การคาดการณ์ดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย โดยใช้มะชีนเลิร์นนิงเทคนิค

ชลรวี หาญลำยวง

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการจัดการมหาบัณฑิต
วิทยาลัยการจัดการ มหาวิทยาลัยมหิดล
พ.ศ. 2562

ลิขสิทธิ์ของมหาวิทยาลัยมหิดล

กิตติกรรมประกาศ

สารนิพนธ์ เรื่อง การคาดการณ์ดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทยโดยใช้มะชีนเลิร์นนิงเทคนิค (SET INDEX FORECASTING THROUGH MACHINE LEARNING TECHNIQUES) สำเร็จลุล่วงและประสบความสำเร็จได้ด้วยดี เนื่องจากได้รับความอนุเคราะห์จากผู้ช่วยศาสตราจารย์ ปิยภัส ธาระวานิช, Ph.D. อาจารย์ที่ปรึกษาสารนิพนธ์ ที่กรุณาสละให้คำปรึกษาและข้อเสนอแนะที่เป็นประโยชน์ต่อการศึกษาวิจัย และรองศาสตราจารย์ ธาตรี จันทรโคลิกา, Ph.D. ที่ได้ให้ข้อเสนอแนะเพิ่มเติมที่เป็นประโยชน์ ขอขอบพระคุณอาจารย์ทุกท่านเป็นอย่างสูง

นอกจากนี้ ผู้วิจัยขอขอบพระคุณคณาจารย์ของวิทยาลัยการจัดการ มหาวิทยาลัยมหิดลทุกท่านที่ได้ให้ความรู้ต่างๆ ซึ่งสามารถนำมาประยุกต์ในสารนิพนธ์ฉบับนี้ รวมถึงกำลังใจและความช่วยเหลือต่างๆ เป็นอย่างดี จากบิดามารดา ครอบครัว และเพื่อนๆ ทุกท่าน ส่งผลให้สารนิพนธ์ฉบับนี้ประสบความสำเร็จลุล่วงไปได้ด้วยดี

ชลรวี หาญลำยวง

สารบัญ

	หน้า
กิตติกรรมประกาศ	**ข**
บทคัดย่อ	**ค**
สารบัญตาราง	**จ**
สารบัญรูปภาพ	**ฉ**
บทที่ 1 บทนำ	**1**
บทที่ 2 ทบทวนวรรณกรรม	**4**
บทที่ 3 วิธีการวิจัย	**7**
3.1 ข้อมูลที่ใช้ในการศึกษา	7
3.2 ตัวแปร	7
3.3 การเตรียมข้อมูล	8
3.4 วิธีการทางสถิติ	10
บทที่ 4 ผลการวิจัย	**11**
4.1 Decision Tree	11
4.2 Artificial Neural Networks	11
4.3 Support Vector Machine	12
4.4 การจำลองการลงทุน	15
บทที่ 5 สรุปผลการศึกษา	**18**
บรรณานุกรม	**20**
ภาคผนวก	**22**
ประวัติผู้วิจัย	**24**

สารบัญตาราง

ตาราง		หน้า
1	งานวิจัยสำคัญที่ใช้ machine learning algorithms ในการคาดการณ์ราคาหุ้นหรือดัชนีตลาดหลักทรัพย์โดยข้อมูลที่นำมาใช้คาดการณ์ที่แตกต่างกัน	6
2	แสดงค่าเฉลี่ย (mean), ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) และค่าทางสถิติในรูปแบบ confusion matrix	13
3	แสดงความแม่นยำ (accuracy) ของแต่ละช่วงเวลาของโมเดลที่ดีที่สุดในแต่ละอัลกอริทึม	14
4	ความแม่นยำเฉลี่ยเทียบตามช่วงเวลา	15
5	การจำลองการลงทุนจากการคาดการณ์ของโมเดลและเกณฑ์มาตรฐาน	17

สารบัญรูปภาพ

ภาพ		หน้า
1	แสดงแผนภาพ Sliding Windows	9
2	แสดงแผนภาพ Expanding Windows	9
3	แสดงกราฟ DT Sliding	14
4	แสดงกราฟ DT Expanding	14
5	แสดงกราฟ ANN Sliding	14
6	แสดงกราฟ ANN Expanding	14
7	แสดงกราฟ SVM Expanding	15
8	แสดงกราฟ SVM Sliding	15

การคาดการณ์ดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทยโดยใช้มะชีนเลิร์นนิงเทคนิค
SET INDEX FORECASTING THROUGH MACHINE LEARNING TECHNIQUES

ชลรวี หาญลำยวง 5850067

กจ.ม.

คณะกรรมการที่ปรึกษาสารนิพนธ์ : ผู้ช่วยศาสตราจารย์ปิยภัสร ธาระวานิช, Ph.D., รองศาสตราจารย์ธาตรี จันทรโคลิกา, Ph.D., ผู้ช่วยศาสตราจารย์ชนินทร์ อยู่เพชร, Ph.D.

บทคัดย่อ

งานศึกษานี้มีวัตถุประสงค์เพื่อคาดการณ์การเปลี่ยนแปลงของดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET Index) ในวันทำการถัดไปโดยใช้เทคนิค machine learning โดยอัลกอริทึม (algorithms) ที่เลือกใช้เป็นแบบ classification ได้แก่ Decision Tree, Artificial Neural Networks และ Support Vector Machine ผู้แต่งใช้ประโยชน์จากความสัมพันธ์ระยะสั้นระหว่างการเปลี่ยนแปลงของดัชนีราคาหุ้นตลาดหลักทรัพย์จากทั่วโลก (global stock market indices) อัตราแลกเปลี่ยน (exchange rates) และราคาสินค้าโภคภัณฑ์ (commodities) ข้อมูลทั้งหมดจะถูกแบ่งสองกลุ่มคือ ชุดข้อมูลที่ใช้ในการสร้างโมเดล (training sets) และชุดข้อมูลที่ใช้ทดสอบ (testing sets) เพื่อวัดประสิทธิภาพของแต่ละโมเดลโดยใช้ความแม่นยำ (accuracy) การคัดเลือกฟีเจอร์ใช้วิธีการฟิลเตอร์ (filter approach) และ forward feature selection

ผลลัพธ์ของการศึกษาพบว่าโมเดลที่คาดการณ์ได้ความแม่นยำ (accuracy) สูงสุดได้เท่ากับ 63.2% โมเดลที่คาดการณ์ได้ความแม่นยำสูงที่สุดของแต่ละอัลกอริทึมมีความแม่นยำใกล้เคียงกันโดยค่าความแม่นยำเฉลี่ยอยู่ที่ 61.63% และส่วนเบี่ยงเบนมาตรฐาน (standard deviation) อยู่ที่ 1.08% สุดท้ายได้นำโมเดลที่มีความแม่นยำที่สุดสร้างแบบจำลองการซื้อขายอย่างง่ายเพื่อศึกษาประสิทธิภาพของอัลกอริธึมเทียบกับเกณฑ์มาตรฐาน (benchmarks)

คำสำคัญ : SET INDEX/ MACHINE LEARNING/ DECISION TREE/ ARTIFICIAL NEURAL NETWORKS/ SUPPORT VECTOR MACHINE

24 หน้า

บทที่ 1
บทนำ

การคาดการณ์แนวโน้มของราคาหลักทรัพย์หรือดัชนีตลาดหลักทรัพย์เป็นหัวข้อที่ได้รับความสนใจและมีการศึกษากันอย่างกว้างขวางต่อเนื่องมาเป็นระยะเวลานาน โดยทั่วไปแล้วมีสองวิธีที่นิยมนำมาใช้ในการวิเคราะห์เพื่อการคาดการณ์แนวโน้มของหลักทรัพย์ คือ การวิเคราะห์พื้นฐานซึ่งอาศัยผลการดำเนินงานที่ผ่านมาและงบการเงินของบริษัทเพื่อใช้การคาดการณ์ อีกวิธีคือการวิเคราะห์ทางเทคนิค โดยพยายามทำความเข้าใจเกี่ยวกับการเปลี่ยนแปลงและรูปแบบการเคลื่อนไหวของราคาในอดีตเพื่อคาดการณ์ราคาในอนาคต จะเห็นว่าข้อมูลส่วนใหญ่ที่ถูกนำมาใช้ในการวิเคราะห์หรือสร้างโมเดลเป็นข้อมูลของแต่บริษัทหรือข้อมูลของดัชนีตลาดหลักทรัพย์ภายในประเทศหรือภายในตลาดเดียวกัน แต่ข้อมูลสำคัญจากภายนอกจากทั่วโลกก็อาจมีความสำคัญเช่นเดียวกัน โดยเฉพาะอย่างยิ่งในโลกยุคปัจจุบัน เทคโนโลยีและการสื่อสารที่รวดเร็วและทันสมัยมากขึ้น ทำให้ข้อมูลจากตลาดทางการเงินและเศรษฐกิจทั่วโลกแทบจะถูกเชื่อมต่อกันในทันที ดังนั้นข่าวความเคลื่อนไหวในวงการการเงิน การเปลี่ยนแปลงของดัชนีตลาดหลักทรัพย์ทั่วโลก สภาวะเศรษฐกิจหรือนโยบายของประเทศที่มีขนาดเศรษฐกิจใหญ่ อาจมีแนวโน้มที่จะมีผลกระทบต่อสินทรัพย์ทางการเงินทั่วโลกรวมถึงในประเทศไทยด้วย อาทิเช่น วิกฤติทางการเงินของสหรัฐอเมริกาส่งผลกระทบต่อเศรษฐกิจและดัชนีตลาดหลักทรัพย์ทั่วโลก วิกฤติหนี้ที่เกิดขึ้นในกรีซทำให้ดัชนีตลาดหลักทรัพย์ตกต่ำในหลายประเทศ ข้อมูลทางเศรษฐกิจการเมืองหรือเหตุการณ์สำคัญในประเทศใดประเทศหนึ่งอาจส่งผลให้เกิดความผันผวนต่อประเทศอื่นทั่วทั้งโลกได้เช่นเดียวกัน เช่น ความกังวลการเลือกตั้งอเมริกาหรือกรณี Brexit เป็นต้น นอกเหนือจากดัชนีตลาดหลักทรัพย์แล้วผลิตภัณฑ์ทางการเงินอื่น อาทิ สินค้าโภคภัณฑ์และอัตราแลกเปลี่ยนสกุลเงินต่างประเทศอาจเป็นข้อมูลหนึ่งที่มีความเชื่อมโยงกับบริษัทหรือดัชนีตลาดหลักทรัพย์ในประเทศ เช่น การแข็งค่าของสกุลเงินต่างประเทศอาจจะส่งผลดีให้ผู้ส่งออก หรือการลดลงอย่างรุนแรงของราคาน้ำมันในช่วงปี 2015-2016 ก็ส่งผลกระทบกับธุรกิจพลังงานซึ่งเป็นกลุ่มบริษัทสำคัญของตลาดหลักทรัพย์แห่งประเทศไทย ปัจจุบันแทบไม่มีตลาดหลักทรัพย์ใดที่เป็นอิสระและไม่ได้รับหรือไม่ส่งผลกระทบกับตลาดอื่น ดังนั้นเราจึงพยายามใช้ดัชนีตลาดหลักทรัพย์สำคัญของโลกจากหลากหลายทวีป อัตรา

แลกเปลี่ยนเทียบไทยบาทรวมถึงราคาสินค้าโภคภัณฑ์เป็นปัจจัยสำหรับการคาดการณ์ดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย โดยเฉพาะอย่างยิ่งกับความสัมพันธ์ระยะสั้นระหว่างการเปลี่ยนแปลงของของสินทรัพย์ทางการเงินในวันทำการก่อนหน้าควรให้ข้อมูลที่มีความสำคัญต่อการเปลี่ยนแปลงของดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET Index) เนื่องจากการเปลี่ยนแปลงดังกล่าวได้สะท้อนถึงความเชื่อมั่นในตลาดที่เกิดขึ้นล่าสุด

Machine learning algorithms ได้รับความสนใจมากขึ้นในระยะเวลาหลายปีที่ผ่านมา เนื่องจากประสิทธิภาพของคอมพิวเตอร์ที่มีประสิทธิภาพมากขึ้นและการพัฒนาของเทคนิคต่างๆ จึงมีการใช้กันอย่างกว้างขวางในหลายสาขาวิชารวมถึงในสาขาการเงินด้วยเช่นกัน เช่น มีการนำมาประยุกต์ใช้ในการวิเคราะห์หลักทรัพย์ทั้งการวิเคราะห์พื้นฐานหรือการวิเคราะห์ทางเทคนิครวมถึงแบบมีการนำมาใช้ควบคู่กัน (hybrid) โดยมีรูปแบบการใช้งานที่หลากหลาย เช่น การสร้างเป็นระบบซื้อขายอัตโนมัติ (trading system) เพื่อช่วยในการคาดการณ์ในสถานการณ์หรือเงื่อนไขต่างๆ แม้ว่าไม่ใช่เรื่องง่ายที่จะใช้โมเดลเพียงอย่างเดียวมาใช้แทนที่ความเชี่ยวชาญของนักลงทุนผู้มีประสบการณ์ แต่ก็สามารถนำมาใช้ควบคู่ร่วมกับความเชี่ยวชาญของนักลงทุนเพื่อช่วยในการตัดสินใจ ดังนั้นโมเดลที่สามารถทำนายอย่างแม่นยำสามารถส่งผลโดยตรงต่อผลกำไรที่สูงขึ้นสำหรับการลงทุนได้ งานศึกษานี้พัฒนาโมเดลซึ่งเน้นด้านความแม่นยำ (accuracy) เพื่อการคาดการณ์การเปลี่ยนแปลงดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET Index) ในวันทำการถัดไป

งานวิจัยนี้ใช้อัลกอริธึม (algorithms) แบบ classification คือ Decision Tree, Artificial Neural Networks และ Support Vector Machine ในการคาดการณ์การเปลี่ยนแปลงของดัชนีตลาดหลักทรัพย์ประเทศไทย (SET index) รายวันโดยใช้ข้อมูลรายวันของดัชนีราคาหุ้นตลาดหลักทรัพย์จากทั่วโลก (global stock market indices) อัตราแลกเปลี่ยนเทียบเงินบาท (exchange rates) และราคาสินค้าโภคภัณฑ์ (commodities) ที่ปิดทำการก่อนตลาดหลักทรัพย์แห่งประเทศไทยเปิดทำการการ ระหว่างวันที่ 2 สิงหาคม 2550 ถึงวันที่ 2 สิงหาคม 2561

ผลการศึกษาพบว่าโมเดลที่คาดการณ์ได้ความแม่นยำ (accuracy) สูงสุดได้เท่ากับ 63.2% และในการสร้างแบบจำลองการซื้อขายอย่างง่ายเพื่อศึกษาประสิทธิภาพของอัลกอริธึมเทียบกับ 2 เกณฑ์มาตรฐาน (benchmarks) โดยแบ่งการซื้อขายเป็นช่วงเวลาละ 50 วัน 10 ช่วงเวลา เกณฑ์มาตรฐาน 1 (benchmark 1) คือ กลยุทธ์ซื้อตอนเริ่มต้นและขายตอนสิ้นสุดของแต่ละช่วงเวลาเพื่อเป็นตัวชี้วันสภาวะตลาดในแต่ละช่วงเวลา และ เกณฑ์มาตรฐาน 2 (benchmark 2) คือ กลยุทธ์ซื้อขายรายวันโดยจะซื้อหากวันนี้ดัชนีตลาดหลักทรัพย์สูงกว่าเมื่อวานและถือเงินสดหากวันนี้ดัชนีตลาดหลักทรัพย์ต่ำกว่าวันทำการ

ที่ผ่านมา เทียบกับการซื้อขายโดยใช้ผลที่ได้จากโมเดลที่มีความแม่นยำสูงที่สุด โดยกลยุทธ์คือ ซื้อหากโมเดลคาดการณ์ว่าดัชนีตลาดหลักทรัพย์ในวันทำการถัดไปจะเป็นบวก (Positive) และถือเงินสด (ไม่ซื้อ) หากโมเดลคาดการณ์ว่าจะเป็นลบ (Negative) สมมติให้ดัชนีตลาดหลักทรัพย์ประเทศไทย (SET index) ถือเป็น 1 หน่วยการลงทุนและในระหว่างการซื้อขายไม่มีปันผล ค่าธรรมเนียมการซื้อขาย ภาษี และไม่มีธุรกรรมขายชอร์ต (short sale) หากมีการซื้อหมายความว่าซื้อในราคาปิดวันนี้และจะขายในราคาปิดของวันทำการถัดไป ผลของแบบจำลองปรากฏว่าผลตอบแทนหากใช้โมเดลในการคาดการณ์ได้ผลตอบแทนมากกว่าทั้งสองเกณฑ์มาตรฐาน 8 ใน 10 ช่วงเวลาและผลตอบแทนจากการใช้แบบจำลองตลอด 10 ช่วงเวลาไม่มีช่วงเวลาใดที่ขาดทุน

บทที่ 2
ทบทวนวรรณกรรม

มีหลายงานศึกษาการที่น่าสนใจในการพยายามนำข้อมูลหลากหลายประเภทมาใช้เพื่อการคาดการณ์แนวโน้มของหลักทรัพย์หรือดัชนีตลาดหลักทรัพย์โดยใช้ machine learning algorithms ตัวอย่างการศึกษาเช่น

Shah (2007) ใช้ข้อมูลทางด้านเทคนิคอย่างเช่น Exponential Moving Average (EMA) ในการศึกษาราคาหุ้นของกูเกิ้ล (GOOG) และ ยาฮู (YHOO) ในประเทศสหรัฐอเมริกาด้วยวิธีการที่หลากหลายเช่น linear regression และ Support Vector Machine (SVM) เป็นต้น

Huang, Nakamori, and Wan (2005) ศึกษาการเปลี่ยนแปลงของทิศทางรายสัปดาห์ของดัชนี NIKKEI 225 โดยพบว่า Support Vector Machine (SVM) มีประสิทธิภาพสูงกว่าเมื่อเทียบกับอัลกอริทึมอื่นๆ

Kara, Boyacioglu, and Baykan (2011) ศึกษาการเปลี่ยนแปลงของทิศทางรายวันของดัชนี Istanbul Stock Exchange (ISE) 100 ผลลัพท์ที่ได้ Artificial Neural Networks (ANN) มีประสิทธิภาพเฉลี่ยสูงกว่า Support Vector Machine (SVM) ที่ 75.74% และ 71.52% ตามลำดับ

Olson and Mossman (2003) มีการใช้ข้อมูลของอัตราส่วนทางบัญชีของบริษัทในประเทศแคนนาดา พบว่าการใช้ classification models มีประสิทธิภาพมากว่า estimation models

Zhu, Wang, Xu, and Li (2008) ศึกษาการใช้ข้อมูลทางเทคนิคอย่างผลตอบแทนของหุ้นและใช้ข้อมูลปริมาณการซื้อขายมาเพิ่มประสิทธิภาพของ Artificial Neural Networks (ANN) ในการคาดการณ์ดัชนี NASDAQ, DJIA และ STI โดยผลคือการใช้ข้อมูลของปริมาณการซื้อขายสามารถเพิ่มประสิทธิภาพของโมเดลในการคาดการณ์ดัชนีในระยะกลางและระยะยาวได้

Tsai and Wang (2009) ศึกษาโดยการใช้ข้อมูลทั้งพื้นฐานและทางด้านเทคนิคร่วมกันในการคาดการณ์ราคาหุ้นในอุตสาหกรรมอิเล็กทรอนิกส์ในไต้หวันโดยใช้ Decision Tree (DT) และ Artificial Neural Networks (ANN) ทั้งยังมีการสร้างโมเดลโดยใช้ทั้งสองอัลกอริทึมเพื่อร่วมกันในการคาดการณ์ซึ่งให้ผลลัพท์ที่ดีกว่า ค่าความถูกต้อง (accuracy) ที่ 77%

Shen, Jiang, and Zhang (2012)

ศึกษาโดยใช้ข้อมูลจากดัชนีตลาดหลักทรัพย์จากทั่วโลกและผลิตภัณฑ์ทางการเงินต่างๆ ในการคาดการณ์ดัชนี NASDAQ, S&P 500 และ DJIA ในวันทำการถัดไปได้ค่าความถูกต้อง (accuracy) 74.4%, 76% และ 77.6% ตามลำดับ โดยใช้ Support Vector Machine (SVM) และ reinforcement learning

สรุปวิธีการแก้ปัญหาที่เลือกใช้ (Algorithms)

2.1 Decision Tree

Decision Tree (DT) Quinlan (1986) เป็นหนึ่งในโมเดล classification ที่เป็นเทคนิคที่เป็นที่นิยมเพราะสามารถเข้าใจได้ง่ายและใช้แก้ปัญหาได้หลากหลาย Decision Tree มีโครงสร้างเหมือนต้นไม้ที่แตกกิ่งออกไป โดยจะมีการสร้างจุดแตกกิ่ง (node) และ กิ่งสาขา (branch) ซึ่งแต่ละกิ่งถูกสร้างสถานการณ์ที่แตกต่างกันออกไป โดยแต่ละ node ถูกตัดสินใจโดยใช้ฟีเจอร์ (features) ในการจำแนก วิธีการนี้เป็นการประมาณค่าแบบไม่ต่อเนื่อง หากใช้ accuracy เป็นตัวชี้วัดในชุดข้อมูลที่ใช้ในการสร้างโมเดล (training sets) มักเกิดปัญหา accuracy ลดลงเมื่อนำโมเดลที่สร้างได้มาใช้กับชุดข้อมูลที่ใช้ทดสอบ (testing sets) ปัญหานี้เกิดขึ้นจากสิ่งที่เรียกว่า overfitting

2.2 Artificial Neural Networks

Artificial Neural Networks (ANN) Rosenblatt (1958) คือเทคนิคที่พัฒนาโดยการจำลองระบบประสาทของมนุษย์เป็นโครงข่ายประสาทเทียมที่จำลองมีความสามารถในการคาดการณ์ได้ดีจากฐานข้อมูลจำนวนมากและใช้การเรียนรู้จากชุดข้อมูลและกลับมาเรียนรู้ใหม่ด้วยวิธี Back propagation สำหรับ hidden layer เป็นที่ถกเถียงถึงจำนวน node ในการสร้างโมเดล โดยเฉพาะหากใช้ node มากเกินกว่าจำนวน feature ที่เราใช้ในการสร้างโมเดล มีงานศึกษาพบกว่าการใช้ 1 hidden layer สามารถทำให้ผลลัพธ์ของการคาดการณ์แม่นยำขึ้น Swales Jr and Yoon (1992)

2.2 Support Vector Machine

Support Vector Machine (SVM) Cortes and Vapnik (1995) เป็น machine learning algorithm ที่พยายามสร้าง hyperplane ที่สามารถจำแนกข้อมูลได้ดีที่สุด โดยมี margin สูงที่สุด และใช้ kernel function เป็นตัวกำหนดวิธีการในการสร้าง hyperplane

ตารางที่ 1 งานวิจัยสำคัญที่ใช้ machine learning algorithms ในการคาดการณ์ราคาหุ้นหรือดัชนีตลาดหลักทรัพย์โดยข้อมูลที่นำมาใช้คาดการณ์ที่แตกต่างกัน

Research	Target	Algorithm	Predictor
Qian and Rasheed (2007)	DJIA index	ANN k-NN DT	Technical analysis
Tsai and Wang (2009)	Stock price in Taiwan	ANN DT	Fundamental and Technical analysis
Patel, Shah, Thakkar, and Kotecha (2015)	Stock price and stock index in India	ANN, SVM random forest naive-Bayes	Technical analysis
Kara et al. (2011)	Istanbul Stock Exchange	ANN SVM	Technical analysis
Shen et al. (2012)	Stock indices in USA	SVM	Global financial markets
Shah (2007)	Google Inc. and Yahoo Inc.	Linear Regression SVM	Technical analysis
Huang et al. (2005)	NIKKEI 225 index	SVM	Macroeconomic variables
Olson and Mossman (2003)	Stock return in Canada	ANN	Fundamental analysis

บทที่ 3
วิธีการวิจัย

3.1 ข้อมูลที่ใช้ในการศึกษา (Data)

การศึกษานี้ใช้ข้อมูลในการคำนวณตัวแปรต่างๆ ทั้งหมดจากฐานข้อมูลของ Thomson Reuters ข้อมูลที่ใช้เป็นข้อมูลรายวันของดัชนีราคาหุ้นตลาดหลักทรัพย์จากทั่วโลก (global stock market indices) อัตราแลกเปลี่ยนเทียบเงินบาท (exchange rates) และราคาสินค้าโภคภัณฑ์ (commodities) ระหว่างวันที่ 2 สิงหาคม 2550 ถึงวันที่ 2 สิงหาคม 2561 เพื่อใช้ในการสร้างโมเดลจาก Decision Tree, Artificial Neural Networks และ Support Vector Machine สำหรับตัดสินใจในการคาดการณ์ทิศทางการเปลี่ยนแปลงในวันทำการถัดไปของดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET Index)

3.2 ตัวแปร (Variables)

3.2.1 ตัวแปรอิสระ (Independent variables)

การเปลี่ยนแปลงของราคาปิดเทียบกับวันทำการก่อนหน้าโดยใช้การเปลี่ยนแปลงของดัชนีราคาหุ้นตลาดหลักทรัพย์จากทั่วโลก (global stock market indices) อัตราแลกเปลี่ยนเทียบเงินบาท (exchange rates) และราคาสินค้าโภคภัณฑ์ (commodities) ที่ปิดทำการก่อนตลาดหลักทรัพย์แห่งประเทศไทยจะเปิดทำการ ตัวแปรอิสระที่ถูกแปลงแล้วจะถูกเรียกว่าฟีเจอร์ (features)

3.2.1.1 ดัชนีราคาหุ้นตลาดหลักทรัพย์จากทั่วโลก (global stock market indices)

อเมริกาเหนือ ได้แก่ Dow Jones Industrial Average (USA), S&P 500 (USA) และ NASDAQ (USA)

อเมริกาใต้ ได้แก่ BOVESA (Brazil) และ MERVAL (Argentina)

ยุโรป ได้แก่ FTSE 100 (U.K.), DAX 30 (Germany) และ CAC 40 (France)

เอเชีย ได้แก่ Shanghai (China), Hang Seng (Hong Kong), TOPIX (Japan), Nikkei 225 (Japan) KOSPI (Korea) และ S&P/ASX 200 (Australia)

เอเชียตะวันออกเฉียงใต้ ได้แก่ BURSA (Malaysia), IDX (Indonesia) และ PSEi (Philippines)

3.2.1.2 อัตราแลกเปลี่ยนเทียบเงินบาท (exchange rates)

United States Dollar (USD), Euro (EUR), Pound Sterling (UKP), Australian Dollar (AUD) และ Japanese Yen (JPY)

3.2.1.3 ราคาสินค้าโภคภัณฑ์ (commodities)

Crude oil, Gold, Silver และ Platinum

3.2.2 ตัวแปรตาม (Dependent variable)

การเปลี่ยนแปลงรายวันของดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET Index) ในลักษณะของทิศทางว่าขึ้นหรือลงไม่ใช่ในรูปเปอร์เซ็นต์ ตัวแปรตามจะถูกแทนค่าด้วย “Positive” หากราคาปิดของดัชนีสูงกว่าราคาปิดของวันทำการก่อนหน้าและถูกแทนค่าด้วย “Negative” หากราคาปิดของดัชนีต่ำกว่าราคาปิดของวันทำการก่อนหน้า

3.3 การเตรียมข้อมูล (Data preparation)

ข้อมูลที่ใช้ในการศึกษานี้เรียงลำดับข้อมูลเรียงเป็นรายวันแบบ time series โดยตัวแปรอิสระแต่ละสินทรัพย์จะใช้ 1 วันทำการย้อนหลังและต้องปิดทำการก่อนตลาดหลักทรัพย์แห่งประเทศไทยจะเปิดทำการ เพื่อใช้เป็นฟีเจอร์ (features) ในการคาดการณ์ทิศทางของตัวแปรตามในวันทำการนั้นๆ ฟีเจอร์ (features) และตัวแปรตามทั้งหมดจะถูกแบ่งเป็นชุดข้อมูลชุดข้อมูลที่ใช้ในการสร้างโมเดล (training sets) และชุดข้อมูลที่ใช้ทดสอบ (testing sets) โดยใช้วิธีการแบ่ง 2 แบบ คือ sliding windows และ expanding windows ชุดข้อมูลทดสอบจะถูกแบ่งออกเป็นชุดที่มีความต่อเนื่องกัน 10 ชุดข้อมูล ชุดข้อมูลละ 50 วัน ดังนั้นจะได้ชุดทดสอบที่แตกต่างกัน 10 ชุดซึ่งคิดเป็นจำนวนข้อมูลประมาณ 17.42% ของข้อมูลทั้งหมด

3.3.1 Sliding windows

ภาพที่ 1 แสดงแผนภาพ Sliding Windows

การแบ่งชุดข้อมูลแบบ sliding windows คือการแบ่งชุดข้อมูลที่ใช้ในการสร้างโมเดล (training sets) ให้มีจำนวนคงที่ เมื่อมีการเรียงข้อมูลแบบ time series ชุดข้อมูลที่ใช้ในการสร้างโมเดลจะถูกเพิ่มด้วยชุดข้อมูลที่ใช้ทดสอบในชุดก่อนหน้าและนำข้อมูลที่เก่าที่สุดออกไปเป็นจำนวนเท่ากับข้อมูลที่เพิ่มมา ข้อดีคือ ชุดข้อมูลที่ใช้ในการสร้างโมเดล (training sets) ข้อมูลเก่าที่อาจจะมีความสำคัญในการสร้างโมเดลน้อยกว่าข้อมูลใหม่ หรือ หมดความสำคัญแล้วจะไม่ถูกนำมาใช้ในการสร้างโมเดล

3.3.2 Expanding windows

ภาพที่ 2 แสดงแผนภาพ Expanding Windows

การแบ่งชุดข้อมูลแบบ expanding windows คือการแบ่งชุดข้อมูลที่ใช้ในการสร้างโมเดล (training sets) ให้มีจำนวนเพิ่มขึ้นด้วยชุดข้อมูลที่ใช้ทดสอบในชุดก่อนหน้า ข้อดีคือ ชุดข้อมูลที่ใช้ในการสร้างโมเดล (training sets) จะมีปริมาณมากขึ้นซึ่งจะส่งผลดีต่อการสร้างบางโมเดลที่ต้องการข้อมูลในการเรียนรู้จำนวนมาก

3.4 วิธีการทางสถิติ (Model and Estimation Method)

3.4.1 การคัดเลือกฟีเจอร์ (features) ด้วยวิธีการแบบฟิลเตอร์ (Filter approach)

ฟีเจอร์ (features) คือตัวแปรที่ใช้ในการสร้างโมเดล ในการคัดเลือกตัวแปรอิสระเพื่อให้มีความเหมาะสมกับวิธีการและเครื่องมือในการสร้างโมเดล เพื่อหาความสัมพันธ์ระหว่างฟีเจอร์ (features) และตัวแปรตามโดยใช้ค่าสหสัมพันธ์ (correlation) ที่อยู่ในรูปค่าสัมบูรณ์ (absolute value) และเรียงลำดับตามค่าสหสัมพันธ์ (correlation) ของแต่ละตัวแปรอิสระ เพื่อคัดเลือกตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามสูงสุด 10 อันดับแรก

3.4.2 การคัดเลือกฟีเจอร์ (features) ด้วยวิธีแรปเปอร์ (wrapper approach) และการสร้างโมเดล

Forward feature selection เป็นการคัดเลือกฟีเจอร์ (features) โดยสร้างแต่ละโมเดล (classification model) ด้วยเซตของฟีเจอร์ที่กำหนดและวัดประสิทธิภาพของโมเดล ในการศึกษานี้ใช้ความแม่นยำ (accuracy) ในการพิจารณาประสิทธิภาพของโมเดล เซตของฟีเจอร์ที่กำหนดเริ่มจากใช้ฟีเจอร์แต่ละฟีเจอร์ในการสร้างโมเดลและวัดประสิทธิภาพแล้วเก็บฟีเจอร์ที่มีความแม่นยำ (accuracy) สูงที่สุดไว้และเลือกฟีเจอร์อื่นเพิ่มทีละฟีเจอร์เพื่อสร้างโมเดลและวัดประสิทธิภาพต่อจนกระทั่งจนถึงจำนวนฟีเจอร์ที่กำหนดหรือความแม่นยำของโมเดลไม่เพิ่มขึ้นจึงหยุดและเลือกเซตของฟีเจอร์ที่ทำให้โมเดลมีประสิทธิภาพที่สุดมาใช้กับชุดข้อมูลที่ใช้ทดสอบ (testing sets)

$$Accuracy = \frac{\text{True Positive} + \text{True Negative}}{Total}$$

ในการสร้างทุกโมเดล (classification model) จากทั้ง 3 วิธีการ (algorithms) คือ Decision Tree, Artificial Neural Networks และ Support Vector Machine จะใช้การแบ่งชุดข้อมูลที่ใช้ในการสร้างโมเดล (training sets) โดยวิธี 5-fold cross validation คือการแบ่งข้อมูลออกเป็น 5 ส่วนโดยแต่ละส่วนมีจำนวนข้อมูลเท่ากัน ใช้ข้อมูลหนึ่งส่วนเป็นตัวทดสอบประสิทธิภาพของโมเดล ทำวนจนครบ 5 ส่วนที่แบ่งไว้ เพื่อหลีกเลี่ยงปัญหาที่โมเดลจดจำรูปแบบของข้อมูลที่ใช้ในการสร้างโมเดล (overfitting)

บทที่ 4
ผลการวิจัย

การประเมินประสิทธิภาพของแต่ละโมเดล สิ่งสนใจที่สุดคือความแม่นยำ (accuracy) ของแต่ละโมเดลและส่วนประกอบคือ ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) แสดงผลออกมาในรูปของ confusion matrix และ receiver operating characteristic (ROC) curve เพื่อร่วมประเมิน โดยค่าความมั่นใจ (confidence) เริ่มต้นถูกกำหนดไว้ที่ 0.5

4.1 Decision Tree

ค่าพารามิเตอร์ที่ทำให้โมเดลมีความแม่นยำที่สุด คือ minimal gain เท่ากับ 10-3 และ confidence เท่ากับ 10-7 ฟีเจอร์ที่ถูกคัดเลือกมาได้แก่ NASDAQ (USA), S&P 500 (USA), MERVAL (Argentina) และ DAX 30 (Germany) การแบ่งชุดข้อมูลแบบ sliding windows ได้ผลลัพธ์คือความแม่นยำเฉลี่ย 63.2% ส่วนการแบ่งชุดข้อมูลแบบ expanding windows ได้ผลลัพธ์คือความแม่นยำเฉลี่ย 61.2%

4.2 Artificial Neural Networks

ค่าพารามิเตอร์ที่ทำให้โมเดลมีความแม่นยำที่สุด คือ momentum เท่ากับ 0.1 และ learning rate เท่ากับ 0.3 ฟีเจอร์ที่ถูกคัดเลือกมาได้แก่ Dow Jones Industrial Average (USA), S&P 500 (USA) และ MERVAL (Argentina) การแบ่งชุดข้อมูลแบบ sliding windows ได้ผลลัพธ์คือความแม่นยำเฉลี่ย 60.2% ส่วนการแบ่งชุดข้อมูลแบบ expanding windows ได้ผลลัพธ์คือความแม่นยำเฉลี่ย 61.2%

4.3 Support Vector Machine

ค่าพารามิเตอร์ที่ทำให้โมเดลมีความแม่นยำที่สุด คือ gamma เท่ากับ 103 และค่า C เท่ากับ 0.1 kernel ใช้เป็น Radial Basis Function (RBF) ฟีเจอร์ที่ถูกคัดเลือกมาได้แก่ S&P 500 (USA), MERVAL (Argentina) และ FTSE 100 (U.K.) การแบ่งชุดข้อมูลแบบ sliding windows ได้ผลลัพธ์คือความแม่นยำเฉลี่ย 61.4% ส่วนการแบ่งชุดข้อมูลแบบ expanding windows ได้ผลลัพธ์คือความแม่นยำเฉลี่ย 62.6%

ในตารางที่ 2 แสดงค่าเฉลี่ย (mean), ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) และ confusion matrix ของโมเดลแสดงผลของโมเดลที่ดีที่สุดในแต่ละอัลกอริทึมตามการแบ่งชุดที่มีความต่อเนื่องกัน 10 ชุดข้อมูล ชุดข้อมูลละ 50 วัน รวม 500 วัน

- Accuracy เป็นค่าเฉลี่ยของ 10 ชุดข้อมูล เพื่อแสดงให้เห็นว่าโดยรวมแล้วโมเดลมีความสามารถคาดการณ์มากน้อยเพียงใด

- Standard Deviation คือ ส่วนเบี่ยงเบนมาตรฐานของ accuracy ของชุดข้อมูลทั้ง 10 ชุดข้อมูล

จากผลที่ได้โมเดล Decision Tree ที่มีการแบ่งข้อมูลแบบ sliding windows มี accuracy สูงที่สุด อยู่ที่ 63.20% และยังมีส่วนเบี่ยงเบนมาตรฐานต่ำที่สุดที่ 4.24% แต่หากมองโดยรวมแล้ว accuracy เฉลี่ยจากทุกอัลกอริธึมอยู่ในช่วง 60.20% - 63.20% แสดงให้เห็นว่าด้วยการใช้ ฟีเจอร์, ปริมาณข้อมูล และการปรับแต่ง (tuning) ค่าพารามิเตอร์ตามการศึกษานี้ แต่ละอัลกอริทึมที่ใช้ให้ค่าความถูกต้องใกล้เคียงกัน

true positive คือ โมเดลคาดการณ์ว่าขึ้นแล้ว set index ขึ้นจริง

true negative คือ โมเดลคาดการณ์ว่าลงแล้ว set index ลงจริง

false positive คือ โมเดลคาดการณ์ว่าขึ้นแต่ set index ลง

false negative คือ โมเดลคาดการณ์ว่าลงแต่ set index ขึ้น

การใช้ accuracy ในการวัดผลเป็นการบอกประสิทธิภาพของโมเดลในภาพรวมซึ่งจะเห็นความใกล้เคียงกันของผลลัพท์ แต่หากพิจารณาจาก confusion matrix (ตารางที่ 2) จะเห็นว่ามีจุดสังเกตุที่เด่นชัดของความสามารถของแต่ละโมเดลและการแบ่งชุดข้อมูลดังนี้

Decision Tree และ Artificial Neural Networks คาดการณ์เป็น Positive มากกว่า Negative เล็กน้อยในทุกการแบ่งชุดข้อมูล แต่ Support Vector Machine มีแนวโน้มในการคาดการณ์เป็น Positive มากกว่า Negative มากกว่าอย่างชัดเจน ซึ่งส่งผลให้มี Recall (True Positive Rate) อัตราส่วนที่โมเดล

สามารถคาดการณ์ได้ว่าขึ้นอย่างถูกต้องต่อจำนวน set index ที่ขึ้นทั้งหมด มีค่าสูงว่าโมเดลอื่น ในทางกลับกันก็มี True Negative Rate ที่ต่ำกว่าโมเดลอื่น

ในการแบ่งชุดข้อมูลแบบ expanding windows ซึ่งทำให้ชุดข้อมูลที่ใช้ในการสร้างโมเดล (training sets) มีจำนวนเพิ่มมากขึ้นเรื่อยๆ สามารถทำให้โมเดลของ Artificial Neural Networks และ Support Vector Machine ที่มีความสามารถในการเรียนรู้จากข้อมูลจำนวนมาก มีค่าเฉลี่ย accuracy ที่สูงกว่าการแบ่งชุดข้อมูลแบบ sliding windows ที่จำนวนของชุดข้อมูลที่ใช้ในการสร้างโมเดล (training sets) คงที่

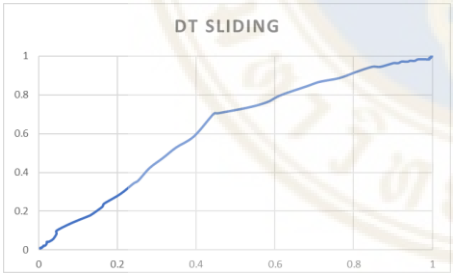
ตารางที่ 2 แสดงค่าเฉลี่ย (mean), ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) และค่าทางสถิติในรูปแบบ confusion matrix

Algorithm	Mean (accuracy)	Standard Deviation	true positives	true negatives	false positives	false negatives
Decision Tree (sliding windows)	63.20%	4.24%	182	134	107	77
Decision Tree (expanding windows)	61.20%	5.59%	161	145	96	98
Artificial Neural Networks (sliding windows)	60.20%	4.47%	163	138	103	96
Artificial Neural Networks (expanding windows)	61.20%	6.27%	159	147	94	100
Support Vector Machine (sliding windows)	61.40%	7.00%	193	114	127	66
Support Vector Machine (expanding windows)	62.60%	5.97%	186	127	114	73

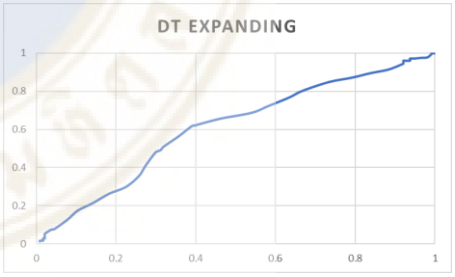
ตารางที่ 3 แสดงความแม่นยำ (accuracy) ของแต่ละช่วงเวลาของโมเดลที่ดีที่สุดในแต่ละอัลกอริทึม

Testing Set	1	2	3	4	5	6	7	8	9	10	Average	SD
DT (Sliding)	60%	64%	68%	60%	60%	72%	60%	60%	62%	66%	63.20%	4.24%
DT (Expanding)	60%	64%	66%	60%	54%	62%	66%	50%	68%	62%	61.20%	5.59%
ANN (Sliding)	54%	58%	66%	60%	56%	64%	64%	54%	62%	64%	60.20%	4.47%
ANN (Expanding)	54%	58%	64%	58%	58%	68%	62%	54%	62%	74%	61.20%	6.27%
SVM (Sliding)	58%	62%	66%	60%	52%	74%	56%	54%	70%	62%	61.40%	7.00%
SVM (Expanding)	58%	62%	68%	60%	56%	70%	58%	56%	72%	66%	62.60%	5.97%

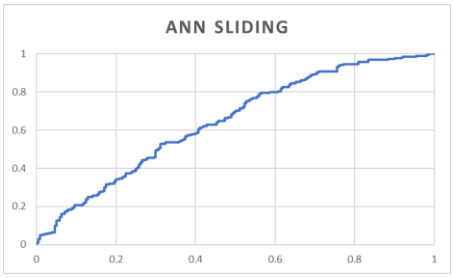
Receiver Operating Characteristic (ROC) curve คือกราฟความสัมพันธ์ระหว่างอัตราส่วนที่โมเดลคาดการณ์ได้ว่าขึ้นอย่างถูกต้องต่อจำนวน set index ที่ขึ้นทั้งหมด หรือ true positive rate (Sensitivity or Recall) ในแกน Y กับอัตราส่วนที่โมเดลคาดการณ์ว่าขึ้นแต่ผิดต่อจำนวน set index ที่ลงทั้งหมด หรือ false positive rate (1 – Specificity) ในแกน X มีประโยชน์ในการใช้เพื่อใช้ในการตัดสินใจในการเลือกค่าความมั่นใจ (confidence) อาทิเช่น หากต้องการเลือก false positive rate น้อยลงเพื่อให้โอกาสที่คาดการณ์ผิดน้อยลงอาจต้องแลกกับ sensitivity ที่น้อยลงคือพลาดบางโอกาสในกรณีที่ set index ขึ้น เป็นต้น



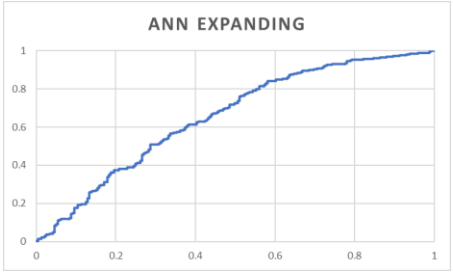
ภาพที่ 3 แสดงกราฟ DT Sliding



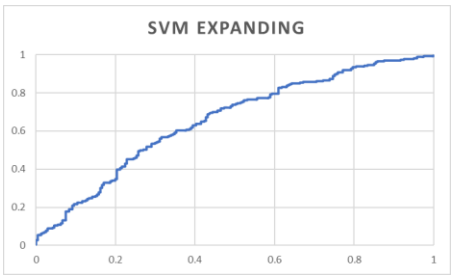
ภาพที่ 4 แสดงกราฟ DT Expanding



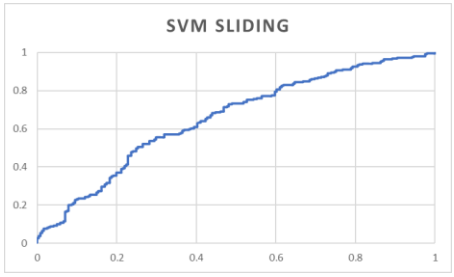
ภาพที่ 5 แสดงกราฟ ANN Sliding



ภาพที่ 6 แสดงกราฟ ANN Expanding



ภาพที่ 7 แสดงกราฟ SVM Expanding



ภาพที่ 8 แสดงกราฟ SVM Sliding

ตารางที่ 4 ความแม่นยำเฉลี่ยเทียบตามช่วงเวลา

Training Day	Accuracy
1 – 10	60.0%
11 – 20	68.2%
21 – 30	55.5%
31 – 40	63.0%
41 - 50	61.5%

4.4 การจำลองการลงทุน (Investing simulation)

ในสร้างแบบจำลองการลงทุนโดยใช้ที่คาดการณ์ได้แม่นยำที่สุดในการคาดการณ์ดัชนีตลาดหลักทรัพย์ประเทศไทยในวันทำการถัดไปเปรียบเทียบกับ 2 เกณฑ์มาตรฐาน (benchmarks) โดยในการจำลองจะแบ่งการลงทุนเป็น 10 ช่วงเวลา ช่วงเวลาละ 50 วันทำการ ดังนั้นมีโอกาสให้โมเดลทำนาย 50 ครั้งต่อหนึ่งช่วงเวลา โดยประเมินผลเป็นกำไร/ขาดทุนทุกๆ 1 ช่วงเวลา สมมติว่าดัชนีตลาดหลักทรัพย์ประเทศไทย (SET index) ถือเป็น 1 หน่วยการลงทุนและในระหว่างการซื้อขายไม่มีปันผล ค่าธรรมเนียมการซื้อขาย ภาษีและไม่มีธุรกรรมขายชอร์ต (short sale) หากมีการซื้อหมายความว่าซื้อในราคาปิดวันนี้และต้องขายออกไปในราคาปิดของวันทำการถัดไป

ในการคำนวณผลตอบแทนในแต่ละช่วงเวลาใช้วิธีการคำนวนผลตอบแทนดังนี้

$\text{Return} = (V_t - V_0) / V_0$

Return คือผลตอบแทนในช่วงเวลานั้น

V_t คือ มูลค่าของหน่วยการลงทุนตอนสิ้นสุดของช่วงเวลา

V_0 คือ มูลค่าของหน่วยการลงทุนตอนเริ่มต้นของช่วงเวลา

เกณฑ์มาตรฐาน 1 (Benchmark 1) คือ กลยุทธ์คือซื้อตอนเริ่มต้นและขายตอนสิ้นสุดของแต่ละช่วงเวลา ดังนั้นกำไรจะขึ้นอยู่กับสภาวะตลาดในแต่ละช่วงเวลา เป็นการแสดงให้เห็นถึงสภาวะตลาดในช่วงเวลานั้น

เกณฑ์มาตรฐาน 2 (Benchmark 2) คือ กลยุทธ์คือซื้อหากวันนี้ดัชนีตลาดหลักทรัพย์สูงกว่าวันทำการที่ผ่านมาและไม่ซื้อ (ถือเงินสด) หากวันนี้ดัชนีตลาดหลักทรัพย์ต่ำกว่าวันทำการที่ผ่านมา เป็นสมมติฐานที่ว่าดัชนีตลาดหลักทรัพย์ในวันทำการถัดไปจะปรับตัวขึ้นหากวันนี้ดัชนีตลาดหลักทรัพย์สูงกว่าวันทำการที่ผ่านมา

Model คือ ใช้การคาดการณ์ที่ได้จากโมเดลที่มีความแม่นยำในการคาดการณ์สูงที่สุด โดยกำหนดความมั่นใจในคาดการณ์ว่าดัชนีตลาดหลักทรัพย์ในวันทำการถัดไปจะเป็นบวก (Positive) มากกว่า 0.5 (confidence Positive > 50%) ดังนั้นกลยุทธ์คือ ซื้อหากโมเดลคาดการณ์ว่าดัชนีตลาดหลักทรัพย์ในวันทำการถัดไปจะเป็นบวก (Positive) และถือเงินสด (ไม่ซื้อ) หากโมเดลคาดการณ์ว่าจะเป็นลบ (Negative)

ผลของการจำลองการลงทุนปรากฏว่าผลตอบแทนหากใช้โมเดลในการคาดการณ์ได้ผลตอบแทนมากกว่าทั้งสองเกณฑ์มาตรฐาน 8 ใน 10 ช่วงเวลา โดยได้ผลตอบแทนเฉลี่ย 3.93% ต่อหนึ่งช่วงเวลา อีกทั้งได้ผลตอบแทนต่อหนึ่งช่วงเวลาเฉลี่ยมากกว่าเกณฑ์มาตรฐาน 1 และ เกณฑ์มาตรฐาน 2 อยู่ 2.83% และ 2.15% ตามลำดับ ผลตอบแทนจากโมเดลไม่มีช่วงเวลาใดเลยที่เป็นลบ

หากคำนวณจุดคุ้มทุน (breakeven) โดยคำนวณจากกำไรของจากการซื้อขายทั้งหมดต่อจำนวนการซื้อขาย ดังนั้นจุดคุ้มทุนจะอยู่ที่ประมาณ 0.068% ต่อครั้ง หรือ 0.1361% ต่อรอบ หมายความว่าหากค่าธรรมเนียมในการซื้อขาย ภาษี หรือ ค่าใช้จ่ายในการซื้อขายต่ำกว่าจุดคุ้มทุนจะสามารถทำกำไรได้จากการใช้การคาดการณ์โดยใช้โมเดล

ตารางที่ 5 การจำลองการลงทุนจากการคาดการณ์ของโมเดลและเกณฑ์มาตรฐาน

Data Set (Period)	BM1	BM2	Model	Model-BM1	Model-BM2
1	-1.65%	7.65%	4.27%	5.93%	-3.38%
2	2.68%	0.70%	3.69%	1.01%	2.99%
3	1.61%	1.22%	3.45%	1.84%	2.23%
4	-0.61%	0.29%	1.60%	2.20%	1.31%
5	-0.08%	-1.27%	1.57%	1.65%	2.84%
6	8.92%	5.96%	11.05%	2.13%	5.09%
7	4.81%	1.99%	4.20%	-0.61%	2.20%
8	1.40%	1.40%	4.79%	3.40%	3.40%
9	-4.60%	0.81%	0.97%	5.57%	0.16%
10	-1.40%	-0.92%	3.74%	5.14%	4.66%
Mean	1.11%	1.78%	3.93%	2.83%	2.15%
SD	3.78%	2.85%	2.82%		
Mean/SD	0.293	0.624	1.393		

บทที่ 5
สรุปผลการศึกษา

งานศึกษานี้มีวัตถุประสงค์เพื่อคาดการณ์การเปลี่ยนแปลงของดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET Index) ในวันทำการถัดไปโดยใช้เทคนิค machine learning โดยใช้ประโยชน์จากความสัมพันธ์ระยะสั้นระหว่างการเปลี่ยนแปลงของดัชนีราคาหุ้นตลาดหลักทรัพย์จากทั่วโลก (global stock market indices) อัตราแลกเปลี่ยน (exchange rates) และราคาสินค้าโภคภัณฑ์ (commodities) เพื่อทดสอบความแม่นยำ (accuracy) ของแต่ละโมเดล ในการศึกษาใช้ข้อมูลระหว่างวันที่ 2 สิงหาคม 2550 ถึงวันที่ 2 สิงหาคม 2561

Machine learning algorithm ที่สร้างโมเดลเพื่อคาดการณ์ได้ความแม่นยำมากที่สุดคือ Decision Tree แต่ในทุกอัลกอริธึมมีข้อสังเกตที่น่าสนใจคือฟีเจอร์ที่ถูกคัดเลือกมาด้วยวิธีการ forward feature selection จะมีฟีเจอร์ S&P500 (USA) และ MERVAL (Argentina) อยู่ด้วยในทุกโมเดล ซึ่งเป็นการแสดงให้เห็นว่าการเปลี่ยนแปลงของดัชนีตลาดหลักทรัพย์ทั้งสองมีความสัมพันธ์กับการเปลี่ยนแปลงของดัชนีตลาดหลักทรัพย์ประเทศไทยในวันทำการถัดไป โดยสาเหตุอาจมาจาก time zone โดยในทวีปอเมริกาเหนือและอเมริกาใต้ตลาดหลักทรัพย์จะปิดทำการก่อนที่ตลาดหลักทรัพย์ประเทศไทยจะเปิดทำการไม่นาน ทำให้นักลงทุนในประเทศไทยสามารถทราบข้อมูลของสองดัชนีปิดของตลาดแล้วของทั้งสองประเทศ ทั้งดัชนีตลาดหลักทรัพย์ในสหรัฐอเมริกาซึ่งเป็นตลาดหลักทรัพย์ที่ใหญ่และมีความสำคัญมาก ส่วนดัชนีตลาดหลักประเทศอาร์เจนตินาก็เป็นตลาดเกิดใหม่คล้ายกับประเทศไทย ซึ่งอาจเป็นผลที่ทำให้เกิดความสัมพันธ์เหล่านี้

ผลการศึกษาพบว่าการใช้ Decision Tree ที่มีการแบ่งชุดข้อมูลแบบ sliding windows สามารถสร้างโมเดลที่คาดการณ์ได้ความแม่นยำเฉลี่ยสูงที่สุดที่ 63.2% และมีส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ต่ำที่สุด 4.24%

แต่หากมองในภาพรวมทุกอัลกอริทึมสามารถคาดการณ์ได้ความแม่นยำเฉลี่ยใกล้เคียงกัน อีกหนึ่งข้อสังเกตคือ Artificial Neural Networks และ Support Vector Machine สามารถคาดการณ์ในการแบ่งชุดข้อมูลแบบ expanding windows มีความแม่นยำเฉลี่ยมากกว่าและมี false positive rate (FPR)

และ false negative rate (FNR) ต่ำกว่าการแบ่งชุดข้อมูลแบบ sliding windows ซึ่งแสดงให้เห็นว่าทั้งสองอัลกอริทึมมีแนวโน้มที่สร้างจะโมเดลที่มีประสิทธิภาพมากขึ้นหากชุดข้อมูลในการสร้างโมเดลมีปริมาณมากขึ้น ผลการศึกษายังพบอีกว่า Artificial Neural Networks มี false positive rate (FPR) ต่ำกว่า false negative rate (FNR) ทั้งการแบ่งข้อมูล 2 รูปแบบแต่ Support Vector Machine มี และ false negative rate (FNR) ต่ำกว่า false positive rate (FPR) ทั้งการแบ่งข้อมูล 2 รูปแบบ

ในงานศึกษานี้ยังมีข้อจำกัดเกี่ยวกับปริมาณข้อมูลที่ใช้ในการทดสอบและการคัดเลือกตัวแปรอิสระ เนื่องด้วยประสิทธิภาพของเครื่องมือ (คอมพิวเตอร์) ที่ใช้ หากต่อไปมีการใช้ปริมาณข้อมูลที่มากขึ้นและเพิ่มเทคนิคในการสร้างโมเดลมากขึ้นก็อาจจะพบผลลัพธ์ที่แตกต่างออกไป

หากมีงานศึกษาเพิ่มเติมในอนาคต การสร้างฟีเจอร์ที่มีความถี่มากขึ้น การใช้ Machine learning algorithm หลายอัลกอริธึมร่วมกันในการสร้างโมเดลก็เป็นหนึ่งสิ่งที่น่าสนใจ รวมไปถึงสร้างโมเดลการใช้ข้อมูลในหลายแง่มุมที่แตกต่างกันมารวมเข้าด้วยกันเพื่อสร้างฟีเจอร์ อาทิเช่น technical analysis, fundamental analysis, global financial market, macroeconomic หรือข้อมูลยุคใหม่ อย่างเช่น ปริมาณการค้นหาชื่อบริษัทหรือชื่อย่อใน search engine ก็น่าสนใจไม่แพ้กัน

บรรณานุกรม

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Kluwer Academic Publishers, 20(3), 273-297.

Huang, W., Nakamori, Y., & Wan, S. Y. (2005). Forecasting stock market movement direction with support vector machine. Computers & Operations Research, 32(10), 2513-2522.

Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Expert Systems with Applications, 38(5), 5311-5319.

Olson, D., & Mossman, C. (2003). Neural network forecasts of Canadian stock returns using accounting ratios. International Journal of Forecasting, 19(3), 453-465.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Systems with Applications, 42(1), 259-268.

Qian, B., & Rasheed, K. (2007). Stock market prediction with multiple classifiers. Applied Intelligence, 26(1), 25-33.

Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81-106.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review, 65(6), 386-408.

Shah, V. H. (2007). Machine learning techniques for stock prediction. Foundations of Machine Learning, 1-19.

Shen, S., Jiang, H., & Zhang, T. (2012). Stock market forecasting using machine learning algorithms. Department of Electrical Engineering, Stanford University, 1-5.

Swales Jr, G. S., & Yoon, Y. (1992). Applying artificial neural networks to investment analysis. Financial Analysts Journal, 48(5), 78-80.

บรรณานุกรม (ต่อ)

Tsai, C. F., & Wang, S. P. (2009). Stock price forecasting by hybrid machine learning techniques. Paper presented at the Proceedings of the International MultiConference of Engineers and Computer Scientists.

Zhu, X., Wang, H., Xu, L., & Li, H. (2008). Predicting stock index increments by neural networks: The role of trading volume under different horizons. Expert Systems with Applications, 34(4), 3043-3054.

ภาคผนวก

ภาคผนวก

แต่ละวิธีการ (algorithms) จะถูกปรับแต่งค่าพา รามิเตอร์ (tuning parameters) เพื่อเพิ่มประสิทธิภาพ (optimize) ให้โมเดลที่ได้มีความแม่นยำ (accuracy) สูงสุด โดยวัดความแม่นยำในทุกโมเดลจากการสร้างโมเดลในช่วงค่าพารามิเตอร์และเงื่อนไขที่กำหนดดังนี้

Decision Tree

- Minimal gain : range [10^{-3} – 10^{3}]
- Confidence : range [10^{-7} – 0.5]
- Maximal depth : 20

Artificial Neural Networks

- Training cycles : 500
- Learning rate : range [0.1 – 0.9]
- Momentum : range [0.1 – 0.9]

Support Vector Machine

- Kernel type : Radial Basis Function(RBF) or Sigmoid
- gamma : range [10^{-3} – 10^{3}]
- C : range [10^{-3} – 10^{3}]