด Overhead จากการสื่อสารระหว่างระบบคลาสสิกและควอนตัม



2. การเพิ่มประสิทธิภาพของวงจรควอนตัม (Circuit Optimization) ควรมีการพัฒนาโครงสร้างวงจรที่มีความลึกต่ำและทนทานต่อ noise เพื่อให้สามารถทำงานบนฮาร์ดแวร์ควอนตัมจริงได้อย่างมีประสิทธิภาพยิ่งขึ้น
3. การออกแบบวิธีเข้ารหัสข้อมูลที่สามารถขยายขนาดได้ (Scalable Encoding) ควรพัฒนาเทคนิคการเข้ารหัสที่เหมาะสมกับข้อมูลที่มีมิติจำนวนมาก โดยไม่เพิ่มจำนวนคิวบิตหรือความซับซ้อนของวงจรจนเกินไป
4. การทดสอบกับชุดข้อมูลขนาดใหญ่บนคอมพิวเตอร์ควอนตัมจริง ควรขยายการทดลองไปยังชุดข้อมูลที่มีขนาดและความซับซ้อนมากขึ้น พร้อมทั้งประเมินประสิทธิภาพบนควอนตัมคอมพิวเตอร์ที่มีอยู่จริง เพื่อศึกษาความสามารถของระบบในบริบทที่ใกล้เคียงการใช้งานจริงมากขึ้น



929906852

บรรณานุกรม

- [1] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, 2018.
- [2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge: Cambridge University Press, 2010.
- [3] P. W. Shor, "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer," *SIAM J. Comput.*, vol. 26, pp. 1484-1509, 1995.
- [4] L. K. Grover, "Quantum Mechanics Helps in Searching for a Needle in a Haystack," *Physical Review Letters*, vol. 79, no. 2, pp. 325-328, 07/14/ 1997, doi: 10.1103/PhysRevLett.79.325.
- [5] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," *Electronics*, vol. 9, no. 8, p. 1295, 2020. [Online]. Available: <https://www.mdpi.com/2079-9292/9/8/1295>.
- [6] D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding," in *ACM-SIAM Symposium on Discrete Algorithms*, 2007.
- [7] I. Kerenidis, J. Landman, A. Luongo, and A. Prakash, "q-means: A quantum algorithm for unsupervised machine learning," *ArXiv*, vol. abs/1812.03584, 2018.
- [8] S. U. Khan, A. J. Awan, and G. Vall-Ilosera, "K-Means Clustering on Noisy Intermediate Scale Quantum Computers," *ArXiv*, vol. abs/1909.12183, 2019.
- [9] C. Gong, Z. Y. Dong, A. B. Gani, and H. Qi, "Quantum k-means algorithm based on trusted server in quantum cloud computing," *Quantum Information Processing*, vol. 20, 2020.
- [10] S. Deshmukh, B. K. Behera, and P. Mulay, "Patient Data Analysis with the Quantum Clustering Method," *Quantum Reports*, vol. 5, no. 1, pp. 138-155, 2023. [Online]. Available: <https://www.mdpi.com/2624-960X/5/1/10>.
- [11] S. S. Kavitha and N. Kaulgud, "Quantum K-means clustering method for detecting heart disease using quantum circuit approach," *Soft Computing*, 2022/05/31 2022, doi: 10.1007/s00500-022-07200-x.
- [12] O. Ouedrhiri, O. Banouar, S. El Hadaj, and S. Raghay, "Intelligent recommender system based on quantum clustering and matrix completion," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 15, p. e6943, 2022, doi:



<https://doi.org/10.1002/cpe.6943>.

- [13] H. Urgelles, P. Picazo-Martínez, and J. F. Monserrat, "Application of Quantum Computing to Accurate Positioning in 6G Indoor Scenarios," in *ICC 2022 - IEEE International Conference on Communications*, 16-20 May 2022 2022, pp. 643-647, doi: 10.1109/ICC45855.2022.9838523.
- [14] A. Poggiali, A. Berti, A. Bernasconi, G. M. Del Corso, and R. Guidotti, "Quantum clustering with k-Means: A hybrid approach," *Theoretical Computer Science*, vol. 992, p. 114466, 2024/04/21/ 2024, doi: <https://doi.org/10.1016/j.tcs.2024.114466>.
- [15] S. DiAdamo, C. O'Meara, G. Cortiana, and J. Bernabé-Moreno, "Practical quantum k-means clustering: Performance analysis and applications in energy grid classification," *IEEE Transactions on Quantum Engineering*, vol. 3, pp. 1-16, 2022.
- [16] H. Gupta, H. Varshney, T. K. Sharma, N. Pachauri, and O. P. Verma, "Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction," *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 3073-3087, 2022/08/01 2022, doi: 10.1007/s40747-021-00398-7.
- [17] K. Gangal. "Heart Disease Dataset (UCI)." <https://www.kaggle.com/datasets/ketangangal/heart-disease-dataset-uci/data> (accessed.
- [18] H. El-Sofany, B. Bouallegue, and Y. M. A. El-Latif, "A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method," *Scientific Reports*, vol. 14, no. 1, p. 23277, 2024/10/07 2024, doi: 10.1038/s41598-024-74656-2.
- [19] A. M. Kuruvilla and N. V. Balaji, "Heart disease prediction system using Correlation Based Feature Selection with Multilayer Perceptron approach," *IOP Conference Series: Materials Science and Engineering*, vol. 1085, no. 1, p. 012028, 2021/02/01 2021, doi: 10.1088/1757-899X/1085/1/012028.





929906852

CU iThesis 6470323821 thesis / recv: 28052568 21:50:52 / seq: 12

ประวัติผู้เขียน

ชื่อ-สกุล	Panuwat Tanapornchinpong
วัน เดือน ปี เกิด	26 March 1997
สถานที่เกิด	Thailand
วุฒิการศึกษา	B.Eng. in Mechanical Engineering, Sirindhorn International Institute of Technology
ที่อยู่ปัจจุบัน	56 Ladprakaw 32, Ladpraw Bangkok 10230
ผลงานตีพิมพ์	-



929906852

CU iThesis 6470323821 thesis / recv: 28052568 21:50:52 / seq: 12