

การเปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์ทิศทางการปิดอีเธอร์เรียม

นายจักรกริช กองจินดา 610510540

รายงานการศึกษาค้นคว้าอิสระประกอบการศึกษาวิชาการศึกษาด้วยตนเอง (208499)

เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

สาขาวิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

ปีการศึกษา 2564



การเปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์ทิศทางการปิดอีเธอร์เรียม

นายจักรกริช กองจินดา 610510540

รายงานการศึกษาค้นคว้าอิสระประกอบการศึกษาวิชาการศึกษาด้วยตนเอง (208499)

เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

สาขาวิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

ปีการศึกษา 2564

## กิตติกรรมประกาศ

การศึกษาครั้งนี้สำเร็จลุล่วงไปด้วยดี ด้วยความกรุณาจากคณาจารย์และผู้ที่เกี่ยวข้องทุกท่าน โดยเฉพาะอย่างยิ่ง อาจารย์ ดร.วีรินทร์ดา วงศ์รินทร์ อาจารย์ที่ปรึกษาหลัก และผู้ช่วยศาสตราจารย์ ดร.มานะชัย รอดชื่น อาจารย์ที่ปรึกษาร่วม ที่คอยให้คำแนะนำ และคำปรึกษาตลอดการทำวิจัย ผู้วิจัยรู้สึกซาบซึ้งต่อความกรุณาของอาจารย์ทั้งสองท่านเป็นอย่างยิ่ง จึงขอขอบพระคุณเป็นอย่างสูง ณ โอกาสนี้

ขอขอบพระคุณอาจารย์ ดร.กมลรัตน์ สุภาวรรณ กรรมการสอบโครงร่างที่กรุณาให้ข้อเสนอแนะ และแนวคิดต่าง ๆ ที่เป็นประโยชน์ต่อการศึกษาครั้งนี้

ขอขอบพระคุณ คณาจารย์ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ทุกท่าน ที่ได้ให้คำแนะนำ และให้ความรู้ทางวิชาการสถิติเพื่อนำมาประยุกต์ใช้ในการศึกษาครั้งนี้จนสำเร็จสมบูรณ์ และขอขอบคุณครอบครัว เพื่อน ตลอดจนบุคคลต่าง ๆ สำหรับความช่วยเหลือ คำปรึกษา คำแนะนำ ตลอดจนเป็นกำลังใจให้เสมอมา

ท้ายที่สุดนี้ ผู้วิจัยหวังเป็นอย่างยิ่งว่าการศึกษาครั้งนี้จะเป็นประโยชน์สำหรับผู้สนใจศึกษาในแนวทางเดียวกัน หากเนื้อหาในรายงานเล่มนี้มีข้อผิดพลาดประการใด ผู้วิจัยต้องขออภัยมา ณ ที่นี้ด้วย

จักรกริช กองจินดา

มีนาคม 2565

**ชื่อเรื่อง:** การเปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์ทิศทางราคาปิดอีเธอเรียม

**ผู้รับผิดชอบโครงการ:** นายจักรกริช กองจินดา                      **รหัสนักศึกษา** 610510540

**อาจารย์ที่ปรึกษาหลัก:** ดร.วีรินทร์ดา วงศ์รินทร์

**อาจารย์ที่ปรึกษาร่วม:** ผศ.ดร.มานะชัย รอดชื่น

### **บทคัดย่อ**

การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์ทิศทางราคาปิดรายวันของเงินสกุลดิจิทัลอีเธอเรียมที่ได้จากวิธีการพยากรณ์ 4 วิธี ได้แก่ วิธีต้นไม้ตัดสินใจ วิธี Random Forest วิธี Extreme Gradient Boosting (XGBoost) และวิธี Extra Tree ในการพยากรณ์ทิศทางราคาปิดล่วงหน้า 1 วัน โดยใช้ข้อมูลตั้งแต่วันที่ 7 สิงหาคม พ.ศ. 2558 ถึงวันที่ 31 มกราคม พ.ศ. 2565 จำนวนทั้งสิ้น 2,370 วัน และทำการแบ่งข้อมูลออกเป็น 5 ชุด ด้วยการตรวจสอบสลับไขว้แบบอนุกรมเวลา (5-Fold Cross-Validation) โดยเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์ ได้แก่ ค่า Accuracy, Recall และ F1 - Score ผลการศึกษพบว่า วิธี Extra Tree เป็นวิธีที่มีประสิทธิภาพในการพยากรณ์ทิศทางราคาปิดอีเธอเรียมมากที่สุด โดยให้ค่า Accuracy เท่ากับ 57.18% ค่า Recall เท่ากับ 57.50% และค่า F1 - Score เท่ากับ 57.50% โดยรองลงมาคือวิธี Random Forest ที่ให้ค่า Accuracy เท่ากับ 47.69% ค่า Recall เท่ากับ 46.50% และค่า F1 - Score ที่ 46.50%

**คำสำคัญ :** อีเธอเรียม, ทิศทางของราคา, ราคาปิด, เงินสกุลดิจิทัล

## สารบัญ

	หน้า
กิตติกรรมประกาศ	ก
บทคัดย่อ	ข
สารบัญ	ค
สารบัญรูป	ค
สารบัญตาราง	ฅ
<b>1 บทนำ</b>	<b>1</b>
1.1 ที่มาและความสำคัญ . . . . .	1
1.2 วัตถุประสงค์ของการวิจัย . . . . .	2
1.3 ประโยชน์ที่คาดว่าจะได้รับ . . . . .	3
1.4 นิยามคำศัพท์ . . . . .	3
<b>2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง</b>	<b>4</b>
2.1 ความรู้พื้นฐานเกี่ยวกับเงินสกุลดิจิทัล . . . . .	4
2.2 เครื่องมือวิเคราะห์เชิงเทคนิค . . . . .	5
2.3 วิธีการพยากรณ์ . . . . .	12
2.4 งานวิจัยที่เกี่ยวข้อง . . . . .	18
<b>3 วิธีการดำเนินการวิจัย</b>	<b>20</b>
3.1 ขอบเขตของการวิจัย . . . . .	20
3.2 กรอบแนวคิดการวิจัย . . . . .	21
3.3 การจัดเตรียมข้อมูล . . . . .	21
3.4 ขั้นตอนการดำเนินการวิจัย . . . . .	26
<b>4 ผลการวิจัย</b>	<b>27</b>
4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น . . . . .	27
4.2 ผลการค้นหาค่าพารามิเตอร์ที่เหมาะสมที่สุดสำหรับตัวแบบพยากรณ์ด้วยวิธี Grid Search	29
4.3 ตารางแสดง Confusion Matrix ของตัวแบบพยากรณ์ . . . . .	31
4.4 ผลการวิเคราะห์ประสิทธิภาพของตัวแบบพยากรณ์ . . . . .	33
4.5 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ . . . . .	36

<b>5</b>	<b>สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ</b>	<b>37</b>
5.1	สรุปผลการวิจัย . . . . .	37
5.2	อภิปรายผล . . . . .	38
5.3	ข้อเสนอแนะ . . . . .	38
	<b>บรรณานุกรม</b>	<b>39</b>

## สารบัญรูป

	หน้า
รูปที่ 2.1 ตัวอย่างแท่งเทียนราคาเปิด ราคาสูงสุด ราคาต่ำสุด และราคาปิด . . . . .	5
รูปที่ 2.2 ตัวอย่างกราฟแท่งเทียนและเครื่องมือ SMA . . . . .	6
รูปที่ 2.3 ตัวอย่างกราฟแท่งเทียนและเครื่องมือ WMA . . . . .	7
รูปที่ 2.4 ตัวอย่างกราฟแท่งเทียนและเครื่องมือ MACD . . . . .	8
รูปที่ 2.5 ตัวอย่างกราฟแท่งเทียนและเครื่องมือ %K . . . . .	9
รูปที่ 2.6 ตัวอย่างกราฟแท่งเทียนและเครื่องมือ RSI . . . . .	10
รูปที่ 2.7 ตัวอย่างกราฟแท่งเทียนและเครื่องมือ %R . . . . .	11
รูปที่ 2.8 ตัวอย่างต้นไม้ตัดสินใจ . . . . .	12
รูปที่ 2.9 ตัวอย่างข้อมูลเพื่อใช้ในการคำนวณ Information Gain . . . . .	14
รูปที่ 2.10 หลักการสร้างตัวแบบ Random Forest . . . . .	15
รูปที่ 2.11 หลักการสร้างตัวแบบ XGBoost . . . . .	16
รูปที่ 2.12 หลักการสร้างตัวแบบ Extra Tree . . . . .	16
รูปที่ 3.1 กราฟแสดงราคาปิดรายวันของเงินสกุลดิจิทัลอีเธอเรียม ตั้งแต่วันที่ 7 สิงหาคม พ.ศ. 2558 ถึงวันที่ 31 มกราคม พ.ศ. 2565 . . . . .	20
รูปที่ 3.2 กรอบแนวคิดการวิจัย . . . . .	21
รูปที่ 3.3 ตัวอย่างการตรวจสอบสลับไขว้ . . . . .	22
รูปที่ 3.4 ตัวอย่างการแบ่งข้อมูลอนุกรมเวลาด้วยวิธีตรวจสอบสลับไขว้ . . . . .	22
รูปที่ 4.1 ร้อยละของราคาปิดอีเธอเรียมรายวันจำแนกตามทิศทางของราคาปิด และชุดข้อมูล	28
รูปที่ 4.2 ต้นไม้ตัดสินใจที่ได้จากตัวแบบ . . . . .	34



## สารบัญตาราง

	หน้า
ตารางที่ 2.1 Confusion Matrix . . . . .	17
ตารางที่ 3.1 การกำหนดช่วงของค่าพารามิเตอร์สำหรับวิธีต้นไม้ตัดสินใจ . . . . .	23
ตารางที่ 3.2 การกำหนดช่วงของค่าพารามิเตอร์สำหรับวิธี Random Forest . . . . .	24
ตารางที่ 3.3 การกำหนดช่วงของค่าพารามิเตอร์สำหรับวิธี XGBoost . . . . .	25
ตารางที่ 3.4 การกำหนดช่วงของค่าพารามิเตอร์สำหรับวิธี Extra Tree . . . . .	25
ตารางที่ 4.1 ค่าสถิติพื้นฐานของตัวแปรที่ใช้ในการศึกษา . . . . .	27
ตารางที่ 4.2 จำนวนข้อมูลทั้งหมดในชุดฝึกสอนและชุดทดสอบ . . . . .	28
ตารางที่ 4.3 ค่าพารามิเตอร์ที่เหมาะสมที่สุดของตัวแบบต้นไม้ตัดสินใจ . . . . .	29
ตารางที่ 4.4 ค่าพารามิเตอร์ที่เหมาะสมที่สุดของตัวแบบ Random Forest . . . . .	29
ตารางที่ 4.5 ค่าพารามิเตอร์ที่เหมาะสมที่สุดของตัวแบบ XGBoost . . . . .	30
ตารางที่ 4.6 ค่าพารามิเตอร์ที่เหมาะสมที่สุดของตัวแบบ Extra Tree . . . . .	30
ตารางที่ 4.7 Confusion Matrix ของตัวแบบต้นไม้ตัดสินใจ . . . . .	31
ตารางที่ 4.8 Confusion Matrix ของตัวแบบ Random Forest . . . . .	32
ตารางที่ 4.9 Confusion Matrix ของตัวแบบ XGBoost . . . . .	32
ตารางที่ 4.10 Confusion Matrix ของตัวแบบ Extra Tree . . . . .	33
ตารางที่ 4.11 ร้อยละของความแม่นยำ ค่าความระลึกลับ และ F1 - Score ของตัวแบบต้นไม้ ตัดสินใจ . . . . .	33
ตารางที่ 4.12 ร้อยละของความแม่นยำ ค่าความระลึกลับ และ F1 - Score ของตัวแบบ Random Forest . . . . .	34
ตารางที่ 4.13 ร้อยละของความแม่นยำ ค่าความระลึกลับ และ F1 - Score ของตัวแบบ XGBoost	35
ตารางที่ 4.14 ร้อยละของความแม่นยำ ค่าความระลึกลับ และ F1 - Score ของตัวแบบ Extra Tree . . . . .	36
ตารางที่ 4.15 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ . . . . .	36

## บทที่ 1

### บทนำ

#### 1.1 ที่มาและความสำคัญ

เงินสกุลดิจิทัล หรือ Cryptocurrency เป็นเงินสกุลใหม่ที่สร้างขึ้นจากกลไกทางคณิตศาสตร์ที่กำหนดจำนวนไว้จำกัด ต้องใช้ระบบคอมพิวเตอร์ถอดรหัสเพื่อนำเงินออกจากกลไก เช่น บิทคอยน์ (Bitcoin: BTC) ไลท์คอยน์ (Litecoin) และอีเธอเรียม (Ethereum: ETH) เป็นต้น เงินสกุลใหม่สร้างขึ้นเพื่อลดการรวมศูนย์ของระบบการชำระเงินผ่านสถาบันการเงินให้สามารถกระจายไปยังผู้ใช้ในเครือข่ายสกุลเงินนั้น ๆ ได้ โดยใช้เทคโนโลยีบล็อกเชน (Blockchain) ติดตามการเคลื่อนไหวของเงินแม้จะไม่มีตัวกลางและสามารถป้องกันการปลอมแปลงได้ด้วย การชำระหรือโอนเงินจึงอยู่แค่ภายในเครือข่าย ซึ่งมีข้อดี คือรวดเร็ว ต้นทุนต่ำ และปลอดภัย แต่ธนาคารกลางส่วนใหญ่ยังไม่รับรองการใช้เงินสกุลดิจิทัลที่เอกชนสร้างขึ้นมาในการชำระหนี้ได้ตามกฎหมาย เงินสกุลดิจิทัลจึงทำหน้าที่ของเงินได้ไม่ครบเนื่องจากยังไม่สามารถนำมาใช้ในการชำระเงินและไม่ถูกใช้เป็นหน่วยกำหนดราคาสินค้าของและมูลค่าของเงินสกุลดิจิทัลเองมีการผันผวนมาก ซึ่งถ้าเป็นเงินสกุลดิจิทัลที่ธนาคารกลางออกใช้ (Central Bank Digital Currency: CBDC) จะมีคุณสมบัติของเงินที่ครบถ้วนเพราะมีมูลค่าที่แน่นอนและใช้แทนเงินสดท้องถิ่นได้ตามกฎหมาย (ฐิติมา ชูเชิด, 2564)

อีเธอเรียม ถูกพัฒนาโดย Vitalik Buterin โปรแกรมเมอร์ที่มีอายุ 19 ปีและเป็นหนึ่งในผู้ร่วมก่อตั้งเว็บไซต์ Bitcoin Magazine ด้วยความที่เขามีความสนใจในเงินสกุลดิจิทัลและบิทคอยน์ เป็นอย่างมากเขาจึงตั้งคำถามว่า เทคโนโลยีบล็อกเชน เบื้องหลังบิทคอยน์ว่ามันไม่ควรจะทำได้แค่เพียงโอนเงินหากันไปมาเท่านั้น เขาเลยได้เริ่มพัฒนาอีเธอเรียมขึ้นมาเป็นบล็อกเชนที่มีเทคโนโลยี Smart Contract ใส่ลงไปด้วยโดยที่ Smart Contract ก็คือ กระบวนการทางดิจิทัลที่กำหนดขั้นตอนการทำธุรกรรมโดยอัตโนมัติไว้ล่วงหน้า โดยไม่ต้องอาศัยตัวกลาง ซึ่งเกิดมาจาก Nick Szabo ที่เป็นผู้เสนอความคิดว่าบล็อกเชนสามารถใช้ในการบันทึกข้อตกลงของสัญญาที่สามารถดำเนินการได้ด้วยตัวเอง ไม่จำเป็นต้องมีคนกลางหรือใช้พนักงานในการมานั่งตรวจสอบเอกสาร โดยทุกอย่างให้คอมพิวเตอร์และโปรแกรมจัดการ และการ Hack ข้อมูลทำได้ยาก (สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์, 2561) ซึ่งสัญญาที่ว่ามันสามารถบันทึกข้อมูล ข้อตกลงสัญญา และสามารถยืนยันตัวตนได้เมื่อเงื่อนไขที่ถูกเขียนไว้ครบถ้วน และกระจายสำเนาไปถึงทุกคนในระบบให้ตรวจสอบกันเอง โดยไม่ต้องมีตัวกลางหรือมีคนกลางมานั่งตรวจสอบความถูกต้องของเอกสาร เมื่อ Smart Contract ต่าง ๆ ถูกรันบนเครือข่ายอีเธอเรียม บล็อกเชนอีเธอเรียมจึงไม่ต่างอะไรกับซูเปอร์คอมพิวเตอร์ขนาดใหญ่เครื่องหนึ่งของโลก การมีสัญญา Smart Contract ทำให้อีเธอเรียมเป็นมากกว่าเงินสกุลดิจิทัลที่ใช้จ่ายและแลกเปลี่ยน เพราะเงินสกุลดิจิทัลอีเธอเรียมถูกใช้เป็นเชื้อเพลิงสำคัญ

ในการทำธุรกรรมต่าง ๆ บนบล็อกเชน (Narisara, 2021)

ในปัจจุบันการลงทุนในเงินสกุลดิจิทัลนับเป็นทางเลือกหนึ่งในการลงทุนที่ได้รับความนิยมอย่างแพร่หลาย เนื่องจากได้รับผลตอบแทนในอัตราที่สูงกว่าการฝากเงินไว้กับธนาคาร แต่ถึงอย่างไรก็ตามการลงทุนในเงินสกุลดิจิทัลนั้นถือว่าเป็นเรื่องยากในการคาดการณ์ว่าสินทรัพย์ที่เลือกลงทุนมีราคาเหมาะสมที่จะลงทุนหรือไม่ เพราะราคาของเงินสกุลดิจิทัลมีความอ่อนไหว ผันผวนไม่แน่นอนและแปรเปลี่ยนไปตามปัจจัยต่าง ๆ นักลงทุนและนักวิเคราะห์จึงต้องอาศัยหลักการและเทคนิคต่าง ๆ ในการวิเคราะห์เพื่อให้ได้รับผลตอบแทนให้ได้มากที่สุด

เทคนิคที่นิยมนำมาวิเคราะห์ข้อมูลเพื่อเป็นแนวทางในการลงทุนคือการพยากรณ์ข้อมูลอนุกรมเวลา และจากงานวิจัยของ Basak et al. (2019) ได้ศึกษาการพยากรณ์ทิศทางราคาหุ้นโดยใช้วิธีการพยากรณ์ 2 แบบ คือ วิธี Random Forest และวิธี Extreme Gradient Boosting (XGBoost) ในหุ้น 2 ตัว ได้แก่ AAPL (Apple Inc) และ FB (Facebook) ผลการศึกษาพบว่าการพยากรณ์ด้วยวิธี Random Forest ในหุ้น AAPL และ FB ให้ค่า Accuracy ที่ร้อยละ 93.02 และร้อยละ 94.76 ตามลำดับ ส่วนวิธี XGBoost ให้ค่า Accuracy ที่ร้อยละ 77.13 และร้อยละ 94.44 ตามลำดับ และงานวิจัยของ Ampomah et al. (2020) พบว่า วิธี Random Forest, วิธี XGBoost และวิธี Extra Tree เป็นวิธีที่นำมาใช้ในการวิเคราะห์ทิศทางของหุ้นได้อย่างเหมาะสมและเป็นวิธีที่ให้ผลลัพธ์เป็นที่น่าพอใจ

ในปัจจุบันการเรียนรู้ของเครื่อง (Machine Learning) กำลังเป็นที่นิยมเป็นอย่างมากในการนำมาพยากรณ์ข้อมูล เนื่องจากประมวลผลได้อย่างรวดเร็ว สามารถหาความสัมพันธ์ในข้อมูลที่ซับซ้อนและตัดสินใจภายใต้ข้อมูลปริมาณมหาศาลได้ (วสันต์ ศิลปะ และคณะ, 2560) และวิธีต้นไม้ตัดสินใจ (Decision Tree) เป็นวิธีหนึ่งในศาสตร์การเรียนรู้ของเครื่อง ซึ่งจำลองการทำงานมาจากต้นไม้และสามารถเรียนรู้ข้อมูลตัวอย่างที่ป้อนให้ แล้วทำการสร้างฟังก์ชันการประมวลผลขึ้นมาใช้ในการแก้ไขปัญหาของข้อมูลนั้น ๆ ได้ ดังนั้นผู้วิจัยจึงสนใจที่จะทำการเปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์ทิศทาง ราคาปิดรายวันของเงินสกุลดิจิทัลโอเอเธอร์ด้วยวิธีต้นไม้ตัดสินใจ วิธี Random Forest วิธี XGBoost และวิธี Extra Tree

## 1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบประสิทธิภาพวิธีการพยากรณ์ทิศทางของราคาปิดโอเอเธอร์ด้วย 4 วิธี ได้แก่ วิธีต้นไม้ตัดสินใจ วิธี Random Forest วิธี XGBoost และวิธี Extra Tree

### 1.3 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ทราบแนวโน้มของทิศทางเงินสกุลดิจิทัล
2. ได้ตัวแบบพยากรณ์ทิศทางของราคาปิดเงินสกุลดิจิทัล
3. เพื่อเป็นแนวทางในการตัดสินใจในการลงทุนในเงินสกุลดิจิทัล

### 1.4 นิยามคำศัพท์

**ราคาปิด (Close Price)** คือ ราคาของเงินสกุลดิจิทัลที่เกิดจากการซื้อขายเป็นรายการสุดท้ายของแต่ละวัน หรือก่อน เวลา 23.59 น.

**เงินสกุลดิจิทัล (Cryptocurrency)** คือ หน่วยข้อมูลทางอิเล็กทรอนิกส์ ที่สร้างขึ้นเพื่อใช้เป็นสื่อกลางในการแลกเปลี่ยนสินค้า บริการ หรือแลกเปลี่ยนกับสินทรัพย์ดิจิทัลอื่น ๆ

**ทิศทางของราคา (Direction)** คือ แนวโน้มของราคาปิดรายวัน โดยจำแนกออกเป็น 2 ทิศทาง ได้แก่ ทิศทางลงหรือคงที่ และทิศทางขึ้น

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในงานศึกษานี้ ผู้วิจัยต้องการศึกษาการเปรียบเทียบประสิทธิภาพของวิธีพยากรณ์ทิศทางราคาปิดอีเธอเรียมที่ได้จากวิธีการพยากรณ์ต่าง ๆ อันได้แก่ วิธีต้นไม้ตัดสินใจ (Decision Tree) วิธี Random Forest วิธี Extreme Gradient Boosting (XGBoost) และวิธี Extra Tree ในการพยากรณ์ทิศทางราคาปิดรายวันของอีเธอเรียมเพื่อให้ได้วิธีการพยากรณ์ที่ดีที่สุด ในบทนี้ผู้วิจัยจึงได้รวบรวมความรู้เกี่ยวกับเงินสกุลดิจิทัล เครื่องมือวิเคราะห์เชิงเทคนิค ทฤษฎีและงานวิจัยที่เกี่ยวข้องจากแหล่งเรียนรู้ต่าง ๆ เพื่อนำมาเป็นแนวทางในการศึกษาวิจัยครั้งนี้

#### 2.1 ความรู้พื้นฐานเกี่ยวกับเงินสกุลดิจิทัล

##### 2.1.1 เงินสกุลดิจิทัล (Cryptocurrency)

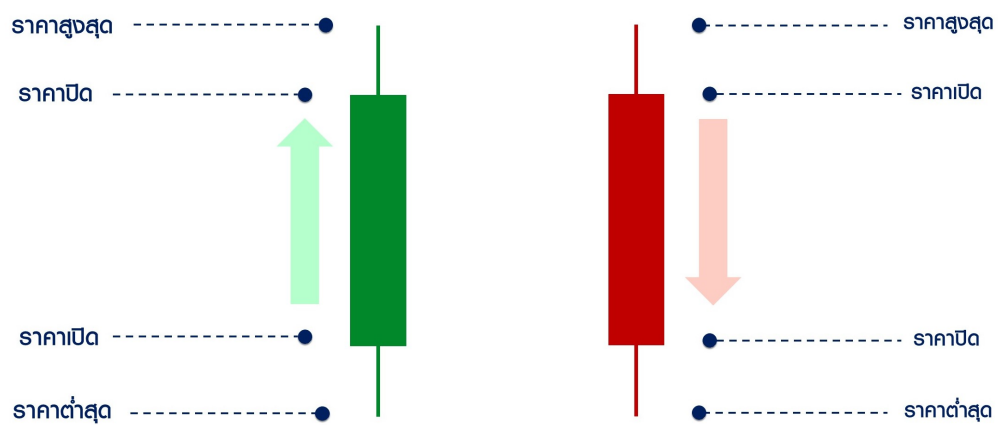
เงินสกุลดิจิทัล คือ สกุลเงินเสมือนจริงที่ใช้การเข้ารหัสเพื่อทำให้มีความปลอดภัยที่ไม่สามารถปลอมแปลง หรือจ่ายซ้ำได้ โดยมีลักษณะคล้ายเงินที่บริหารจัดการ จัดเก็บ และแลกเปลี่ยนบนระบบคอมพิวเตอร์แบบดิจิทัลโดยการจัดการผ่านอินเทอร์เน็ต ซึ่งมีการเปิดตัวในปี พ.ศ. 2552 จากบุคคลที่รู้จักกันในนามแฝงว่า ซาโตชิ นากาโมโต ซึ่งสกุลเงินดิจิทัลแรกที่ถูกกำเนิดคือ บิทคอยน์ (Bitcoin) จากความสำเร็จของบิทคอยน์จึงทำให้มีการพัฒนาเงินสกุลดิจิทัลตัวอื่น ๆ เพิ่มมากขึ้น เช่น Binance Coin (BNB), Litecoin (LTC) และอีเธอเรียม (Ethereum: ETH) เป็นต้น (วีรพงษ์ ชุตินันท์, 2564)

##### 2.1.2 อีเธอเรียม (Ethereum)

อีเธอเรียม เป็นเงินสกุลดิจิทัลหนึ่งที่ถูกพัฒนาขึ้นในปี พ.ศ. 2556 โดยโปรแกรมเมอร์ชาวรัสเซีย ชื่อว่า Vitalik Buterin ซึ่งมีเทคโนโลยีบล็อกเชนเป็นพื้นฐานเช่นเดียวกับบิทคอยน์ โดยอีเธอเรียมมีจุดเด่นสำคัญ คือ Smart Contract ซึ่งเป็นชุดคำสั่งที่ดำเนินการได้เองโดยอัตโนมัติเมื่อครบตามสัญญาหรือข้อตกลงที่ถูกโปรแกรมไว้ โดยผู้ใช้งานสามารถตรวจสอบข้อมูลย้อนหลังได้ ทำให้เกิดความโปร่งใสในการทำธุรกรรม (จิรายุส ทรัพย์ศรีโสภา, 2564) ได้ด้วยตัวเอง ไม่จำเป็นต้องมีคนกลาง หรือใช้พนักงานในการมานั่งตรวจสอบเอกสาร โดยทุกอย่างให้คอมพิวเตอร์และโปรแกรมจัดการ และการ Hack ข้อมูลทำได้ยากเนื่องจาก Smart Contract ถูกรันบนเครือข่ายอีเธอเรียมซึ่งเป็นเครือข่ายแบบบล็อกเชน และบล็อกเชนอีเธอเรียมเปรียบเสมือนซูเปอร์คอมพิวเตอร์ขนาดใหญ่เครื่องหนึ่งของโลกการมีสัญญา Smart Contract จึงทำให้อีเธอเรียมเป็นมากกว่าเงินสกุลดิจิทัลที่ใช้จ่ายและแลกเปลี่ยนเพราะเงินสกุลดิจิทัลอีเธอเรียมถูกใช้เป็นเครื่องมือสำคัญในการทำธุรกรรมต่าง ๆ บนเครือข่ายบล็อกเชน (Narisara, 2021)

ราคาของเงินสกุลดิจิทัลจะประกอบไปด้วย ราคาเปิด (Open Price) ราคาสูงสุด (High Price) ราคาต่ำสุด (Low Price) และราคาปิด (Close Price) ดังรายละเอียดต่อไปนี้

1. ราคาเปิด (Open Price) หมายถึง ราคาซื้อขายแรกที่เกิดขึ้นเป็นรายการแรกในช่วงเวลา 00.00 น. ถึง 23.59 น.
2. ราคาสูงสุด (High Price) หมายถึง ราคาสูงสุดที่มีการซื้อขายในระหว่างวัน
3. ราคาต่ำสุด (Low Price) หมายถึง ราคาต่ำสุดที่มีการซื้อขายในระหว่างวัน
4. ราคาปิด (Close Price) หมายถึง ราคาสุดท้ายที่เกิดขึ้นจากการซื้อขายรายการสุดท้ายของวันตั้งแต่วันที่ 00.00 น. ถึง 23.59 น.



รูปที่ 2.1: ตัวอย่างแท่งเทียนราคาเปิด ราคาสูงสุด ราคาต่ำสุด และราคาปิด  
(Advisor, 2017)

## 2.2 เครื่องมือวิเคราะห์เชิงเทคนิค

ในหัวข้อนี้จะกล่าวถึงตัวแปรอิสระที่ใช้ในการพยากรณ์แนวโน้มทิศทางราคาปิดของเงินสกุลดิจิทัล อีเธอเรียม ซึ่งจะประกอบด้วยเครื่องมือวิเคราะห์เชิงเทคนิค (Technical Analysis Method) จำนวน 6 เครื่องมือ ดังต่อไปนี้

### 2.2.1 ค่าเฉลี่ยเคลื่อนที่อย่างง่าย (Simple Moving Average: SMA)

ค่าเฉลี่ยเคลื่อนที่อย่างง่ายเป็นเครื่องมือวิเคราะห์เชิงเทคนิครูปแบบหนึ่งสามารถคำนวณจากการหาค่าเฉลี่ยของข้อมูลโดยให้น้ำหนักของข้อมูลทุกตัว (Elena, 2021)

$$SMA_t = \frac{C_t + C_{t-1} + \dots + C_{t-9}}{10}$$

เมื่อ  $C_t$  คือราคาปิด ณ วันที่  $t$ ;  $t = 10, 11, \dots, n$  เมื่อ  $n$  แทนจำนวนข้อมูล



รูปที่ 2.2: ตัวอย่างกราฟแท่งเทียนและเครื่องมือ SMA (tradingview, 2022)

## 2.2.2 ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนัก (Weighted Moving Average: WMA)

ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนักเป็นเครื่องมือวิเคราะห์เชิงเทคนิครูปแบบหนึ่ง โดยข้อมูลที่เกิดล่าสุดจะได้รับการถ่วงน้ำหนักมากกว่าข้อมูลในอดีต และข้อมูลตัวถัดไปก็จะได้รับการถ่วงน้ำหนักลดลงไปตามลำดับ โดยทั่วไปจะใช้ค่าเริ่มต้นคือ 10 วัน (Elena, 2021)

$$WMA_t = \frac{kC_t + (k-1)C_{t-1} + \dots + C_{t-9}}{10 + 9 + 8 + \dots + 1}$$

เมื่อ  $k$  คือจำนวนวันย้อนหลังที่ใช้



รูปที่ 2.3: ตัวอย่างกราฟแท่งเทียนและเครื่องมือ WMA (tradingview, 2022)



### 2.2.3 ค่าเฉลี่ยเคลื่อนที่ร่วมทาง/แยกทาง (Moving Average Convergence/ Divergence: MACD)

ค่าเฉลี่ยเคลื่อนที่ร่วมทาง/แยกทาง เป็นเครื่องมือวิเคราะห์เชิงเทคนิคที่สร้างขึ้นและพัฒนาโดย Gerald Appel ในปี ค.ศ. 1979 ซึ่งเป็นเครื่องมือที่เคลื่อนที่ไปในทิศทางเดียวกับราคา (Trend Following) ที่มีแนวคิดมาจากการสร้างเส้นค่าเฉลี่ยเคลื่อนที่ (Moving Average: MA) ขึ้นพร้อม ๆ กัน 2 เส้น แล้วนำมาวิเคราะห์ลักษณะที่เกิดขึ้นระหว่างเส้นค่าเฉลี่ยเคลื่อนที่ทั้ง 2 เส้นโดยที่เส้นค่าเฉลี่ยเคลื่อนที่เส้นแรกจะใช้ระยะเวลาในการคำนวณยาวกว่าเส้นค่าเฉลี่ยเคลื่อนที่เส้นที่สอง

$$EMA_t = EMA_{t-1} + SF(C_t - EMA_{t-1})$$

เมื่อ  $EMA_t$  คือค่าของ  $EMA$  ณ วันที่  $t$

$EMA_{t-1}$  คือค่าของ  $EMA$  ณ วันก่อนหน้า เมื่อ  $t$  คือ วันปัจจุบัน

$SF$  คือค่าของตัวถ่วงน้ำหนักหาได้จาก  $\frac{2}{k+1}$  เมื่อ  $k$  คือจำนวนวันย้อนหลังที่ใช้โดยสูตรในการคำนวณค่า MACD เป็นดังนี้

$$MACD_t = EMA_t(12) - EMA_t(26)$$

เมื่อ  $EMA_t(12)$  คือค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนักจำนวน 12 วัน

$EMA_t(26)$  คือค่าเฉลี่ยแบบถ่วงน้ำหนักจำนวน 26 วัน



รูปที่ 2.4: ตัวอย่างกราฟแท่งเทียนและเครื่องมือ MACD (tradingview, 2022)

## 2.2.4 ดัชนีการแกว่ง (Stochaic Oscillator: %K)

ดัชนีการแกว่ง เป็นเครื่องมือวิเคราะห์เชิงเทคนิคประเภทโมเมนตัม (Momentum) ถูกพัฒนาโดย George Lane ในปี ค.ศ. 1950 สำหรับศึกษาความสัมพันธ์ระหว่างการเคลื่อนไหวของราคาในช่วงเวลาหนึ่งกับราคาปิด ซึ่งค่าดัชนีการแกว่งจะมีค่าตั้งแต่ 0-100 โดยสามารถคำนวณได้ดังนี้

$$\%K = \frac{C_t - L_{14}}{H_{14} - L_{14}} \times 100$$

เมื่อ  $H_{14}$  คือราคาสูงสุดในช่วง 14 วันที่ผ่านมา

$L_{14}$  คือราคาต่ำสุดในช่วง 14 วันที่ผ่านมา



รูปที่ 2.5: ตัวอย่างกราฟแท่งเทียนและเครื่องมือ %K (tradingview, 2022)

## 2.2.5 ดัชนีกำลังสัมพัทธ์ (Relative Strength Index: RSI)

ดัชนีกำลังสัมพัทธ์ หรือ RSI เป็นเครื่องมือวิเคราะห์เชิงเทคนิคประเภทโมเมนตัม (Momentum) ใช้สำหรับวัดการแกว่งตัวของราคาว่ามีภาวะการซื้อมากเกินไป (Overbought) หรือ การขายมากเกินไป (Oversold) โดยมีค่าตั้งแต่ 0 - 100 ซึ่งถูกพัฒนาโดย Welles Wilder ในช่วงปี ค.ศ. 1978 โดยค่ามาตรฐานที่ใช้กันทั่วไปจะอยู่ที่ค่า RSI เท่ากับ 30 และ 70 หากค่า RSI อยู่ในระดับที่ต่ำกว่า 30 จะถือว่าอยู่ในภาวะการขายมากเกินไปและหากค่า RSI มากกว่า 70 จะถือว่าอยู่ใน ภาวะการซื้อมากเกินไป ค่า RSI สามารถคำนวณได้ ดังนี้

$$RSI = 100 - \frac{100}{1 + RS}$$

$$RS = \frac{AverageGain}{AverageLoss}$$

โดย *AverageGain* คือค่าเฉลี่ยของราคาปิดที่เปลี่ยนแปลงเพิ่มขึ้น 14 วัน

*AverageLoss* คือค่าเฉลี่ยของราคาปิดที่เปลี่ยนแปลงลดลง 14 วัน



รูปที่ 2.6: ตัวอย่างกราฟแท่งเทียนและเครื่องมือ RSI (tradingview, 2022)

## 2.2.6 Williams Percentage Range: %R

เป็นเครื่องมือวิเคราะห์เชิงเทคนิคประเภท Momentum ใช้เพื่อหาช่วงการแกว่งของราคาในช่วงเวลาที่ระบุว่ามีภาวะการซื้อมากเกินไป หรือ การขายมากเกินไป โดยมีค่าในช่วงระหว่าง -100 ถึง 0 ซึ่งเครื่องมือ %R ถูกพัฒนาขึ้นโดย Larry R. Williams ในปี ค.ศ. 1987 โดยค่ามาตรฐานที่ใช้กันทั่วไปเท่ากับ -20 และ -80 หากค่า %R มากกว่า -20 จะถือว่ามีภาวะการซื้อมากเกินไป และหากค่า %R อยู่ต่ำกว่า -80 จะถือว่ามีภาวะการขายมากเกินไป โดยค่า %R สามารถคำนวณได้ ดังนี้

$$\%R = \frac{H_{14} - C_t}{H_{14} - L_{14}} \times (-100)$$



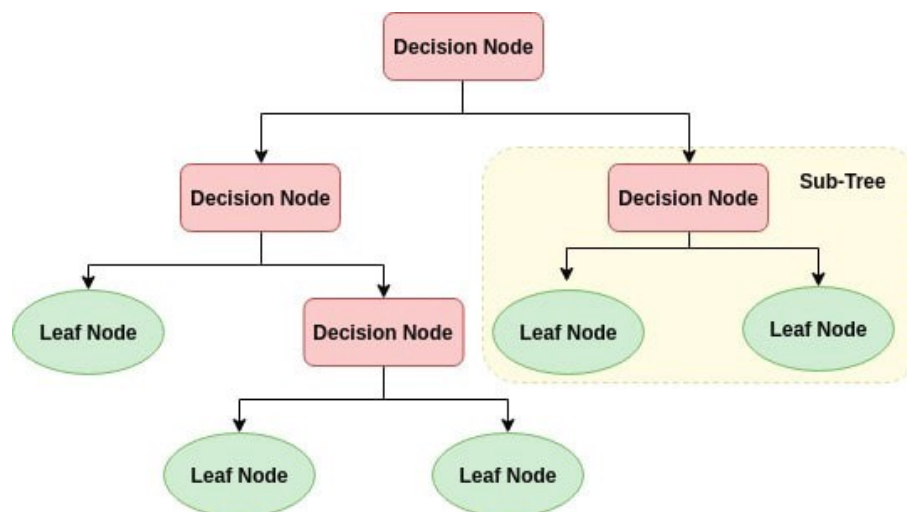
รูปที่ 2.7: ตัวอย่างกราฟแท่งเทียนและเครื่องมือ %R (tradingview, 2022)

## 2.3 วิธีการพยากรณ์

### 2.3.1 ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจ คือแบบจำลองทางคณิตศาสตร์เพื่อการหาทางเลือกที่ดีที่สุด โดยการนำข้อมูลมาสร้างตัวแบบการพยากรณ์เพื่อจำแนกประเภทข้อมูล (Classification Data) ซึ่งจะใช้กฎในรูปแบบ "ถ้า (เงื่อนไข) แล้ว (ผลลัพธ์)" มาประกอบการสร้างต้นไม้ตัดสินใจ โดยมีโครงสร้างประกอบด้วย

1. โหนด (Node) คือคุณสมบัติต่าง ๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทางใดโดยโหนดที่อยู่สูงสุดหรือโหนดแรกเรียกว่า โหนดราก (Root Node)
2. กิ่ง (Branch) คือคุณสมบัติของโหนดที่แตกออกมาโดยจำนวนของกิ่งจะเท่ากับคุณสมบัติของโหนด
3. ใบ (Leaf) คือกลุ่มของผลลัพธ์ในการจำแนกข้อมูล



รูปที่ 2.8: ตัวอย่างต้นไม้ตัดสินใจ  
(Chalach, 2019)

การสร้างต้นไม้ตัดสินใจเป็นการสร้างในลักษณะจากบนลงล่าง (Top-Down) โดยเริ่มจากการเลือกคุณสมบัติที่ดีที่สุดในการแยกกลุ่มมาเป็นโหนดราก เมื่อข้อมูลผ่านการแบ่งที่โหนดรากแล้วก็จะหาคุณสมบัติที่ดีที่สุดรองจากโหนดรากของข้อมูลที่ถูกแบ่งออกมาแล้วสร้างโหนดใบต่อไป และจะวนการสร้างในโหนดใบ และต้นไม้ย่อยของแต่ละกิ่งไปเรื่อย ๆ จนกว่าข้อมูลที่ผ่านการแบ่งแยกนั้นจะจัดอยู่ในกลุ่มเดียวกัน โดยแสดงขั้นตอนการสร้างต้นไม้ตัดสินใจได้ดังนี้

1. ต้นไม้ตัดสินใจเริ่มต้นโดยมีโหนดเพียงโหนดเดียว
2. ถ้าข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกันแล้ว ให้โหนดนั้นเป็นใบ
3. ถ้าในโหนดมีข้อมูลหลายกลุ่มปะปนกันอยู่ จะต้องวัดค่า Information Gain ของแต่ละคุณสมบัติ เพื่อที่จะใช้เป็นเกณฑ์ในการคัดเลือกคุณสมบัติ ที่มีความสามารถในการแบ่งแยกข้อมูลออกเป็นกลุ่มต่าง ๆ ได้ดีที่สุด โดยคุณสมบัติที่มีค่า Information Gain มากที่สุดจะถูกเลือกให้เป็นตัวทดสอบหรือคุณสมบัติที่ใช้ในการตัดสินใจ
4. กิ่งของต้นไม้ ถูกสร้างขึ้นจากค่าต่าง ๆ ที่เป็นไปได้ของโหนดทดสอบ และข้อมูลจะถูกแบ่งออกตามกิ่งต่าง ๆ ที่สร้างขึ้น
5. ทำการวนซ้ำเพื่อหาคุณสมบัติที่มีค่ามากที่สุด สำหรับข้อมูลที่ถูกแบ่งแยกออกมาในแต่ละกิ่งเพื่อนำคุณสมบัตินี้มาสร้างเป็นโหนดตัดสินใจต่อไป โดยที่คุณสมบัติที่ถูกเลือกมาเป็นโหนดแล้วจะไม่ถูกเลือกมาอีกสำหรับโหนดในระดับต่อ ๆ ไป
6. ทำการวนซ้ำเพื่อแบ่งข้อมูลและแตกกิ่งของต้นไม้ไปเรื่อย ๆ โดยการวนซ้ำจะสิ้นสุดก็ต่อเมื่อไม่สามารถสร้างโหนดตัดสินใจได้อีก

การคำนวณค่า Information Gain ต้นไม้ตัดสินใจเป็นโครงสร้างที่ใช้แสดงกฎที่ได้จากเทคนิคการจำแนกประเภทข้อมูล โดยต้นไม้ตัดสินใจจะมีลักษณะคล้ายโครงสร้างต้นไม้ที่แต่ละโหนดแสดงคุณสมบัติในการสร้างต้นไม้ตัดสินใจ ปัญหาสำคัญที่ต้องพิจารณาคือ ควรจะตัดสินใจเลือกคุณสมบัติใดมาทำหน้าที่เป็นโหนดรากในแต่ละขั้นตอนของการสร้างต้นไม้และต้นไม้ย่อย (Subtree) ของต้นไม้ตัดสินใจเกณฑ์ที่ใช้ช่วยประกอบการเลือกคุณสมบัติคือการคำนวณค่ามาตรฐานเกน (Gain Criterion) ซึ่งเป็นค่าที่บ่งบอกว่าคุณสมบัตินั้นสามารถจำแนกกลุ่มของข้อมูลได้ดีเพียงใด โดย ทดลองเลือกแต่ละคุณสมบัติที่เป็นไปได้จากชุดข้อมูลฝึกสอนมาทำหน้าที่เป็นโหนดราก ถ้าคุณสมบัติใดให้ค่าเกนสูงสุด แสดงว่าคุณสมบัตินั้นสามารถจำแนกกลุ่มของข้อมูลได้ดีที่สุด การใช้ค่า Information Gain จะช่วยลดจำนวนครั้งของการทดสอบในการจำแนกข้อมูลอีกทั้งยังรับประกันว่า ต้นไม้ตัดสินใจที่ได้ไม่มีความซับซ้อนมากเกินไป (ขจรศักดิ์ ศรีอ่อน, 2552) ซึ่งสามารถคำนวณได้จากสมการดังนี้

$$InformationGain(S, A) = E(S) - E(A)$$

$$E(S) = \sum_{i=1}^c -p_i \log_2(p_i)$$

โดย  $S$  คือ คุณสมบัติที่นำมาวัดค่าเอนโทรปี (Entropy)

$p_i$  คือ สัดส่วนของจำนวนสมาชิกในกลุ่ม  $i; i = 1, 2, \dots, c$  กับจำนวนสมาชิกทั้งหมดของกลุ่มตัวอย่าง เมื่อ  $c$  แทนจำนวนกลุ่มของข้อมูล

ID	Outlook	Temperature	Humidity	Windy	Play
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

รูปที่ 2.9: ตัวอย่างข้อมูลเพื่อใช้ในการคำนวณ Information Gain (Pacharawongsakda, 2015)

จากรูปที่ 2.9 สามารถคำนวณค่า  $E(S)$  ได้ดังนี้  $E(S) = -P(Play = yes) \times \log_2 P(Play = yes) - P(Play = no) \times \log_2 P(Play = no)$

$$= -\left[\frac{9}{14} \times \log_2\left(\frac{9}{14}\right) + \frac{5}{14} \times \log_2\left(\frac{5}{14}\right)\right]$$

$$= 0.94$$

ซึ่งสามารถคำนวณค่าเอนโทรปีของ Outlook = overcast, Outlook = rainy และ Outlook = sunny ได้ดังนี้

$$Entropy(Outlook = overcast) = -P(Play = yes) \times \log_2 P(Play = yes) - P(Play = no) \times \log_2 P(Play = no)$$

$$= -\left[\frac{4}{4} \times \log_2\left(\frac{4}{4}\right) + \frac{0}{4} \times \log_2\left(\frac{0}{4}\right)\right]$$

$$= 0$$

ในส่วนของเอนโทรปีของ Outlook = rainy และ Outlook = sunny ก็คำนวณเช่นเดียวกัน ดังนั้นค่า Information Gain จะคำนวณได้จาก

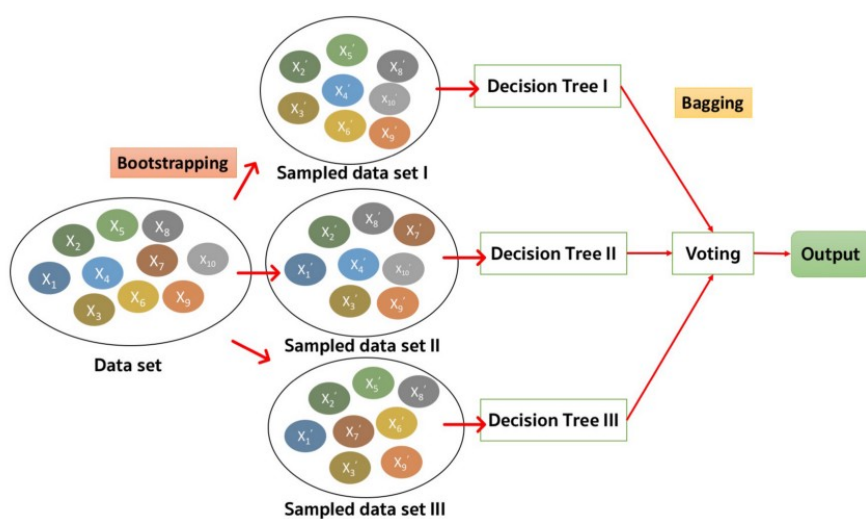
$$InformationGain(Outlook) = E(S) - [P(Outlook = overcast) \times Entropy(Outlook = overcast) + P(Outlook = rainy) \times Entropy(Outlook = rainy) + P(Outlook = sunny) \times Entropy(Outlook = sunny)]$$

$$= 0.94 - [0.29 \times 0 + 0.36 \times 0.97 + 0.36 \times 0.97]$$

$$= 0.25$$

### 2.3.2 Random Forest

เป็นเทคนิคที่ทำการสุ่มเลือกตัวอย่างและคุณสมบัติออกมาจากชุดข้อมูลหลาย ๆ ชุด จากนั้นนำเอาชุดของคุณสมบัติเหล่านั้นมาสร้างตัวแบบด้วยวิธีต้นไม้ตัดสินใจหลาย ๆ ต้น (Ensemble Of Decision Trees) โดยสร้างจากการเลือกตัวอย่างสุ่มแบบใส่คืน (Random Sampling With Replacement) เพื่อนำมาสร้างตัวแบบต้นไม้ตัดสินใจโดยแต่ละต้นที่มีลักษณะที่ไม่ซ้ำกัน โดยแต่ละตัวแบบจะมีการทำนายผลลัพธ์ซึ่งผลจากการทำนายของตัวแบบต้นไม้ตัดสินใจแต่ละต้น จะทำการโหวตเลือกผลการทำนายที่ได้รับการโหวตมากที่สุดวิธีการนี้เรียกว่า Bagging หรือ Bootstrap (ภูริพัทธ์ ทองคำ, 2559) ดังตัวอย่างในรูปที่ 2.10



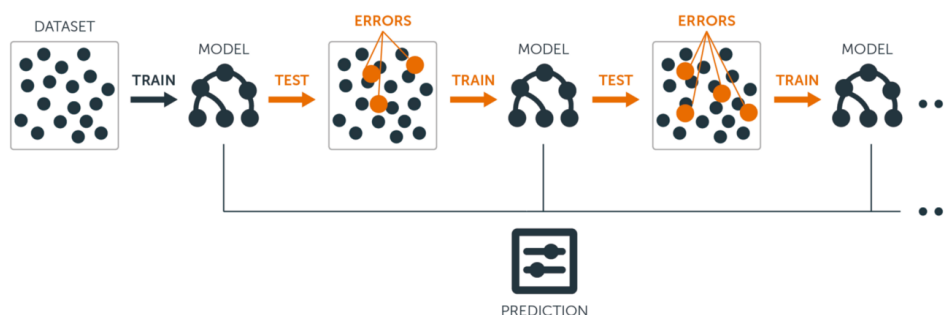
รูปที่ 2.10: หลักการสร้างตัวแบบ Random Forest (Witchapong, 2018b)

### 2.3.3 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting หรือ XGBoost มีพื้นฐานมาจาก Gradient Boosting ถูกจัดอยู่ในกลุ่มของการจำแนกประเภทแบบ Boosting ประเภทหนึ่งที่ได้รับการพัฒนาให้มีประสิทธิภาพมากขึ้น จึงทำให้สามารถทำงานได้รวดเร็ว และสามารถจัดการหน่วยการทำงานย่อยแบบ Multi-thread ได้ดี และการทำให้ตัวแปรมีค่าอยู่ในระดับปกติ (Regularization) เป็นการป้องกันการเข้ากันได้เกินไป (Over Fit) ผลการเรียนรู้จะได้เป็นตัวแบบที่นำเอาต้นไม้ตัดสินใจมาสร้างให้เกิดการเรียนรู้แบบต่อเนื่องต่อกัน การเรียนรู้จากการสร้างต้นไม้ตัดสินใจหลาย ๆ ต้น จะทำให้เกิดการเรียนรู้โดยสนใจจากค่าความคลาดเคลื่อน หรือเกิดการเรียนรู้จากความคลาดเคลื่อนของต้นไม้ตัดสินใจที่สร้างได้ก่อนหน้านี้ทำให้ความแม่นยำของการพยากรณ์แม่นยำมากขึ้นเรื่อย ๆ เมื่อการเรียนรู้ของตัวแบบต้นไม้ตัดสินใจดำเนินอย่างต่อเนื่องกัน จนมีความลึกมากพอหรือไม่มีรูปแบบอื่น ๆ ให้เรียนรู้เพิ่มแล้วระบบจะหยุดการเรียนรู้ ตัวแบบที่ได้จาก



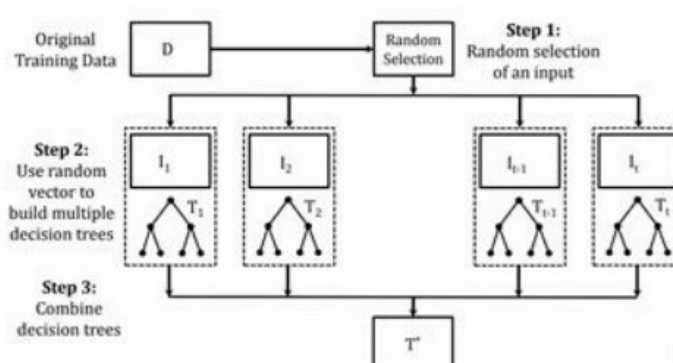
วิธี XGBoost จะใช้ตัวแบบหลาย ๆ ตัวแบบมาประกอบกันเป็นตัวแบบที่มีความซับซ้อนขึ้น (ศิริลักษณ์ หล่อพันธุ์มณี และคณะ, 2562)



รูปที่ 2.11: หลักการสร้างตัวแบบ XGBoost (Chukamphaeng, 2018)

### 2.3.4 Extra Tree

Extra Tree เป็นเทคนิคที่ใช้ในการจำแนกประเภทข้อมูลอีกเทคนิคหนึ่งที่มีพื้นฐานมาจาก Random Forest แต่จะทำการสุ่มเลือกข้อมูล โดยเลือกตัวอย่างสุ่มแบบไม่ใส่คืน (Random Sampling Without Replacement) เมื่อสุ่มจนได้จำนวนชุดข้อมูลตามที่ต้องการเรียบร้อยแล้วทำการนำชุดข้อมูลที่ได้ทั้งหมดไปทำการเรียนรู้เพื่อสร้างตัวแบบพยากรณ์ โดยผลการทำนายของต้นไม้ตัดสินใจแต่ละต้นจะทำการโหวตเลือกผลการทำนายที่ได้รับการโหวตมากที่สุด ดังรูปที่ 2.12



รูปที่ 2.12: หลักการสร้างตัวแบบ Extra Tree (Chakrabarty and Biswas, 2020)

### 2.3.5 เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ (Criteria)

การวัดประสิทธิภาพของตัวแบบพยากรณ์ในการทำนายทิศทางราคาปิดจะใช้ค่า Accuracy, Recall และ F1-Score ซึ่งหาได้จากการพิจารณา Confusion Matrix ดังตารางที่ 2.1

ตารางที่ 2.1: Confusion Matrix

ค่าพยากรณ์	ค่าจริง	
	0 : ทิศทางลงหรือคงที่	1 : ทิศทางขึ้น
0 : ทิศทางลงหรือคงที่	True Positive (TP)	False Positive (FP)
1 : ทิศทางขึ้น	False Negative (FN)	True Negative (TN)

โดยที่เกณฑ์ในการเปรียบเทียบที่ได้จาก Confusion Matrix ได้แก่

- True Positive (TP) คือ จำนวนข้อมูลที่พยากรณ์ถูกว่าเป็นจริง และมีค่าเป็นจริง ซึ่งในการวิจัยครั้งนี้ หมายถึง จำนวนข้อมูลที่พยากรณ์ถูกว่าเป็นทิศทางลงหรือคงที่ และมีทิศทางลงหรือคงที่จริง
- True Negative (TN) คือ จำนวนข้อมูลที่พยากรณ์ว่าเป็นเท็จ และมีค่าเป็นเท็จ ซึ่งในการวิจัยครั้งนี้ หมายถึง จำนวนข้อมูลที่พยากรณ์ถูกว่าเป็นทิศทางขึ้น และมีทิศทางขึ้น
- False Positive (FP) คือ จำนวนข้อมูลที่พยากรณ์ว่าเป็นจริง แต่มีค่าเป็นเท็จ ซึ่งในการวิจัยครั้งนี้ หมายถึง จำนวนข้อมูลที่พยากรณ์ว่าเป็นทิศทางลงหรือคงที่ แต่มีทิศทางขึ้น
- False Negative (FN) คือ จำนวนข้อมูลที่พยากรณ์ว่าเป็นเท็จ แต่มีค่าเป็นจริง ซึ่งในการวิจัยครั้งนี้ หมายถึง จำนวนข้อมูลที่พยากรณ์ว่าเป็นทิศทางขึ้น แต่มีทิศทางลงหรือคงที่

จากตารางที่ 2.1 สามารถวัดประสิทธิภาพของตัวแบบในการพยากรณ์ทิศทางราคาปิดจากเกณฑ์ดังนี้

1. Accuracy เป็นการวัดความถูกต้องของตัวแบบพยากรณ์ ซึ่งพิจารณารวมทุกกลุ่ม โดยคำนวณจาก

$$\frac{TP + TN}{TP + TN + FP + FN}$$

2. Recall เป็นการวัดความระลึกของตัวแบบพยากรณ์ ซึ่งพิจารณาแยกทีละกลุ่ม โดยคำนวณจาก

$$\frac{TP}{TP + FN}$$

3. F1-Score คือค่าเฉลี่ยแบบ harmonic ระหว่างค่า Precision และค่า Recall โดยคำนวณจาก

$$F1 = 2 \times \left( \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \right)$$

โดยที่  $\text{precision} = \frac{TP}{TP+FP}$  คือค่าความแม่นยำของตัวแบบพยากรณ์

## 2.4 งานวิจัยที่เกี่ยวข้อง

Elena (2021) ได้ศึกษาการพยากรณ์แนวโน้มราคาปิดรายวันของหุ้น OMXS30 ของประเทศสวีเดน โดยใช้วิธี XGBoost มีตัวแปรที่ใช้ในการพยากรณ์แนวโน้มราคาปิดรายวันของหุ้นดังกล่าว ได้แก่ SMA, WMA, %K, RSI, %R และคุณลักษณะที่มาจากข่าวการเงิน ซึ่งผลการศึกษาพบว่าได้ค่า Accuracy เท่ากับร้อยละ 73 และเมื่อเพิ่มคุณลักษณะที่มาจากข่าวทางการเงิน ทำให้ได้ค่า Accuracy เพิ่มขึ้นเล็กน้อยเป็นร้อยละ 73.71

Oliveira (2021) ได้ศึกษาการพยากรณ์แนวโน้มราคาปิดรายชั่วโมงของเงินสกุลดิจิทัลบิทคอยน์ เพื่อเปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์ ได้แก่ วิธีต้นไม้ตัดสินใจ วิธี Support Vector Machine (SVM) และวิธี Long Short-Term Memory Neural Network (LSTM) โดยใช้ข้อมูลตั้งแต่วันที่ 23 กุมภาพันธ์ ค.ศ. 2018 ถึงวันที่ 1 พฤศจิกายน ค.ศ. 2020 และทำการแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ ข้อมูลชุดฝึกสอนซึ่งเก็บรวบรวมตั้งแต่วันที่ 23 กุมภาพันธ์ ค.ศ. 2018 ถึงวันที่ 19 เมษายน ค.ศ. 2020 คิดเป็นร้อยละ 80 ของข้อมูลทั้งหมด และข้อมูลชุดทดสอบคือข้อมูลส่วนที่เหลือ ซึ่งผลการศึกษาพบว่าวิธี SVM ให้ค่า Accuracy มากกว่าวิธี LSTM โดยมีค่า Accuracy เท่ากับร้อยละ 59.43

Ampomah et al. (2020) ได้ศึกษาการพยากรณ์ทิศทางราคาหุ้น ในการพยากรณ์หุ้น 8 ตัว ได้แก่ Bank of America Corporation (BAC), Exxon Mobil Corporation (XOM), S&P 500, Microsoft Corporation (MSFT), Dow Jones Industrial Average (DJIA), CarMax Corporation (KMX), Tata Steel Limited (TATASTEEL) และ HCL Technologies Limited (HCLTECH) โดยใช้ข้อมูลที่ตั้งแต่วันที่ 1 มกราคม ค.ศ. 2005 ถึงวันที่ 30 ธันวาคม ค.ศ. 2019 ด้วยวิธี Random Forest วิธี XGBoost วิธี Bagging วิธี AdaBoost วิธี Extra Trees และวิธี Voting ผลการศึกษาพบว่า วิธี Extra Trees ให้ค่า Accuracy ดีที่สุดที่ร้อยละ 83.75

Castoe (2020) ได้ศึกษาการพยากรณ์ทิศทางราคาหุ้นรายวันของ AAPL ล่วงหน้า 30, 60 และ 90 วัน ด้วยวิธี Random Forest และ Quantile Regression Forest โดยใช้ข้อมูลตั้งแต่วันที่ 12 ธันวาคม ค.ศ. 1980 ถึงวันที่ 1 สิงหาคม ค.ศ. 2020 ซึ่งมีตัวแปรที่ใช้คือ MACD, RSI, ROC (Price Rate of Change), %K, %R, CCI (Commodity Channel Index) และ DIX (Disparity Index) ผลการศึกษาพบว่าวิธี Random Forest ให้ค่า Accuracy ในการพยากรณ์ล่วงหน้า 30, 60 และ 90 วัน เป็นร้อยละ 94.58 ร้อยละ 94.08 และร้อยละ 94.19 ตามลำดับ ส่วนวิธี Quantile Regression Forest ให้ค่าเฉลี่ยของ

ร้อยละความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error: MAPE) ในการพยากรณ์ล่วง  
หน้า 30, 60 และ 90 วัน เป็นร้อยละ 0.04 ร้อยละ 0.06 และร้อยละ 0.09 ตามลำดับ

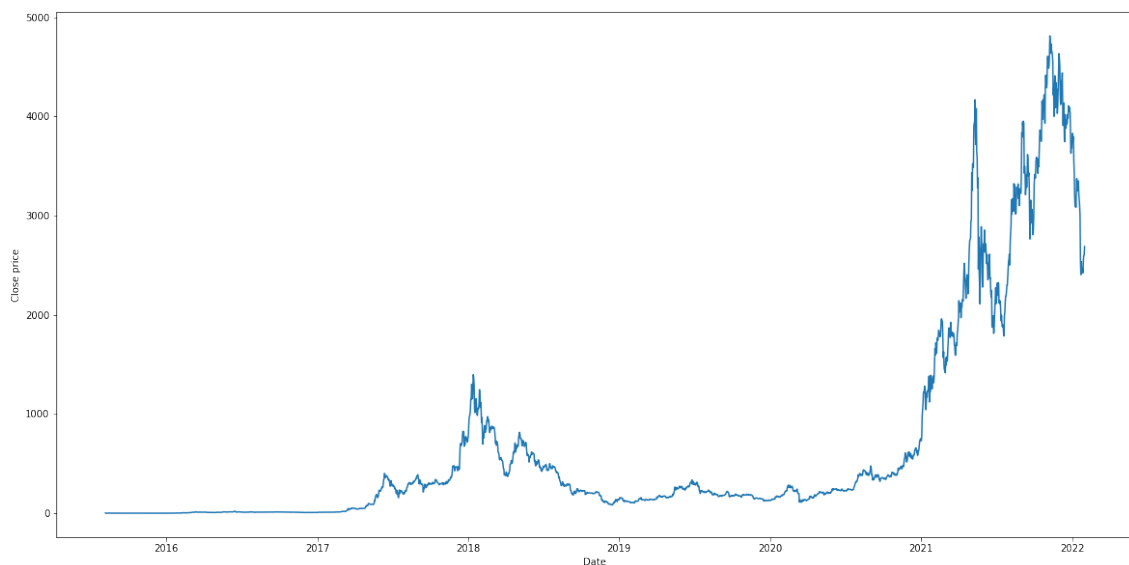
### บทที่ 3

#### วิธีการดำเนินการวิจัย

การ ศึกษา ครั้ง นี้ มี วัตถุประสงค์ เพื่อ เปรียบ เทียบ ประสิทธิภาพ ของ ตัว แบบ พยากรณ์ ที่ ได้ จาก วิธี ต้นไม้ ตัดสินใจ วิธี Random Forest วิธี XGBoost และวิธี Extra Tree ในการ พยากรณ์ ราคา ปิด รายวัน ของ เงิน สกุล ดิจิทัล อีเธอเรียม โดย ในการ ดำเนิน การ ศึกษา ครั้ง นี้ ได้ ใช้ ชุด ข้อมูล จริง จำนวน 2,370 วัน ตั้งแต่วันที่ 7 สิงหาคม พ.ศ. 2558 ถึงวันที่ 31 มกราคม พ.ศ. 2565 ซึ่ง เก็บ รวบรวม มา จาก เว็บไซต์ [www.coinmarketcap.com](http://www.coinmarketcap.com) โดย วัด ประสิทธิภาพ ของ ตัว แบบ ด้วย ค่า Accuracy, Recall และ F1-Score ซึ่ง ทำ การ วิเคราะห์ ข้อมูล ทั้งหมด ด้วย ภาษา ไพธอน (Python) มี รายละเอียด การ ดำเนิน การ ศึกษา ดัง นี้

#### 3.1 ขอบเขตของการวิจัย

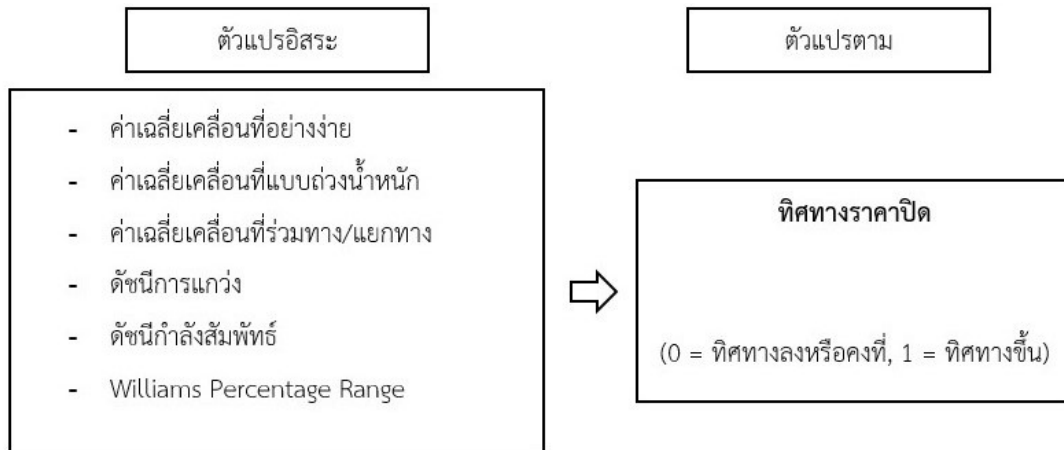
ข้อมูล ที่ นำ มา ใช้ ใน การ ศึกษา คือ ราคา ต่ำ สุด (Low Price) ราคา สูง สุด (High Price) และ ราคา ปิด รายวัน ของ เงิน สกุล ดิจิทัล อีเธอเรียม (Close Price) ซึ่ง เก็บ รวบรวม จาก เว็บไซต์ [www.coinmarketcap.com](http://www.coinmarketcap.com) ตั้งแต่วันที่ 7 สิงหาคม พ.ศ. 2558 ถึงวันที่ 31 มกราคม พ.ศ. 2565 รวม เป็น จำนวน 2370 วัน ดัง แสดง ใน รูปที่ 3.1 ทั้งนี้ จะ นำ ข้อมูล ราคา ปิด รายวัน ของ เงิน สกุล ดิจิทัล อีเธอเรียม ไป แปลง ให้ อยู่ ใน รูปแบบ ทิศทาง ของ ราคา ปิด



รูปที่ 3.1: กราฟแสดงราคาปิดรายวันของเงินสกุลดิจิทัลอีเธอเรียม ตั้งแต่วันที่ 7 สิงหาคม พ.ศ. 2558 ถึงวันที่ 31 มกราคม พ.ศ. 2565

### 3.2 กรอบแนวคิดการวิจัย

การศึกษานี้ใช้ตัวแปรอิสระทั้งหมด 6 ตัวแปร และสามารถเขียนเป็นกรอบแนวคิดการวิจัยได้ดังนี้



รูปที่ 3.2: กรอบแนวคิดการวิจัย

### 3.3 การจัดเตรียมข้อมูล

#### 3.3.1 การแปลงค่าข้อมูล (Data Transformation)

การแปลงค่าของราคาปิดรายวันของสกุลเงินดิจิทัลอีเธอเรียม ซึ่งเป็นตัวแปรเชิงปริมาณให้เป็นตัวแปรเชิงคุณภาพที่มีค่าของข้อมูลที่เป็นไปได้ 2 ค่า หรือเรียกว่า ตัวแปรไบนารี (Binary Variable) โดยมีฟังก์ชันในการแปลงดังนี้

$$Y_t = \begin{cases} 1; & \text{ถ้า } \frac{C_t}{C_{t-1}} > 1 \\ 0; & \text{อื่น ๆ} \end{cases}$$

โดย 0 หมายถึง ทิศทางลงหรือคงที่

1 หมายถึง ทิศทางขึ้น

#### 3.3.2 การตรวจสอบสลับไขว้ (Cross-Validation)

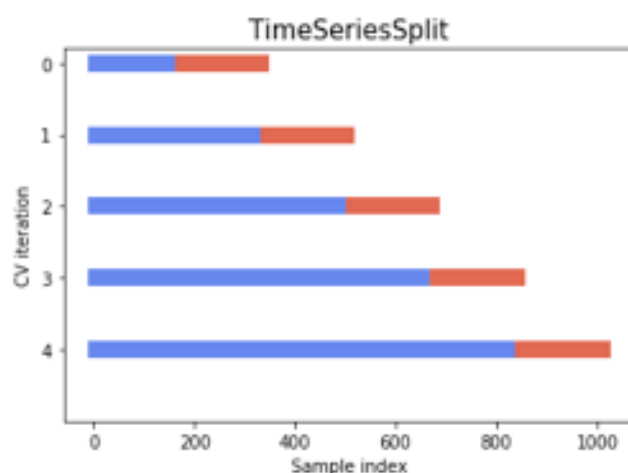
วิธีการตรวจสอบสลับไขว้ เป็นการแบ่งข้อมูลแบบสุ่มออกเป็นจำนวนกลุ่ม (k) ตามที่ต้องการ และจะมีการแบ่งข้อมูลออกเป็นส่วนย่อย ๆ จำนวน k ส่วน และจะเก็บข้อมูลจำนวน 1 ส่วนย่อย เพื่อนำไปใช้ในการทดสอบประสิทธิภาพของวิธีการพยากรณ์ส่วนข้อมูลที่เหลือจะนำมาสร้างตัวแบบพยากรณ์ และจะทำไปเรื่อย ๆ จนกว่าข้อมูลถูกใช้ทดสอบจนหมดทุกส่วน เช่น ถ้าแบ่งข้อมูลออกเป็น 5 กลุ่ม จะมีการทดสอบจำนวน 5 รอบ โดยรอบที่ 1 เป็นการเอาข้อมูลในกลุ่มที่ 1 ออกเพื่อใช้ในการทดสอบ และใช้กลุ่มที่ 2 ถึง 5 ในการสร้างตัวแบบพยากรณ์ ในรอบที่ 2 เป็นการเอาข้อมูลกลุ่มที่ 2 ออกเพื่อใช้ในการทดสอบและใช้ข้อมูล

กลุ่มที่ 1, 3 ถึง 5 ในการสร้างตัวแบบพยากรณ์ทำจนครบตามจำนวนกลุ่มข้อมูลที่กำหนดไว้ ดังรูปที่ 3.3



รูปที่ 3.3: ตัวอย่างการตรวจสอบสลับไขว้ (Tefera, 2020)

แต่เนื่องจากข้อมูลที่ผู้วิจัยนำมาใช้นั้นเป็นข้อมูลอนุกรมเวลา (Time Series Data) จึงไม่สามารถใช้การตรวจสอบสลับไขว้แบบปกติได้ จึงได้ทำการตรวจสอบสลับไขว้แบบอนุกรมเวลาโดยผู้วิจัยได้แบ่งข้อมูลออกเป็น 5 กลุ่มย่อย ซึ่งจะมีการทดสอบจำนวน 5 รอบ โดยรอบที่ 1 จะเป็นการแบ่งข้อมูลส่วนที่ 2 เพื่อใช้ในการทดสอบประสิทธิภาพในการพยากรณ์ และนำส่วนที่ 1 ใช้ในการเรียนรู้ ในรอบที่ 2 เป็นการนำข้อมูลส่วนที่ 3 ออกเพื่อใช้ในการทดสอบและใช้ข้อมูลส่วนที่ 1 และ 2 ในการเรียนรู้ทำจนครบ 5 รอบตามที่กำหนดไว้



รูปที่ 3.4: ตัวอย่างการแบ่งข้อมูลอนุกรมเวลาด้วยวิธีตรวจสอบสลับไขว้ (Soumya, 2020)

จากรูปที่ 3.4 ทำการแบ่งข้อมูลโดยใช้ไลบรารี TimeSeriesSplit จาก sklearn จำนวน 5 ส่วน (5-Fold) จะทำให้ได้ชุดข้อมูลสำหรับการเรียนรู้จำนวน 1980 วัน คิดเป็นร้อยละ 82 ของข้อมูลทั้งหมด

และชุดข้อมูลสำหรับทดสอบ 390 วัน คิดเป็นร้อยละ 18 ของข้อมูลทั้งหมด

### 3.3.3 Grid Search

Grid Search เป็นเทคนิคการปรับแต่งที่พยายาม ค้นหา ค่าที่เหมาะสมที่สุดของพารามิเตอร์ของตัวแบบเป็นระบบ โดยวิธีการคือ กำหนดค่าของแต่ละพารามิเตอร์ที่ต้องการจะทดสอบ จากนั้นทำการฝึกสอนตัวแบบในทุก ๆ ชุดของค่าพารามิเตอร์ที่เป็นไปได้ โดยผู้วิจัยต้องเลือกค่าพารามิเตอร์ที่เหมาะสมไว้ล่วงหน้า ซึ่งแต่ละตัวแบบพยากรณ์ผู้วิจัยได้กำหนดพารามิเตอร์ไว้ ดังตารางต่อไปนี้

#### 1. วิธีต้นไม้ตัดสินใจ

ในการทำ Grid Search สำหรับวิธีต้นไม้ตัดสินใจ จะพิจารณา ค่าพารามิเตอร์เพื่อ ค้นหา ค่าพารามิเตอร์ที่เหมาะสมจากพารามิเตอร์ 4 ตัว ซึ่งในการวิจัยครั้งนี้จะกำหนดช่วงของค่าพารามิเตอร์ที่จะทำการค้นหา ดังตารางที่ 3.1

ตารางที่ 3.1: การกำหนดช่วงของค่าพารามิเตอร์สำหรับวิธีต้นไม้ตัดสินใจ

พารามิเตอร์	ช่วงของค่าพารามิเตอร์	ค่าเริ่มต้น
ccp_alpha	[ 0.001 , 0.1 ]	0
max_depth	[ 2 , 6 ]	None
min_samples_split	[ 2 , 6 ]	2
min_samples_leaf	[ 2 , 6 ]	1

โดยที่ ccp\_alpha คือพารามิเตอร์ความซับซ้อนที่ใช้สำหรับการตัดแต่งกิ่งต้นไม้

max\_depth คือจำนวนชั้นของต้นไม้สูงสุด

min\_samples\_split คือจำนวนตัวอย่างขั้นต่ำที่จำเป็นในการแยกโหนดภายใน

min\_samples\_leaf คือจำนวนตัวอย่างขั้นต่ำที่อยู่ทีโหนดปลายสุด

จากตารางที่ 3.1 ผู้วิจัยทำการกำหนดช่วงของค่าพารามิเตอร์ในการค้นหาต่าง ๆ ดังนี้ค่า ccp\_alpha จะเริ่มค้นหาตั้งแต่ 0.001 ถึง 0.1 ค่า max\_depth เริ่มค้นหาตั้งแต่ 2 ถึง 6 ค่า min\_samples\_split จะเริ่มค้นหาตั้งแต่ 2 ถึง 6 และค่า min\_samples\_leaf จะเริ่มค้นหาตั้งแต่ 2 ถึง 6



## 2. Random Forest

ในการทำ Grid Search สำหรับวิธี Random Forest จะพิจารณาค่าพารามิเตอร์ เพื่อค้นหาค่าพารามิเตอร์ที่เหมาะสม จากพารามิเตอร์ 5 ตัว ซึ่งในการวิจัยครั้งนี้จะกำหนดช่วงของค่าพารามิเตอร์ที่จะทำการค้นหา ดังตารางที่ 3.2

ตารางที่ 3.2: การกำหนดช่วงของค่าพารามิเตอร์สำหรับวิธี Random Forest

พารามิเตอร์	ช่วงของค่าพารามิเตอร์	ค่าเริ่มต้น
n_estimators	[ 50 , 150 ]	100
ccp_alpha	[ 0.001 , 0.1 ]	0
max_depth	[ 2 , 6 ]	None
min_samples_split	[ 2 , 4 ]	2
min_samples_leaf	[ 1 - 4 ]	1

โดยที่ n\_estimators คือจำนวนต้นไม้ในตัวอย่าง

จากตารางที่ 3.2 ผู้วิจัยทำการกำหนดช่วงของค่าพารามิเตอร์ในการค้นหาต่าง ๆ ดังนี้ ค่า n\_estimators จะเริ่มค้นหาตั้งแต่ 50 ถึง 200 ค่า ccp\_alpha จะเริ่มค้นหาตั้งแต่ 0.01 ถึง 0.1 ค่า max\_depth จะเริ่มค้นหาตั้งแต่ 2 ถึง 6 ค่า min\_samples\_split จะเริ่มค้นหาตั้งแต่ 2 ถึง 4 และค่า min\_samples\_leaf จะเริ่มค้นหาตั้งแต่ 1 ถึง 4

## 3. XGBoost

ในการทำ Grid Search สำหรับวิธี XGBoost จะพิจารณาค่าพารามิเตอร์ เพื่อค้นหาค่าพารามิเตอร์ที่เหมาะสม จากพารามิเตอร์ 4 ตัว ซึ่งในการวิจัยครั้งนี้จะกำหนดช่วงของค่าพารามิเตอร์ที่จะทำการค้นหา ดังตารางที่ 3.3

ตารางที่ 3.3: การกำหนดช่วงของค่าพารามิเตอร์สำหรับวิธี XGBoost

พารามิเตอร์	ช่วงของค่าพารามิเตอร์	ค่าเริ่มต้น
learning_rate	[0.001 , 0.3 ]	0.3
n_estimators	[ 50 , 150 ]	100
max_depth	[ 2 , 6 ]	6
gamma	[ 0.001 , 0.2 ]	0

โดยที่ learning\_rate คือ อัตราการเรียนรู้

gamma คือ Regularisation parameter

จากตารางที่ 3.3 ผู้วิจัยทำการ กำหนดช่วง ของค่าพารามิเตอร์ในการ ค้นหาต่าง ๆ ดังนี้ ค่า learning\_rate จะเริ่ม ค้นหาตั้งแต่ 0.01 ถึง 0.3 ค่า n\_estimators จะเริ่ม ค้นหาตั้งแต่ 50 ถึง 150 ค่า max\_depth จะเริ่ม ค้นหาตั้งแต่ 2 ถึง 6 และ ค่า gamma จะเริ่ม ค้นหาตั้งแต่ 0.01 ถึง 0.2

#### 4. Extra Tree

ในการทำ Grid Search สำหรับวิธี Extra Tree จะพิจารณา ค่าพารามิเตอร์ เพื่อ ค้นหา ค่าพารามิเตอร์ที่เหมาะสม จากพารามิเตอร์ 5 ตัว ซึ่งในการวิจัยครั้งนี้จะกำหนดช่วงของค่าพารามิเตอร์ที่จะทำการค้นหา ดังตารางที่ 3.4

ตารางที่ 3.4: การกำหนดช่วงของค่าพารามิเตอร์สำหรับวิธี Extra Tree

พารามิเตอร์	ช่วงของค่าพารามิเตอร์	ค่าเริ่มต้น
n_estimators	[ 50 , 150 ]	100
ccp_alpha	[ 0.001 , 0.1 ]	0
max_depth	[ 2 , 6 ]	None
min_samples_split	[ 2 , 4 ]	2
min_samples_leaf	[ 1 , 4 ]	1

จากตารางที่ 3.4 ผู้วิจัยทำการ กำหนดช่วง ของค่าพารามิเตอร์ในการ ค้นหาต่าง ๆ ดังนี้ ค่า n\_estimators จะเริ่ม ค้นหาตั้งแต่ 50 ถึง 200 ค่า ccp\_alpha จะเริ่ม ค้นหาตั้งแต่ 0.01 ถึง 0.1 ค่า max\_depth จะเริ่ม ค้นหาตั้งแต่ 2 ถึง 6 ค่า min\_samples\_split จะเริ่ม ค้นหาตั้งแต่ 2 ถึง 4 และ

ค่า `min_samples_leaf` จะเริ่มค้นหาตั้งแต่ 1 ถึง 4

### 3.4 ขั้นตอนการดำเนินการวิจัย

1. ศึกษาและค้นคว้างานวิจัยที่เกี่ยวข้องกับการพยากรณ์ทิศทางราคาหุ้น
2. ศึกษาทฤษฎีที่เกี่ยวข้องกับงานวิจัย คือ วิธีต้นไม้ตัดสินใจ วิธี Random Forest วิธี XGBoost และวิธี Extra Tree
3. ศึกษาแพ็คเกจ `crypto2` ในภาษา R เพื่อใช้ในการเก็บรวบรวมข้อมูลราคาปิดรายวันของเงินสกุลดิจิทัลอีเธอเรียมจากเว็บไซต์ [www.coinmarketcap.com](http://www.coinmarketcap.com)
4. ศึกษาไลบรารีต่าง ๆ ในภาษา Python ได้แก่ `pandas` เพื่อใช้ในการจัดการข้อมูล และไลบรารี `Sklearn` เพื่อใช้สำหรับการแบ่งข้อมูลและวิเคราะห์ข้อมูล
5. คำนวณเครื่องมือวิเคราะห์เชิงเทคนิคจำนวน 6 เครื่องมือ ได้แก่ SMA, WMA, MACD, %K, RSI, และ %R และทำการแปลงข้อมูลตัวแปรตาม
6. ทำการแบ่งข้อมูลออกเป็น 5 ชุดด้วยไลบรารี `TimeSeriesSplit` จาก `sklearn` และในแต่ละชุดแบ่งเป็นร้อยละ 82 สำหรับการสร้างตัวแบบพยากรณ์ส่วนที่เหลืออีกร้อยละ 18 สำหรับทดสอบตัวแบบพยากรณ์ ลักษณะการแบ่งข้อมูล 2 ชุด ได้แก่ ชุดฝึกสอนจำนวน 1980 วัน และชุดทดสอบจำนวน 390 วัน
7. ทำการค้นหาค่าพารามิเตอร์ที่ดีที่สุดของแต่ละวิธีการพยากรณ์ด้วย Grid Search เพื่อนำมาสร้างตัวแบบพยากรณ์ทิศทางราคาปิดอีเธอเรียม
8. ทำการสร้างตัวแบบ 4 ตัวแบบพยากรณ์ด้วยวิธีต้นไม้ตัดสินใจ วิธี Random Forest วิธี XGBoost และวิธี Extra Tree โดยใช้ตัวแปรอิสระเป็นเครื่องมือวิเคราะห์เชิงเทคนิคจำนวน 6 ตัว ย้อนหลัง 1 วัน เพื่อพยากรณ์ทิศทางราคาปิดอีเธอเรียม
9. ทำการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ที่ได้จากวิธีการพยากรณ์ทั้ง 3 วิธี ด้วยค่า Accuracy, Recall และ F1 - Score
10. สรุปและอภิปรายผล

## บทที่ 4

### ผลการวิจัย

การศึกษานี้ได้ทำการเปรียบเทียบประสิทธิภาพวิธีการพยากรณ์ทิศทางราคาปิดของเงินสกุลดิจิทัล อีเธอเรียม ด้วย วิธี ต้นไม้ ตัดสินใจ วิธี Random Forest วิธี XGBoost และ วิธี Extra Tree ผลการวิเคราะห์ข้อมูลแสดงได้ดังนี้

#### 4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

ในส่วนนี้จะเป็นการนำเสนอผลการวิเคราะห์ข้อมูลเบื้องต้นของตัวแปรที่ใช้ในการวิจัย สำหรับค่าสถิติพื้นฐานที่ใช้ในการวิเคราะห์ข้อมูล ประกอบด้วย ค่าต่ำสุด (Minimum) ค่าสูงสุด (Maximum) ค่าเฉลี่ยเลขคณิต (Mean) ค่ามัธยฐาน (Median) และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ของตัวแปรที่ใช้ในการศึกษาสามารถวิเคราะห์ได้ดังตารางที่ 4.1

ตารางที่ 4.1: ค่าสถิติพื้นฐานของตัวแปรที่ใช้ในการศึกษา

ตัวแปร	ค่าสูงสุด	ค่าต่ำสุด	ค่าเฉลี่ย	ค่ามัธยฐาน	ส่วนเบี่ยงเบนมาตรฐาน
ราคาปิด (Close)	4812.09	0.44	660.81	222.49	1068.21
ค่าเฉลี่ยเคลื่อนที่อย่างง่าย (SMA)	4655.83	0.52	655.90	219.62	1062.45
ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนัก (WMA)	4669.87	0.50	657.50	220.20	1063.84
ค่าเฉลี่ยเคลื่อนที่รวมทาง/แยกทาง (MACD)	454.41	-323.65	9.30	0.37	67.62
ค่าดัชนีการแกว่ง (%K)	100.00	0.00	54.53	54.93	29.25
ค่าดัชนีกำลังสัมพันธ์ (RSI)	93.47	15.69	53.14	52.41	14.61
ค่า Williams Percentage Range (%R)	0.00	-100.00	-45.47	-45.07	29.25

จากตารางที่ 4.1 ราคาปิดรายวันของเงินสกุลดิจิทัลอีเธอเรียม มีค่าสูงสุดอยู่ที่ 4812.09 USD ค่าต่ำสุดอยู่ที่ 0.44 USD ค่าเฉลี่ยอยู่ที่ 660.81 USD และส่วนเบี่ยงเบนมาตรฐานอยู่ที่ 1068.21 USD

ค่าเฉลี่ยเคลื่อนที่อย่างง่าย มีค่าสูงสุดอยู่ที่ 4655.83 ค่าต่ำสุดอยู่ที่ 0.52 ค่าเฉลี่ยอยู่ที่ 655.90 และส่วนเบี่ยงเบนมาตรฐานอยู่ที่ 1062.45

ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนัก มีค่าสูงสุดอยู่ที่ 4669.87 ค่าต่ำสุดอยู่ที่ 0.50 ค่าเฉลี่ยอยู่ที่

657.50 และส่วนเบี่ยงเบนมาตรฐานอยู่ที่ 1063.84

ค่าเฉลี่ยเคลื่อนที่ร่วมทาง/แยกทาง มีค่าสูงสุดอยู่ที่ 454.41 ค่าต่ำสุดอยู่ที่ -323.65 ค่าเฉลี่ยอยู่ที่ 9.30 และส่วนเบี่ยงเบนมาตรฐานอยู่ที่ 67.62

ค่าดัชนีการแกว่ง มีค่าสูงสุดอยู่ที่ 100.00 ค่าต่ำสุดอยู่ที่ 0.00 ค่าเฉลี่ยอยู่ที่ 54.53 และส่วนเบี่ยงเบนมาตรฐานอยู่ที่ 29.25

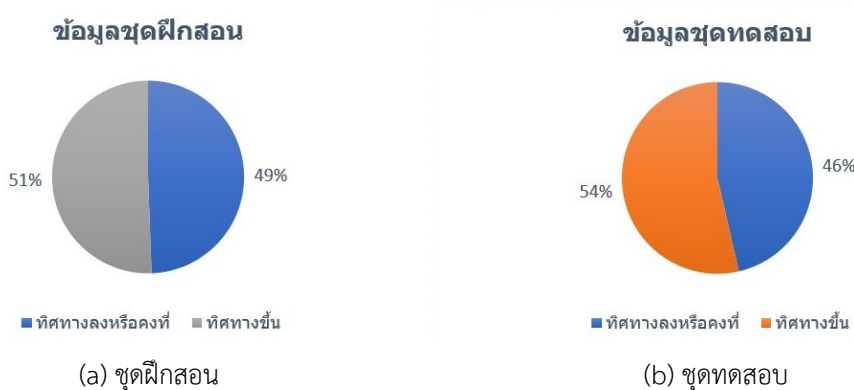
ค่าดัชนีกำลังสัมพันธ์ มีค่าสูงสุดอยู่ที่ 93.47 ค่าต่ำสุดอยู่ที่ 15.69 ค่าเฉลี่ยอยู่ที่ 53.14 และส่วนเบี่ยงเบนมาตรฐานอยู่ที่ 14.61

ค่า Williams Percentage Range มีค่าสูงสุดอยู่ที่ 0.00 ค่าต่ำสุดอยู่ที่ -100.00 ค่าเฉลี่ยอยู่ที่ -45.47 และส่วนเบี่ยงเบนมาตรฐานอยู่ที่ 29.25

ตารางที่ 4.2: จำนวนข้อมูลทั้งหมดในชุดฝึกสอนและชุดทดสอบ

	ทิศทางลงหรือคงที่ (0)	ทิศทางขึ้น (1)	ผลรวมทั้งหมด
จำนวนข้อมูลชุดฝึกสอน	966	989	1955
จำนวนข้อมูลชุดทดสอบ	181	209	390

จากตารางที่ 4.2 แสดงค่าสถิติพื้นฐานของข้อมูลทิศทางราคาปิดรายวันของเงินสกุลดิจิทัลอีเธอเรียมทั้งหมด 2,345 ค่า พบว่าในข้อมูลชุดฝึกสอนมีทิศทางลงหรือคงที่จำนวน 966 ค่า และมีทิศทางขึ้น 989 ค่า ส่วนข้อมูลชุดทดสอบมีทิศทางลงหรือคงที่จำนวน 181 ค่า และมีทิศทางขึ้น 209 ค่า



รูปที่ 4.1: ร้อยละของราคาปิดอีเธอเรียมรายวันจำแนกตามทิศทางของราคาปิด และชุดข้อมูล

จากรูปที่ 4.1 พบว่าในข้อมูลชุดฝึกสอนมีทิศทางลงหรือคงที่คิดเป็นร้อยละ 51 และทิศทางขึ้นคิดเป็นร้อยละ 49 ส่วนข้อมูลในชุดทดสอบมีทิศทางลงหรือคงที่ร้อยละ 54 และทิศทางขึ้นร้อยละ 46

## 4.2 ผลการค้นหาค่าพารามิเตอร์ที่เหมาะสมที่สุดสำหรับตัวแบบพยากรณ์ด้วยวิธี Grid Search

### 4.2.1 วิธีต้นไม้ตัดสินใจ

จากการค้นหาค่าพารามิเตอร์ที่เหมาะสมสำหรับวิธีต้นไม้ตัดสินใจ โดยพิจารณาจากค่า Accuracy ที่มีค่าสูงสุด ได้ค่าพารามิเตอร์ที่เหมาะสมดังตารางที่ 4.3

ตารางที่ 4.3: ค่าพารามิเตอร์ที่เหมาะสมที่สุดของตัวแบบต้นไม้ตัดสินใจ

พารามิเตอร์	ค่าที่เหมาะสมที่สุด
ccp_alpha	0.001
max_depth	2
min_samples_split	2
min_samples_leaf	2

จากตารางที่ 4.3 พบว่าค่าพารามิเตอร์ที่ดีที่สุดของตัวแบบพยากรณ์ต้นไม้ตัดสินใจ คือค่า ความซับซ้อนที่ใช้สำหรับการตัดแต่งกิ่งต้นไม้ (ccp\_alpha) เป็น 0.001 ค่า จำนวนชั้นของต้นไม้สูงสุด (max\_depth) เป็น 2 ค่า จำนวนตัวอย่างขั้นต่ำที่ใช้ในการแยกโหนด (min\_samples\_split) เป็น 2 และค่า จำนวนตัวอย่างขั้นต่ำของโหนดปลายสุด (min\_samples\_leaf) เป็น 2

### 4.2.2 Random Forest

จากการค้นหาค่าพารามิเตอร์ที่เหมาะสมสำหรับวิธี Random Forest โดยพิจารณาจากค่า Accuracy ที่มีค่าสูงสุด ได้ค่าพารามิเตอร์ที่เหมาะสมดังตารางที่ 4.4

ตารางที่ 4.4: ค่าพารามิเตอร์ที่เหมาะสมที่สุดของตัวแบบ Random Forest

พารามิเตอร์	ค่าที่เหมาะสมที่สุด
n_estimators	90
ccp_alpha	0.004
max_depth	4
min_samples_split	2
min_samples_leaf	1

จากตารางที่ 4.4 พบว่าค่าพารามิเตอร์ที่ดีที่สุดของตัวแบบ Random Forest คือค่า จำนวนต้นไม้ในตัวแบบ (n\_estimators) เป็น 90 ค่า ccp\_alpha เป็น 0.004 ค่า max\_depth เป็น 4 ค่า min\_samples\_split เป็น 2 และค่า min\_samples\_leaf เป็น 1

### 4.2.3 XGBoost

จากการค้นหาค่าพารามิเตอร์ที่เหมาะสมสำหรับวิธี XGBoost โดยพิจารณาจากค่า Accuracy ที่มีค่าสูงที่สุด ได้ค่าพารามิเตอร์ที่เหมาะสมดังตารางที่ 4.5

ตารางที่ 4.5: ค่าพารามิเตอร์ที่เหมาะสมที่สุดของตัวแบบ XGBoost

พารามิเตอร์	ค่าที่เหมาะสมที่สุด
n_estimators	140
learning_rate	0.001
max_depth	2
gamma	0.001

จากตารางที่ 4.5 พบว่าค่าพารามิเตอร์ที่ดีที่สุดของตัวแบบ XGBoost คือค่า n\_estimators เป็น 140 ค่า อัตราการเรียนรู้ (learning\_rate) เป็น 0.001 ค่า max\_depth เป็น 2 และค่า gamma เป็น 0.001

### 4.2.4 Extra Tree

จากการค้นหาค่าพารามิเตอร์ที่เหมาะสมสำหรับวิธี Extra Tree โดยพิจารณาจากค่า Accuracy ที่มีค่าสูงที่สุด ได้ค่าพารามิเตอร์ที่เหมาะสมดังตารางที่ 4.6

ตารางที่ 4.6: ค่าพารามิเตอร์ที่เหมาะสมที่สุดของตัวแบบ Extra Tree

พารามิเตอร์	ค่าที่เหมาะสมที่สุด
n_estimators	60
ccp_alpha	0.001
max_depth	5
min_samples_split	2
min_samples_leaf	2

จากตารางที่ 4.6 พบว่าค่าพารามิเตอร์ที่ดีที่สุดของตัวแบบ Extra Tree คือค่า n\_estimators เป็น 60 ค่า ccp\_alpha เป็น 0.001 ค่า max\_depth เป็น 5 ค่า min\_samples\_split เป็น 2 และค่า min\_samples\_leaf เป็น 2

### 4.3 ตารางแสดง Confusion Matrix ของตัวแบบพยากรณ์

#### 4.3.1 วิธีต้นไม้ตัดสินใจ

จากตัวแบบพยากรณ์ที่ได้จากวิธีต้นไม้ตัดสินใจ สามารถสร้างตาราง Confusion Matrix ที่มีค่า Accuracy สูงที่สุดได้ ดังตารางที่ 4.7

ตารางที่ 4.7: Confusion Matrix ของตัวแบบต้นไม้ตัดสินใจ

ทำนาย	ค่าจริง		รวม	ร้อยละความถูกต้องของการทำนาย
	ทิศทางลงหรือคงที่	ทิศทางขึ้น		
ทิศทางลงหรือคงที่	53	78	131	33.59
ทิศทางขึ้น	128	131	259	66.41
รวม	181	209	390	47.18

จากตารางที่ 4.7 พบว่าจำนวนข้อมูลที่ใช้ทดสอบมีทั้งหมด 390 ค่า และทิศทางลงหรือคงที่สามารถทำนายได้อย่างถูกต้องร้อยละ 33.59 ทิศทางขึ้นสามารถทำนายได้อย่างถูกต้องร้อยละ 66.41 และสามารถทำนายความถูกต้องโดยรวมได้ร้อยละ 47.18

#### 4.3.2 Random Forest

จากตัวแบบพยากรณ์ที่ได้จากวิธี Random Forest สามารถสร้างตาราง Confusion Matrix ที่มีค่า Accuracy สูงที่สุดได้ ดังตารางที่ 4.8



ตารางที่ 4.8: Confusion Matrix ของตัวแบบ Random Forest

ทำนาย	ค่าจริง		รวม	ร้อยละความถูกต้องของการทำนาย
	ทิศทางลงหรือคงที่	ทิศทางขึ้น		
ทิศทางลงหรือคงที่	55	78	133	34.10
ทิศทางขึ้น	126	131	257	65.90
รวม	181	209	390	47.69

จากตารางที่ 4.8 พบว่าจำนวนข้อมูลที่ใช้ทดสอบมีทั้งหมด 390 ค่า และทิศทางลงหรือคงที่สามารถทำนายได้อย่างถูกต้องร้อยละ 34.10 ทิศทางขึ้นสามารถทำนายได้อย่างถูกต้องร้อยละ 65.90 และสามารถทำนายความถูกต้องโดยรวมได้ร้อยละ 47.69

#### 4.3.3 XGBoost

จากตัวแบบพยากรณ์ที่ได้จากวิธี XGBoost สามารถสร้างตาราง Confusion Matrix ที่มีค่า Accuracy สูงที่สุดได้ ดังตารางที่ 4.9

ตารางที่ 4.9: Confusion Matrix ของตัวแบบ XGBoost

ทำนาย	ค่าจริง		รวม	ร้อยละความถูกต้องของการทำนาย
	ทิศทางลงหรือคงที่	ทิศทางขึ้น		
ทิศทางลงหรือคงที่	150	197	347	88.97
ทิศทางขึ้น	31	12	43	11.03
รวม	181	209	390	41.54

จากตารางที่ 4.9 พบว่าจำนวนข้อมูลที่ใช้ทดสอบมีทั้งหมด 390 ค่า และทิศทางลงหรือคงที่สามารถทำนายได้อย่างถูกต้องร้อยละ 88.97 ทิศทางขึ้นสามารถทำนายได้อย่างถูกต้องร้อยละ 11.03 และสามารถทำนายความถูกต้องโดยรวมได้ร้อยละ 41.54

#### 4.3.4 Extra Tree

จากตัวแบบพยากรณ์ที่ได้จากวิธี Extra Tree สามารถสร้างตาราง Confusion Matrix ที่มีค่า Accuracy สูงที่สุดได้ ดังตารางที่ 4.7

ตารางที่ 4.10: Confusion Matrix ของตัวแบบ Extra Tree

ทำนาย	ค่าจริง		รวม	ร้อยละความถูกต้องของการทำนาย
	ทิศทางลงหรือคงที่	ทิศทางขึ้น		
ทิศทางลงหรือคงที่	110	96	206	52.82
ทิศทางขึ้น	71	113	184	47.18
รวม	181	209	390	57.18

จากตารางที่ 4.10 พบว่าจำนวนข้อมูลที่ใช้ทดสอบมีทั้งหมด 390 ค่า และทิศทางลงหรือคงที่สามารถทำนายได้อย่างถูกต้องร้อยละ 52.82 ทิศทางขึ้นสามารถทำนายได้อย่างถูกต้องร้อยละ 47.18 และสามารถทำนายความถูกต้องโดยรวมได้ร้อยละ 57.18

#### 4.4 ผลการวิเคราะห์ประสิทธิภาพของตัวแบบพยากรณ์

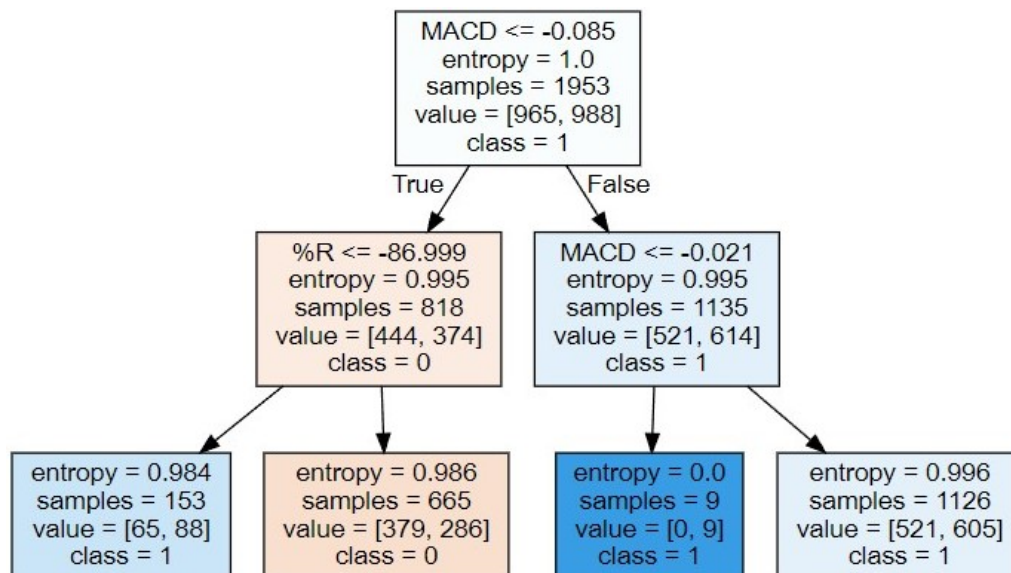
##### 4.4.1 วิธีต้นไม้ตัดสินใจ

จากตัวแบบพยากรณ์ที่ได้จากวิธีต้นไม้ตัดสินใจ สามารถสร้างตารางร้อยละของความแม่นยำ (Precision) ร้อยละของค่าความระลึก (Recall) และร้อยละของ F1 - Score ได้ ดังตารางที่ 4.11

ตารางที่ 4.11: ร้อยละของความแม่นยำ ค่าความระลึก และ F1 - Score ของตัวแบบต้นไม้ตัดสินใจ

เกณฑ์ที่ใช้เปรียบเทียบ	ทิศทางราคาปิดอีเธอเรียม		ร้อยละโดยรวม
	ทิศทางลงหรือคงที่	ทิศทางขึ้น	
	ร้อยละ	ร้อยละ	
Precision	40	51	45.50
Recall	29	63	46.00
F1 - Score	34	56	45.00

จากตารางที่ 4.11 พบว่า ค่า Precision ของทิศทางลงหรือคงที่คิดเป็นร้อยละ 40 ค่า Recall คิดเป็นร้อยละ 29 และค่า F1 - Score คิดเป็นร้อยละ 34 ส่วนทิศทางขึ้นค่า Precision คิดเป็นร้อยละ 51 ค่า Recall คิดเป็นร้อยละ 63 และค่า F1 - Score คิดเป็นร้อยละ 56



รูปที่ 4.2: ต้นไม้ตัดสินใจที่ได้จากตัวแบบ

จากรูปที่ 4.2 ตัวแปรอิสระที่นำมาสร้างต้นไม้ตัดสินใจมี 2 ตัวแปร คือ ค่าเฉลี่ยเคลื่อนที่ร่วมทาง/แยกทาง (MACD) และค่า Williams Percentage Range (%R)

#### 4.4.2 Random Forest

จากตัวแบบพยากรณ์ที่ได้จากวิธี Random Forest สามารถสร้างตารางร้อยละของความแม่นยำ ร้อยละของค่าความระลึกลับ และร้อยละของ F1 - Score ได้ ดังตารางที่ 4.12

ตารางที่ 4.12: ร้อยละของความแม่นยำ ค่าความระลึกลับ และ F1 - Score ของตัวแบบ Random Forest

เกณฑ์ที่ใช้เปรียบเทียบ	ทิศทางการค้าปิดหรือเปิด		ร้อยละโดยรวม
	ทิศทางลงหรือคงที่	ทิศทางขึ้น	
	ร้อยละ	ร้อยละ	
Precision	43	51	47.00
Recall	36	57	46.50
F1 - Score	39	54	46.50

จากตารางที่ 4.12 พบว่า ค่า Precision ของทิศทางลงหรือคงที่คิดเป็นร้อยละ 43 ค่า Recall คิดเป็นร้อยละ 36 และค่า F1 - Score คิดเป็นร้อยละ 39 ส่วนทิศทางขึ้นค่า Precision คิดเป็นร้อยละ 51 ค่า Recall คิดเป็นร้อยละ 57 และค่า F1 - Score คิดเป็นร้อยละ 54

#### 4.4.3 XGBoost

จากตัวแบบพยากรณ์ที่ได้จากวิธี XGBoost สามารถสร้างตารางร้อยละของความแม่นยำ ร้อยละของค่าความระลึกลับ และร้อยละของ F1 - Score ได้ ดังตารางที่ 4.13

ตารางที่ 4.13: ร้อยละของความแม่นยำ ค่าความระลึกลับ และ F1 - Score ของตัวแบบ XGBoost

เกณฑ์ที่ใช้เปรียบเทียบ	ทิศทางราคาปิดโอเฮโอเรียม		ร้อยละโดยรวม
	ทิศทางลงหรือคงที่	ทิศทางขึ้น	
	ร้อยละ	ร้อยละ	
Precision	43	28	35.50
Recall	83	6	44.50
F1 - Score	57	10	33.50

จากตารางที่ 4.13 พบว่า ค่า Precision ของทิศทางลงหรือคงที่คิดเป็นร้อยละ 43 ค่า Recall คิดเป็นร้อยละ 83 และค่า F1 - Score คิดเป็นร้อยละ 57 ส่วนทิศทางขึ้นค่า Precision คิดเป็นร้อยละ 28 ค่า Recall คิดเป็นร้อยละ 6 และค่า F1 - Score คิดเป็นร้อยละ 10

#### 4.4.4 Extra Tree

จากตัวแบบพยากรณ์ที่ได้จากวิธี Extra Tree สามารถสร้างตารางร้อยละของความแม่นยำ ร้อยละของค่าความระลึกลับ และร้อยละของ F1 - Score ได้ ดังตารางที่ 4.14

ตารางที่ 4.14: ร้อยละของความแม่นยำ ค่าความระลึกลับ และ F1 - Score ของตัวแบบ Extra Tree

เกณฑ์ที่ใช้เปรียบเทียบ	ทิศทางราคาปิดอีเธอเรียม		ร้อยละโดยรวม
	ทิศทางลงหรือคงที่	ทิศทางขึ้น	
	ร้อยละ	ร้อยละ	
Precision	53	61	57.00
Recall	61	54	57.50
F1 - Score	57	58	57.50

จากตารางที่ 4.14 พบว่า ค่า Precision ของทิศทางลงหรือคงที่คิดเป็นร้อยละ 53 ค่า Recall คิดเป็นร้อยละ 61 และค่า F1 - Score คิดเป็นร้อยละ 57 ส่วนทิศทางขึ้นค่า Precision คิดเป็นร้อยละ 61 ค่า Recall คิดเป็นร้อยละ 54 และค่า F1 - Score คิดเป็นร้อยละ 58

#### 4.5 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์

จากการสร้างตัวแบบพยากรณ์ทิศทางราคาปิดอีเธอเรียมด้วยวิธีต้นไม้ตัดสินใจ วิธี Random Forest วิธี XGBoost และวิธี Extra Tree ได้ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ดังต่อไปนี้

ตารางที่ 4.15: ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์

เกณฑ์ที่ใช้เปรียบเทียบ	ต้นไม้ตัดสินใจ	Random Forest	XGBoost	Extra Tree
Accuracy (ร้อยละ)	47.18	47.69	41.54	<b>57.18</b>
Precision (ร้อยละ)	46.50	47.00	35.50	<b>57.00</b>
Recall (ร้อยละ)	46.00	46.50	44.50	<b>57.50</b>
F1 - Score (ร้อยละ)	45.00	46.50	33.50	<b>57.50</b>

จากตารางที่ 4.15 แสดงการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ พบว่าตัวแบบ Extra Tree ที่มีค่าพารามิเตอร์ จำนวนต้นไม้ในตัวแบบ คือ 60 ค่าความซับซ้อนที่ใช้สำหรับการตัดแต่งกิ่งต้นไม้ คือ 0.001 ค่าจำนวนชั้นของต้นไม้สูงสุด คือ 5 ค่าจำนวนตัวอย่างขั้นต่ำที่ใช้แยกโหนดภายใน คือ 2 และค่าจำนวนตัวอย่างขั้นต่ำที่โหนดปลายสุด คือ 2 เป็นตัวแบบพยากรณ์ที่ดีที่สุดโดยมีค่า Accuracy ร้อยละ 57.18

## บทที่ 5

### สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

จากการศึกษาการพยากรณ์ทิศทางราคาปิดรายวันของเงินสกุลดิจิทัลอีเธอเรียม เพื่อเปรียบเทียบประสิทธิภาพวิธีการพยากรณ์ ได้แก่ วิธีต้นไม้ตัดสินใจ วิธี Random Forest วิธี XGBoost และวิธี Extra Tree โดยใช้ข้อมูลราคาปิดย้อนหลังรายวันที่เก็บรวบรวมมาจากเว็บไซต์ [www.coinmarketcap.com](http://www.coinmarketcap.com) แล้วทำการวิเคราะห์ข้อมูลด้วยภาษาไพธอน สามารถสรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะต่าง ๆ สำหรับงานวิจัยในครั้งนี้ ดังนี้

#### 5.1 สรุปผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีการพยากรณ์ทิศทางราคาปิดรายวันของเงินสกุลดิจิทัลอีเธอเรียมที่ได้จากวิธีการพยากรณ์ ได้แก่ วิธีต้นไม้ตัดสินใจ วิธี Random Forest วิธี XGBoost และวิธี Extra Tree ในการพยากรณ์ทิศทางราคาปิดของเงินสกุลดิจิทัลอีเธอเรียม โดยข้อมูลที่ใช้เป็นข้อมูลราคาปิดรายวัน จำนวน 2370 วัน ทำการแบ่งข้อมูลออกเป็น 2 ชุดด้วยวิธีการตรวจสอบสลับไขว้แบบอนุกรมเวลาจำนวน 5 ส่วน (5-Fold Cross-Validation) จะทำให้ได้ข้อมูล ชุดแรกที่ใช้ในการสร้างตัวแบบจำนวน 1980 วัน คิดเป็นร้อยละ 82 ของข้อมูลทั้งหมด และข้อมูลชุดที่สองที่ใช้ในการตรวจสอบตัวแบบจำนวน 390 วัน คิดเป็นร้อยละ 18 ของข้อมูลทั้งหมด โดยเกณฑ์ที่ใช้ในการตัดสินประสิทธิภาพของวิธีการพยากรณ์ คือค่า Accuracy, Recall และค่า F1 - Score

จากการศึกษาพบว่า วิธีต้นไม้ตัดสินใจ มีค่า Accuracy เท่ากับร้อยละ 47.18 ค่า Recall เท่ากับร้อยละ 46.00 และค่า F1 - Score เท่ากับร้อยละ 45.00 ซึ่งประกอบไปด้วยค่าพารามิเตอร์ ค่าความซับซ้อนที่ใช้สำหรับการตัดแต่งกิ่งต้นไม้ (ccp\_alpha) คือ 0.001 ค่าจำนวนชั้นของต้นไม้สูงสุด (max\_depth) คือ 2 ค่าจำนวนตัวอย่างขั้นต่ำที่ใช้ในการแยกโหนด (min\_samples\_split) คือ 2 และค่าจำนวนตัวอย่างขั้นต่ำของโหนดปลายสุด (min\_samples\_leaf) คือ 2 วิธีการพยากรณ์ Random Forest ให้ค่า Accuracy เท่ากับร้อยละ 47.69 ค่า Recall เท่ากับร้อยละ 46.50 และค่า F1 - Score เท่ากับร้อยละ 46.50 ซึ่งประกอบไปด้วยค่าพารามิเตอร์ จำนวนต้นไม้ในตัวแบบ (n\_estimators) คือ 90 ค่าความซับซ้อนที่ใช้สำหรับการตัดแต่งกิ่งต้นไม้ คือ 0.004 ค่าจำนวนชั้นของต้นไม้สูงสุด คือ 4 ค่าจำนวนตัวอย่างขั้นต่ำที่ใช้ในการแยกโหนด คือ 2 และค่าจำนวนตัวอย่างขั้นต่ำของโหนดปลายสุด คือ 1 วิธีการพยากรณ์ XGBoost ให้ค่า Accuracy เท่ากับร้อยละ 41.54 ค่า Recall เท่ากับร้อยละ 44.50 และค่า F1 - Score เท่ากับร้อยละ 33.50 ซึ่งประกอบไปด้วยค่าพารามิเตอร์ ค่าจำนวนชั้นของต้นไม้สูงสุด คือ 140 ค่าอัตราการเรียนรู้ (learning\_rate) คือ 0.001 ค่าจำนวนชั้นของต้นไม้สูงสุด คือ 2 และค่า gamma คือ 0.001 และวิธีการ

พยากรณ์ Extra Tree ให้ค่า Accuracy เท่ากับร้อยละ 57.18 ค่า Recall เท่ากับ ร้อยละ 57.50 และค่า F1 - Score เท่ากับร้อยละ 57.50 ซึ่งประกอบไปด้วยค่าพารามิเตอร์ ค่าจำนวนต้นไม้ในต้นไม้ คือ 60 ค่าความซับซ้อนที่ใช้สำหรับการตัดแต่งกิ่งต้นไม้ คือ 0.001 ค่าจำนวนชั้นของต้นไม้สูงสุด คือ 5 ค่าจำนวนตัวอย่างขั้นต่ำที่ใช้ในการแยกโหนด คือ 2 และค่าจำนวนตัวอย่างขั้นต่ำของโหนดปลายสุด คือ 2 ซึ่งในการวิจัยครั้งนี้วิธีการพยากรณ์ทิศทางราคาปิดที่แม่นยำที่สุดจะเป็นวิธีการที่ให้ค่า Accuracy, Recall และ F1 - Score สูงที่สุด ดังนั้นวิธีการพยากรณ์ Extra Tree จึงเป็นวิธีการพยากรณ์ทิศทางราคาปิดรายวันของอีเธอเรียมที่สามารถพยากรณ์ได้แม่นยำที่สุด

## 5.2 อภิปรายผล

จากผลการศึกษาพบว่า ตัวแบบที่พยากรณ์ทิศทางราคาปิดรายวันของเงินสกุลดิจิทัลอีเธอเรียมล่วงหน้า 1 วัน ได้แม่นยำที่สุด คือวิธี Extra Tree ซึ่งสอดคล้องกับ งานวิจัยของ (Ampomah et al., 2020) ที่ได้ทำการศึกษาการเปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์ในหุ้น 8 ตัว ได้แก่ Bank of America Corporation (BAC), Exxon Mobil Corporation (XOM), S&P 500, Microsoft Corporation (MSFT), Dow Jones Industrial Average (DJIA), CarMax Corporation (KMX), Tata Steel Limited (TATASTEEL) และ HCL Technologies Limited (HCLTECH) โดยใช้ ข้อมูล ที่ ตั้งแต่ วันที่ 1 มกราคม ค.ศ. 2005 ถึงวันที่ 30 ธันวาคม ค.ศ. 2019 ด้วยวิธีการพยากรณ์ Random Forest, XGBoost, Bagging, AdaBoost, Extra Trees และ Voting ที่พบว่าวิธีการพยากรณ์ Extra Tree นั้นสามารถพยากรณ์ได้แม่นยำที่สุด

## 5.3 ข้อเสนอแนะ

1. ในการวิจัยครั้งถัดไปอาจจะเพิ่มตัวแปรตัวอิสระอื่น ๆ ที่นอกเหนือจากที่ใช้ในงานวิจัยครั้งนี้ เช่น เครื่องมือวิเคราะห์เชิงเทคนิค และคุณลักษณะที่มาจากข่าวการเงิน เป็นต้น
2. ในการวิจัยครั้งถัดไปควรขยายช่วงของพารามิเตอร์ในแต่ละวิธีให้กว้างกว่านี้อาจจะทำให้ตัวแบบที่ได้มีประสิทธิภาพมากยิ่งขึ้น
3. ในการวิจัยครั้งถัดไปอาจจะเปลี่ยนจากราคาปิดรายวันเป็นราคาปิดรายชั่วโมง
4. ในการวิจัยครั้งถัดไปควรศึกษาวิธีการพยากรณ์อื่น ๆ เพิ่มเติม เช่น วิธี Quantile Regression Forest เป็นต้น
5. ในการวิจัยครั้งถัดไปอาจจะพยากรณ์ล่วงหน้ามากกว่า 1 วัน

## บรรณานุกรม

- ขจรศักดิ์ ศรีอ่อน. (2552). การทำนายสาเหตุของเหตุการณ์กระแสไฟฟ้าขัดข้องโดยใช้เทคนิคการทำเหมืองข้อมูลในระบบจำหน่ายของการไฟฟ้าส่วนภูมิภาค เขต 1 ภาคกลาง. [วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ], มหาวิทยาลัยเกษตรศาสตร์.
- จิรายุส ทรัพย์ศรีโสภา . (2564). รู้จัก ethereum เหรียญอันดับ 2 รองจาก bitcoin. สืบค้นเมื่อวันที่ 1 พฤศจิกายน 2564. จาก <https://moneyandbanking.co.th/article/the-guru/ethereum-cryptocurrency-mb472-aug2021-230864>.
- ฐิติมา ชูเชิด. (2564). สกุลเงินดิจิทัล ใกล้ตัวเราแค่ไหน. สืบค้นเมื่อวันที่ 3 ธันวาคม 2564. จาก [https://www.bot.or.th/Thai/ResearchAndPublications/articles/Pages/Article\\_01Feb2019.aspx](https://www.bot.or.th/Thai/ResearchAndPublications/articles/Pages/Article_01Feb2019.aspx).
- ภูริพัทธ์ ทองคำ. (2559). อัลกอริทึมแบบรวมสำหรับการเลือกคุณสมบัติของข้อมูล. [วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ], มหาวิทยาลัยธรรมศาสตร์.
- วีรพงษ์ ชุตินันท์. (2564). Crypto series : ตอนที่ 1 กำเนิด bitcoin. สืบค้นเมื่อวันที่ 1 ธันวาคม 2564. จาก <https://www.trinitythai.com/th/Knowledge/PublishArticles/66>.
- วสันต์ ศิลปะ, สมพร ปันโนชา, และบำรุง พ่วงเกิด. (2560). การทำนายราคาหลักทรัพย์โดยใช้โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ โดยใช้ตัวแบบอนุกรมเวลาฟังก์ชันแบบไม่เชิงเส้น. *การประชุมนำเสนอผลงานวิจัยระดับบัณฑิตศึกษา*, 12:1508--1518.
- ศิริลักษณ์ หล่อพันธุ์มณี, ศิริพร ฉิมพลี, วิชชา ฉิมพลี, วัชรกรรณ์ เนตรหาญ และณรงค์ฤทธิ์ ภิรมย์นง. (2562). การพยากรณ์รูปแบบของการกระทำความผิดซ้ำของเด็กและเยาวชน. *วารสารนวัตกรรมการพัฒนาศาสตร์เพื่อการพัฒนาอย่างยั่งยืน*, 1:80--96.
- สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (2561). Smart contract คืออะไร. สืบค้นเมื่อ 1 ธันวาคม 2564. จาก <https://www.etda.or.th/th/Useful-Resource/knowledge-sharing/articles/What-is-Smart-Contract.aspx>.
- อฤดี พันธนา. (2557). การพยากรณ์ความต้องการสายพานรถยนต์และสายพานอุตสาหกรรมของลูกค้า. [วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ], มหาวิทยาลัยบูรพา.
- Advisor, D. (2017). มือใหม่...เริ่มอ่านแท่งเทียน. สืบค้นเมื่อ 30 ธันวาคม 2564. จาก <https://knowledge.bualuang.co.th/knowledge-base/candlesticks/>.



- Ampomah, E. K., Qin, Z., and Nyame, G. (2020). Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement. *Information*, 11(6):332.
- Basak, S., Kar, S., Saha, S., Khaidem, L., and Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47:552--567.
- Castoe, M. (2020). Predicting stock market price direction with uncertainty using quantile regression forest. [master's thesis], Uppsala University.
- Chakrabarty, N. and Biswas, S. (2020). Navo minority over-sampling technique (nmote): A consistent performance booster on imbalanced datasets. *Journal of Electronics and Informatics*, 2:96--136.
- Chalach, M. (2019). ทำ decision tree ด้วย scikit-learn. สืบค้นเมื่อ 11 ธันวาคม 2564. จาก <https://medium.com/equinox-blog/decision-tree-101-scikit-learn-afed3bf414a2>.
- Chukamphaeng, N. (2018). ทำไมใครๆก็ใช้ xgboost กันจัง. สืบค้นเมื่อ 8 มกราคม 2565. จาก <https://medium.com/@nutorbitx/ทำไมใครๆก็ใช้-xgboost-กันจัง-a775b53cc1eb>.
- Elena, P. (2021). Predicting the movement direction of omxs30 stock index using xgboost and sentiment analysis. [doctoral dissertation], Blekinge Institute of Technology.
- Forex (2016). Moving average คืออะไรมีวิธีใช้อย่างไร. สืบค้นเมื่อ 10 ธันวาคม 2564. จาก <https://www.forexinthai.com/1302/>
- Narisara, S. (2021). Ethereum คืออะไร. สืบค้นเมื่อ 10 ธันวาคม 2564. จาก <https://thematter.co/futureverse/futureword-ethereum/161241>.
- Oliveira, H. (2021). Bitcoin price movement direction prediction : a comparison between Decision Tree SVM and LSTM. [master's thesis], University Católica Portuguesa.
- Pacharawongsakda, E. (2015). การคัดเลือก feature (feature selection) ด้วยวิธี information gain. สืบค้นเมื่อ 22 ธันวาคม 2564. จาก <https://th.linkedin.com/pulse/การคัดเลือก-feature-selection-ด้วยวิธี-information-gain-pacharawongsakda>.
- Pagon, G. (2018). Confusion matrix เครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนายใน machine learning. สืบค้นเมื่อ 1 มกราคม 2565. จาก <https://medium.com/@pagongatchalee/>

- confusion-matrix-เครื่องมือ สำคัญ ใน การ ประเมิน ผลลัพธ์ ของ การ ทำนาย-ในmachine-learning-fba6e3f9508cc.
- Soumya, S. (2020). Cross validation in time series. สืบค้นเมื่อ 5 ธันวาคม 2564. จาก <https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce48cc>.
- Tefera, L. (2020). k-fold cross-validation. สืบค้นเมื่อ 17 ธันวาคม 2564. จาก <https://medium.com/@lidetsal/k-fold-cross-validation-cf0bbb6fa37b>.
- tradingview (2022). สืบค้นเมื่อ 2 กุมภาพันธ์ 2565. จาก <https://www.tradingview.com/chart/>.
- Witchapong, D. (2018a). รู้จัก decision tree, random forest, และ xgboost — part 1. สืบค้นเมื่อ 22 มกราคม 2565. จาก <https://medium.com/@witchapongdaroontham/รู้จัก-decision-tree-random-forrest-และ-xgboost-part-1-cb49c4ac1315>.
- Witchapong, D. (2018b). เจาะลึก random forest !!! — part 2 of “รู้จัก decision tree, random forest, และ xgboost!!!”. สืบค้นเมื่อ 22 มกราคม 2565. จาก <https://medium.com/@witchapongdaroontham/เจาะ-ลึก-random-forest-part-2-of-รู้จัก-decision-tree-random-forest-และ-xgboost-79b9f41a1c1c>.