

การประมาณต้นทุนงบประมาณก่อสร้างโดยวิธี Machine Learning
และข้อมูลการเปิดเผยราคากลางของรัฐในอดีต

Construction Cost Estimating for Budgeting Purposes
Using Machine Learning and Government's Reference Price Data

นายปวิษฐา อภิวัฒน์ผลอง

นายธีรภัทร กล้าการไธ

นายธีรทัต บรรณาลัย

อาจารย์ที่ปรึกษา : รศ.ดร.นคร กกแก้ว

ภาควิชาวิศวกรรมโยธา คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย จ.กรุงเทพฯ

บทคัดย่อ

อุตสาหกรรมก่อสร้างเป็นภาคส่วนสำคัญที่ส่งผลต่อการพัฒนาเศรษฐกิจของประเทศ โดยเฉพาะการก่อสร้างภาครัฐ ซึ่งมีบทบาทในการพัฒนาโครงสร้างพื้นฐาน ปัจจุบันภาครัฐมีการปรับปรุงกระบวนการให้มีความโปร่งใสมากขึ้น เช่น การคำนวณราคากลาง การเปิดเผยราคากลาง และการประมูลผ่านระบบ E-bidding อย่างไรก็ตาม การของบประมาณจากหน่วยงานรัฐส่วนใหญ่มักมุ่งเน้นการแสดงความพร้อมด้านเอกสาร มากกว่าการพิจารณาความคุ้มค่าทางเศรษฐศาสตร์ของโครงการ ส่งผลให้บางโครงการที่ได้รับงบประมาณอาจไม่มีประสิทธิภาพหรือไม่สามารถใช้งานได้จริงในระยะยาว ด้วยเหตุนี้ การนำข้อมูลราคากลางและ BOQ ที่เผยแพร่ผ่านระบบ E-bidding มาใช้ร่วมกับเทคนิค Machine Learning เพื่อพัฒนาสมการประมาณต้นทุนงบประมาณก่อสร้าง จึงเป็นแนวทางที่มีศักยภาพในการเพิ่มประสิทธิภาพการจัดสรรงบประมาณของภาครัฐอย่างเหมาะสมและคุ้มค่า

คำสำคัญ : งบประมาณก่อสร้าง, Machine Learning, ราคากลาง

Abstract

The construction industry plays a vital role in national economic development and serves as an indicator of economic growth. Public sector construction significantly contributes to infrastructure development. In recent years, transparency in government construction processes has improved through practices such as standard cost estimation, public disclosure of pricing, and E-bidding systems. However, many government agencies tend to focus on demonstrating project readiness through detailed design documents and cost estimates, rather than evaluating the project's economic viability. As a result, some funded projects may be inefficient, underutilized, or difficult to maintain in the long term. To address this issue, the utilization of historical government cost data and BOQs, combined with Machine Learning techniques, presents a promising approach to developing cost estimation models for budget planning. This method can enhance the efficiency and effectiveness of public budget allocation.

Key words: Construction Budget, Machine Learning, Construction Cost Estimate

1. บทนำ

อุตสาหกรรมก่อสร้างถือเป็นหนึ่งในภาคการผลิตที่มีบทบาทสำคัญต่อการพัฒนาเศรษฐกิจของประเทศ ทั้งในด้านการลงทุน การจ้างงาน และการขับเคลื่อนอุตสาหกรรมเกี่ยวเนื่อง โดยเฉพาะในส่วนของการก่อสร้างภาครัฐที่มีผลต่อการขยายตัวทางเศรษฐกิจผ่านการพัฒนาโครงสร้างพื้นฐาน กระบวนการจัดทำงบประมาณก่อสร้างของภาครัฐจึงมีความสำคัญอย่างยิ่ง โดยในปัจจุบันรัฐได้เปิดเผยข้อมูลราคากลางและดำเนินการประมูลงานผ่านระบบ E-bidding ซึ่งก่อให้เกิดโอกาสในการนำข้อมูลเหล่านี้มาใช้ประโยชน์ในการปรับปรุงกระบวนการวางแผนงบประมาณให้มีประสิทธิภาพมากยิ่งขึ้น

อย่างไรก็ตาม การจัดทำค่าของในปัจจุบันยังขาดการพิจารณาความคุ้มค่าทางเศรษฐศาสตร์ในระยะยาว ทำให้บางโครงการอาจไม่ก่อให้เกิดประโยชน์สูงสุดต่อประเทศ การประยุกต์ใช้เทคโนโลยี Machine Learning เข้ากับข้อมูลราคากลางที่เผยแพร่โดยภาครัฐจึงเป็นแนวทางหนึ่งในการเพิ่มความแม่นยำของการประมาณต้นทุน ลดความซับซ้อนในการจัดทำเอกสาร และลดเวลาในการจัดสรรงบประมาณ โดยยังสามารถใช้ในการคาดการณ์ต้นทุนของโครงการในอนาคตได้อย่างรวดเร็วและมีประสิทธิภาพ

โครงการนี้จึงมุ่งเน้นการพัฒนาโมเดลการประมาณต้นทุนงบประมาณก่อสร้าง โดยอาศัยข้อมูลราคากลางในอดีตร่วมกับเทคโนโลยี Machine Learning เพื่อส่งเสริมให้การดำเนินโครงการก่อสร้างภาครัฐเป็นไปอย่างคุ้มค่า โปร่งใส และสามารถสร้างประโยชน์ต่อการพัฒนาประเทศในระยะยาวได้อย่างแท้จริง [1]

2. แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 อุตสาหกรรมก่อสร้างไทย

อุตสาหกรรมก่อสร้างเป็นอุตสาหกรรมการผลิตที่สำคัญของประเทศ เป็นอุตสาหกรรมที่นำไปสู่การพัฒนาประเทศ ส่งผลให้เกิดการกระจายตัวของเม็ดเงินทั้งในด้านเงินทุนในการก่อสร้าง การจ้างงาน รวมไปถึงอุตสาหกรรมต่าง ๆ ที่เกี่ยวข้องกับกระบวนการก่อสร้าง อีกทั้งธุรกิจก่อสร้างยังเป็นธุรกิจที่เติบโตไปพร้อมกับความเจริญของประเทศ หากเศรษฐกิจของประเทศรุ่งเรือง อุตสาหกรรมก่อสร้างก็จะเติบโตควบคู่กันไป [1]

2.2 โครงการก่อสร้างภาครัฐ

โครงการก่อสร้างภาครัฐ (Infrastructure) เป็นกลไกสำคัญภายใต้นโยบายการคลัง ซึ่งครอบคลุมการบริหารจัดการรายรับและรายจ่ายของภาครัฐ โดยการใช้จ่ายด้านโครงสร้างพื้นฐาน เช่น การก่อสร้างถนนหรือระบบคมนาคม มีบทบาทในการกระตุ้นการขยายตัวทางเศรษฐกิจในช่วงที่รัฐบาลดำเนินนโยบายการคลังแบบขาดดุล อย่างไรก็ตาม การดำเนินนโยบายดังกล่าวจำเป็นต้องพิจารณาขนาดของการ

ขาดดุลและแนวทางในการแก้ไขอย่างรอบคอบ เพื่อรักษาเสถียรภาพทางเศรษฐกิจในระยะยาว [2]

2.3 การประมาณราคาก่อสร้าง

การประมาณราคาก่อสร้างคือการคำนวณงบประมาณโครงการโดยใช้ข้อมูลปริมาณงานจากการถอดแบบ เพื่อประเมินค่าใช้จ่ายทั้งทางตรงและทางอ้อม ราคาที่ได้เรียกว่า “ราคากลาง” ซึ่งเป็นราคาประมาณการใกล้เคียงความเป็นจริง แต่ไม่ใช่ราคาจริง เนื่องจากมีปัจจัยหลายประการที่ส่งผลให้เกิดความคลาดเคลื่อน หากดำเนินการโดยผู้เชี่ยวชาญ ราคากลางควรมีความคลาดเคลื่อนไม่เกินร้อยละ 10 และสามารถใช้เป็นข้อมูลอ้างอิงในการจัดจ้างภาครัฐได้อย่างมีประสิทธิภาพ [3]

2.3.1 การประมาณราคาก่อสร้างในช่วง Conceptual Design

การประมาณราคาก่อสร้างในขั้นแนวคิดการออกแบบ (Conceptual Design) หรือ Rough Order of Magnitude (ROM) Estimate เป็นกระบวนการประเมินงบประมาณเบื้องต้นของโครงการก่อนมีแบบก่อสร้างที่สมบูรณ์ โดยอาศัยข้อมูลคร่าว ๆ เช่น พื้นที่ใช้สอยและประเภทอาคาร วิธีที่นิยมได้แก่ การคำนวณราคาต่อหน่วยพื้นที่ การอ้างอิงจากโครงการที่คล้ายกัน การใช้ต้นทุนตามฟังก์ชันของอาคาร และการปรับราคาด้วยดัชนีค่าก่อสร้าง ทั้งนี้เพื่อนำไปใช้ประกอบการตัดสินใจในระยะเริ่มต้นของโครงการอย่างมีประสิทธิภาพ [9]

2.3.2 การประมาณราคาแบบ Detailed Design

การประมาณราคาในระยะ Detailed Design เป็นขั้นตอนที่ใช้แบบร่างทางสถาปัตยกรรมซึ่งมีรายละเอียดชัดเจนมากขึ้น เช่น แปลนพื้น รูปด้าน และข้อมูลวัสดุเบื้องต้น เพื่อนำมาถอดปริมาณงานและประเมินค่าใช้จ่ายเบื้องต้น โดยมีความแม่นยำอยู่ในช่วง $\pm 15\%$ ถึง $\pm 25\%$ วิธีที่ใช้ ได้แก่ การถอดปริมาณเบื้องต้นตามหมวดงาน การใช้ราคาต่อหน่วยจากฐานข้อมูลที่เชื่อถือได้ และการรวมต้นทุนพร้อมค่าดำเนินการและกำไรของผู้รับเหมา ทั้งนี้เพื่อวัตถุประสงค์ในการอนุมัติงบประมาณหรือวางแผนทางการเงินก่อนเข้าสู่ขั้นตอนออกแบบรายละเอียด [9]

2.4 หลักการที่ใช้ในการคำนวณปริมาณวัสดุ

2.4.1 การถอดแบบ

การแยกงานก่อสร้างออกเป็นปริมาณงานย่อยๆ แล้วบันทึกลงในแบบฟอร์มมาตรฐานเดียวกัน

2.4.2 การใส่ราคาวัสดุ

การหาราคาวัสดุของงานย่อยแต่ละประเภท จากแหล่งข้อมูลต่างๆ เพื่อนำมาบันทึกลงในแบบฟอร์ม

2.4.3. การใส่ค่าแรงงาน

การหาราคาวัสดุของงานย่อยแต่ละประเภท จากแหล่งข้อมูลต่างๆ ซึ่งอาจเป็นราคาต่อหน่วยหรือชุด เพื่อนำมาบันทึกลงในแบบฟอร์ม

2.4.4. การใส่ค่าใช้จ่ายในการดำเนินการ

การกำหนดค่าใช้จ่าย ซึ่งมาในรูปแบบตาราง Factor F [4]

2.5 แบบจำลองโมเดลที่ใช้ในการประมาณงบประมาณ

ก่อสร้าง

2.5.1 K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) เป็นอัลกอริทึมการเรียนรู้แบบมีผู้สอนและไม่ใช้พารามิเตอร์ ใช้ข้อมูล K จุดที่ใกล้ที่สุดในการทำนาย โดยนับเสียงข้างมาก (classification) หรือหาค่าเฉลี่ย (regression) จำเป็นต้องปรับมาตราส่วนข้อมูลและเลือกค่า K ที่เหมาะสมเพื่อลด overfitting หรือ underfitting โมเดลนี้เป็น lazy learning ที่ใช้ข้อมูลขณะทำนาย [6]

2.5.2. Random Forest

Random Forest เป็นแบบจำลองในกลุ่ม Ensemble Learning ที่พัฒนาจาก Decision Tree โดยสร้างต้นไม้หลายต้นจากการสุ่มข้อมูล (Bootstrap Sampling) และสุ่มเลือกคุณลักษณะบางส่วนเพื่อลดความสัมพันธ์ระหว่างต้นไม้ การทำนายจะใช้วิธีเฉลี่ยผล (regression) หรือโหวตเสียงข้างมาก (classification) เพื่อเพิ่มความแม่นยำและลดปัญหา overfitting [7]

2.5.3. XGBoost

XGBoost (Extreme Gradient Boosting) เป็นอัลกอริทึมที่พัฒนาจาก Gradient Boosting ใช้การสร้างต้นไม้การตัดสินใจหลายต้นโดยเรียนรู้จากข้อผิดพลาดของต้นก่อนหน้า เพื่อปรับปรุงความแม่นยำ จุดเด่นคือความเร็วในการประมวลผลและการจัดการข้อมูลที่มีโครงสร้างได้ดี การทำนายใช้การรวมผลของต้นไม้ทั้งหมดเพื่อลดความผิดพลาดและลด overfitting [8]

2.5.4 Linear regression

Linear Regression เป็นอัลกอริทึมพื้นฐานสำหรับหาความสัมพันธ์เชิงเส้นระหว่างตัวแปรต้นและตัวแปรตาม โดยหาสมการเส้นตรงที่ลดความคลาดเคลื่อนรวมให้น้อยที่สุด โมเดลประเมินด้วยค่า R-squared, MAE, RMSE และอาจใช้ Cross-Validation เพื่อลด overfitting [10]

2.5.5. Ridge Regression

Ridge Regression เป็นการถดถอยเชิงเส้นที่เพิ่มการปรับค่าปกติ (regularization) เพื่อลด overfitting โดยเพิ่มพจน์ลงโทษ $\lambda \sum \beta^2$ ในฟังก์ชันการสูญเสีย เพื่อควบคุมขนาด

สัมประสิทธิ์ ข้อมูลควรผ่านการ standardized ก่อนใช้งาน โมเดลประเมินด้วย R-squared, RMSE, MAE และเลือก λ ที่เหมาะสมด้วย Cross-Validation [10]

2.5.6. Lasso Regression

Lasso Regression เป็นการถดถอยเชิงเส้นที่เพิ่มการปรับค่าปกติ (regularization) ด้วยพจน์ลงโทษ $\lambda \sum |\beta|$ ทำให้สามารถลดสัมประสิทธิ์บางตัวเป็นศูนย์ จึงช่วยเลือกตัวแปรที่สำคัญ ข้อมูลควร standardized ก่อนใช้งาน โมเดลประเมินด้วย R-squared, RMSE, MAE และใช้ Cross-Validation เลือกค่า λ ที่เหมาะสม [10]

2.5.7 Ensemble models

Ensemble Models เป็นเทคนิคที่รวมผลลัพธ์จากหลายโมเดลเพื่อเพิ่มความแม่นยำและความเสถียร โดยมีหลักการลด bias และ variance ผ่านวิธีสำคัญ เช่น Bagging (ฝึกหลายโมเดลบนข้อมูลสุ่ม เช่น Random Forest), Boosting (สร้างโมเดลลำดับเพื่อแก้ error เช่น XGBoost), และ Stacking (รวมโมเดลต่างชนิดด้วย meta-model) ข้อดีคือเพิ่มความแม่นยำ ลด overfitting แต่มีข้อจำกัดด้านทรัพยากรและความซับซ้อน [11]

2.6 การจัดสรรงบประมาณในการก่อสร้างภาครัฐ

2.6.1 งบประมาณแสดงรายการ (Line-Item Budgeting)

ควบคุมปัจจัยนำเข้า เพื่อไม่ให้มีการใช้จ่าย เกินไปจากที่กำหนด

2.6.2 งบประมาณแบบแสดงผลงาน (Performance Budgeting)

ต้องบริหารงานอย่างมีประสิทธิภาพ กำหนดจากการวิเคราะห์อย่างรอบคอบ

2.6.3 งบประมาณแสดงแผนงาน (Program Budgeting)

การนำแผนงานระยะยาว มาใช้ในการกำหนดงบประมาณ โดยต้องเชื่อมโยงกับรัฐบาล

2.6.4 งบประมาณแบบมุ่งเน้นผลงาน (Performance-Based Budgeting: PBB)

กำหนดเป้าหมาย กลยุทธ์ขององค์กรเป็นกฎเกณฑ์ [9]

3. วัตถุประสงค์และขอบเขตการวิจัย

การวิจัยนี้ใช้ข้อมูลจากโครงการก่อสร้างภาครัฐจำนวน 100 โครงการในปีงบประมาณ พ.ศ. 2567 ซึ่งเผยแพร่ผ่านเว็บไซต์ของกรมบัญชีกลาง กระทรวงการคลัง ข้อมูลที่นำมาใช้ประกอบด้วยคุณลักษณะทางกายภาพของอาคาร เช่น พื้นที่ก่อสร้าง จำนวนชั้น พื้นที่ต่อชั้น และวัสดุหลักที่ใช้ในการ

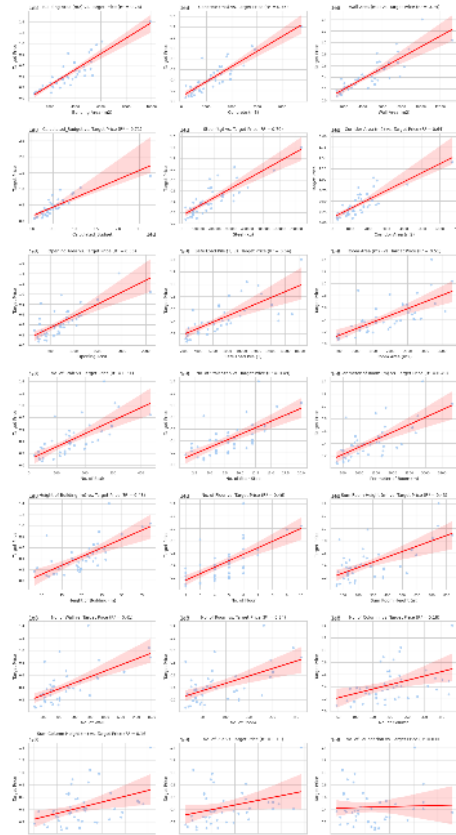
ก่อสร้าง เป็นต้น แบบจำลองการพยากรณ์จะถูกพัฒนาโดยใช้เทคนิค Machine Learning โดยเน้นการเปรียบเทียบประสิทธิภาพของแบบจำลองในด้านความแม่นยำและความสามารถในการอธิบายตัวแปรต้นที่ส่งผลต่องบประมาณ

4. ผลการดำเนินงานวิจัย

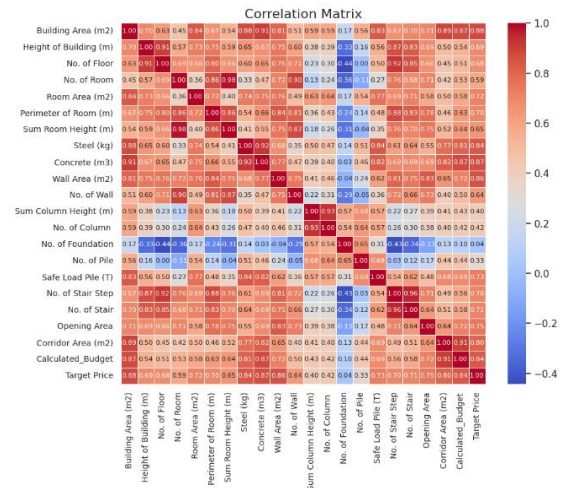
4.1 การวิเคราะห์ข้อมูลภาพรวมที่นำเข้ามาใช้ในการวิจัย

ในการประยุกต์ใช้แบบจำลองที่นำเสนอ คณะผู้วิจัยได้รวบรวมข้อมูลโครงการก่อสร้างอาคารภาครัฐจำนวน 100 โครงการจากระบบการจัดซื้อจัดจ้างทั่วประเทศ โดยใช้ราคากลางจากเอกสารประมาณการค่าก่อสร้าง (BOQ) และขอบเขตของงาน (TOR) เป็นตัวแปรตาม ข้อมูลที่รวบรวมมีจำนวน 21 ตัวแปรซึ่งเป็นคุณลักษณะของโครงการ เช่น ขนาดพื้นที่ จำนวนชั้น และวัสดุก่อสร้าง

เบื้องต้นได้วิเคราะห์ข้อมูลจาก 50 โครงการแรกเพื่อตรวจสอบความสัมพันธ์ระหว่างตัวแปรต่าง ๆ กับราคากลาง โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์ (R^2) ในการประเมินระดับความสัมพันธ์ จากนั้นจึงคัดเลือกตัวแปรที่มีผลต่อราคากลางอย่างมีนัยสำคัญมาใช้ร่วมกับข้อมูลอีก 50 โครงการสำหรับเตรียมชุดข้อมูลในการฝึกและทดสอบแบบจำลอง Machine Learning ต่อไป



รูปที่ 1 ค่าความสัมพันธ์ระหว่างตัวแปรและราคาโครงการ

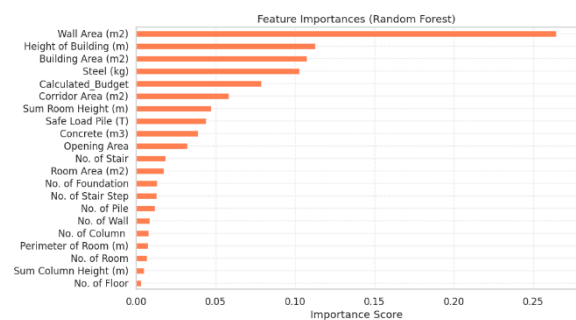


รูปที่ 2 Correlation Matrix

จากผลการวิเคราะห์ พบว่า ค่าความสัมพันธ์ที่สูงกว่า 0.75 ขึ้นไป ถือว่ามีความสัมพันธ์กันในระดับสูงตาม ซึ่งบ่งชี้ว่าตัวแปรเหล่านี้มีแนวโน้มเคลื่อนไหวไปในทิศทางเดียวกันอย่างชัดเจน

4.2 การวิเคราะห์ความสำคัญของตัวแปร (Feature Importance)

การวิเคราะห์ความสำคัญของตัวแปร (Feature Importance) ช่วยให้สามารถระบุได้ว่าตัวแปรใดมีผลต่อการพยากรณ์ราคากลางของโครงการก่อสร้างมากหรือน้อยเพียงใด และยังสามารถนำข้อมูลนี้ไปใช้ในการเลือกตัวแปรที่เหมาะสมเพื่อพัฒนาแบบจำลองในขั้นตอนถัดไป



รูปที่ 3 ผลการวิเคราะห์ Feature Importances

4.3 การคัดเลือกตัวแปรและการเพิ่มความหลากหลายของข้อมูล (Data Engineering)

4.3.1 ผลการคัดเลือกตัวแปร

ตารางที่ 1 ข้อมูลในการประมาณงบประมาณการก่อสร้าง

ชื่อข้อมูล (Feature)	คำอธิบาย
Building Area	พื้นที่ใช้สอยของโครงการทั้งหมด (ตร.ม.)
Area per floor	พื้นที่ใช้สอยของโครงการ ต่อ (ตร.ม./ชั้น)
No. of floor	จำนวนชั้นทั้งหมดของโครงการ
Steel	น้ำหนักเหล็กรวมทั้งโครงการ (กิโลกรัม)
Concrete	ปริมาตรของคอนกรีตที่ใช้ทั้งโครงการ (ลบ.ม.)
Area of wall	พื้นที่ของกำแพงทั้งโครงการ (ตร.ม.)
Pile	น้ำหนักที่เสาเข็มทั้งหมดรับได้ (ตัน)
Step of stair	จำนวนขั้นบันได (ขั้น)
Budget	ราคาประมาณของโครงการที่ได้จากการนำราคาพื้นที่ต่อ ตร.ม. ของสมาคมผู้ประเมินค่าทรัพย์สินแห่งประเทศไทย มาใช้ (บาท)

4.3.2 การใช้ราคาประมาณการก่อสร้างต่อหน่วยในกระบวนการสร้างแบบจำลอง

บัญชีราคามาตรฐานต่อหน่วย วิศวกรรมสิ่งปลูกสร้าง
สมาคมผู้ประเมินค่าทรัพย์สินแห่งประเทศไทย พ.ศ.2566 - 2567

รายการ	หน่วย	ค่า	ปริมาณ	ราคาต่อหน่วย
1. ปูนซีเมนต์				
1001 ซีเมนต์ปอร์ตแลนด์ 42.5	ตร.ม.	11,300	13,000	14,600
1002 ซีเมนต์ปอร์ตแลนด์ 42.5 (สีเทา)	ตร.ม.	13,500	14,100	15,600
1003 ซีเมนต์ปอร์ตแลนด์ 2 ชั้น	ตร.ม.	10,100	12,300	14,100
2. ทราย				
2004 ทรายสะอาดสีน้ำตาล	ตร.ม.	12,200	13,900	15,800
2005 ทรายสะอาดสีน้ำตาล 2 ชั้น (สีเทา)	ตร.ม.	11,300	13,100	15,100
2006 ทรายสะอาดสีน้ำตาล 2-4 ชั้น (สีเทา)	ตร.ม.	23,400	26,500	31,800
2007 ทรายสะอาดสีน้ำตาล 3-4 ชั้น (สีเทา)	ตร.ม.	12,400	14,450	17,500
2008 ทรายสะอาดสีน้ำตาล 2-3 ชั้น	ตร.ม.	10,300	12,200	14,000
3. วัสดุเสริมคอนกรีต				
3009 วัสดุเสริมคอนกรีต 2 ชั้น	ตร.ม.	9,400	11,350	13,900
4. วัสดุเสริมคอนกรีต				
4010 วัสดุเสริมคอนกรีต 1 ชั้น (สีเทา)	ตร.ม.	14,400	15,600	17,200
4011 วัสดุเสริมคอนกรีต 2 ชั้น (สีเทา)	ตร.ม.	21,700	24,900	28,600
4012 วัสดุเสริมคอนกรีต 3 ชั้น (สีเทา)	ตร.ม.	13,100	13,700	15,300

รูปที่ 4 ตารางข้อมูลราคากลางมาตรฐานจากบัญชีราคามาตรฐานต่อหน่วยวิศวกรรมสิ่งปลูกสร้าง

ในการสร้างแบบจำลอง Machine Learning เพื่อพยากรณ์งบประมาณก่อสร้างเบื้องต้น พบว่าข้อมูลจากโครงการจริงมีความหลากหลายและไม่สมบูรณ์ โดยเฉพาะการขาดข้อมูลประเภทอาคารซึ่งเป็นปัจจัยสำคัญ จึงมีการนำ “ราคาก่อสร้างต่อหน่วยมาตรฐาน” จากแหล่งอ้างอิงภาครัฐหรือวิชาชีพมาใช้เป็นฟีเจอร์เสริม เพื่อช่วยสะท้อนบริบทของประเภทและช่วงเวลาอย่างเหมาะสม ด้วยเหตุผลหลัก ดังนี้

4.3.2.1 ทำหน้าที่ Normalization ตามประเภทและเวลา

ราคาต่อหน่วยสะท้อนต้นทุนพื้นฐานตามประเภทอาคารและปีงบประมาณ ช่วยให้โมเดลเข้าใจบริบทของข้อมูลได้แม้ในกรณีที่ไม่มีรายละเอียดครบถ้วน และลดความเหลื่อมล้ำด้านเวลาในการรวบรวมข้อมูล

4.3.2.2 ทดแทนฟีเจอร์ “ประเภทอาคาร” ที่ไม่มีในชุดข้อมูล

การใส่ประเภทอาคารเป็นข้อมูลเชิงข้อความโดยตรงทำให้โมเดลมีประสิทธิภาพลดลง จากปัญหา class imbalance

และความไม่ชัดเจนของการจัดกลุ่ม การใช้ราคาต่อหน่วยที่แยกตามประเภทจึงเป็นการ encode ประเภทอาคารในเชิงตัวเลขทางอ้อมได้อย่างมีประสิทธิภาพกว่า

4.3.2.3 ช่วยลด noise และเพิ่มความเสถียรของแบบจำลอง
ราคาต่อหน่วยมาตรฐานทำหน้าที่เป็น anchor จากความรู้ผู้เชี่ยวชาญ ช่วยให้แบบจำลองไม่พยากรณ์ค่าที่หลุดจากความเป็นจริง โดยเฉพาะในระยะที่ข้อมูลยังมีความไม่แน่นอนสูง เช่น Conceptual Design

4.3.2.4 เพิ่มความยืดหยุ่นในการนำไปใช้งานจริง

เมื่อผู้ใช้งานไม่มีข้อมูลเชิงลึก เช่น แบบก่อสร้างหรือรายการวัสดุ ฟีเจอร์นี้ช่วยให้โมเดลสามารถพยากรณ์ได้แม่นยำในระดับที่ใช้งานจริงได้ทันที

แม้ราคาต่อหน่วยจะไม่ใช่ข้อมูลราคาจริงจากโครงการ แต่ถือเป็นองค์ความรู้จากผู้เชี่ยวชาญที่ช่วยเสริมความแม่นยำให้แบบจำลองในสภาพที่ข้อมูลมีข้อจำกัด การเลือกใช้ฟีเจอร์นี้จึงเป็นแนวทางหนึ่งในการแปลงความรู้ทางวิศวกรรมเชิงประสบการณ์ให้เป็นข้อมูลเชิงปริมาณที่ระบบสามารถเรียนรู้และใช้งานได้จริง

4.3.3 กลุ่มข้อมูล Conceptual Stage

ข้อมูลที่ใช้ในระยะ Conceptual Stage เป็นข้อมูลพื้นฐานซึ่งสามารถหาได้ตั้งแต่ขั้นตอนการวางแผนแนวคิดของโครงการ โดยประกอบด้วย

- Building Area (m²) พื้นที่ใช้สอยรวมของโครงการ
- Area_per_floor พื้นที่ใช้สอยต่อชั้น
- No. of floor จำนวนชั้นของอาคาร
- Budgetราคาก่อสร้างที่ประเมินจากราคาต่อหน่วยของปีงบประมาณนั้น ๆ

ข้อมูลทั้ง 4 รายการนี้ถูกนำมาใช้ในการประมาณต้นทุนงบประมาณการก่อสร้าง โดยมีตัวแปรเป้าหมายคือ REAL (ราคากลางในการก่อสร้าง) จุดเด่นของชุดข้อมูล Conceptual คือสามารถใช้ได้ตั้งแต่ระยะต้นของโครงการและมักมีอยู่ในเอกสารเบื้องต้นของโครงการภาครัฐ

4.3.3 กลุ่มข้อมูล Detailed Stage

ข้อมูลในระยะ Detailed Stage จะมีรายละเอียดเชิงวิศวกรรมที่มากขึ้น มักได้จากการออกแบบเบื้องต้น BOQ หรือเอกสารการประมูล ประกอบด้วย

- Building Area (m²) พื้นที่ใช้สอยรวมของโครงการ
- Area_per_floor พื้นที่ต่อชั้น
- No. of floor จำนวนชั้นของอาคาร

- Steel (kg) ปริมาณเหล็กที่ใช้
- Concrete (m³) ปริมาตรคอนกรีตที่ใช้
- Area_of_wall พื้นที่ผนัง
- Pile (T) น้ำหนักบรรทุกของเสาเข็ม
- Step_of_stair จำนวนขั้นบันได
- Budget ราคาต่อหน่วยที่ Normalize แล้ว

ข้อมูลในกลุ่มนี้ถูกนำไปสร้างโมเดล Detailed Design Stage ซึ่งจะถูกเปรียบเทียบกับโมเดลจาก Conceptual Stage โดยมีจุดประสงค์เพื่อประเมินว่า โมเดลที่ใช้ข้อมูลพื้นฐานเพียงเล็กน้อยใน Conceptual Stage สามารถพยากรณ์งบประมาณได้ใกล้เคียงกับโมเดลที่ใช้ข้อมูลเชิงลึกจาก Detailed Stage หรือไม่ และเพียงพอสำหรับการใช้งานจริงหรือไม่

4.3.5 การเพิ่มความหลากหลายของข้อมูล (Data Engineering)

ในการศึกษานี้ มีการสร้างฟีเจอร์ใหม่ (Feature Engineering) สำหรับข้อมูลสองช่วง ได้แก่ Conceptual และ Detailed เพื่อเพิ่มความหลากหลายและความเข้าใจเชิงลึกเกี่ยวกับโครงการก่อสร้าง โดยในช่วง Conceptual ได้สร้างฟีเจอร์จากข้อมูลพื้นฐาน เช่น พื้นที่ใช้สอย จำนวนชั้น และงบประมาณ ได้แก่ Area_x_Floor, Budget_per_floor, Area_per_budget, Floor_to_area_ratio และ Area_ratio_per_floor เพื่อสะท้อนขนาดและต้นทุนโดยรวมของโครงการ แม้ยังไม่มีข้อมูลโครงสร้างละเอียด

ส่วนช่วง Detailed ซึ่งมีรายละเอียดมากขึ้น เช่น ปริมาณวัสดุและองค์ประกอบโครงสร้าง ได้มีการสร้างฟีเจอร์ขึ้นสูง เช่น Structure_Density, Steel_to_Concrete_Ratio, Total_Stair_Length, Wall_Area_per_Floor, Pile_per_m2, Budget_per_m2, Area_Growth_per_Floor, High_Rise, Material_Intensity_per_Floor และ Budget_Efficiency เพื่อสะท้อนมิติด้านวัสดุ ประสิทธิภาพการใช้ทรัพยากร และความซับซ้อนของอาคาร ฟีเจอร์เหล่านี้ช่วยเพิ่มศักยภาพของโมเดลในการพยากรณ์งบประมาณอย่างแม่นยำ ครอบคลุมทั้งอาคารทั่วไป อาคารสูง และโครงการที่มีข้อมูลจำกัดในช่วงออกแบบเบื้องต้น.

4.4 ผลการเปรียบเทียบประสิทธิภาพของแต่ละเครื่องมือ Machine Learning พื้นฐาน ในการทำนายงบประมาณราคาก่อสร้างเพื่อหาแบบจำลองที่เหมาะสมที่สุด

4.4.1 ผลลัพธ์ Base Model ในช่วง Conceptual Design Phase

ตารางที่ 2 ผลลัพธ์ Base Model ในช่วง Conceptual Phase

แบบจำลอง	R ²	MAE	RMSE	Explained Variance
Linear Regression	0.877	10,692,042.64	10,692,042.64	0.88
Lasso Regression	0.877	10,690,358.99	15,475,919.05	0.88
Ridge Regression	0.895	9,273,575.95	14,283,289.11	0.90
XGBoost	0.905	9,274,286.45	13,602,149.86	0.91
KNN	0.901	8,685,118.07	13,876,670.62	0.90
Random Forest	0.896	10,259,708.13	14,227,893.06	0.90

จากผลการประเมินประสิทธิภาพของแบบจำลองการทำนายพบว่า XGBoost เป็นแบบจำลองที่ให้ผลลัพธ์ดีที่สุด โดยมีค่าค่าสัมประสิทธิ์การกำหนด (R²) เท่ากับ 0.905 และค่า Explained Variance เท่ากับ 0.91 ซึ่งสะท้อนถึงความสามารถในการอธิบายความแปรปรวนของข้อมูลเป้าหมายได้สูงสุดเมื่อเทียบกับแบบจำลองอื่น นอกจากนี้ ยังมีค่า RMSE ต่ำที่สุดที่ 13,602,149.86 แสดงถึงความคลาดเคลื่อนเชิงรากที่สองต่ำที่สุด อย่างไรก็ตาม ค่า MAE ของ XGBoost อยู่ที่ 9,274,286.45 ซึ่งแม้จะไม่สูงมาก แต่ยังเป็นรองแบบจำลอง KNN เล็กน้อย

สำหรับแบบจำลอง K-Nearest Neighbors (KNN) แม้ว่าค่า R² จะอยู่ที่ 0.901 ซึ่งต่ำกว่า XGBoost เล็กน้อย แต่ให้ค่า MAE ต่ำที่สุดในบรรดาทุกโมเดล คือ 8,685,118.07 สะท้อนว่าแบบจำลองนี้มีความคลาดเคลื่อนเฉลี่ยในเชิงแนวตรงต่ำที่สุด อย่างไรก็ตาม ค่า RMSE ของ KNN ยังสูงกว่า XGBoost เล็กน้อย ซึ่งอาจบ่งชี้ว่ามีบางกรณีที่ผลการทำนายเบี่ยงเบนไปจากค่าจริงในระดับสูง

เมื่อพิจารณาแบบจำลองเชิงเส้น ได้แก่ Linear Regression, Lasso Regression และ Ridge Regression พบว่า Ridge Regression มีความแม่นยำสูงสุดในกลุ่มนี้ โดยมีค่า R² เท่ากับ 0.895 และค่า RMSE ต่ำกว่า Linear และ Lasso อย่างชัดเจน อย่างไรก็ตาม แบบจำลองทั้งสามยังคงมีประสิทธิภาพต่ำกว่าแบบจำลองเชิงไม่เชิงเส้น เช่น XGBoost, KNN และ Random Forest

ในส่วนของการ Random Forest แม้ว่าจะมีค่า R² เท่ากับ 0.896 ซึ่งใกล้เคียงกับ Ridge Regression และค่า RMSE อยู่ที่ 14,227,893.06 ซึ่งอยู่ในระดับปานกลาง แต่ค่า MAE ค่อนข้างสูงเมื่อเทียบกับแบบจำลองอื่น ๆ แสดงให้เห็นถึงความไม่สม่ำเสมอของความคลาดเคลื่อนในการทำนาย

4.4.2 ผลลัพธ์ Base Model ในช่วง Detailed Design Phase

ตารางที่ 3 ผลลัพธ์ Base Model ในช่วง Detailed Design Phase

แบบจำลอง	R ²	MAE	RMSE	Explained Variance
Linear Regression	0.873	11,172,131.42	15,751,870.25	0.87
Lasso Regression	0.873	11,165,430.94	15,744,018.56	0.87
Ridge Regression	0.920	7,907,611.97	12,484,208.42	0.92
XGBoost	0.890	9,418,200.7	14,662,537.82	0.89
KNN	0.849	12,629,708.53	17,149,930.66	0.87
Random Forest	0.888	9,090,327.1	14,757,037.59	0.89

โมเดล Ridge Regression ให้ผลการทำนายที่ดีที่สุดในกลุ่ม โดยมีค่า $R^2 = 0.92$, MAE = 7,907,611.97 และ RMSE = 12,484,208.42 แสดงให้เห็นถึงความสามารถในการอธิบายความแปรปรวนของข้อมูลได้มากที่สุดและให้ค่าความคลาดเคลื่อนต่ำที่สุดเมื่อเปรียบเทียบกับโมเดลอื่น ๆ

โมเดล XGBoost และ Random Forest แสดงผลการทำนายที่ตรงลงมา โดยมีค่า $R^2 = 0.89$ และ 0.888 ตามลำดับ ซึ่งแสดงถึงความสามารถในการจัดการกับข้อมูลที่ซับซ้อนและไม่เป็นเชิงเส้นได้ดี โดยที่ XGBoost ให้ RMSE ต่ำกว่าเล็กน้อย ขณะที่ Random Forest มีค่า MAE ที่ต่ำกว่า

ในส่วนโมเดลเชิงเส้นอย่าง Linear Regression และ Lasso Regression มีประสิทธิภาพใกล้เคียงกัน โดยให้ค่า $R^2 = 0.873$ และ RMSE ประมาณ 15.75 ล้านบาท แสดงถึงความแม่นยำในระดับปานกลาง และเหมาะกับข้อมูลที่มีความสัมพันธ์เชิงเส้นชัดเจน

และโมเดล K-Nearest Neighbors (KNN) ให้ผลลัพธ์ที่แม่นยำน้อยที่สุด โดยมีค่า $R^2 = 0.849$, MAE = 12,629,708.53 และ RMSE = 17,149,930.66 ซึ่งแสดงถึงข้อจำกัดของโมเดลนี้ในการประมวลผลข้อมูลจำนวนมากหรือมีมิติสูง

4.5 การวิเคราะห์การทำ Ensemble Model ที่ช่วยพัฒนาประสิทธิภาพในการวิเคราะห์ข้อมูล

4.5.1 ในขั้น Conceptual Design Phase

ตารางที่ 4 Model Comparison ของแบบจำลอง Ensemble ต่างๆ ของ Conceptual Design Phase

แบบจำลอง	R^2	MAE	RMSE	Explained Variance
Ensemble Averaging	0.905035	8,511,293.10	13,595,085.72	0.905108
Weighted Averaging	0.905208	8,505,787.99	13,582,681.12	0.905282
Bagged_RF_KNN_XGBoost	0.909139	9,045,578.82	13,298,096.71	0.909168
Bagged_All	0.905035	8,511,293.10	13,595,085.72	0.905108

โมเดล Bagged_RF_KNN_XGBoost มีผลการทำนายที่ดีที่สุด โดยให้ค่า $R^2 = 0.909139$, MAE = 9,045,578.82 และ RMSE = 13,298,096.71 ซึ่งแสดงถึงความสามารถในการอธิบายความแปรปรวนของข้อมูลได้มากที่สุดเมื่อเปรียบเทียบกับโมเดลอื่น ๆ ในส่วนของโมเดล Weighted Averaging ให้ค่าความคลาดเคลื่อนเฉลี่ยน้อยที่สุด โดยที่สังเกตว่าค่า (MAE = 8,505,787.99) แต่ค่า R^2 ยังต่ำกว่าโมเดลแบบ Bagged โมเดล Ensemble Averaging และ Bagged_All ให้ค่าประเมินที่ใกล้เคียงกัน และยังมีความแม่นยำในระดับสูงเช่นกันโดยสังเกตได้จากค่า R^2

4.5.2 ในขั้น Detailed Design Phase

ตารางที่ 5 Model Comparison ของแบบจำลอง Ensemble ต่างๆ ของ Detailed Design Phase

แบบจำลอง	R^2	MAE	RMSE	Explained Variance
Ensemble Averaging	0.912428	8,511,757.04	13,055,146.72	0.914195
Weighted Averaging	0.912578	8,504,508.14	13,043,961.31	0.914297
Bagged_RF_KNN_XGBoost	0.895823	8,934,071.50	14,239,189.63	0.902987
Bagged_All	0.912428	8,511,757.04	13,055,146.72	0.914195

โมเดลที่ให้ผลดีที่สุดในระยะนี้คือ Weighted Averaging โดยมีค่า $R^2 = 0.912578$ และค่า RMSE ต่ำที่สุด

โมเดล Ensemble Averaging และ Bagged_All ให้ค่าผลลัพธ์เท่ากันทุกตัวชี้วัด ($R^2 = 0.912428$, RMSE = 13,055,146.72) ซึ่งอาจบอกได้ว่าโมเดล 2 ตัวนี้ กับชุดข้อมูลที่นำมาทดสอบ ให้ผลที่ใกล้เคียงกัน ในขณะที่ Bagged_RF_KNN_XGBoost มีผลลัพธ์ด้อยกว่ารูปแบบอื่นเล็กน้อยในระยะนี้ โดยมีค่า R^2 เท่ากับ 0.895823 และค่า RMSE สูงที่สุด

4.6 ผลการเปรียบเทียบประสิทธิภาพของแต่ละเครื่องมือ Machine Learning ในการทำนายงบประมาณราคาก่อสร้างเพื่อหาแบบจำลองที่เหมาะสมที่สุด

4.6.1 การประเมินผลแบบจำลองในช่วง Conceptual Phase

ตาราง 6 Model Comparison ของแบบจำลอง Conceptual Design Phase

	แบบจำลอง	R^2	MAE	RMSE	Explained Variance
Ensemble Model	Ensemble Averaging	0.905035	8,511,293.10	13,595,085.72	0.905108
	Weighted Averaging	0.905208	8,505,787.99	13,582,681.12	0.905282
	Bagged_RF_KNN_XGBoost	0.909139	9,045,578.82	13,298,096.71	0.909168
	Bagged_All	0.905035	8,511,293.10	13,595,085.72	0.905108
Base Model	Linear Regression	0.877	10,692,042.64	10,692,042.64	0.88
	Lasso Regression	0.877	10,690,358.99	15,475,919.05	0.88
	Ridge Regression	0.895	9,273,575.95	14,283,289.11	0.90
	XGBoost	0.905	9,274,286.45	13,602,149.86	0.91
	KNN	0.901	8,685,118.07	13,876,670.62	0.90
	Random Forest	0.896	10,259,708.13	14,227,893.06	0.90

การทดลองนี้มีเป้าหมายเพื่อเปรียบเทียบประสิทธิภาพของแบบจำลอง Machine Learning หลายประเภทในการพยากรณ์งบประมาณก่อสร้าง โดยใช้ตัวชี้วัดหลัก ได้แก่ R^2 , MAE, RMSE และ Explained Variance เพื่อหาโมเดลที่แม่นยำและเหมาะสมที่สุดกับการใช้งานจริง โดยแบ่งแบบจำลองออกเป็น 2 กลุ่ม คือ Ensemble Model (เช่น Bagged_RF_KNN_XGBoost, Ensemble Averaging, Weighted Averaging) และ Base Model (เช่น Linear

Regression, Ridge Regression, XGBoost, KNN) ผลการทดลองพบว่า Bagged_RF_KNN_XGBoost ให้ค่า R^2 สูงสุดที่ 0.909 และมี MAE กับ RMSE อยู่ในระดับแข่งขันได้ ส่วน Weighted Averaging มีค่า MAE ต่ำสุดที่ 8.5 ล้านบาท แสดงถึงความแม่นยำในเชิงสัมบูรณ์ โมเดลเดี่ยวอย่าง XGBoost และ KNN ก็มี R^2 สูง (0.905 และ 0.901 ตามลำดับ) เกือบเคียงกับแบบรวมผล ในขณะที่ Linear และ Lasso Regression มีประสิทธิภาพต่ำที่สุด ($R^2 = 0.877$, $MAE > 10$ ล้านบาท) แสดงข้อจำกัดของโมเดลเชิงเส้นเมื่อเจอกับข้อมูลที่ซับซ้อน

โดยสรุป กลุ่ม Ensemble Model โดยเฉพาะ Bagging และ Averaging Techniques ให้ผลแม่นยำและเสถียรกว่าโมเดลเดี่ยว ทั้งในด้าน R^2 และค่าคลาดเคลื่อน ทำให้โมเดล Weighted Averaging และ Bagged_RF_KNN_XGBoost เป็นตัวเลือกที่เหมาะสมที่สุดสำหรับการนำไปใช้งานจริง โดยมีความสมดุลระหว่างความแม่นยำ ความยืดหยุ่น และการรองรับข้อมูลที่หลากหลายในอนาคต

4.6.2 การประเมินผลแบบจำลองในช่วง Detailed Design Phase

ตาราง 7 Model Comparison ของแบบจำลอง Detailed Design Phase

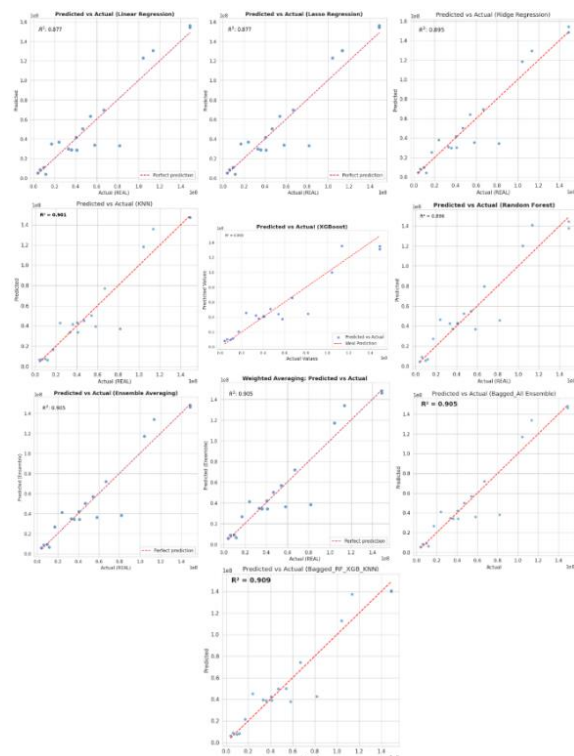
	แบบจำลอง	R^2	MAE	RMSE	Explained Variance
Ensemble Model	Ensemble Averaging	0.912428	8,511,757.04	13,055,146.72	0.914195
	Weighted Averaging	0.912578	8,504,508.14	13,043,961.31	0.914297
	Bagged_RF_KNN_XGBoost	0.895823	8,934,071.50	14,239,189.63	0.902987
	Bagged_All	0.912428	8,511,757.04	13,055,146.72	0.914195
Base Model	Linear Regression	0.873	11,172,131.42	15,751,870.25	0.87
	Lasso Regression	0.873	11,165,430.94	15,744,018.56	0.87
	Ridge Regression	0.920	7,907,611.97	12,484,208.42	0.92
	XGBoost	0.890	9,418,200.7	14,662,537.82	0.89
	KNN	0.849	12,629,708.53	17,149,930.66	0.87
	Random Forest	0.888	9,090,327.1	14,757,037.59	0.89

การทดสอบในช่วง Detailed Design มีเป้าหมายเพื่อประเมินความสามารถของแบบจำลองที่คัดเลือกไว้จาก Conceptual Design ภายใต้อายุที่ซับซ้อนขึ้น โดยวัดด้วย R^2 , MAE, RMSE และ Explained Variance เพื่อวัดความแม่นยำ เสถียรภาพ และความสามารถในการ generalize ผลการทดลองพบว่า Ridge Regression มีประสิทธิภาพสูงสุดในทุกตัวชี้วัด ($R^2 = 0.920$, $MAE \approx 7.91$ ล้าน, $RMSE \approx 12.48$ ล้าน, $Explained\ Variance = 0.92$) สะท้อนความสามารถในการจัดการกับข้อมูลที่ซับซ้อน แม้จะเป็นโมเดลเชิงเส้น แต่การใช้ regularization ช่วยลด overfitting และจัดการ multicollinearity ได้ดี ทำให้มี

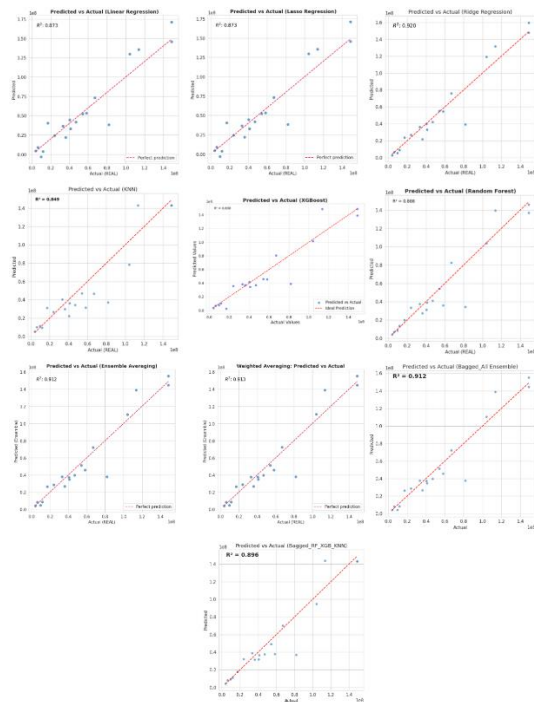
เสถียรภาพและความแม่นยำสูงสุด

กลุ่ม Ensemble Models เช่น Weighted Averaging และ Bagged_All แม้จะให้ผลดี ($R^2 \approx 0.912$, $MAE \approx 8.5$ ล้าน) แต่ยังเป็นรอง Ridge Regression ในทุกตัวชี้วัด โดยเฉพาะ MAE และ RMSE ที่ยังสูงกว่า แสดงว่าแม้การรวมโมเดลช่วยเพิ่มความเสถียร แต่ไม่ได้เสริมความแม่นยำเหนือกว่าโมเดลเดี่ยวที่ปรับจูนมาอย่างดี

Bagged_RF_KNN_XGBoost มี $R^2 = 0.896$ และ $MAE > 8.9$ ล้าน ยังเป็นรองทั้ง Ridge และ Ensemble Averaging ส่วน Linear และ Lasso Regression ยังคงประสิทธิภาพต่ำ ($R^2 = 0.873$, $MAE > 11$ ล้าน) ไม่สามารถรับมือกับความซับซ้อนของข้อมูลในช่วงนี้ได้สุดท้าย KNN แสดงผลลัพธ์แย่ที่สุด ($MAE \approx 12.63$ ล้าน, $RMSE \approx 17.15$ ล้าน) แสดงข้อจำกัดในการ generalize ข้อมูลที่ซับซ้อนหรือมีจำนวนจุดข้อมูลมากสรุปได้ว่า Ridge Regression เหมาะสมที่สุดในบริบทของ Detailed Design ด้วยความแม่นยำ เสถียรภาพ และความสามารถในการจัดการข้อมูลที่ซับซ้อน โดยให้ผลเหนือกว่าแบบจำลองทุกกลุ่ม รวมถึง Ensemble Model ซึ่งเน้นย้ำว่า การเลือกและปรับจูนโมเดลเดี่ยวที่เหมาะสมกับบริบทข้อมูล อาจให้ผลดีกว่าการรวมโมเดลหลายตัวแบบไม่เฉพาะเจาะจง



รูปที่ 4 กราฟแสดงผล Predicted Budget กับ Actual Budget ใน Conceptual Phase



รูปที่ 5 กราฟแสดงผล Predicted Budget กับ Actual Budget ใน Detailed Phase

4.7 วิเคราะห์ผลการเปรียบเทียบ Ensemble Model และชุดความละเอียดของข้อมูล

การเปรียบเทียบประสิทธิภาพของแบบจำลอง Machine Learning ระหว่างช่วง Conceptual Design และ Detailed Design แสดงให้เห็นความแตกต่างในด้านผลลัพธ์ที่ได้จากข้อมูลที่มีระดับความละเอียดต่างกัน

ในช่วง Conceptual Design ซึ่งใช้ข้อมูลเบื้องต้น เช่น พื้นที่ใช้สอย จำนวนชั้น และประเภทอาคาร พบว่าแบบจำลองในกลุ่ม Ensemble Models ให้ผลลัพธ์ที่แม่นยำและเสถียรอย่างชัดเจน โดยเฉพาะ Bagged_RF_KNN_XGBoost ที่ให้ค่า $R^2 = 0.909139$ และ Weighted Averaging ซึ่งมี MAE ต่ำสุดที่ 8.5 ล้านบาท ขณะที่ XGBoost และ KNN ก็มี R^2 สูงถึง 0.905 และ 0.901 ตามลำดับ

ในทางกลับกัน ช่วง Detailed Design ซึ่งข้อมูลมีความละเอียดมากขึ้น พบว่าแบบจำลอง Ridge Regression แสดงประสิทธิภาพดีที่สุด ด้วยค่า $R^2 = 0.920$, MAE = 7.9 ล้านบาท และ RMSE = 12.48 ล้านบาท ซึ่งเหนือกว่าทุกแบบจำลองทั้ง Base และ Ensemble รวมถึงค่า Explained Variance ที่สอดคล้องกันที่ 0.92

Ensemble Models ในช่วง Detailed เช่น Ensemble Averaging, Weighted Averaging, และ Bagged_All มีค่า R^2 เฉลี่ยประมาณ 0.912 และ MAE อยู่ที่ราว 8.5 ล้านบาท ซึ่งแม้จะมีความแม่นยำสูง แต่ยังคงเป็นรอง Ridge Regression ในทุกตัวชี้วัด

โมเดล Bagged_RF_KNN_XGBoost ในช่วง Detailed Design ให้ค่า R^2 ที่ลดลงเหลือ 0.895823 และ MAE สูงกว่า 8.9 ล้านบาท ซึ่งต่ำกว่าผลลัพธ์ในช่วง Conceptual Design

แบบจำลองเชิงเส้นพื้นฐาน เช่น Linear Regression และ Lasso Regression มีค่า R^2 เท่ากันที่ 0.873 ทั้งในช่วง Conceptual และ Detailed แต่มี MAE สูงกว่า 11 ล้านบาทในช่วงหลัง สะท้อนข้อจำกัดในการจัดการกับข้อมูลที่ซับซ้อนมากขึ้นและ K-Nearest Neighbors (KNN) เป็นโมเดลที่ให้ผลลัพธ์ต่ำสุดในช่วง Detailed โดยมี MAE สูงถึง 12.63 ล้านบาท และ RMSE สูงสุดที่ 17.15 ล้านบาท

5. สรุปผลการศึกษา

การศึกษานี้มีวัตถุประสงค์เพื่อประเมินประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning) ในการทำนายราคาก่อสร้างในสองช่วงของการออกแบบ ได้แก่ Conceptual Design Phase และ Detailed Design Phase โดยเปรียบเทียบระหว่างแบบจำลองเดี่ยว เช่น Linear Regression, Ridge, Lasso, XGBoost, KNN, Random Forest กับแบบจำลองแบบผสม (Ensemble Models) ได้แก่ Averaging, Weighted Averaging และ Bagging

ในช่วง Conceptual Design Phase แบบจำลอง XGBoost ให้ผลลัพธ์ที่ดีที่สุดในกลุ่มเดี่ยว โดยมีค่า R^2 เท่ากับ 0.905 และ RMSE เท่ากับ 13.60 ล้านบาท แสดงถึงศักยภาพในการเรียนรู้ความสัมพันธ์เชิงซับซ้อนของข้อมูลที่ยังหายอบอยู่ในช่วงต้นของการออกแบบ ขณะที่แบบจำลอง Bagged_RF_KNN_XGBoost ซึ่งเป็นการผสมผสานแบบ Bagging ของ XGBoost, KNN และ Random Forest ให้ประสิทธิภาพสูงสุดในทุกตัวชี้วัด โดยมีค่า R^2 เท่ากับ 0.909 และ RMSE ลดลงเหลือ 13.30 ล้านบาท ทั้งนี้ การวิเคราะห์เพิ่มเติมพบว่าแบบจำลองเหล่านี้ให้ความแม่นยำในการทำนายราคาสูงในช่วงโครงการที่มีมูลค่าระหว่าง 1 ล้านบาทถึง 70 ล้านบาท โดยราคาที่เกินกว่า 80 ล้านบาทจะเริ่มมีแนวโน้มแปรผันจากกราฟเส้นตรง แม้จะยังเกาะกลุ่มในเทรนเดิม แต่แสดงให้เห็นถึงความผิดพลาดที่เพิ่มขึ้นเล็กน้อย ซึ่งยืนยันถึงประโยชน์ของการรวมจุดแข็งของแต่ละแบบจำลองเพื่อเพิ่มเสถียรภาพและความแม่นยำ เหมาะสมอย่างยิ่งสำหรับการขออนุมัติงบประมาณในขั้นต้นของโครงการ

ในช่วง Detailed Design Phase ซึ่งข้อมูลมีลักษณะเป็นระบบระเบียบและมีรายละเอียดมากขึ้น แบบจำลอง Ridge Regression ให้ผลลัพธ์ดีที่สุดในกลุ่มเดี่ยว โดยมีค่า R^2 เท่ากับ 0.920 และ RMSE ต่ำสุดที่ 12.48 ล้านบาท แสดงถึงประสิทธิภาพของโมเดลเชิงเส้นที่มีการ regularization ที่เหมาะสมกับข้อมูลที่มีความเสถียร แต่ในส่วนแบบจำลองแบบผสมอย่าง

Bagged_RF_KNN_XGBoost ยังคงมีประสิทธิภาพดีใกล้เคียงกัน โดยมี R^2 เท่ากับ 0.909 และ RMSE เท่ากับ 13.30 ล้านบาท ซึ่งแม้จะไม่ต่ำที่สุด แต่ก็แสดงถึงเสถียรภาพของการพยากรณ์ได้อย่างต่อเนื่องในทุกช่วงของข้อมูล

โดยสรุป การเลือกใช้แบบจำลองที่เหมาะสมกับลักษณะของข้อมูลในแต่ละช่วงของการออกแบบมีผลอย่างชัดเจนต่อความแม่นยำในการทำนายราคา โดย XGBoost เหมาะกับช่วง Conceptual ที่ข้อมูลยังหายาก ส่วน Ridge Regression เหมาะกับช่วง Detailed ที่ข้อมูลมีลักษณะเชิงเส้นมากขึ้น อย่างไรก็ตาม การใช้ Ensemble Model โดยเฉพาะ Bagged_RF_KNN_XGBoost แสดงให้เห็นถึงศักยภาพในการเพิ่มความเสถียร ลดความคลาดเคลื่อน และให้ผลการทำนายที่มีความแม่นยำครอบคลุมทั้งช่วงราคาต่ำและสูง จึงเป็นแนวทางที่เหมาะสมในการประเมินราคาก่อสร้างเพื่อการวางแผนงบประมาณในทุกระดับของการออกแบบ

6. ข้อเสนอแนะและแนวทางการพัฒนา

ปริญญานิพนธ์ฉบับนี้ถือเป็นการศึกษาหนึ่งทางเลือกในการพยากรณ์งบประมาณการก่อสร้าง โดยเลือกใช้ฐานข้อมูลจากเว็บไซต์กรมบัญชีกลาง ซึ่งถ้าหากต้องการเพิ่มความแม่นยำและความหลากหลายของข้อมูลให้กับแบบจำลองสามารถทำได้โดยการเก็บจำนวนข้อมูลเพิ่มขึ้นที่ใช้ในการทดลอง ราคาต่อหน่วยที่ใช้เป็นข้อมูลเฉลี่ยรายปี ไม่สะท้อนความผันผวนของเศรษฐกิจ เช่น เงินเฟ้อหรือราคาราคาวัสดุในช่วงเวลาจริง อาจทำให้ผลพยากรณ์คลาดเคลื่อนจากต้นทุนปัจจุบัน ควรปรับพีเจอร์ราคาต่อหน่วยให้สัมพันธ์กับดัชนีเศรษฐกิจรายปี

7. กิตติกรรมประกาศ

ปริญญานิพนธ์เรื่องนี้สามารถสำเร็จลุล่วงได้เนื่องจากได้รับความช่วยเหลือและความอนุเคราะห์จาก รองศาสตราจารย์ ดร.นคร กนกแก้ว ผู้เป็นอาจารย์ที่ปรึกษา รวมถึงอาจารย์ประจำภาควิชาโยธาทุกท่านที่ได้ให้คำแนะนำและแนะแนวทางเพื่อปรับปรุง และพัฒนาปริญญานิพนธ์ฉบับนี้

ขอขอบพระคุณข้อมูลจากหน่วยงานกรมบัญชีกลางที่เป็นข้อมูลในการศึกษาวิเคราะห์ และการสร้างแบบจำลองพยากรณ์งบประมาณ เพื่อจัดทำปริญญานิพนธ์ฉบับนี้เป็นอย่างสูง

8. เอกสารอ้างอิง

- [1] อโนทัย แสงธรรมธร. (2542). การจัดการธุรกิจรับเหมาก่อสร้างอำเภอเมือง จังหวัดนครสวรรค์. การค้นคว้าแบบอิสระ บริหารธุรกิจมหาบัณฑิต มหาวิทยาลัยเชียงใหม่.

- [2] สุภาวดี บุคย์เพชร. (2553). การศึกษาปัจจัยทางเศรษฐกิจที่มีผลต่อมูลค่าการซื้อขายหลักทรัพย์ของนักลงทุนรายย่อยในตลาดหลักทรัพย์แห่งประเทศไทย. การศึกษาค้นคว้าอิสระบริหารธุรกิจมหาบัณฑิต สาขาวิชาการเงิน, มหาวิทยาลัยหอการค้าไทย.
- [3] พิภพ สุนทรสมัย. (2549). การประมาณราคาก่อสร้าง. กรุงเทพฯ : ส.ส.ท
- [4] สถาบันฝึกอบรมโยธาไทย. (2560). แนวทาง วิธีปฏิบัติ และรายละเอียดการถอดแบบคำนวณราคากลางงานก่อสร้าง
- [5] ระพล นามวงศ์ และ นาถ สุขศิลป์. (2565). การจัดสรรงบประมาณในพื้นที่เขตพัฒนาเศรษฐกิจพิเศษนครพนมที่มี ผลต่อการขยายตัวของ การก่อสร้างอาคาร. วารสารสถาปัตยกรรม การออกแบบและการก่อสร้าง คณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคาม, 4(1). 131-132.
- [6] DataCamp. "K-Nearest Neighbors (KNN) Classification with scikit-learn." DataCamp, February 20, 2023. [online]. Available: <https://www.datacamp.com/tutorial/k-nearest-neighbor-classification-scikit-learn>. Accessed April 25, 2025.
- [7] GeeksforGeeks. "Random Forest Algorithm in Machine Learning." GeeksforGeeks, January 16, 2025. [online]. Available: <https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/>. Accessed April 25, 2025.
- [8] IBM. "What is XGBoost?" IBM, May 9, 2024. [online]. Available: <https://www.ibm.com/think/topics/xgboost>. Accessed April 25, 2025.
- [9] Comptroller General's Department. "หลักเกณฑ์และวิธีการกำหนดราคากลางงานก่อสร้าง." กรมบัญชีกลาง, May 5, 2023. [online]. Available: <https://www.yotathai.com/passadu/rakaklang-update-2565>. Accessed May 6, 2025.
- [10] IBM. "What is Regression?" IBM, June 27, 2023. [online]. Available: <https://www.ibm.com/topics/linear-regression>. Accessed May 6, 2025.
- [11] IBM. "What is Ensemble Learning?" IBM, July 5, 2023. [online]. Available: <https://www.ibm.com/topics/ensemble-learning>. Accessed May 6, 2025.