การจัดการวัตถุดิบคงคลังของร้านอาหาร โดยใช้ Proximal Policy Optimization

ณัฐวัฒน์ เอกธรรมนิตย์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (การวิเคราะห์ธุรกิจและวิทยาการข้อมูล)
คณะสถิติประยุกต์
สถาบันบัณฑิตพัฒนบริหารศาสตร์
2563

การจัดการวัตถุดิบคงคลังของร้านอาหาร โดยใช้ Proximal Policy Optimization

ณัฐวัฒน์ เอกธรรมนิตย์

คณะสถิติประยุกต์

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ดร.วรพล พงษ์เพ็ชร)

คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาแล้วเห็นสมควรอนุมัติให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต (การวิเคราะห์ธุรกิจและวิทยาการข้อมูล)

..... ประธานกรรมการ

(ดร.ธนชาตย์ ฤทธิ์บำรุง)

..... กรรมการ

(ดร.วรพล พงษ์เพ็ชร)

..... กรรมการ

(ดร.เอกรัฐ รัฐกาญจน์)

..... กรรมการ

(ดร.รัฐศิลป์ รานอกภานุวัชร์)

..... คณบดี

(ผู้ช่วยศาสตราจารย์ ดร.ปราโมทย์ ลือนาม)

_____/_____/_____

บทคัดย่อ

ชื่อวิทยานิพนธ์	การจัดการวัตถุดิบคงคลังของร้านอาหาร โดยใช้ Proximal Policy Optimization
ชื่อผู้เขียน	ณัฐวัฒน์ เอกธรรมนิตย์
ชื่อปริญญา	วิทยาศาสตรมหาบัณฑิต (การวิเคราะห์ธุรกิจและวิทยาการข้อมูล)
ปีการศึกษา	2563

งานวิจัยนี้มีวัตถุประสงค์ในการจัดทำขึ้นเพื่อศึกษาหลักการ Proximal Policy Optimization ของกระบวนการ Reinforcement Learning ในการสร้างแบบจำลองการพยากรณ์จำนวนการสั่งวัตถุดิบของร้านอาหาร เนื่องมาจากปัญหาการสั่งวัตถุดิบของร้านอาหารในแต่ละวันเกิดความคลาดเคลื่อนจากจำนวนการใช้วัตถุดิบจริงสูง วัตถุดิบที่เหลือจึงกลายเป็นขยะอาหาร ทำให้เกิดการหมักและก่อตัวเป็นก๊าซมีเทนลอยขึ้นไปทำลายโอโซนในชั้นบรรยากาศ โดยแบ่งการศึกษาออกเป็น 2 รูปแบบหลักๆ คือ 1. แบบจำลองแบบหนึ่งแอตทริบิวต์ 2. แบบจำลองแบบหลายแอตทริบิวต์ และชุดข้อมูลจำลองที่ใช้มาจากการจำลองโดยอาศัยหลักการจำลองข้อมูลแบบแจกแจงปกติ ส่วนการประเมินประสิทธิภาพของแบบจำลองจะใช้เครื่องมือ 3 อย่าง คือ F-Statistics ,R-Square และRMSE ในการศึกษาครั้งนี้แต่ละแบบจำลองจะมีการเรียนรู้ทั้งหมด 12 ล้านไทม์สเต็ป ผลจากการศึกษาในงานวิจัยนี้ พบว่า ในการเรียนรู้ของแบบจำลองทั้งหมด 12 ล้านไทม์สเต็ป แบบจำลองแบบหลายแอตทริบิวต์ จะเกิดการลู่เข้าหาค่าที่เหมาะสมของแบบจำลอง ได้เร็วกว่าแบบจำลองแบบหนึ่งแอตทริบิวต์ และประสิทธิภาพของค่าความถูกต้องในการพยากรณ์จำนวนการสั่งวัตถุดิบ อยู่ที่ร้อยละ 82 ของค่าการพยากรณ์ทั้งหมด จากงานวิจัยนี้ทำให้ทราบถึงแนวทางในการนำหลักการ Proximal Policy Optimization ไปสร้างแบบจำลองการพยากรณ์จำนวนการสั่งวัตถุดิบ ให้สามารถพยากรณ์จำนวนวัตถุดิบที่จะสั่ง ได้ใกล้เคียงกับจำนวนวัตถุดิบที่ใช้จริงมากที่สุด และสามารถลดจำนวนขยะอาหารให้มีจำนวนเหลือน้อยลง

ABSTRACT

Title of Thesis	Proximal Policy Optimization on Casual Restaurant Raw Material Stock
Author	Nutthawat Ekthammanit
Degree	Master of Science (Business Analytics and Data Science)
Year	2020

This research focuses on the Proximal Policy Optimization Algorithm (PPO) of Reinforcement Learning to make a forecasting model of raw material stock in restaurants. The restaurant's daily raw material stock ordered and the number of raw material stock used fluctuates daily. The unused raw material stock is left as wasted material. It caused fermentation and produced methane gas that rises to destroy ozone into the atmosphere. This research investigated a One-Attribute Model and a Multi-Attribute Model. The dataset used in this research is synthetic data that use the normal distribution theory to make it. The model's performance was assessed using F-statistics, R-Square, and RMSE. We trained each model trained 12 million timesteps. The result showed that the Multi-Attribute Model would converge to the value optimization faster than the One-Attribute Model. We found that both models' accuracy is about 82 percent. This research demonstrated that the PPO Algorithm could be utilized to make an effective raw material forecasting tool.

กิตติกรรมประกาศ

วิทยานิพนธ์เรื่องการจัดการวัตถุดิบคงคลังของร้านอาหาร โดยใช้ Proximal Policy Optimization สำเร็จลุล่วงไปด้วยดีตามความหวังของผู้เขียนได้ เนื่องจากความกรุณาช่วยเหลือจาก อ.ดร.วรพล พงษ์เพ็ชร ที่ได้ให้คำปรึกษา ข้อแนะนำ ตลอดจนตรวจทานและแก้ไขข้อบกพร่องในวิทยานิพนธ์ฉบับนี้จนเสร็จสมบูรณ์

ขอขอบคุณท่านคณะกรรมการสอบวิทยานิพนธ์ซึ่งได้แก่ อ.ดร.ธนชาตย์ ฤทธิ์บำรุง ประธานกรรมการสอบ อ.ดร.เอกรัฐ รัฐกาญจน์ กรรมการสอบและ อ.ดร.รัฐศิลป์ รานอกภานุวัชร์ กรรมการผู้ทรงคุณวุฒิภายนอก ที่ได้กรุณาให้ข้อเสนอแนะ และให้แนวคิดต่างๆที่เป็นประโยชน์กับวิทยานิพนธ์ฉบับนี้ให้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบคุณกองบริการการศึกษา คณะสถิติประยุกต์ ที่ได้กรุณาให้ความช่วยเหลือในเรื่องเอกสารต่างๆมาตลอดกระบวนการ

ขอขอบคุณบิดา มารดา และครอบครัวของผู้เขียนที่คอยให้การสนับสนุนส่งเสริม และเป็นกำลังใจที่สำคัญให้กับผู้เขียนมาโดยตลอด

ณัฐวัฒน์ เอกธรรมนิตย์

กรกฎาคม 2564

สารบัญ

หน้า

สารบัญตาราง

หน้า

สารบัญภาพ

หน้า

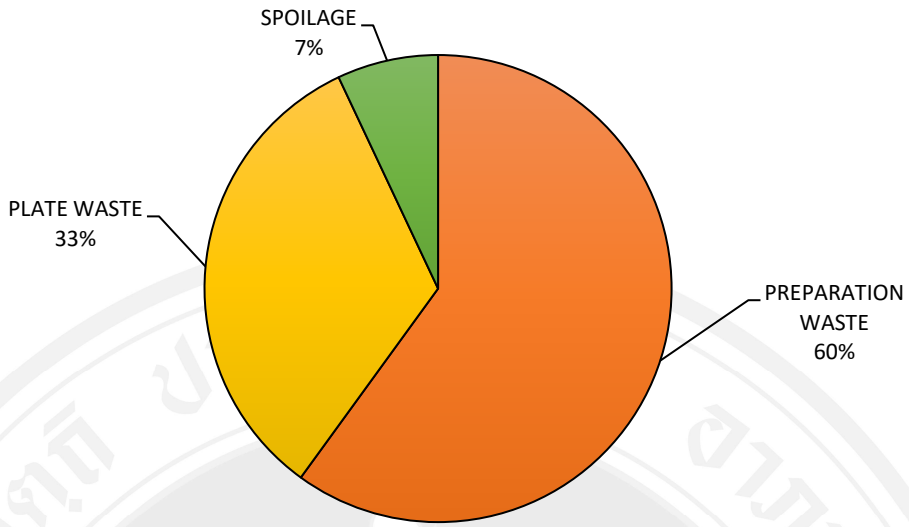
บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันหลายประเทศได้ประสบปัญหาขยะที่เกิดจากอาหารขึ้นเป็นจำนวนมาก โดยจากสถิติในประเทศนิวซีแลนด์ ขยะอาหารที่เกิดจากร้านอาหารมีมากถึง 24,372 ตันต่อปี (WasteMINZ, 2018) ซึ่งจะแบ่งประเภทการเกิดขยะอาหารได้ออกเป็น 3 แหล่งใหญ่ๆ คือ

1) ขยะอาหารที่เกิดจากขั้นตอนการเตรียมอาหาร (ร้อยละ 60 ของน้ำหนักขยะอาหารต่อปี) เช่น ผักที่ถูกตัดแต่ง, อาหารที่มีข้อผิดพลาดไม่สามารถจำหน่ายได้
2) ขยะอาหารที่เกิดจากถาดอาหารหรือจานอาหารของลูกค้า (ร้อยละ 33 ของน้ำหนักขยะอาหารต่อปี) เช่น เศษอาหารหรืออาหารที่ทางลูกค้ารับประทานไม่หมด
3) ขยะอาหารที่เกิดจากการที่ร้านอาหารสั่งซื้อวัตถุดิบมาเกินจำนวนที่ต้องใช้ ทำให้วัตถุดิบเกิดการเน่าเสีย และต้องทิ้งเป็นขยะ (ร้อยละ 7 ของน้ำหนักขยะอาหารต่อปี)



ภาพที่ 1.1 กราฟจำแนกประเภทของแหล่งขยะอาหาร

โดยขยะอาหารเหล่านี้เป็นส่วนหนึ่งของต้นเหตุที่ทำให้เกิดปัญหาใหญ่ๆบนโลก เช่น ปัญหาภาวะเรือนกระจก ที่ขยะอาหารที่ถูกฝังกลบก่อให้เกิดก๊าซมีเทนไปทำลายชั้นโอโซนของโลก ที่เปรียบเสมือนเกราะป้องกันรังสีอัลตราไวโอเลตจากดวงอาทิตย์ ทำให้เมื่อไม่มีชั้นโอโซนคอยกรองรังสีอัลตราไวโอเลต อุณหภูมิของโลกจึงเพิ่มสูงขึ้น และปัญหาการขาดแคลนอาหาร ซึ่งจากงานศึกษาของ Food and Agriculture Organization of the United Nations ขยะอาหารคิดเป็น1 ใน3 ของจำนวนอาหารทั้งหมดผลิตได้ คิดเป็นน้ำหนักของขยะอาหารที่ถูกทิ้งไปเป็นจำนวน 1,300 ล้านตัน ทำให้มีผู้ที่ขาดแคลนอาหารเป็นจำนวน 820 ล้านคน (FAO, 2019)

1.2 วัตถุประสงค์

1) เพื่อสร้างโมเดลการพยากรณ์จำนวนการสั่งวัตถุดิบ
2) เพื่อลดจำนวนวัตถุดิบที่เหลือในแต่ละวัน

1.3 ขอบเขตของการศึกษา

งานวิจัยนี้จัดทำขึ้นเพื่อสร้างโมเดลการพยากรณ์จำนวนการสั่งวัตถุดิบ โดยใช้ Proximal Policy Optimization Algorithm ของกระบวนการ Reinforcement Learning

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1) โมเดลการพยากรณ์ที่ได้สามารถมาช่วยในเรื่องของการลดต้นทุนให้กับทางบริษัท
2) โมเดลการพยากรณ์ที่ได้สามารถช่วยลดจำนวนขยะอาหาร ที่เกิดจากวัตถุดิบคงคลังเน่าเสียก่อนถูกนำไปใช้ ทำให้สามารถช่วยแก้ปัญหาการเกิดก๊าซมีเทนจากขยะอาหารที่ถูกฝังกลบ ไปทำลายชั้นโอโซนของโลก ที่ทำให้เกิดเป็นปัญหาภาวะโลกร้อน
3) โมเดลการพยากรณ์ที่ได้สามารถพยากรณ์ได้แม่นยำ จึงไม่เกิดการสั่งวัตถุดิบมากเกินความจำเป็น ทำให้ช่วยลดปัญหาการขาดแคลนอาหารของมนุษย์บนโลก

บทที่ 2

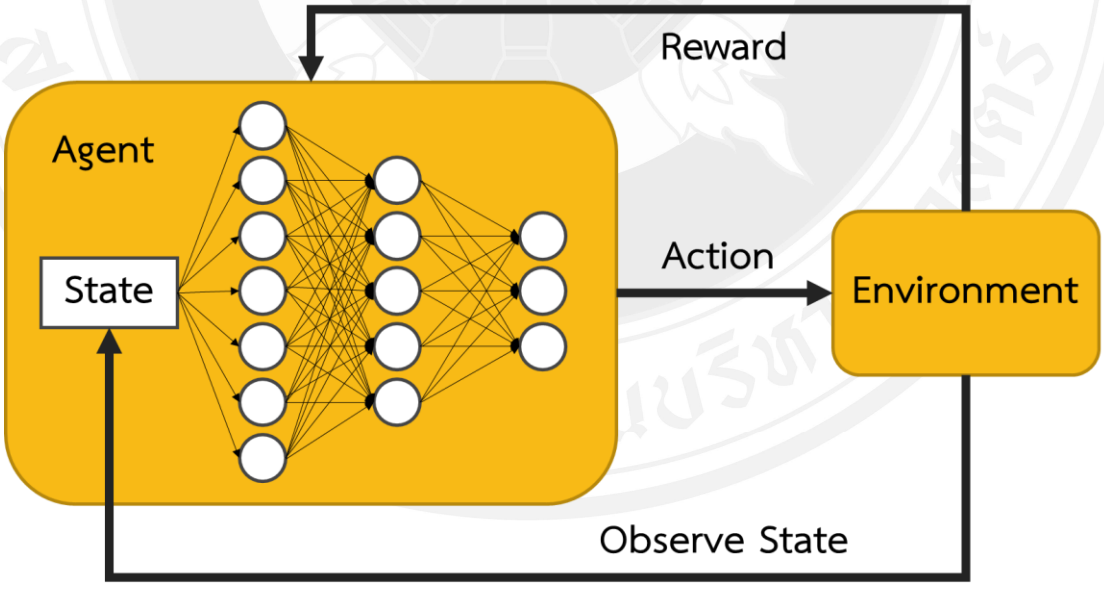
วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎี Reinforcement Learning

Reinforcement Learning คือ ศาสตร์หนึ่งของการทำ Machine Learning โดยจะใช้โรบอท (Agent) มาช่วยในการพิจารณาตัดสินใจ ว่าในสภาพแวดล้อม ณ ขณะนั้น ควรจะต้องตัดสินใจทำสิ่งใดที่ให้ได้ผลลัพธ์ที่ดีที่สุด โดยการเรียนรู้ของโรบอทจะเรียนรู้จากการลองผิดลองถูก และจากการลองผิดลองถูกนี้จะมีการให้คะแนน (Reward) เพื่อเป็นตัวใช้ในการบ่งบอกให้แก่โรบอทว่าควรจะต้องปรับปรุงการตัดสินใจไปในทิศทางใด เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด (AI, 2019; Hill et al., 2018; King, 2019; Schulman, Wolski, Dhariwal, Radford, & Klimov, 2017)

2.1.1 องค์ประกอบหลักของระบบ Reinforcement Learning

ระบบ Reinforcement Learning จะมีองค์ประกอบหลักๆ 5 ส่วน ดังนี้



ภาพที่ 2.1 ระบบ Reinforcement Learning

1) Agent คือ โรบอทที่มีหน้าที่ตัดสินใจว่าจะทำอะไรในสถานการณ์ที่เกิดขึ้น ณ ขณะนั้น โดยการตัดสินใจในแต่ละครั้งจะอ้างอิงจากประสบการณ์การลองผิดลองถูกในสถานการณ์ที่เคยเกิดขึ้นมาก่อนหน้า แล้วได้ผลลัพธ์หรือคะแนน (Reward) ออกมาดี
2) Action คือ การกระทำของโรบอททั้งหมดที่เป็นไปได้ เช่น งานวิจัยครั้งนี้การจะทำให้การสั่งวัตถุดิบออกมาได้ค่าที่เหมาะสม จะมีการกระทำที่เป็นไปได้ทั้งหมด 2 อย่าง คือ เพิ่มจำนวนสั่งวัตถุดิบและลดจำนวนการสั่งวัตถุดิบ
3) Environment คือ สภาพแวดล้อมที่กำหนดให้โรบอทเรียนรู้และต้องทำการตัดสินใจว่าจะทำอย่างไรในสถานการณ์ ณ ขณะนั้น เพื่อให้ได้ค่า Reward ที่สูงที่สุด
4) State คือ ประสบการณ์การเรียนรู้ของโรบอท ว่าถ้าเกิดสถานการณ์แบบนี้ควรจะต้องตัดสินใจไปในทิศทางอย่างไร เช่น ตัวอย่าง Hidden Markov Model จะประกอบด้วยสถานการณ์ที่แดดออก และสถานการณ์ที่ฝนตก โดยจะมีความน่าจะเป็นที่จะเกิดสถานการณ์ต่างๆ
5) Reward คือ คะแนนที่ระบบประเมินผลให้กับการกระทำครั้งนั้นๆที่โรบอทได้ตัดสินใจทำลงไป ค่าคะแนนรางวัลนี้จะเป็นสิ่งบ่งบอกให้โรบอททราบว่าการกระทำกับสภาพแวดล้อมแบบที่เกิดขึ้นนั้น ควรจะกระทำร่วมกันหรือไม่ โดยโรบอทต้องตัดสินใจกระทำในสิ่งที่ได้ค่าคะแนนรางวัลออกมาสูงที่สุด

2.1.2 ประเภทของระบบ Reinforcement Learning

ระบบ Reinforcement Learning แบ่งออกเป็นประเภทหลักๆ 2 ประเภท ดังนี้

1) Model-Free Reinforcement Learning คือ รูปแบบของการเรียนรู้จากประสบการณ์ของการลองผิดลองถูกที่ผ่านมาในอดีต ทำให้ใช้เวลานานในการเรียนรู้ ระบบจะรู้เพียงแค่ว่าสิ่งใดที่ควรทำหรือสิ่งใดที่ไม่ควรทำ โดย Model-Free RL จะเหมาะสมกับข้อมูลที่ไม่มีกฎกติกาการเรียนรู้มารองรับ และModel-Free RL สามารถแบ่งย่อยได้เป็น 2 ประเภท คือ Policy Optimization และ Q-Learning
2) Model-Based Reinforcement Learning คือ รูปแบบของการเรียนรู้ที่มีกฎกติกามาช่วยในการจำกัดขอบเขตของการตัดสินใจให้แคบลง ทำให้ใช้เวลาในการเรียนรู้ที่สั้น เช่น กระบวนการเรียนรู้ของเกมโกะ ที่มีกฎการเดินหมากมาช่วยในการจำกัดขอบเขตการเรียนรู้ โดย Model-Based RL สามารถแบ่งย่อยได้เป็น 2 ประเภท คือ Learn the Model และ Given the Model

2.1.3 ประเภทของแนวทางการพยากรณ์ในระบบ Reinforcement Learning

ประเภทของแนวทางการพยากรณ์ในระบบ Reinforcement Learning แบ่งได้ออกเป็น 2 ประเภท คือ Exploration เป็นการที่ระบบจะทำการค้นหาแนวทางใหม่ เพื่อให้ได้แนวทางการแก้ปัญหา หรือได้ผลลัพธ์ที่ดีกว่าเดิม กับExploitation เป็นการที่ระบบจะนำค่าที่ได้จากการเรียนรู้ แล้วระบบได้จดจำไว้ มาใช้ในการตัดสินใจแก้ไขปัญหา หรือเพื่อใช้ในการพยากรณ์ให้ได้ผลลัพธ์ที่ดี (Yogeswaran & Ponnambalam, 2012)

2.2 ทฤษฎี Proximal Policy Optimization (PPO)

Proximal Policy Optimization (Schulman et al., 2017) เป็นหลักการที่ได้รวมจุดเด่นของ Advantage Actor Critic (A2C) ในเรื่องของการประมวลผลแบบ Multiple Workers ทำให้สามารถประมวลผลได้รวดเร็ว (Mnih et al., 2016) และจุดเด่นของ Trust Region Policy Optimization (TRPO) ที่มีการกำหนดขอบเขตของการอัพเดต Policy ใหม่ เพื่อป้องกันการเกิดการ Divergent ออกจากค่า Optimization (Schulman, Levine, Abbeel, Jordan, & Moritz, 2015) และอีกจุดเด่นของกระบวนการ Proximal Policy Optimization คือ การอัพเดต Policy ของ Stochastic Gradient Ascent แบบ Multiple Epochs ซึ่งทำให้โอกาสที่ข้อมูล Outlier จะส่งผลกระทบกับโมเดลเกิดขึ้นได้ยากยิ่งขึ้น (Xu, Qian, Li, & Jin, 2021)

โดยกระบวนการ Proximal Policy Optimization จะมีส่วนที่เพิ่มเติมขึ้นมาจากระบบ Reinforcement Learning ทั่วไป คือตัว Critic ที่จะทำหน้าที่เป็นผู้ช่วยของ Actor ในการประเมินผลประสิทธิภาพของ Action ที่ Actor ได้กระทำต่อ Environment ด้วยค่า Advantage Function ดังสมการที่ 2.1 ที่จะเป็นการหาค่าผลต่างระหว่างค่า Discounted Rewards (G_t) ที่ได้จากการคำนวณค่า Reward ที่ t แบบ Multiple Epochs ดังสมการที่ 2.2 โดยที่ค่า k คือ ลำดับที่ของ Epoch และ ค่า γ คือค่า Discount Factor ซึ่งในงานวิจัยนี้ได้ใช้ค่า 0.99 ตามงานวิจัย Proximal Policy Optimization (Schulman et al., 2017) กับค่า Value Function ที่จะได้จากการคาดคะเนของ Value Function Neural Network ซึ่งถ้าค่าของ Advantage Function ออกมาเป็นบวกแสดงว่า Action ที่กระทำต่อ State นั้นเป็น Action ที่ดีต่อ State นั้น และจะมีการปรับค่าความน่าจะเป็นที่จะใช้ Action ดังกล่าวเมื่อเกิด State นั้นเพิ่มขึ้น ในทางตรงกันข้าม ถ้าค่าของ Advantage Function ออกมาเป็นลบ จะมีการลดความน่าเป็นที่จะใช้ Action ดังกล่าวเมื่อเกิด Stateนั้น

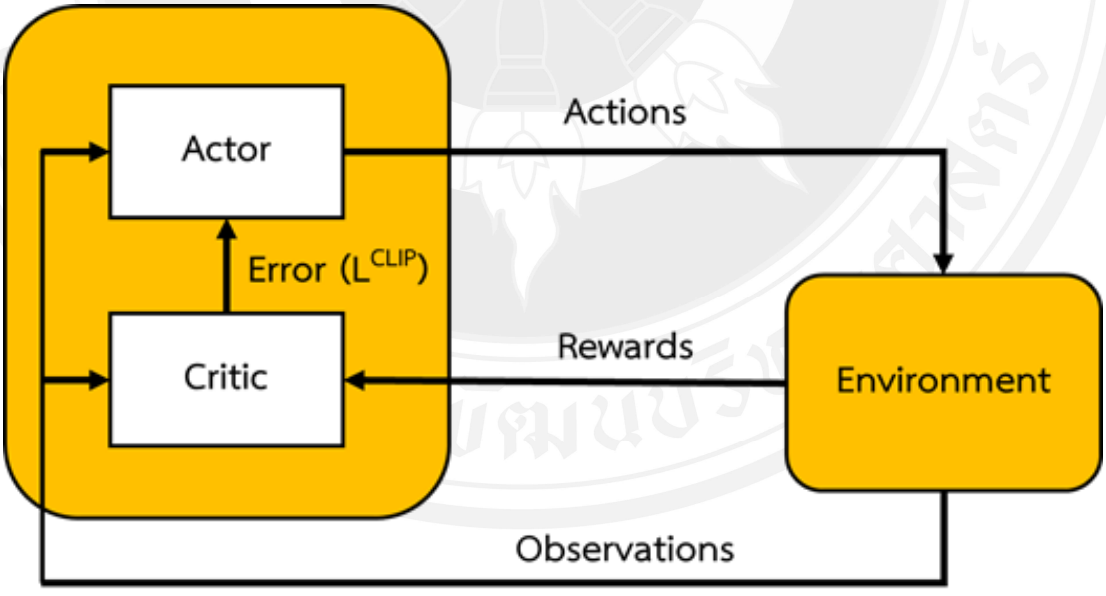
$$\hat{\mathrm{A}}_t = G_t - Value\ Function \tag{2.1}$$

$$G_t = \sum_{k=0}^{x} \gamma^k R_{t+k+1} \tag{2.2}$$

ต่อมาจะเป็นการคำนวณค่า Objective Function (L^{CLIP}) ที่ใช้ในการอัพเดต Policy ด้วยวิธีการ Clip ที่จะช่วยในเรื่องการจำกัดกรอบของโมเดลในการอัพเดต Policy ไม่ให้เกิดกรณีการอัพเดตค่าที่มีการกระโดดมาเกินไป เพื่อป้องกันผลกระทบที่จะเกิดจากการเจอข้อมูลที่เป็น Outlier โดย Objective Function (Fujita & Maeda, 2018) ดังในสมการที่ 2.3 จะเป็นการเปรียบเทียบค่า Minimum ระหว่างค่าอัตราส่วนของค่า Policy ใหม่ ($\pi_\theta(a_t|s_t)$) ต่อค่า Policy เดิม ($\pi_{\theta_{old}}(a_t|s_t)$) และอัตราส่วนของค่า Policy ใหม่ กับ Policy เดิม แบบที่ Clip ค่าให้อยู่ระหว่าง 1-ε กับ 1+ε โดยที่ ε กำหนดค่าตามงานวิจัย Proximal Policy Optimization (Schulman et al., 2017) ให้มีค่าอยู่ที่ 0.2 ซึ่งเป็นค่าที่ให้ประสิทธิภาพออกมาสูงที่สุด โดยเมื่อได้ค่า Minimum ของอัตราส่วนระหว่างค่า Policy ใหม่ กับ Policy เก่า จะนำค่าดังกล่าวไปคูณกับค่า Advantage Function ที่ได้คำนวณมาก่อนหน้า หลังจากนั้นนำค่า L^{CLIP} ที่ได้ไปอัพเดต Gradient Ascent ของระบบใหม่

$$L^{CLIP}(\theta) = \hat{E}[min(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t] \quad (2.3)$$

เมื่อ $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$



ภาพที่ 2.2 Proximal Policy Optimization Diagram

ตารางที่ 2.1 คำนิยามของ Hyperparameter

Hyperparameter	Definition
$\hat{A}_t$	Advantage Function คือค่าที่แสดงให้เห็นประสิทธิภาพการพยากรณ์ของโมเดล เมื่อเทียบกับค่าที่คาดหวังไว้ที่ได้จาก Value Function
G_t	Discounted Rewards คือค่าที่ได้จากการนำ Reward ในแต่ละการเรียนรู้มาคูณกับ Discount Factor เพื่อเป็นการถ่วงน้ำหนักให้การเรียนรู้ในแต่ละครั้ง โดยจะให้น้ำหนักกับ Reward ครั้งแรกๆมากกว่า
γ	Discount Factor เป็นเฟกเตอร์ที่ใช้ในการถ่วงน้ำหนักให้กับการคำนวณ Discounted Rewards
R	Reward เป็นค่าคะแนนที่ได้จากการเรียนรู้โมเดลในแต่ละครั้ง
L^{CLIP}	Objective Function เป็นการใช้ Expectation เปรียบเทียบค่าระหว่างอัตราส่วนของค่า Policy ใหม่ ต่อค่า Policy เก่า กับอัตราส่วนของค่า Policy ใหม่ ต่อค่า Policy เก่าที่ถูกคลิปค่า และเลือกค่าที่น้อยที่สุด
$\hat{E}$	Expectation เป็นค่าความคาดหวังที่จะเปรียบเทียบค่าและเลือกค่าที่ดีที่สุดตรงตามเงื่อนไขที่กำหนดไว้
r_t	r_t คืออัตราส่วนระหว่างค่าของ Policy ใหม่ ต่อค่า Policy เก่า
ε	ε คือเฟกเตอร์ที่ใช้สำหรับการกำหนดค่าที่จะใช้ในการคลิปค่า
$\pi_\theta(a_t\|s_t)$	$\pi_\theta(a_t\|s_t)$ คือค่า Policy ของ Action ณ State ที่ t

ตารางที่ 2.2 ค่าของ Hyperparameter

Hyperparameter	Value
γ	0.99
ε	0.20

2.3 ทฤษฎีการพยากรณ์ (Forecasting)

2.3.1 ความหมายของการพยากรณ์

การพยากรณ์ คือ การทำนายผลลัพธ์หรือสิ่งที่จะเกิดในอนาคต โดยอาศัยการพิจารณาจากเหตุการณ์ที่เคยเกิดขึ้นมาในอดีต ซึ่งจะพิจารณาจากปัจจัยรอบข้างที่ส่งผลกระทบต่อค่าที่จะพยากรณ์ แล้วนำค่าที่ได้มาคำนวณ เพื่อพิจารณาหาลักษณะการเกิด แล้วจดจำใช้ในการพยากรณ์ครั้งต่อไป (Makridakis, Wheelright, & Hyndman, 2000; ไฮเซอร์, 2551)

2.3.2 ประเภทของการพยากรณ์

2.3.2.1 การพยากรณ์เชิงคุณภาพ

คือ การพยากรณ์แบบใช้ความรู้และประสบการณ์ที่ผ่านมาในอดีตมาเป็นเกณฑ์ในการตัดสินใจ สามารถแบ่งย่อยได้ 4 ประเภท ดังนี้

1) การพยากรณ์จากความคิดเห็นของผู้เชี่ยวชาญ หรือผู้บริหาร โดยจะมีการผสมผสานกระบวนการทางสถิติมาประยุกต์ร่วม เพื่อประกอบการพยากรณ์
2) การพยากรณ์จากความคิดเห็นของฝ่ายขาย โดยจะใช้ประสบการณ์จากการทำงานของพนักงานฝ่ายขายมาประกอบการพยากรณ์
3) การพยากรณ์จากข้อมูลการสำรวจตลาดที่ได้จากลูกค้า โดยอาจจะใช้แบบสอบถามมาเป็นเครื่องมือช่วยในการเก็บความคิด ความต้องการของลูกค้า และปัจจัยที่จะทำให้ลูกค้าตัดสินใจใช้บริการ
4) การพยากรณ์โดยอาศัยวิธีแบบเดลฟาย คือ วิธีการที่ประกอบด้วยหน้าที่ 3 ส่วน คือ ผู้ตัดสินใจจะเป็นกลุ่มคนที่มีความชำนาญ มีความรู้ความสามารถในด้านที่จะทำการพยากรณ์ เพื่อรับหน้าที่ตัดสินใจ ต่อมาคือทีมงานจะเป็นกลุ่มคนที่ทำหน้าที่จัดเก็บ และเตรียมข้อมูล เพื่อส่งต่อให้กับผู้ตัดสินใจ ใช้ในการประกอบการตัดสินใจ และส่วนสุดท้าย ผู้ตอบคำถาม คือกลุ่มของลูกค้าที่จะให้ข้อมูล โดยอาจจะผ่านแบบสอบถาม เพื่อให้ผู้ตัดสินใจทราบถึงความต้องการของกลุ่มลูกค้า และสามารถตอบสนองความต้องการนั้นได้อย่างตรงประเด็น

2.3.2.2 พยากรณ์เชิงปริมาณ

คือ การพยากรณ์ที่ใช้การคำนวณทางคณิตศาสตร์ หรือหลักทางสถิติมาช่วยเป็นเกณฑ์ในการประกอบตัดสินใจร่วมกับข้อมูลในอดีตที่ผ่านมา สามารถแบ่งได้เป็น แบบอนุกรมเวลา คือ การพยากรณ์ที่มีการนำตัวแปรของเวลา มาเป็นสิ่งที่ใช้ในการคำนวณทางคณิตศาสตร์ หรือใช้ประกอบกับหลักทางสถิติ สามารถแบ่งย่อยได้ 5 ประเภท ดังนี้

1) อนุกรมเวลาแบบตรงตัว คือ การพยากรณ์จากการใช้ค่าจริงในช่วงเวลาก่อนหน้า มาเป็นค่าพยากรณ์

$$F_t = A_{t-1}$$

โดย

F_t คือ ค่าพยากรณ์ในช่วงเวลาที่ต้องการ

A_{t-1} คือ ค่าจริงในช่วงเวลาก่อนหน้า

2) อนุกรมเวลาแบบค่าเฉลี่ยเคลื่อนที่อย่างง่าย คือ การพยากรณ์ที่ใช้ค่าจริงในช่วงเวลาก่อนหน้า n ช่วง มาเฉลี่ยเป็นค่าพยากรณ์

$$F_t = \frac{A_{t-1} + A_{t-2} + \ldots + A_{t-n}}{n}$$

โดย

F_t คือ ค่าพยากรณ์ในช่วงเวลาที่ต้องการ

$A_{t-1}, A_{t-2}, \ldots, A_{t-n}$ คือ ค่าจริงในช่วงเวลาก่อนหน้า

n คือ จำนวนของค่าจริงในช่วงเวลาก่อนหน้าที่จะนำมาคำนวณค่าเฉลี่ย

3) อนุกรมเวลาแบบค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก คือ การพยากรณ์ที่ใช้ค่าจริงในช่วงเวลาก่อนหน้า n ช่วง มาถ่วงน้ำหนักแล้วหาค่าเฉลี่ย ซึ่งจะได้เป็นค่าพยากรณ์

$$F_t = \frac{w_{t-1}A_{t-1} + w_{t-2}A_{t-2} + \ldots + w_{t-n}A_{t-n}}{\sum w}$$

โดย

F_t คือ ค่าพยากรณ์ในช่วงเวลาที่ต้องการ

$A_{t-1}, A_{t-2}, \ldots, A_{t-n}$ คือ ค่าจริงในช่วงเวลาก่อนหน้า

w คือ ค่าถ่วงน้ำหนักในแต่ละช่วงเวลา

4) อนุกรมเวลาการปรับเรียบแบบเอ็กซ์โปเนนเชียล

$$F_t = F_{t-1} + \alpha(A_{t-1} - F_{t-1})$$

โดย

F_t คือ ค่าพยากรณ์ในช่วงเวลาที่ต้องการ

A_{t-1} คือ ค่าจริงในช่วงเวลาก่อนหน้า

F_{t-1} คือ ค่าพยากรณ์ในช่วงเวลาก่อนหน้า

α คือ ค่าปรับเรียบ โดย $0 < \alpha < 1$

5) อนุกรมเวลาแบบคาดคะเนแนวโน้ม คือ การพยากรณ์จากแนวโน้มของสมการเส้นตรง ที่ได้จากการหาสมการเส้นตรงของข้อมูลในอดีต ที่มีค่าความคลาดเคลื่อนน้อยที่สุด

$$Y_t = a + bt$$

โดย

Y_t คือ ค่าพยากรณ์ในช่วงเวลาที่ต้องการ

a คือ ค่าของ Y_t เมื่อ $t = 0$

b คือ ค่าความชันของเส้นตรง

t คือ ช่วงเวลาที่ต้องการ

และการพยากรณ์เชิงปริมาณอีกแบบหนึ่งคือแบบปัจจัยสาเหตุ คือ การพยากรณ์ที่อาศัยความสัมพันธ์ของตัวแปรอิสระที่ส่งผลกระทบกับค่าที่ต้องการพยากรณ์ โดยจะต้องมีความสัมพันธ์แบบสมการเส้นตรง

$$\hat{Y} = a + bx$$

โดย

$\hat{Y}$ คือ ค่าพยากรณ์ในช่วงเวลาที่ต้องการ

a คือ ค่าคงที่ของจุดตัดบนแกน Y

b คือ ค่าความชันของเส้นตรง

x คือ ค่าของตัวอิสระที่ไม่ใช่เวลา

2.4 งานวิจัยที่เกี่ยวข้อง

งานวิจัยของ (Hu, Chen, & McCain, 2004) ได้ทำการวิจัยเกี่ยวกับการวางแผนและจัดการในการพยากรณ์ระยะสั้นของร้านอาหารในคาสิโน โดยในงานวิจัยได้ใช้เครื่องมือในการสร้างโมเดล คือ Naïve model, Single Moving Average model, Double Moving Average model, Single Exponential Smoothing model, Holt-Winters method และ Regression model ซึ่งเครื่องมือเหล่านี้เป็นเครื่องมือในด้านสถิติในเรื่องของไทม์ซีรี่ส์ โดยในงานวิจัยนี้ได้วัดผลลัพธ์ทางประสิทธิภาพของการพยากรณ์จากค่า MAPE (Mean Absolute Percentage Error) และค่า RMSPE (Root Mean Square Percentage Error) ซึ่งผลลัพธ์ที่สามารถพยากรณ์ได้ค่าความถูกต้องมากที่สุด จะเป็นของโมเดลแบบ Double Moving Average model ที่มีค่าของ MAPE อยู่ที่ 6.68% และค่าของ RMSPE อยู่ที่ 8.92% โดยในส่วนสุดท้ายของงานวิจัย (Hu et al., 2004) ได้แนะนำถึงการค้นหาหลักการสร้างโมเดลพยากรณ์รูปแบบใหม่ มาประยุกต์รวมกับหลักการพยากรณ์รูปแบบเดิมเพื่อให้มีประสิทธิภาพในการพยากรณ์ที่ดีขึ้น

ต่อมากลุ่มนักวิจัยชาวญี่ปุ่น (Tanizaki, Hoshino, Shimmura, & Takenaka, 2019) ได้นำ Machine Learning และหลักทางสถิติ มาใช้ในการพยากรณ์จำนวนความต้องสินค้าของร้านอาหาร โดยได้ทำทั้งหมด 4 วิธี คือ Bayesian Linear Regression, Boosted Decision Tree Regression, Decision Forest Regression และ Stepwise โดยข้อมูลที่ใช้เป็นข้อมูลจากข้อมูลจริงจากเครื่อง POS ของในแต่ละสาขา และใช้การวัดผลลัพธ์ของประสิทธิภาพการพยากรณ์จากค่าของ R-Square ซึ่งผลลัพธ์ประสิทธิภาพการพยากรณ์ที่ได้ของวิธี Bayesian Linear Regression, Decision Forest Regression และ Stepwise ให้ค่าอัตราการพยากรณ์ที่ไม่แตกต่างกัน แต่จะมีวิธี Boosted Decision Tree Regression ที่จะได้ค่าอัตราการพยากรณ์ที่ต่ำเล็กน้อย จึงนำมาสู่การนำงานวิจัยมาปรับปรุง และต่อยอดในปีต่อมา

ในปีต่อมากลุ่มนักวิจัยชาวญี่ปุ่น (Tanizaki, Hoshino, Shimmura, & Takenaka, 2020) ได้ทำงานวิจัยเกี่ยวกับการจัดการร้านอาหาร โดยเน้นการพยากรณ์จากความต้องการของลูกค้า ที่อาศัย Machine Learning มาสร้างโมเดลการพยากรณ์ โดยในงานวิจัยได้ใช้กระบวนการ Random Forest และ Random Forest Regression โดยอาศัยชุดข้อมูลจากเครื่อง POS ของในแต่ละสาขา เหมือนงานวิจัยก่อนหน้า (Tanizaki et al., 2019) ซึ่งผลลัพธ์ที่ได้จากการพยากรณ์จากสินค้าคงคลังมีค่าอัตราการพยากรณ์อยู่ที่ 71% ซึ่ง Takashi Tanizaki กล่าวว่าเป็นค่าความถูกต้องที่ไม่สูง

และในปี 2020 งานวิจัยของ (Geevers, 2020) ได้นำกระบวนการ Deep Reinforcement Learning (DRL) มาประยุกต์ใช้กับการจัดการสินค้าคงคลัง โดยใช้หลักการของ Q-Learning ซึ่งพบกับข้อจำกัดของตาราง Q-Table ที่ทำหน้าที่จัดเก็บค่า Q-Value ที่เมื่อมีค่าเกิน 60 ล้านเซลล์จะไม่สามารถเก็บเพิ่มได้อีก Geevers จึงตัดสินใจเปลี่ยนมาใช้หลักการของ Proximal Policy Optimization ซึ่งให้ผลลัพธ์ที่ออกมาจากการนำไปทดสอบกับบอร์ดเกม Beer game, Divergent และ CBC โดยเปรียบเทียบค่า Total costs ระหว่างค่าที่ได้จาก Proximal Policy Optimization กับค่าที่ได้จากการ Benchmark ซึ่งผลลัพธ์ที่ได้ Proximal Policy Optimization ให้ค่า Total costs ที่น้อยกว่าการ Benchmark ทั้ง 2 เกมส์ คือ Beer game และ Divergent

จากงานวิจัยที่ได้กล่าวมาข้างต้น จึงนำมาสู่แรงจูงใจในการค้นหาหลักการสร้างโมเดลพยากรณ์รูปแบบใหม่ และได้พบกับกระบวนการ Reinforcement Learning ที่มีจุดเด่นในเรื่องของการเรียนรู้ที่ได้แนวคิดมาจากหลักการเรียนรู้ในการลองผิดลองถูก ในแต่ละสถานการณ์ของมนุษย์ ทำให้เกิดเป็นการเรียนรู้ที่ต่อเนื่อง และสามารถพัฒนาผลลัพธ์ที่ได้จากการเรียนรู้ให้มีประสิทธิภาพเพิ่มสูงขึ้นได้อย่างต่อเนื่อง ซึ่งจะแตกต่างกับการเรียนรู้แบบ Supervise Learning ที่จะมีการเรียนรู้รูปแบบเฉพาะจากชุดข้อมูลที่นำมาให้โมเดลเรียนรู้เท่านั้น แล้วนำผลลัพธ์ที่ได้ไปปรับใช้กับข้อมูลชุดใหม่ (Sutton & Barto, 2018) ดังนั้นในงานวิจัยนี้จึงใช้กระบวนการ Reinforcement Learning มาเป็นเครื่องมือในการสร้างโมเดลการพยากรณ์ โดยอาศัยหลักการ Proximal Policy Optimization ที่มีจุดเด่นในเรื่องของการนำ stochastic gradient ascent ของหลาย epoch มาเป็นตัวอัพเดต policy ของระบบ และอีกหนึ่งจุดเด่นจะเป็นในเรื่องของระบบ Trust Region ที่มีความเสถียรและน่าเชื่อถือ และสุดท้ายเป็นเรื่องของการเขียนโค้ดที่สามารถเขียนได้ง่าย (Schulman et al., 2017)

บทที่ 3

วิธีการวิจัย

3.1 ชุดข้อมูล

ชุดข้อมูลที่ใช้ในงานวิจัยครั้งนี้จะเป็นชุดข้อมูลจำลอง (Synthetic Data) ที่อาศัยหลักการจำลองข้อมูลแบบแจกแจงปกติ (Normal Distribution) เนื่องมาจากชุดข้อมูลจริงที่ได้มาจากเครื่อง POS ในแต่ละสาขาของบริษัท เอ เป็นชุดข้อมูลที่มีความแปรปรวนที่สูงและไม่มีแบบฟอร์มการกระจายตัวของข้อมูลที่แน่นอน โดยผู้วิจัยได้ทำการคัดกรองข้อมูลที่เป็น Outlier และข้อมูลส่วนที่ไม่เกี่ยวข้องออก แต่ชุดข้อมูลยังคงมีความแปรปรวนที่สูงอยู่ ผู้วิจัยจึงตัดสินใจใช้ชุดข้อมูลจำลอง โดยใช้เลือกใช้ไลบรารี่ของ Numpy ในส่วนคำสั่งของการ Random แบบ Normal Distribution มาเป็นคำสั่งที่ใช้ในการจำลอง โดยในชุดข้อมูลจำลองจะมีการแยกพฤติกรรมการใช้วัตถุดิบเบื้องต้นเป็นแบบความต้องการวัตถุดิบในวันธรรมดา และความต้องการใช้วัตถุดิบในวันหยุด โดยค่าเฉลี่ยที่ใช้ในการจำลองชุดข้อมูล นำมาจากค่าเฉลี่ยของชุดข้อมูลจริงที่ได้จากเครื่อง POS และกำหนดค่าส่วนเบี่ยงเบนมาตรฐานอยู่ที่ 2 และเนื่องจากในการจำลองชุดข้อมูลชุดนี้มีการแบ่งการจำลองเป็นชุดข้อมูลของพฤติกรรมการใช้วัตถุดิบแบบความต้องการวัตถุดิบในวันธรรมดา และความต้องการใช้วัตถุดิบในวันหยุด ทำให้เมื่อนำมารวมกันเป็นชุดข้อมูลจำลองที่จะนำมาใช้ จึงมีการกระจายตัวของข้อมูลเป็นแบบ Normal Distribution 2 ชุด คือ ชุดการใช้วัตถุดิบในวันธรรมดา และการใช้วัตถุดิบในวันหยุด ซึ่งทำให้ชุดข้อมูลจำลองมีความใกล้เคียงกับชุดข้อมูลจริงมากยิ่งขึ้น

โดยชุดข้อมูลจำลองจะแบ่งย่อยตามโมเดลที่จะทำการเรียนรู้ต่อออกเป็น 2 ประเภท คือ แบบ One-Attribute ที่จะประกอบด้วย Attribute ของการใช้วัตถุดิบในแต่ละวันเพียง Attribute เดียว และแบบ Multi-Attribute ที่จะประกอบด้วย Attribute 4 ตัว คือ Attribute ของการใช้วัตถุดิบในแต่ละวัน ที่เป็นข้อมูลชุดเดียวกับแบบ One-Attribute ต่อมาจะเป็น Attribute ของความชื้นสัมพัทธ์ในบรรยากาศกับค่าฝุ่นละออง PM2.5 ที่อาจส่งผลต่อการเก็บรักษาวัตถุดิบ และสุดท้ายจะเป็น Attribute ของอัตราการเกิดฝนที่อาจส่งผลต่อการเข้าใช้บริการที่สาขาของลูกค้า

3.2 เครื่องมือที่ใช้ในการวิจัย

งานวิจัยนี้จะใช้ชุดข้อมูลมาประกอบกับการใช้เทคนิค Reinforcement Learning เพื่อให้โรบอทเรียนรู้ในการหาแนวทางการตัดสินใจที่ให้ได้ค่าการสั่งวัตถุดิบที่สอดคล้องกับการขายอาหารในแต่ละสภาพแวดล้อมมากที่สุด โดยงานวิจัยนี้จะใช้ Proximal Policy Optimization เป็นอัลกอริทึมในการเรียนรู้ของ Reinforcement Learning

3.3 แผนการดำเนินการวิจัย

งานวิจัยนี้เป็นการสร้างโมเดลการพยากรณ์ค่าการสั่งวัตถุดิบ โดยใช้หลักการ Proximal Policy Optimization ของกระบวนการ Reinforcement Learning โดยในงานวิจัยจะแบ่งการศึกษาออกเป็น 2 รูปแบบ คือ 1. แบบจำลองแบบหนึ่งแอตทริบิวต์ 2. แบบจำลองแบบหลายแอตทริบิวต์ และทำการเปรียบเทียบประสิทธิภาพของการพยากรณ์และ ความรวดเร็วในการเรียนรู้ ซึ่งจะทำการประเมินผลประสิทธิภาพจากเครื่องมือการประเมิน 3 ด้าน คือ F-Statistics ,R-Square และRMSE

3.4 กระบวนการวิจัย

ผู้วิจัยได้วางกระบวนการวิจัย เพื่อให้งานวิจัยสามารถดำเนินการได้ตามวัตถุประสงค์ โดยมีขั้นตอนการดำเนินงานดังนี้

3.4.1 การสร้างชุดข้อมูลสังเคราะห์ (Synthetic Data)

1) การเก็บรวบรวมหาค่าเฉลี่ยของความต้องการในการใช้วัตถุดิบ และค่าสูงสุดและต่ำสุดของแอตทริบิวต์อื่นเพิ่มเติม เช่น ค่าสูงสุดและค่าต่ำสุดของอัตราความชื้นสัมพัทธ์ , ค่าสูงสุดและค่าต่ำสุดของค่าฝุ่นละออง PM2.5 , ค่าสูงสุดและค่าต่ำสุดของอัตราการที่จะเกิดฝน โดยผู้วิจัยได้ทำการเก็บรวบรวมค่าข้อมูลจาก 2 เว็บไซต์ คือ https://www.wunderground.com และ http://berkeleyearth.lbl.gov เพื่อนำข้อมูลดังกล่าวมาหาค่าเฉลี่ยของแต่ละตัวแปร สำหรับใช้ในการสร้างชุดข้อมูลสังเคราะห์

Daily Observations

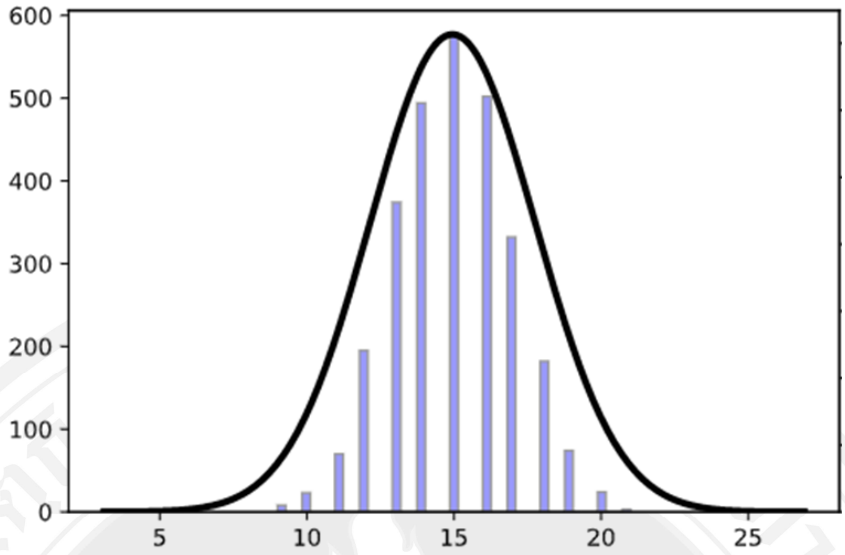
Time	Temperature (° F)			Dew Point (° F)			Humidity (%)			Wind Speed (mph)			Pressure (Hg)		
Jan	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min
1	84	77.0	70	68	63.5	61	78	63.6	51	12	6.6	0	30.1	30.0	29.9
2	86	78.1	72	66	63.0	61	69	59.9	45	12	7.5	3	30.1	30.0	29.9
3	81	77.5	73	64	63.1	61	69	60.8	54	15	8.5	5	30.1	30.0	29.9
4	82	78.8	75	68	65.7	64	73	64.7	58	12	8.2	2	30.0	29.9	29.9
5	82	79.0	77	72	68.9	66	78	71.3	65	12	7.3	2	30.0	29.9	29.9
6	91	81.1	75	75	73.3	72	94	78.4	55	8	3.5	0	30.0	29.9	29.9
7	90	83.1	75	75	73.8	72	94	75.3	59	12	5.8	0	30.0	29.9	29.8
8	88	82.1	77	79	76.6	73	100	84.0	70	10	5.0	0	29.9	29.9	29.8
9	88	80.9	75	79	75.8	73	100	85.6	66	10	3.5	0	30.0	29.9	29.8
10	90	82.8	75	79	74.5	72	100	77.8	55	8	2.8	0	30.0	29.9	29.8
11	91	84.3	77	77	73.6	70	94	72.1	49	12	3.3	0	30.0	29.9	29.8
12	93	85.6	77	77	73.2	68	94	68.2	44	10	4.0	0	30.0	29.9	29.8

ภาพที่ 3.1 ตัวอย่างข้อมูลจากเว็บไซต์ Wunderground

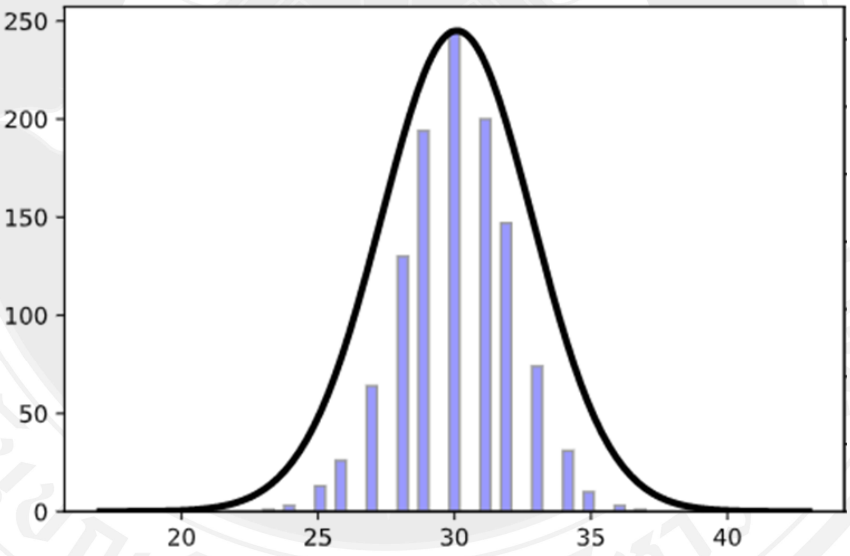
```
% Country: Thailand
% Region: Thailand
% Population: NaN
% Max Latitude: 20.4632
% Max Longitude: 105.6394
% Min Latitude: 5.61
% Min Longitude: 97.3456
% Year, Month, Day, UTC Hour, PM2.5, PM10_mask, Retrospective
2016    3       3       8       53.99   1.00    0
2016    3       3       9       53.95   1.00    0
2016    3       3       10      54.10   1.00    0
2016    3       3       11      53.93   1.00    0
2016    3       3       12      41.23   1.00    0
2016    3       3       13      42.44   1.00    0
2016    3       3       14      38.70   1.00    0
2016    3       3       15      39.44   1.00    0
2016    3       3       16      39.35   1.00    0
2016    3       3       17      54.49   1.00    0
2016    3       3       19      61.38   1.00    0
2016    3       3       20      57.11   1.00    0
2016    3       3       21      51.03   1.00    0
2016    3       3       22      45.63   1.00    0
2016    3       3       23      41.83   1.00    0
```

ภาพที่ 3.2 ตัวอย่างข้อมูลจากเว็บไซต์ Berkeleyearth

2) การจำลองข้อมูลโดยอาศัยหลักการการกระจายตัวแบบปกติของข้อมูล (Normal Distribution) โดยเริ่มจากการใช้ไลบรารี่ Numpy และเลือกใช้คำสั่ง numpy.random.normal(Mean, Standard Derivation, Size) โดยในข้อมูลสังเคราะห์นี้จะมีการจำลองข้อมูลความต้องการของวัตถุดิบออกเป็น 2 แบบหลักๆ คือ ความต้องการในการใช้วัตถุดิบแบบวันธรรมดา และความต้องการในการใช้วัตถุดิบแบบวันหยุด ดังนั้นค่าเฉลี่ยที่จะป้อนให้คำสั่งจึงมี 2 ค่า คือ ความต้องการในการใช้วัตถุดิบแบบวันธรรมดามีค่าเท่ากับ 15 และความต้องการในการใช้วัตถุดิบแบบวันหยุดมีค่าเท่ากับ 30 ส่วนค่าส่วนเบี่ยงเบนมาตราฐาน (Standard Derivation) กำหนดให้มีค่าเท่ากับ 2 และขนาดของจำนวนที่จะจำลองข้อมูลเท่ากับ 4,000 โดยแบ่งเป็นชุดข้อมูลสำหรับเรียนรู้ (Training Set) จำนวน 3,000 record และส่วนที่เหลืออีก 1,000 record กำหนดให้เป็นชุดข้อมูลในการทดสอบ (Test Set)



ภาพที่ 3.3 ตัวอย่างการกระจายตัวของข้อมูลความต้องการในการใช้วัตถุดิบแบบวันธรรมดา



ภาพที่ 3.4 ตัวอย่างการกระจายตัวของข้อมูลความต้องการในการใช้วัตถุดิบแบบวันหยุด

3) ข้อมูลสังเคราะห์ที่ได้จะนำมาแบ่งเป็นชุดข้อมูลแบบหนึ่งแอตทริบิวต์ ที่เป็นชุดข้อมูลที่มีเฉพาะค่าของความต้องการในการใช้วัตถุดิบในแต่ละวัน และชุดข้อมูลแบบหลายแอตทริบิวต์ ที่จะมีข้อมูลของค่าความชื้นสัมพัทธ์ ค่าค่าฝุ่นละออง PM2.5 และค่าอัตราการเกิดฝนมารวมกับค่าของความต้องการในการใช้วัตถุดิบในแต่ละวัน

Day	Qty
Monday	15
Tuesday	18
Wednesday	17
Thursday	15
Friday	16
Saturday	30
Sunday	33
Monday	16
Tuesday	18
Wednesday	15
Thursday	16
Friday	13
Saturday	28
Sunday	29

ภาพที่ 3.5 ตัวอย่างชุดข้อมูลสังเคราะห์ แบบหนึ่งแอตทริบิวต์

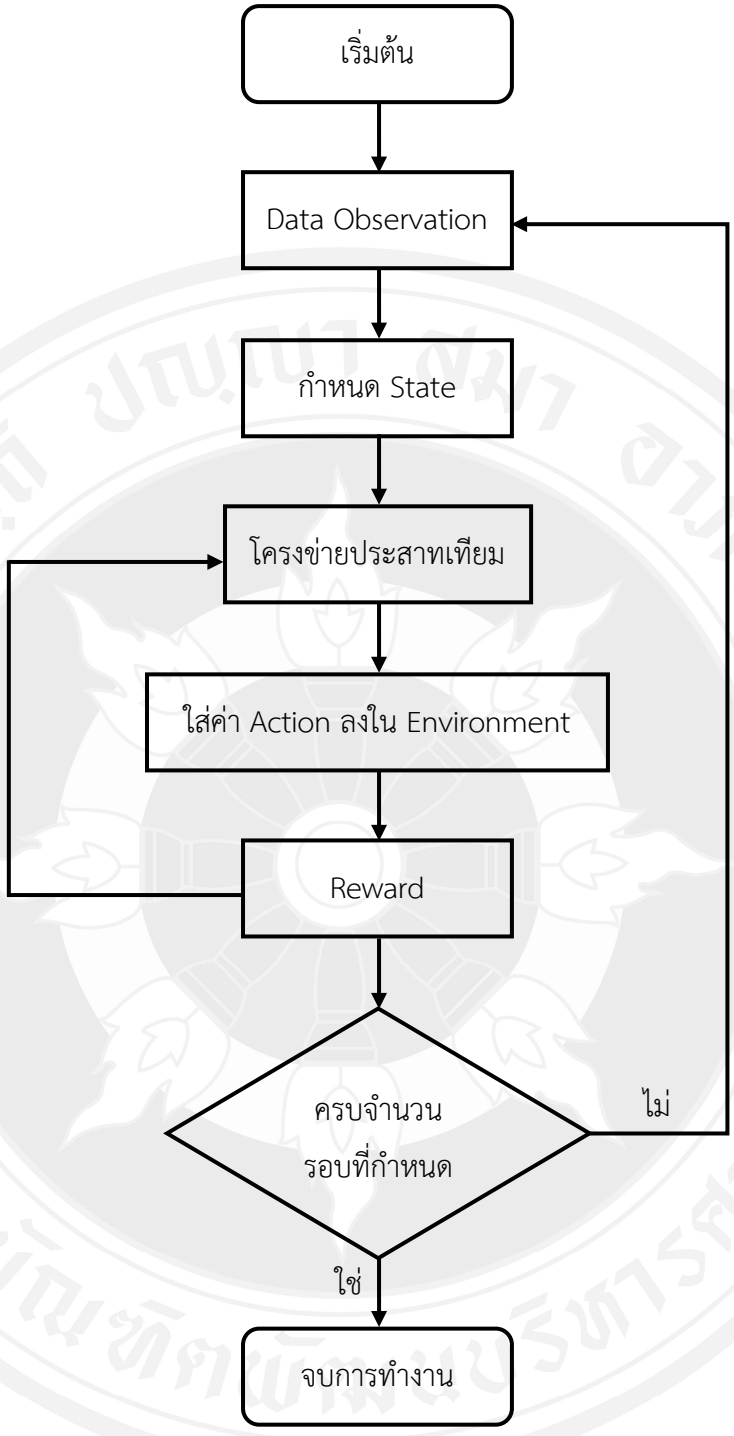
Day	Qty	Humidity	PM25	Rainy
Monday	15	73	98	69
Tuesday	18	70	86	60
Wednesday	17	71	90	63
Thursday	15	73	98	69
Friday	16	72	94	66
Saturday	30	58	38	24
Sunday	33	55	26	15
Monday	16	72	94	66
Tuesday	18	70	86	60
Wednesday	15	73	98	69
Thursday	16	72	94	66
Friday	13	75	106	75
Saturday	28	60	46	30
Sunday	29	59	42	27

ภาพที่ 3.6 ตัวอย่างชุดข้อมูลสังเคราะห์ แบบหลายแอตทริบิวต์

3.4.2 การสร้างระบบ Reinforcement Learning

ระบบ Reinforcement Learning ที่ได้สร้างขึ้นมาประกอบด้วยองค์ประกอบหลักๆ 5 ส่วน ด้วยกัน ดังนี้

1) ส่วนของโรบอทที่จะต้องกำหนดโครงข่ายประสาทเทียม (Neuron Network) ที่ใช้ในการรับค่าข้อมูลที่ได้ Observation มา แล้วป้อนให้กับโครงข่ายประสาทเทียม เพื่อหาค่าน้ำหนักของโครงข่ายประสาทเทียม แล้วได้ผลลัพธ์ออกมาเป็น Action ที่เหมาะสมที่จะกระทำในสถานการณ์ดังกล่าว
2) ส่วนของการนำเข้าข้อมูล (Observation Data) เป็นการเก็บรวบรวมปัจจัยสภาพแวดล้อมที่จะป้อนให้กับโรบอท เพื่อใช้ในการเรียนรู้และทำการตัดสินใจ
3) ส่วนของ Action ที่จะทำหน้าที่นำค่า Action ที่ได้จากโครงข่ายประสาทเทียม มากระทำต่อสภาพแวดล้อมที่ได้ Observation มา เพื่อใช้ในการประเมินผลประสิทธิภาพจากค่า Reward
4) ส่วนของ State ที่ทำหน้าที่เก็บรวบรวมข้อมูลที่ได้จากการเรียนรู้ว่าใน State ที่มีสภาพแวดล้อมแบบที่มีข้อมูลป้อนเข้ามา ควรจะต้องตัดสินใจอย่างไร ที่จะทำให้ได้ผลลัพธ์ออกมาดีที่สุด และนำไปใช้เมื่อมี State ในแบบเดียวกันที่เกิดขึ้นในอนาคต เพื่อให้สามารถตัดสินใจในการออก Action ของโรบอทได้อย่างถูกต้อง และรวดเร็ว
5) ส่วนของ Reward ที่จะเป็นตัววัดผลว่าการตัดสินใจในการกระทำสิ่งหนึ่งลงไปในสถานการณ์ที่ได้ได้รับมามีความเหมาะสมอยู่ในระดับใด โดยในงานวิจัยนี้จะวัดผลของ Reward ที่เกิดขึ้นจากผลการกำไรที่เหลือ เมื่อถูกหักคะแนนจากบทลงโทษในสถานการณ์ที่ตัดสินใจสั่งวัตถุดิบขาดหรือเกินจากวัตถุดิบที่ใช้จริง



ภาพที่ 3.7 แผนผังสรุปการดำเนินงานของระบบ Reinforcement Learning

3.4.3 หลักการคิด Reward ของระบบการจัดการวัตถุดิบคงคลัง

กระบวนการคิดค่า Reward ในงานวิจัยฉบับนี้จะเริ่มจากการคิดหาค่าผลกำไร โดยการนำค่า N ที่เป็นจำนวนสินค้าที่ขายได้ มาคูณกับกำไรต่อชิ้นของสินค้า จากนั้นนำค่าที่ได้ไปหักลบกับค่าของบทลงโทษที่คำนวณจากการนำ ค่า n ที่เป็นค่าของจำนวนชิ้นที่สั่งวัตถุดิบขาดหรือเกินจำนวน คูณกับค่าเฟกเตอร์ของบทลงโทษ โดยค่าเฟกเตอร์ที่ในงานวิจัยฉบับนี้ใช้จะแบ่งออกเป็น 2 อย่าง คือ ถ้าจำนวนชิ้นที่สั่งวัตถุดิบขาดจะให้ค่าเฟกเตอร์บทลงโทษไว้ที่ 1.25 และถ้าจำนวนชิ้นที่สั่งวัตถุดิบเกินจะให้ค่าเฟกเตอร์บทลงโทษไว้ที่ 1 เนื่องจากการสั่งวัตถุดิบขาดจะทำให้เกิดการสูญเสียโอกาสในการขายสินค้าผู้วิจัยจึงกำหนดให้มีค่าเฟกเตอร์บทลงโทษที่สูงกว่า กรณีที่สั่งวัตถุดิบเกิน

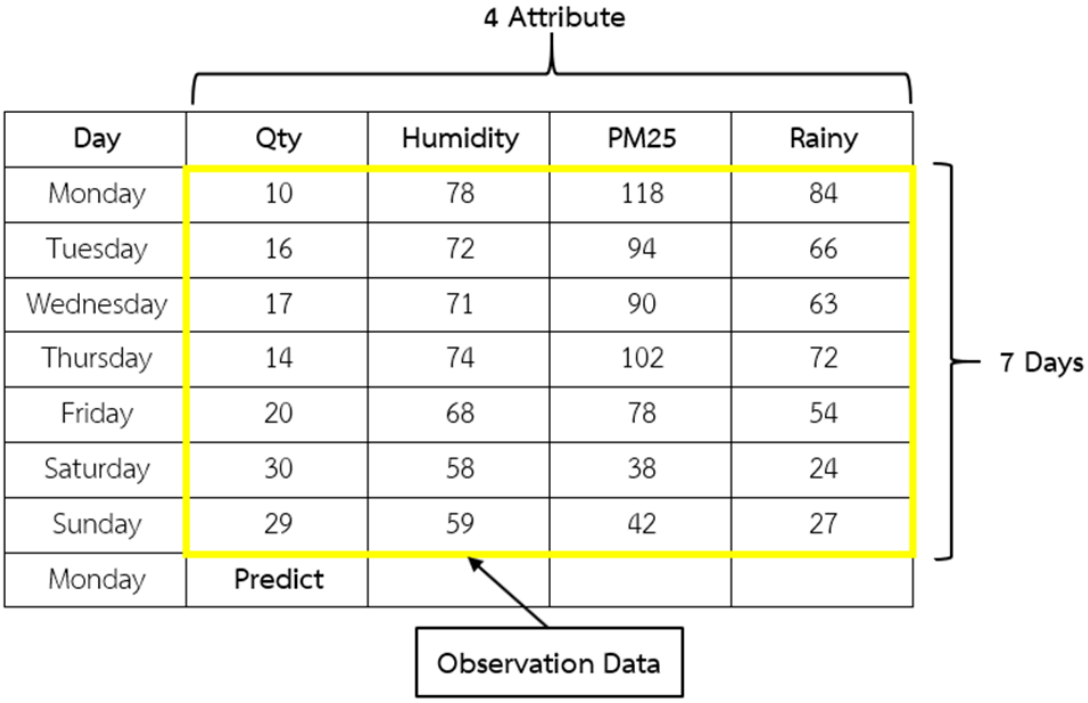
จากนั้นนำค่าผลกำไรที่หักค่าของบทลงโทษในสมการที่ 3.1 มาหารด้วยค่าผลกำไรจริงที่ควรจะได้ คือ ค่า N ที่เป็นจำนวนสินค้าที่ขายได้ มาคูณกับกำไรต่อชิ้นของสินค้า จะได้ออกมาเป็นค่า Reward ดังสมการที่ 3.2 จากนั้นจะทำการ Normalize ค่า Reward ที่ได้ให้อยู่ในช่วงระหว่าง -1 ถึง 1 เนื่องจากระบบในงานวิจัยนี้ใช้ Activation function tanh ที่มีจุดเด่นในเรื่องของการ Classification ระหว่าง 2 สิ่ง โดยในงานวิจัยนี้จะเป็นการแยกระหว่าง การสั่งวัตถุดิบเพิ่มขึ้น หรือลดจำนวนการสั่งวัตถุดิบลง

$$Profit = (N\ x\ Profit\ per\ each) - (n\ x\ Penalty) \quad (3.1)$$

$$Reward = Profit \div (N\ x\ Profit\ per\ each) \quad (3.2)$$

3.4.4 การ Observation ชุดข้อมูลให้ระบบ Reinforcement Learning เรียนรู้

ระบบ Reinforcement Learning มีการ Observation ชุดข้อมูลเพื่อใช้ในการเรียนรู้ของโมเดลการพยากรณ์ในแต่ละรอบการเรียนรู้ โดยจากภาพที่ 3.8 Observation Data ที่ใช้ป้อนให้กับระบบเรียนรู้ในแต่ละรอบจะแบ่งเป็นการป้อนตัวแอตทริบิวต์ ดังในภาพที่ 3.8 จะเป็นการป้อนทั้งหมดเป็นจำนวน 4 แอตทริบิวต์ และในแต่ละแอตทริบิวต์จะมีการป้อนชุดข้อมูลย้อนหลังกลับไปเป็นเวลา 7 วัน นับตั้งแต่วันที่ต้องการพยากรณ์ ทำให้ในแต่ละรอบการเรียนรู้จะมีการป้อนตัวแปรให้กับระบบทั้งหมดเป็นจำนวน 28 ตัวแปร ดังกรอบสีเหลืองในภาพที่ 3.8



ภาพที่ 3.8 ตัวอย่างชุดข้อมูล Observation ที่ระบบ Reinforcement Learning ใช้สำหรับเรียนรู้ในแต่ละรอบการเรียนรู้

3.4.5 การประเมินผลประสิทธิภาพโมเดลการพยากรณ์

เครื่องมือที่ใช้สำหรับการประเมินประสิทธิภาพโมเดลการพยากรณ์ โดยสามารถแบ่งได้ออกเป็นการประเมินประสิทธิภาพใน 3 ด้าน คือ

1) การประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า F-Statistics จะเป็นการวัดผลประสิทธิภาพโมเดลการพยากรณ์ในด้านของความแม่นยำของโมเดล โดยจะเป็นการวัดประสิทธิภาพโมเดลการพยากรณ์จากค่าของความคลาดเคลื่อนที่โมเดลการพยากรณ์ได้ทำการพยากรณ์ออกมา แล้วยังสามารถอยู่ภายใต้ค่านัยสำคัญที่กำหนดไว้
2) การประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า R-Square เป็นการวัดผลประสิทธิภาพโมเดลการพยากรณ์ ในเรื่องของการที่ค่าตัวแปรที่ป้อนให้กับโมเดลการพยากรณ์ เพื่อใช้ในการเรียนรู้ สามารถอธิบายค่าที่พยากรณ์ออกมาได้ เป็นจำนวนกี่เปอร์เซนต์ หรือหมายความว่าค่าอินพุตที่ป้อนให้กับโมเดลการพยากรณ์ ใช้ในการเรียนรู้ มีความสอดคล้องกับค่าที่พยากรณ์ออกมาได้ เป็นจำนวนกี่เปอร์เซนต์
3) การประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า Root Mean Square Error เป็นการวัดผลประสิทธิภาพโมเดลการพยากรณ์ ในเรื่องของค่าความผิดพลาดโดยเฉลี่ยของโมเดลที่ได้ทำการพยากรณ์ค่าออกมา

บทที่ 4

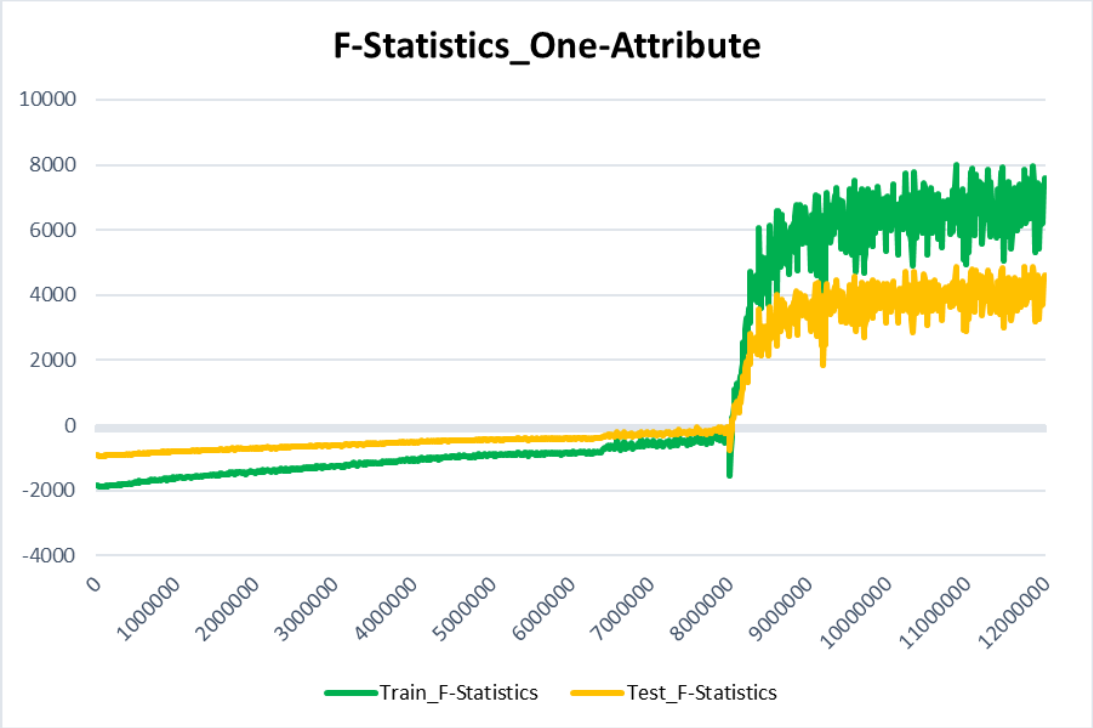
ผลการศึกษา

งานวิจัยครั้งนี้ใช้การประมวลผลการเรียนรู้โมเดล ภายใต้ระบบ Google Cloud Platform โดยใช้เครื่องแบบ n2-standard-4 มีหน่วยประมวลผลกลางจำนวน 4 cores หน่วยความจำหลัก (RAM) 16GB ระบบปฏิบัติการ Debian 9 โดยงานวิจัยนี้จะทำการเรียนรู้โมเดล 2 รูปแบบ จำนวนรูปแบบละ 12 ล้าน timesteps โดยรูปแบบแรกเป็นโมเดลแบบหนึ่งแอตทริบิวต์ ใช้เวลาในการเรียนรู้ทั้งหมด 11 ชั่วโมง ส่วนรูปแบบที่สองเป็นโมเดลแบบหลายแอตทริบิวต์ ใช้เวลาในการเรียนรู้ทั้งหมด 12 ชั่วโมง โดยทั้งสองรูปแบบจะป้อนข้อมูลแบบ Time-Series ย้อนหลัง 7 วัน ทุก Attribute ของแต่ละโมเดล เพื่อใช้ในการพยากรณ์ค่าวัตถุดิบ โดยโมเดลแบบหนึ่งแอตทริบิวต์ จะป้อนเฉพาะค่าวัตถุดิบที่ใช้จริงย้อนหลัง 7 วัน เพื่อใช้เป็นปัจจัยให้อัลกอริทึม Proximal Policy Optimization ใช้ในการเรียนรู้และสร้างเป็นโมเดลออกมา ส่วนโมเดลแบบหลายแอตทริบิวต์ จะมีการเพิ่ม Attribute อีก 3 ตัวคือ Humidity, PM2.5, Rainy ที่มีความสัมพันธ์กับ Attribute ของค่าวัตถุดิบที่ใช้ แล้วเรียนรู้ด้วยกระบวนการของอัลกอริทึม Proximal Policy Optimization

4.1 แบบจำลองแบบหนึ่งแอตทริบิวต์

การประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า F-Statistics ของโมเดลแบบหนึ่งแอตทริบิวต์ที่มีการวัดผลจากการเรียนรู้ของโมเดลทุกๆ 10,000 timesteps โดยในงานวิจัยนี้จะกำหนดค่าความเชื่อมั่นที่ 95% (α = 0.05) และในแต่ละครั้งจะทดสอบโมเดลด้วยข้อมูลจำนวนทั้งหมด 1,000 records ซึ่งเมื่อทำการเปิดตารางค่า F จะได้ค่า $F_{0.05,1,998} \approx 3.85$ นั่นหมายความว่าเมื่อใดที่ค่า $F > F_{0.05,1,998} \approx 3.85$ แสดงว่าโมเดลที่ได้ ณ Timestep นั้นมีความแม่นยำที่อยู่ภายใต้ค่าความเชื่อมั่นที่ 95%

โดยจากภาพที่ 4.1 ในช่วงแรกของการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า F-Statistics ของโมเดลแบบหนึ่งแอตทริบิวต์ จะมีค่า F-Statistics ที่ต่ำกว่า 3.85 ซึ่งหมายความว่าค่าการพยากรณ์ที่ได้จากโมเดล ณ Timestep นั้น ยังมีความแม่นยำที่ต่ำกว่าค่าความเชื่อมั่นที่ 95% ที่ได้ตั้งเอาไว้ และเมื่อทำการเทรนโมเดลจนมาถึง Timestep ที่ 8,040,000 โมเดลที่ได้มีค่า $F > F_{0.05,1,998}$ และใน Timestep ที่ได้เทรนต่อจากนั้นมีค่า F-Statistics ที่สูงขึ้นตามลำดับ ซึ่งหมายความว่าโมเดลการพยากรณ์ที่ได้ตั้งแต่ Timestep ที่ 8,040,000 เป็นต้นไปมีค่าความแม่นยำที่อยู่ภายใต้ค่าความเชื่อมั่นที่ 95% ที่ได้กำหนดไว้

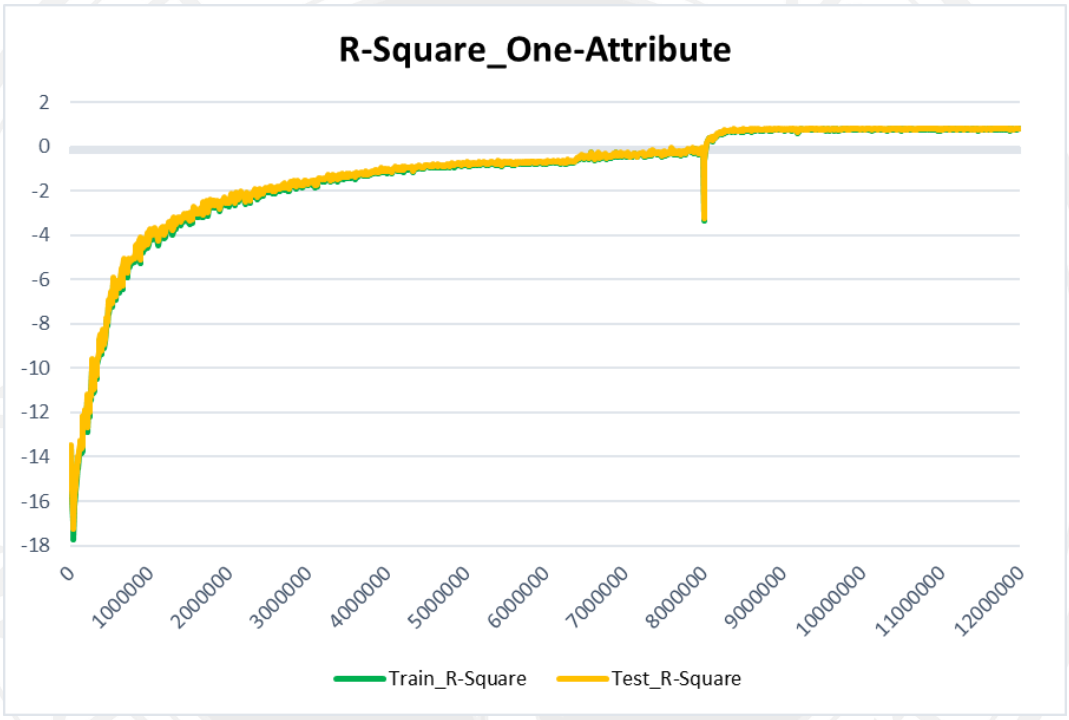


ภาพที่ 4.1 กราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า F-Statistics ของโมเดลแบบหนึ่งแอตทริบิวต์

ตารางที่ 4.1 ตัวอย่างผลการประเมินค่า F-Statistics ของโมเดลแบบหนึ่งแอตทริบิวต์

Timestep	F-Statistics	
	Train	Test
500,000	-1,754.81	-874.32
1,000,000	-1,608.65	-796.98
1,500,000	-1,548.79	-764.94
2,000,000	-1,455.52	-717.67
2,500,000	-1,328.88	-651.29
3,000,000	-1,278.39	-625.67
3,500,000	-1,178.91	-574.25
4,000,000	-1,089.33	-529.11
4,500,000	-956.00	-459.84
5,000,000	-905.11	-433.10
5,500,000	-873.04	-419.70
6,000,000	-889.77	-428.90
6,500,000	-625.63	-288.61
7,000,000	-569.69	-259.98
7,500,000	-395.68	-171.58
8,000,000	-309.38	-129.34
8,500,000	3,697.59	2,161.89
9,000,000	6,071.10	3,582.54
9,500,000	6,246.48	3,664.78
10,000,000	7,018.45	4,354.31
10,500,000	6,840.65	4,088.46
11,000,000	4,939.97	2,903.90
11,500,000	6,094.62	3,709.22
12,000,000	7,596.15	4,619.27

การประเมินประสิทธิภาพโมเดลพยากรณ์ด้วยค่า R-Square ของโมเดลแบบหนึ่งแอตทริบิวต์ ซึ่งค่า R-Square จะมีค่าอยู่ระหว่าง 0 – 1 โดยจากภาพที่ 4.2 กราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า R-Square ของโมเดลแบบหนึ่งแอตทริบิวต์ ในช่วงแรกของกราฟค่า R-Square จะมีค่าต่ำกว่า 0 ซึ่งหมายความว่าค่าตัวแปรที่ป้อนให้กับโมเดล ยังไม่สามารถอธิบายค่าที่พยากรณ์ออกมาได้จนมาถึงในช่วง Timestep ที่ 8,040,000 ค่า R-Square มีค่ามากกว่า 0 และมีค่าสูงขึ้นเรื่อยๆจน R-Square มีค่าเท่ากับ 0.82 ใน Timestep ที่ 12 ล้าน นั่นหมายความว่าค่าตัวแปรที่ป้อนให้กับโมเดล สามารถอธิบายค่าที่พยากรณ์ออกมาได้ 82%

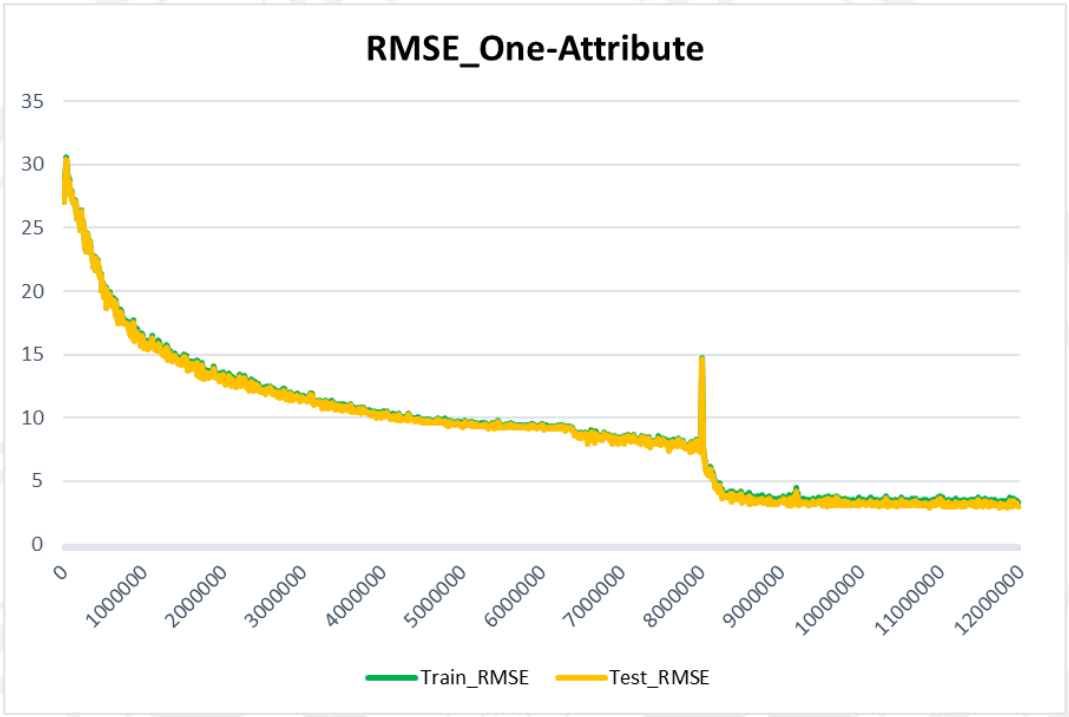


ภาพที่ 4.2 กราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า R-Square ของโมเดลแบบหนึ่งแอตทริบิวต์

ตารางที่ 4.2 ตัวอย่างผลการประเมินค่า R-Square ของโมเดลแบบหนึ่งแอตทริบิวต์

Timestep	R-Square	
	Train	Test
500,000	-7.22	-7.07
1,000,000	-4.13	-3.96
1,500,000	-3.45	-3.28
2,000,000	-2.68	-2.56
2,500,000	-1.99	-1.88
3,000,000	-1.78	-1.68
3,500,000	-1.44	-1.36
4,000,000	-1.20	-1.13
4,500,000	-0.92	-0.85
5,000,000	-0.83	-0.77
5,500,000	-0.78	-0.73
6,000,000	-0.80	-0.75
6,500,000	-0.46	-0.41
7,000,000	-0.40	-0.35
7,500,000	-0.25	-0.21
8,000,000	-0.18	-0.15
8,500,000	0.65	0.68
9,000,000	0.75	0.78
9,500,000	0.76	0.79
10,000,000	0.78	0.81
10,500,000	0.77	0.80
11,000,000	0.71	0.74
11,500,000	0.75	0.79
12,000,000	0.79	0.82

การประเมินประสิทธิภาพโมเดลพยากรณ์ด้วยค่า RMSE ของโมเดลแบบหนึ่งแอตทริบิวต์ โดย RMSE เป็นการวัดค่าความผิดพลาดโดยเฉลี่ยของโมเดล จากภาพที่ 4.3 กราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า RMSE ของโมเดลแบบหนึ่งแอตทริบิวต์ ในช่วงแรกของการเทรนมีค่า RMSE ที่สูงและค่าจะลดลงเรื่อยๆจนมาถึง Timestep ที่ 8,170,000 ค่า RMSE ที่ได้มีค่าต่ำกว่า 5 และมีค่าลดลงเรื่อยๆจนใน Timestep ที่ 12 ล้าน มีค่าเท่ากับ 3 ซึ่งหมายความว่า ค่าความผิดพลาดโดยเฉลี่ยของค่าการพยากรณ์ที่โมเดลได้พยากรณ์ออกมาอยู่ที่ ±3 ชิ้น



ภาพที่ 4.3 กราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า RMSE ของโมเดลแบบหนึ่งแอตทริบิวต์

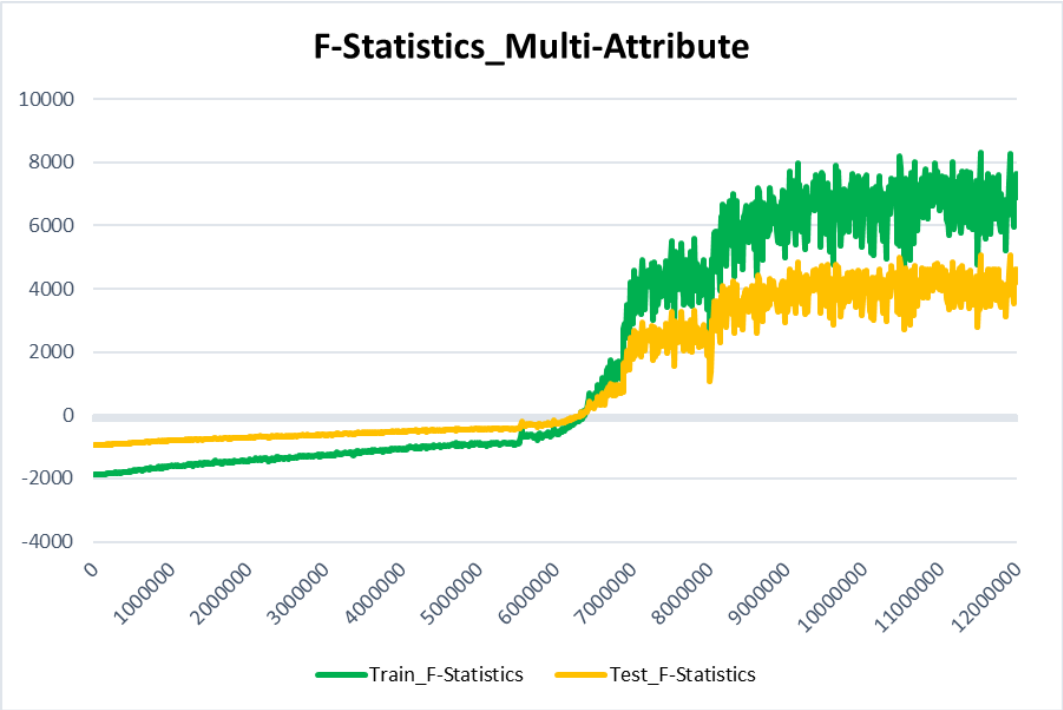
ตารางที่ 4.3 ตัวอย่างผลการประเมินค่า RMSE ของโมเดลแบบหนึ่งแอตทริบิวต์

Timestep	RMSE	
	Train	Test
500,000	20.27	20.18
1,000,000	16.02	15.83
1,500,000	14.92	14.70
2,000,000	13.57	13.41
2,500,000	12.22	12.05
3,000,000	11.79	11.63
3,500,000	11.05	10.90
4,000,000	10.49	10.37
4,500,000	9.79	9.68
5,000,000	9.56	9.44
5,500,000	9.43	9.33
6,000,000	9.50	9.41
6,500,000	8.53	8.43
7,000,000	8.37	8.26
7,500,000	7.90	7.81
8,000,000	7.69	7.62
8,500,000	4.19	3.99
9,000,000	3.52	3.32
9,500,000	3.48	3.29
10,000,000	3.33	3.07
10,500,000	3.36	3.15
11,000,000	3.80	3.59
11,500,000	3.51	3.27
12,000,000	3.23	2.99

4.2 แบบจำลองแบบหลายแอตทริบิวต์

การประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า F-Statistics ของโมเดลแบบหลายแอตทริบิวต์ที่มีการวัดผลจากการเรียนรู้ของโมเดลทุกๆ 10,000 timesteps โดยกำหนดค่าความเชื่อมั่นที่ 95% (α = 0.05) และทำการทดสอบโมเดลด้วยข้อมูลจำนวนทั้งหมด 1,000 records ซึ่งเมื่อทำการเปิดตารางค่า F จะได้ค่า $F_{0.05,1,998} \approx 3.85$ เช่นเดียวกับการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า F-Statistics ของโมเดลแบบหนึ่งแอตทริบิวต์

โดยจากภาพที่ 4.4 ในครึ่งแรกของเทรนมีผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า F-Statistics ของโมเดลแบบหลายแอตทริบิวต์ จะมีค่า F-Statistics ที่ต่ำกว่า 3.85 และเมื่อทำการเทรนโมเดลมาถึง Timestep ที่ 6,360,000 โมเดลที่ได้มีค่า $F > F_{0.05,1,998}$ และใน Timestep ที่ได้เทรนต่อจากนั้นมีค่า F-Statistics ที่สูงขึ้นตามลำดับ ซึ่งหมายความว่าโมเดลการพยากรณ์ที่ได้ตั้งแต่ Timestep ที่ 6,360,000 เป็นต้นไปมีค่าความแม่นยำที่อยู่ภายใต้ค่าความเชื่อมั่นที่ 95% ที่ได้กำหนดไว้ ซึ่งเมื่อเปรียบเทียบกับโมเดลการพยากรณ์แบบหนึ่งแอตทริบิวต์ จะเห็นได้ว่าโมเดลการพยากรณ์แบบหลายแอตทริบิวต์ สามารถสร้างโมเดลที่อยู่ภายใต้ค่าความเชื่อมั่นที่ 95% ได้เร็วกว่าโมเดลการพยากรณ์แบบหนึ่งอตทริบิวต์

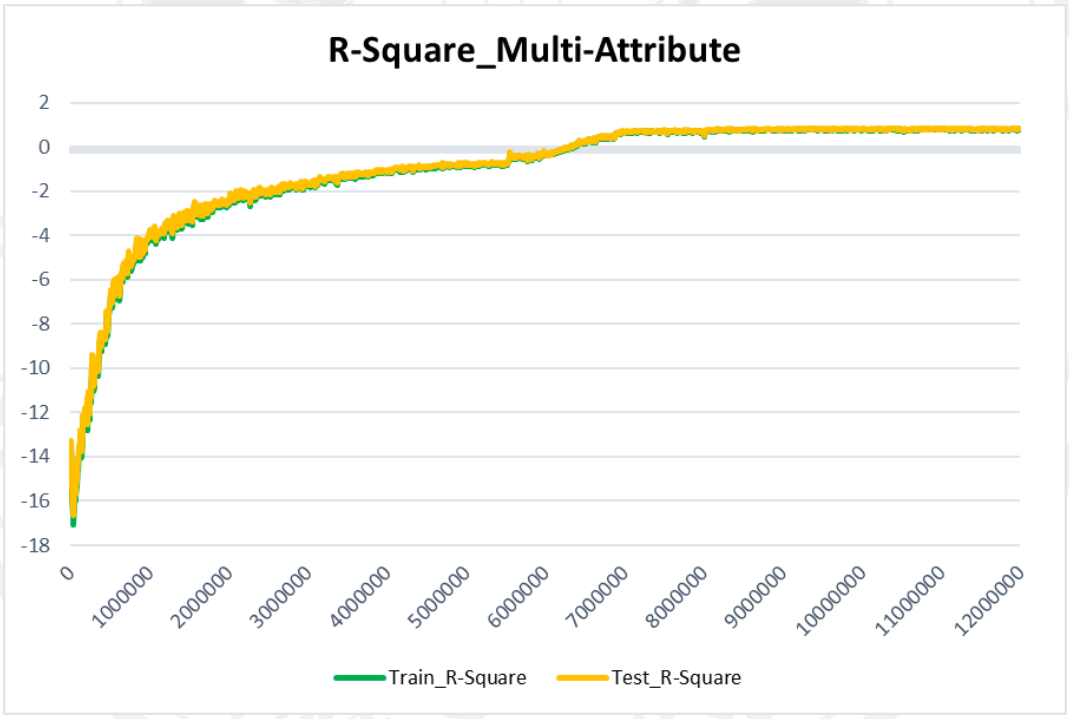


ภาพที่ 4.4 กราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า F-Statistics ของโมเดลแบบหลายแอตทริบิวต์

ตารางที่ 4.4 ตัวอย่างผลการประเมินค่า F-Statistics ของโมเดลแบบหลายแอตทริบิวต์

Timestep	F-Statistics	
	Train	Test
500,000	-1,758.29	-875.64
1,000,000	-1,596.97	-791.39
1,500,000	-1,552.17	-767.81
2,000,000	-1,434.03	-705.93
2,500,000	-1,348.46	-662.18
3,000,000	-1,286.19	-629.49
3,500,000	-1,172.17	-571.55
4,000,000	-1,086.92	-527.79
4,500,000	-1,003.28	-485.82
5,000,000	-904.18	-432.67
5,500,000	-853.17	-407.40
6,000,000	-582.08	-265.22
6,500,000	290.49	204.88
7,000,000	2,914.49	1,780.74
7,500,000	4,865.18	2,832.12
8,000,000	3,726.79	2,146.12
8,500,000	6,444.24	3,890.07
9,000,000	7,035.16	4,211.53
9,500,000	6,780.11	4,152.94
10,000,000	6,144.72	3,696.06
10,500,000	6,339.31	3,805.25
11,000,000	6,474.65	3,985.32
11,500,000	6,217.34	3,659.18
12,000,000	7,636.38	4,619.27

การประเมินประสิทธิภาพโมเดลพยากรณ์ด้วยค่า R-Square ของโมเดลแบบหลายแอตทริบิวต์ ซึ่งค่า R-Square จะมีค่าอยู่ระหว่าง 0 – 1 โดยจากภาพที่ 4.5 กราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า R-Square ของโมเดลแบบหลายแอตทริบิวต์ ในครึ่งแรกของกราฟค่า R-Square จะมีค่าต่ำกว่า 0 และเมื่อเทรนโมเดลการพยากรณ์ต่อมาจนถึงในช่วง Timestep ที่ 6,360,000 ค่า R-Square มีค่ามากกว่า 0 และมีค่าสูงขึ้นเรื่อยๆจน R-Square มีค่าเท่ากับ 0.82 ใน Timestep ที่ 12 ล้าน นั่นหมายความว่าค่าตัวแปรที่ป้อนให้กับโมเดล สามารถอธิบายค่าที่พยากรณ์ออกมาได้ 82% ซึ่งเมื่อเปรียบเทียบกับโมเดลการพยากรณ์แบบหนึ่งแอตทริบิวต์ จะเห็นได้ว่าโมเดลการพยากรณ์ทั้ง 2 แบบใน Timestep ที่ 12 ล้าน มีค่า R-Square ที่เท่ากันที่ 82% แต่โมเดลการพยากรณ์ทั้ง 2 แบบจะแตกต่างกันที่โมเดลการพยากรณ์แบบหลายแอตทริบิวต์ สามารถให้ค่าการประเมินประสิทธิภาพจากค่า R-Square ที่ 82% ได้เร็วกว่าโมเดลการพยากรณ์แบบหนึ่งแอตทริบิวต์

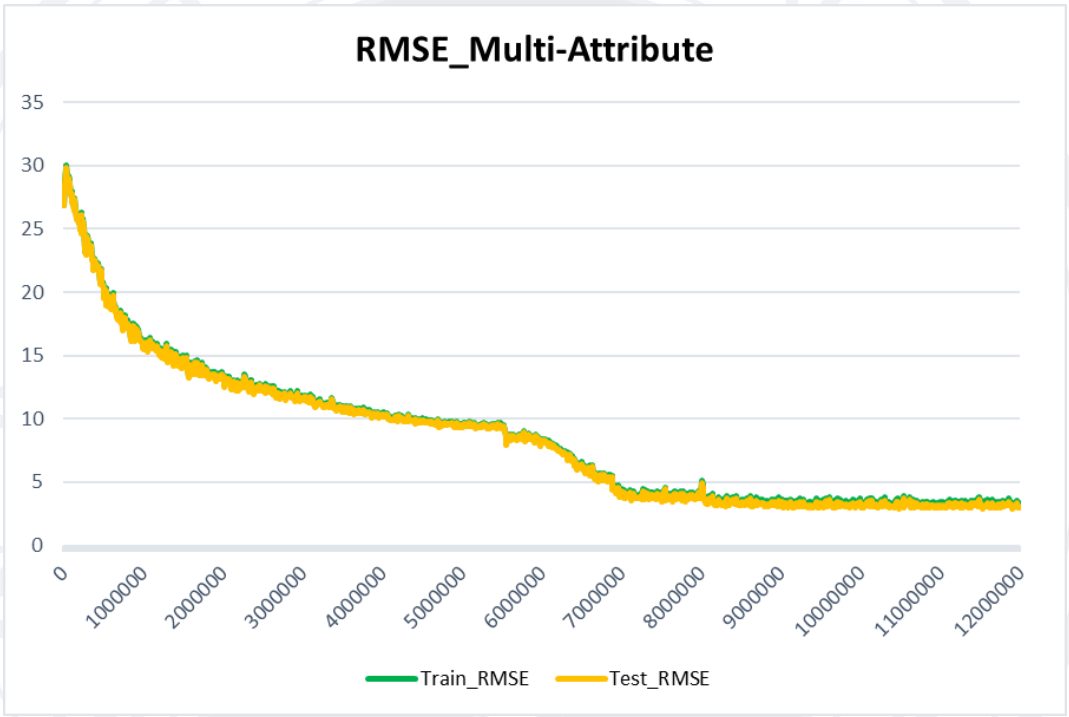


ภาพที่ 4.5 กราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า R-Square ของโมเดลแบบหลายแอตทริบิวต์

ตารางที่ 4.5 ตัวอย่างผลการประเมินค่า R-Square ของโมเดลแบบหลายแอตทริบิวต์

Timestep	R-Square	
	Train	Test
500,000	-7.34	-7.16
1,000,000	-3.98	-3.83
1,500,000	-3.48	-3.34
2,000,000	-2.54	-2.42
2,500,000	-2.08	-1.97
3,000,000	-1.81	-1.71
3,500,000	-1.42	-1.34
4,000,000	-1.19	-1.12
4,500,000	-1.01	-0.95
5,000,000	-0.83	-0.77
5,500,000	-0.75	-0.69
6,000,000	-0.41	-0.36
6,500,000	0.13	0.17
7,000,000	0.59	0.64
7,500,000	0.71	0.74
8,000,000	0.65	0.68
8,500,000	0.76	0.80
9,000,000	0.78	0.81
9,500,000	0.77	0.81
10,000,000	0.75	0.79
10,500,000	0.76	0.79
11,000,000	0.76	0.80
11,500,000	0.76	0.79
12,000,000	0.79	0.82

การประเมินประสิทธิภาพโมเดลพยากรณ์ด้วยค่า RMSE ของโมเดลแบบหลายแอตทริบิวต์ โดย RMSE เป็นการวัดค่าความผิดพลาดโดยเฉลี่ยของโมเดล จากภาพที่ 4.6 กราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า RMSE ของโมเดลแบบหลายแอตทริบิวต์ ในช่วงแรกของการเทรนมีค่า RMSE ที่สูงและค่าจะลดลงเรื่อยๆจนมาถึง Timestep ที่ 6,890,000 ค่า RMSE ที่ได้เริ่มมีค่าต่ำกว่า 5 และเมื่อเทรนโมเดลการพยากรณ์ต่อค่า RMSE จะลดลงต่อเรื่อยๆจนถึง Timestep ที่ 12 ล้าน ค่า RMSE มีค่าเท่ากับ 3 ซึ่งหมายความว่า ค่าความผิดพลาดโดยเฉลี่ยของค่าการพยากรณ์ที่โมเดลได้พยากรณ์ออกมาอยู่ที่ ±3 ชิ้น



ภาพที่ 4.6 กราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า RMSE ของโมเดลแบบหลายแอตทริบิวต์

ตารางที่ 4.6 ตัวอย่างผลการประเมินค่า RMSE ของโมเดลแบบหลายแอตทริบิวต์

Timestep	RMSE	
	Train	Test
500,000	20.42	20.29
1,000,000	15.79	15.62
1,500,000	14.97	14.79
2,000,000	13.31	13.13
2,500,000	12.40	12.25
3,000,000	11.85	11.69
3,500,000	11.00	10.87
4,000,000	10.47	10.35
4,500,000	10.02	9.92
5,000,000	9.56	9.44
5,500,000	9.34	9.24
6,000,000	8.40	8.29
6,500,000	6.61	6.47
7,000,000	4.51	4.26
7,500,000	3.82	3.63
8,000,000	4.18	4.00
8,500,000	3.44	3.21
9,000,000	3.33	3.11
9,500,000	3.37	3.13
10,000,000	3.50	3.28
10,500,000	3.46	3.24
11,000,000	3.43	3.18
11,500,000	3.49	3.29
12,000,000	3.22	2.99

บทที่ 5

สรุปผลการวิจัย

5.1 สรุปผลการวิจัย

งานวิจัยครั้งนี้มีวัตถุประสงค์ที่จะศึกษา Proximal Policy Optimization Algorithm ของกระบวนการ Reinforcement Learning เพื่อสร้างโมเดลการพยากรณ์จำนวนการสั่งวัตถุดิบของร้านอาหาร โดยชุดข้อมูลที่ใช้ในงานวิจัยนี้จะเป็นชุดข้อมูลจำลอง (Synthetic Data) ที่อาศัยหลักการจำลองข้อมูลแบบแจกแจงปกติ (Normal Distribution) เนื่องมาจากชุดข้อมูลจริงที่ได้มาจากเครื่อง POS ของบริษัท เอ เป็นชุดข้อมูลที่มีความแปรปรวนที่สูงและไม่มีแบบฟอร์มการกระจายตัวของข้อมูลที่แน่นอน ผู้วิจัยจึงตัดสินใจใช้ชุดข้อมูลจำลอง โดยในชุดข้อมูลจำลองจะมีการแยกพฤติกรรมการใช้วัตถุดิบเบื้องต้นเป็นแบบความต้องการวัตถุดิบในวันธรรมดา และความต้องการใช้วัตถุดิบในวันหยุด โดยค่าเฉลี่ยที่ใช้ในการจำลองชุดข้อมูล นำมาจากค่าเฉลี่ยของชุดข้อมูลจริงที่ได้จากเครื่อง POS และกำหนดค่าส่วนเบี่ยงเบนมาตรฐานอยู่ที่ 2

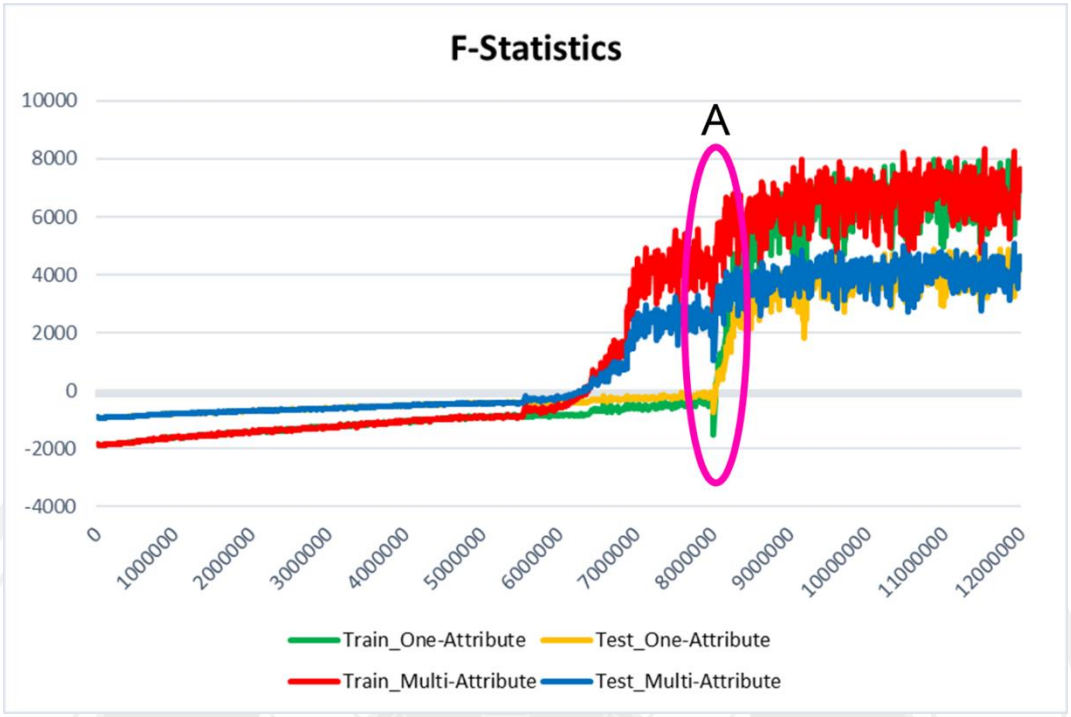
ในการศึกษาครั้งนี้แบ่งการศึกษาออกเป็นโมเดลการพยากรณ์แบบหนึ่งแอตทริบิวต์ และโมเดลการพยากรณ์แบบหลายแอตทริบิวต์ โดยทำการเรียนรู้ทั้งหมด 12,000,000 timesteps โดยเครื่องมือที่ใช้ในการประเมินประสิทธิภาพโมเดลพยากรณ์แบ่งออกเป็น 3 ค่า คือ ค่า F-Statistics เป็นการวัดผลประสิทธิภาพโมเดลการพยากรณ์ในด้านของความแม่นยำของโมเดล ที่อยู่ภายใต้ค่าความเชื่อมั่นที่กำหนด, ค่า R-Square เป็นการวัดผลประสิทธิภาพโมเดลการพยากรณ์ ในเรื่องของค่าตัวแปรที่ป้อนให้กับโมเดลการพยากรณ์ สามารถอธิบายค่าที่พยากรณ์ออกมาได้ เป็นจำนวนกี่เปอร์เซนต์ และค่า RMSE เป็นการวัดผลประสิทธิภาพโมเดลการพยากรณ์ ในเรื่องของค่าความผิดพลาดโดยเฉลี่ยของโมเดลพยากรณ์

โดยจากการประเมินประสิทธิภาพโมเดลพยากรณ์แบบหนึ่งแอตทริบิวต์ และโมเดลการพยากรณ์แบบหลายแอตทริบิวต์ จากชุดข้อมูล 1,000 ค่าที่ใช้สำหรับการประเมินประสิทธิภาพโมเดลพยากรณ์ ได้ค่า F-Statistics ที่มีค่าความแม่นยำอยู่ภายใต้ค่าความเชื่อมั่นที่ 95% ค่า R-Square ที่ใช้สำหรับประเมินประสิทธิภาพของตัวแปรอินพุต สามารถอธิบายค่าที่พยากรณ์ออกมาได้ 82% และมีค่า RMSE หรือค่าความคลาดเคลื่อนโดยเฉลี่ยของโมเดลอยู่ที่ ±3 ชิ้น

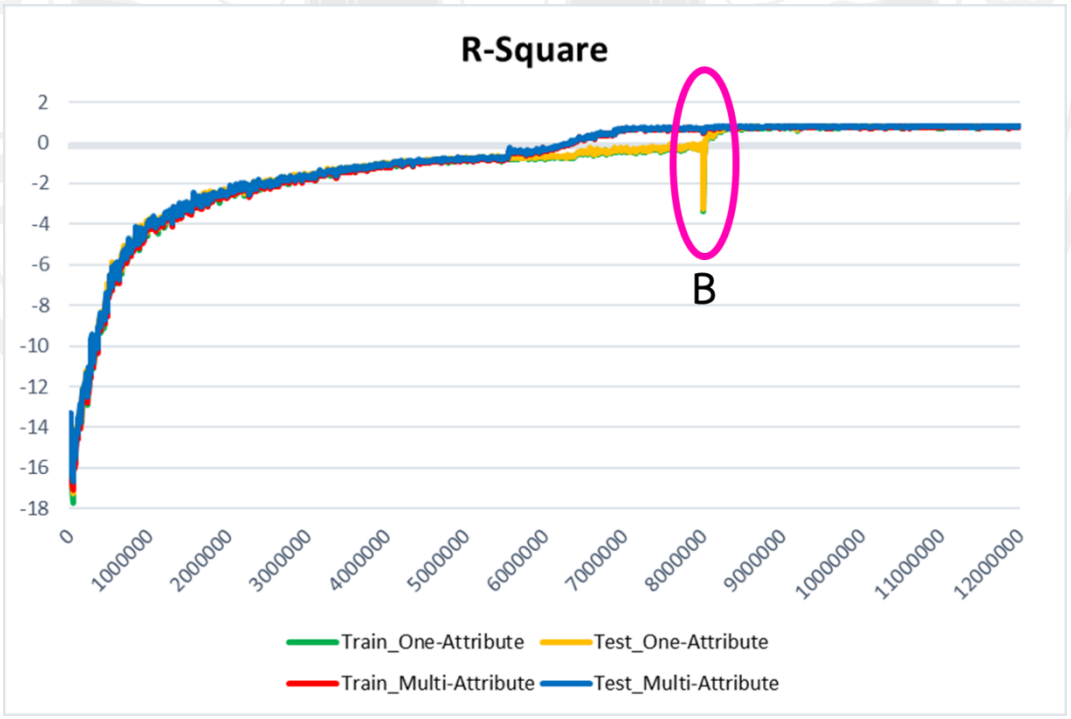
5.2 อภิปรายผลการวิจัย

ในการศึกษาครั้งนี้แบ่งการศึกษาออกเป็นโมเดลการพยากรณ์แบบหนึ่งแอตทริบิวต์ และโมเดลการพยากรณ์แบบหลายแอตทริบิวต์ โดยมีการประเมินประสิทธิภาพโมเดลพยากรณ์จากค่า F-Statistics, ค่า R-Square และค่า RMSE โดยการวัดผลประสิทธิภาพโมเดลการพยากรณ์จะใช้ชุดข้อมูลจำลองที่แยกจากชุดข้อมูลที่ใช้เรียนรู้จำนวน 1,000 records และพบว่าโมเดลการพยากรณ์แบบหนึ่งแอตทริบิวต์ และโมเดลการพยากรณ์แบบหลายแอตทริบิวต์ใน Timestep ที่ 12 ล้าน ให้ผลลัพธ์จากการประเมินประสิทธิภาพโมเดลพยากรณ์ที่ใกล้เคียงกัน แต่จะแตกต่างกันที่โมเดลพยากรณ์แบบหลายแอตทริบิวต์ จะเกิดการลู่เข้า (Convergence) หาค่าที่เหมาะสมของโมเดลพยากรณ์ได้เร็วกว่าการเรียนรู้ของโมเดลพยากรณ์แบบหนึ่งแอตทริบิวต์

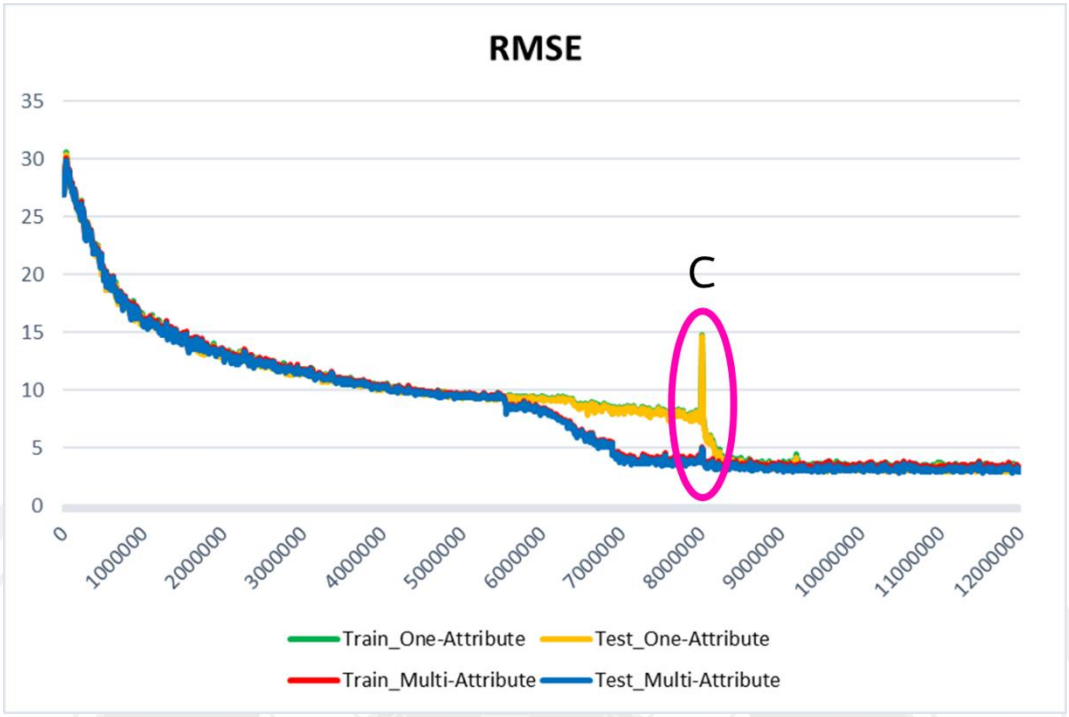
และจากกราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ ในภาพที่ 5.1, ภาพที่ 5.2 และภาพที่ 5.3 ในวงสีชมพูที่จุด A, B และ C สาเหตุที่ค่าใน Timestep ดังภาพมีประสิทธิภาพการพยากรณ์ที่ลดลงอย่างกระทันหัน เนื่องมาจากคุณสมบัติของระบบ Reinforcement Learning ที่มีการตัดสินใจแบบ Exploration คือการที่ระบบจะทำการค้นหาแนวทางใหม่ในการตัดสินใจ และการตัดสินใจแบบ Exploitation ที่จะใช้แนวทางที่เคยตัดสินใจในอดีตกลับมาใช้ใหม่ เพราะฉะนั้นจากกราฟในภาพทั้งสามที่มีการลดลงของประสิทธิภาพอย่างกระทันหัน มีสาเหตุมาจากการที่ระบบ Reinforcement Learning ใช้การตัดสินใจแบบ Exploration แล้วได้แนวทางการตัดสินใจที่มีประสิทธิภาพลดลง ทำให้ระบบ Reinforcement Learning กลับไปใช้การตัดสินใจแบบ Exploitation เพื่อให้โมเดลพยากรณ์กลับไปยังค่าการตัดสินใจก่อนหน้า ที่ให้แนวทางการตัดสินใจที่มีประสิทธิภาพ และทำการเทรนโมเดลพยากรณ์ต่อให้เข้าสู่ค่าที่เหมาะสม



ภาพที่ 5.1 จุดสังเกตในกราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า F-Statistics



ภาพที่ 5.2 จุดสังเกตในกราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า R-Square



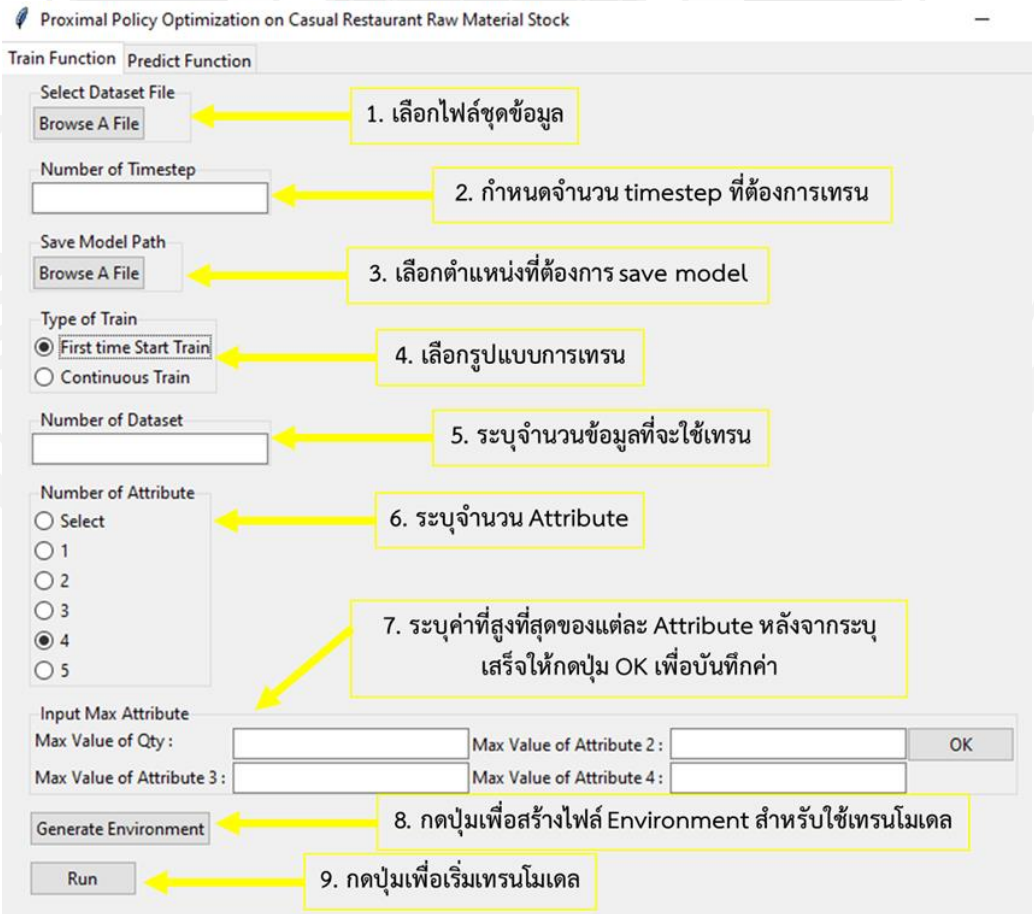
ภาพที่ 5.3 จุดสังเกตในกราฟผลการประเมินประสิทธิภาพโมเดลการพยากรณ์ด้วยค่า RMSE

5.3 การนำโมเดลการพยากรณ์การจัดการวัตถุดิบคงคลังไปใช้งาน

การนำระบบการจัดการวัตถุดิบคงคลังของร้านอาหาร โดยใช้ Proximal Policy Optimization มาประยุกต์ใช้งานจริง ผู้วิจัยได้จัดทำตัวอินเตอร์เฟสของระบบขึ้นมา เพื่อให้ผู้ที่จะนำไปใช้งานต่อสามารถเรียนรู้และใช้งานได้ง่ายยิ่งขึ้น โดยอินเตอร์เฟสที่ผู้วิจัยได้สร้างขึ้น สามารถแบ่งได้ออกเป็น 2 ส่วนหลักๆดังนี้

5.3.1 ส่วนของอินเตอร์เฟสชุดคำสั่งการเรียนรู้ข้อมูล

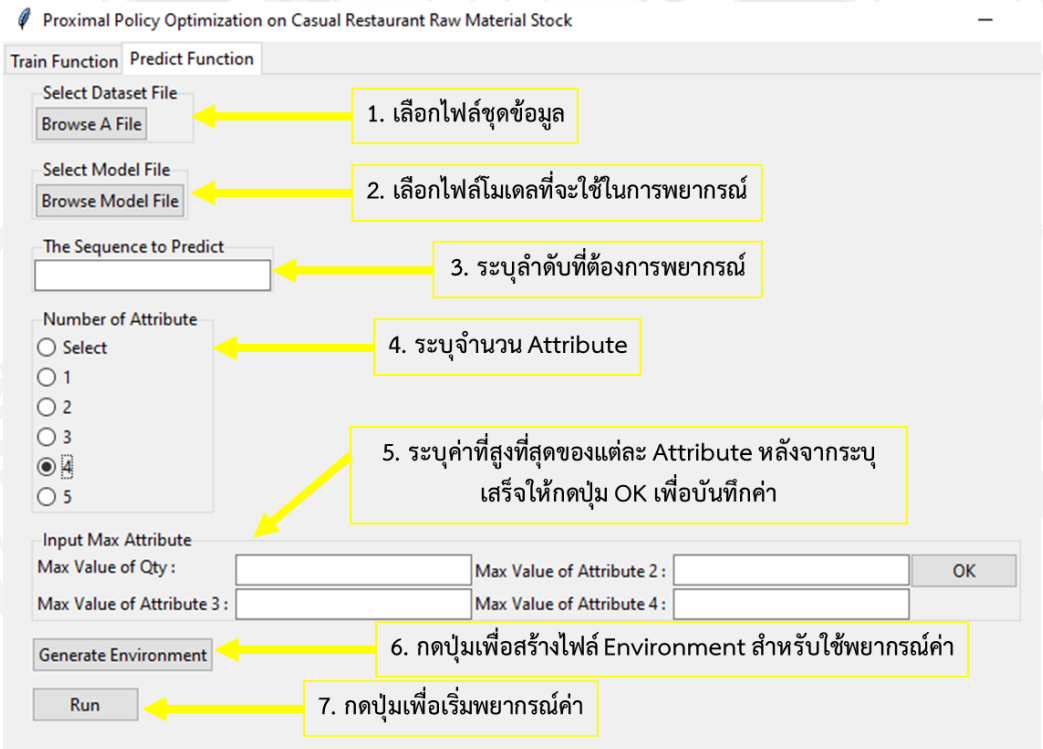
ส่วนของชุดคำสั่งการเรียนรู้ข้อมูล เป็นอินเตอร์เฟสที่อำนวยความสะดวกให้ผู้ที่นำไปใช้งานสามารถกำหนดค่าตัวแปรที่ต้องการ สำหรับการเทรนโมเดลได้อย่างครบถ้วนทุกตำแหน่งในตัวโค้ดต้นฉบับ และเป็นการป้องกันการกำหนดค่าที่ไม่ครบถ้วน ซึ่งทำให้เกิดข้อผิดพลาดในการเทรนโมเดลตามมา โดยในการใช้ตัวอินเตอร์เฟสจะมีลำดับขั้นตอนตามดังภาพที่ 5.4



ภาพที่ 5.4 อินเตอร์เฟสของชุดคำสั่งการเรียนรู้ข้อมูล

5.3.2 ส่วนของอินเตอร์เฟสชุดคำสั่งการพยากรณ์ข้อมูล

ส่วนของอินเตอร์เฟสชุดคำสั่งการพยากรณ์ข้อมูล เป็นอินเตอร์เฟสที่ทำให้ผู้ใช้งานสามารถนำไฟล์โมเดลการพยากรณ์ที่ได้จากการเทรนชุดข้อมูลมาก่อน มาใช้ในการพยากรณ์โดยมีลำดับขั้นตอนดังภาพที่ 5.5 โดยเริ่มจากการเลือกชุดข้อมูลที่จะให้ระบบ Observation เพื่อใช้ในการคำนวณค่าที่จะพยากรณ์ออกมาอย่างน้อยย้อนหลัง 7 วัน ก่อนวันที่จะทำการพยากรณ์ ต่อมาจะเป็นการเลือกไฟล์โมเดลพยากรณ์ที่ได้จากการเรียนรู้มาก่อนหน้า หลังจากนั้นจะเป็นการระบุลำดับของวันที่ผู้ใช้งานต้องการพยากรณ์ ขั้นตอนถัดมาจะเป็นการระบุจำนวนแอตทริบิวต์และค่าสูงสุดของแต่ละแอตทริบิวต์ที่ผู้ใช้งานได้กำหนดในตอนเทรนโมเดล หลังจากกำหนดข้อมูลทุกอย่างเรียบร้อย ให้ผู้ใช้งานทำการสร้าง Environment โดยการกดปุ่มที่กำหนดไว้ และขั้นตอนสุดท้ายคือการกดที่ปุ่ม Run เพื่อเริ่มการคำนวณค่าที่ต้องการพยากรณ์



ภาพที่ 5.5 อินเตอร์เฟสของชุดคำสั่งการพยากรณ์ข้อมูล

5.4 ข้อเสนอแนะงานวิจัยในอนาคต

1) งานวิจัยครั้งนี้มีการใช้ชุดข้อมูลจำลองที่มีการกระจายตัวของข้อมูลแบบปกติ ในงานวิจัยในอนาคตอาจศึกษาชุดข้อมูลที่มีการกระจายตัวในแบบอื่น
2) งานวิจัยครั้งนี้ในชุดข้อมูลจำลองยังมีความ Imbalance ระหว่างข้อมูลในส่วนของความต้องการวัตถุดิบในวันธรรมดากับข้อมูลในส่วนของความต้องการวัตถุดิบในวันหยุด งานวิจัยในอนาคตอาจใช้ชุดข้อมูลที่มีความ Balance ระหว่างข้อมูลในส่วนของความต้องการวัตถุดิบในวันธรรมดากับข้อมูลในส่วนของความต้องการวัตถุดิบในวันหยุด

บรรณานุกรม

AI, S. (2019, 18 February 2019). Reinforcement Learning algorithms — an intuitive overview. Retrieved from https://medium.com/@SmartLabAI/reinforcement-learning-algorithms-an-intuitive-overview-904e2dff5bbc

FAO. (2019). The State of Food and Agriculture 2019. Moving forward on food loss and waste reduction. Retrieved from http://www.fao.org/3/ca6030en/ca6030en.pdf

Fujita, Y., & Maeda, S. (2018). *Clipped Action Policy Gradient.* Paper presented at the ICML.

Geevers, K. (2020). *Deep Reinforcement Learning in Inventory Management.* Retrieved from http://essay.utwente.nl/85432/

Hill, A. a. R., Antonin and Ernestus, M. a. G., Adam and Kanervisto, A. a. T., Rene and Dhariwal, P. a. H., Christopher and Klimov, O. a. N., Alex and Plappert, M. a. R., . . . Szymon and Wu, Y. (2018). Stable Baselines. *GitHub repository*. Retrieved from https://github.com/hill-a/stable-baselines

Hu, C., Chen, M., & McCain, S.-L. C. (2004). Forecasting in Short-Term Planning and Management for a Casino Buffet Restaurant. *Journal of Travel & Tourism Marketing, 16*(2-3), 79-98. doi:10.1300/J073v16n02_07

King, A. (2019, 10 April 2019). Create custom gym environments from scratch — A stock market example. Retrieved from https://towardsdatascience.com/creating-a-custom-openai-gym-environment-for-stock-trading-be532be3910e

Makridakis, S., Wheelright, S., & Hyndman, R. (2000). *Manual of Forecasting: Methods and Applications*.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., . . . Kavukcuoglu, K. (2016). Asynchronous Methods for Deep Reinforcement Learning. *ArXiv, abs/1602.01783*.

Schulman, J., Levine, S., Abbeel, P., Jordan, M. I., & Moritz, P. (2015). Trust Region Policy Optimization. *ArXiv, abs/1502.05477*.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization Algorithms. *ArXiv, abs/1707.06347*.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*: A Bradford Book.

Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2019). Demand forecasting in restaurants using machine learning and statistical analysis. *Procedia CIRP, 79*, 679-683. doi:10.1016/j.procir.2019.02.042

Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2020). Restaurants store management based on demand forecasting. *Procedia CIRP, 88*, 580-583. doi:10.1016/j.procir.2020.05.101

WasteMINZ. (2018). FOOD WASTE IN THE CAFE & RESTAURANT SECTOR IN NEW ZEALAND. Retrieved from http://www.wasteminz.org.nz/wp-content/uploads/2018/10/New-Zealand-cafe-and-resturant-food-waste-WasteMINZ-2018.pdf

Xu, Y., Qian, Q., Li, H., & Jin, R. (2021). Why Does Multi-Epoch Training Help? *ArXiv, abs/2105.06015*.

Yogeswaran, M., & Ponnambalam, S. G. (2012). Reinforcement learning: exploration–exploitation dilemma in multi-agent foraging task. *OPSEARCH, 49*(3), 223-236. doi:10.1007/s12597-012-0077-2

ไฮเซอร์, เ. (2551). การจัดการการผลิตและการปฏิบัติการ = *Operations management* (พิมพ์ครั้งที่ 4. ed.). กรุงเทพฯ: กรุงเทพฯ : เพียร์สันเอ็ดดูเคชั่นอินโดไชน่า.

ปญฺญา สมา อาภา
สถาบันบัณฑิตพัฒนบริหารศาสตร์

ประวัติผู้เขียน

ชื่อ-นามสกุล นายณัฐวัฒน์ เอกธรรมนิตย์

ประวัติการศึกษา วิศวกรรมศาสตรบัณฑิต (วิศวกรรมอุตสาหการ)
มหาวิทยาลัยเกษตรศาสตร์
ปีที่สำเร็จการศึกษา พ.ศ. 2557

ประสบการณ์การทำงาน พ.ศ. 2557-2560
วิศวกรออกแบบ
บริษัท อีเอสพี เอเซียน เซ็นเตอร์ จำกัด
เขตวัฒนา กรุงเทพมหานคร 10110

พ.ศ. 2557-2560
วิศวกรระบบ
บริษัท อีเอสพี เอเซียน เซ็นเตอร์ จำกัด
เขตวัฒนา กรุงเทพมหานคร 10110