

Analysis of factors associated with inpatient reimbursement outcomes under the universal coverage scheme in Thailand

Veeradate Chalermopolprapa

Nakhonpathom Hospital, Nakhonpathom

Abstract

Hospitals under Thailand's Universal Coverage Scheme (UC) face financial challenges due to discrepancies between Diagnosis-Related Group (DRG) reimbursements and actual treatment costs. This study aimed to identify factors associated with hospital financial outcomes and evaluate the performance of Machine Learning (ML) models in predicting profit or loss for Nakhon Pathom Hospital. A retrospective analysis of 67,115 inpatient records (fiscal years 2023–2024) compared Random Forest and Extreme Gradient Boosting (XGBoost) algorithms. The Random Forest model demonstrated the best performance (Accuracy 84.5%, Receiver Operating Characteristic Area Under Curve (ROC AUC) 0.788). Length of Stay (LOS) was the most influential factor, followed by drug and non-drug costs. The model demon-

strated strong capability in identifying loss cases. (True Negative Rate 95.8%). These findings indicate that Machine Learning, particularly the Random Forest algorithm, can serve as an effective decision-support tool for hospital financial management, enabling proactive risk mitigation and improved resource utilization within Thailand's national reimbursement system.

Keywords: *Machine learning; Diagnosis-related groups; Healthcare reimbursement; Length of stay; Hospital management*

Received: 29 September 2025, Revised: 25 October 2025, Accepted: 30 November 2025

Correspondence: Veeradate Chalermopolprapa, Nakhonpathom Hospital, 196 Tesa Rd, Tambon Phra Prathom Chedi, Mueang Nakhon Pathom District, Nakhon Pathom 73000, Thailand, Tel.: 034 240 000, email: veeradate@moph.go.th

การวิเคราะห์ปัจจัยที่สัมพันธ์กับผลลัพธ์การชดเชยค่ารักษาพยาบาลผู้ป่วยใน ภายใต้สิทธิหลักประกันสุขภาพแห่งชาติ

วีระเดช เฉลิมพลประภา

โรงพยาบาลนครปฐม นครปฐม

บทคัดย่อ

โรงพยาบาลที่ให้บริการภายใต้ระบบหลักประกันสุขภาพแห่งชาติ (Universal Coverage Scheme: UC) ของประเทศไทยเผชิญความท้าทายทางการเงินจากค่าชดเชยตามกลุ่มวินิจฉัยโรคร่วม (Diagnosis-Related Group: DRG) ที่อาจไม่สอดคล้องกับต้นทุนการรักษาที่แท้จริง การศึกษานี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยที่สัมพันธ์กับผลลัพธ์ทางการเงินของโรงพยาบาลนครปฐม และประเมินประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning) ในการพยากรณ์ผลกำไรหรือขาดทุน ทำการวิเคราะห์ข้อมูลย้อนหลังของผู้ป่วยในสิทธิ UC จำนวน 67,115 รายการ (ปีงบประมาณ 2566–2567) โดยเปรียบเทียบแบบจำลอง Random Forest และ Extreme Gradient Boosting (XGBoost) พบว่าแบบจำลอง Random Forest ให้ผลดีที่สุด (Accuracy 84.5%, Receiver Operating Characteristic Area Under Curve: ROC AUC 0.788) โดย ระยะเวลาวันนอน (Length of Stay: LOS)

เป็นปัจจัยสำคัญที่สุด รองลงมาคือค่าใช้จ่ายด้านยา และเวชภัณฑ์ ที่มีค่าใช้จ่ายแบบจำลองสามารถทำนายกรณีขาดทุนได้อย่างแม่นยำสูง (True Negative Rate 95.8%) ผลลัพธ์ชี้ให้เห็นว่า เทคโนโลยี Machine Learning โดยเฉพาะแบบจำลอง Random Forest สามารถใช้เป็นเครื่องมือสนับสนุนการตัดสินใจเชิงกลยุทธ์ของโรงพยาบาล เพื่อบริหารความเสี่ยงด้านการเงินและเพิ่มประสิทธิภาพการใช้ทรัพยากรได้อย่างยั่งยืน

คำสำคัญ: การเรียนรู้ของเครื่อง; กลุ่มวินิจฉัยโรคร่วม; การชดเชยค่ารักษาพยาบาล; ระยะเวลาวันนอน; การบริหารจัดการโรงพยาบาล

วันที่รับต้นฉบับ: 29 กันยายน 2568, วันที่แก้ไข: 25 ตุลาคม 2568, วันที่ตอบรับ: 30 พฤศจิกายน 2568

บทนำ

ปัจจุบันสำนักงานหลักประกันสุขภาพแห่งชาติ (สปสช.) ใช้กลุ่มวินิจฉัยโรคร่วมไทยและน้ำหนักสัมพัทธ์ (Thai Diagnosis Related Group, DRG and RW) เป็นเครื่องมือการชดเชยค่าบริการ (Reimbursement) ผู้ป่วยในสิทธิหลักประกันสุขภาพแห่งชาติ (UC) ภายใต้งบประมาณแบบภาพรวม มีการปรับค่าน้ำหนักสัมพัทธ์ (RW) เป็น Adjusted Relative Weight (AdjRW) โดยปรับตามวันนอน (Length of Stay: LOS) ภาวะแทรกซ้อน และกรณีที่ใช้จ่ายสูงผิดปกติ (Outlier case) ทำให้การชดเชยค่าบริการสะท้อนต้นทุนของโรงพยาบาลได้ดีขึ้น อย่างไรก็ตาม จากการศึกษาพบว่าต้นทุนการรักษาต่อ 1 AdjRW ของโรงพยาบาลยังคงสูงกว่าอัตราการชดเชยที่กำหนด^[1] ทำให้โรงพยาบาลมีความเสี่ยงที่จะขาดทุนค่อนข้างสูง

เพื่อให้โรงพยาบาลได้รับการชดเชยที่ใกล้เคียงต้นทุน จำเป็นต้องมีการดูแลรักษาตามมาตรฐานควบคู่กับการใช้ทรัพยากร

อย่างสมเหตุสมผล และมีระบบการจัดการข้อมูลที่มีคุณภาพ ซึ่งประกอบด้วยการบันทึกข้อมูลการรักษาที่ครบถ้วน (Care) การให้รหัสทางการแพทย์ที่ถูกต้อง (Code) และการเบิกจ่ายที่ตรงตามเงื่อนไข (Claim)^[2] แต่ในทางปฏิบัติ กระบวนการให้รหัสมีความซับซ้อนและเกี่ยวข้องกับบุคลากรหลายฝ่าย ซึ่งอาจนำไปสู่ข้อผิดพลาดและส่งผลกระทบต่อรายได้ของโรงพยาบาล ในปีงบประมาณ 2567 โรงพยาบาลนครปฐมได้รับการชดเชย 58.94% ของค่าใช้จ่ายในการรักษา ซึ่งนำไปสู่การสูญเสียรายได้ของโรงพยาบาลจำนวนมาก การศึกษานี้จึงมุ่งเน้นการวิเคราะห์ความสัมพันธ์ระหว่างค่าใช้จ่ายในการรักษาที่เกิดขึ้นจริงกับเงินชดเชยที่ได้รับ โดยไม่ได้พิจารณาถึงข้อผิดพลาดที่อาจเกิดจากระบบจัดการข้อมูล

ปัจจุบันเทคโนโลยีการวิเคราะห์ข้อมูลและ MACHINE LEARNING (ML) ถูกนำมาใช้อย่างแพร่หลายในวงการสาธารณสุขเพื่อทำนายค่าใช้จ่ายและวิเคราะห์ปัจจัยที่ส่งผลต่อผลลัพธ์ทางการเงิน การศึกษาในต่างประเทศ เช่น งานของ S. KULKARNI และคณะ (2021) ได้ใช้ ML เพื่อพยากรณ์ค่าใช้จ่ายและระบุปัจจัยสำคัญในการรักษาผู้ป่วยใน^[3] สำหรับประเทศไทยมีงานวิจัยของ W. THONGPETH และคณะ (2021)

ผู้สนับสนุนประสานงาน: วีระเดช เฉลิมพลประภา โรงพยาบาลนครปฐม 196 ถ.เทศบาล ต.พระปฐมเจดีย์ อ.เมืองนครปฐม จ.นครปฐม 73000, โทร.: 034 240 000, email: veeradate@moph.go.th

ที่พบว่าแบบจำลอง RANDOM FOREST มีประสิทธิภาพดีที่สุดในการทำนายค่าใช้จ่ายผู้ป่วยโรคเรื้อรัง [4] และงานของ ว. ศิวประภากร (2021) ที่ใช้เทคนิค DECISION TREE เพื่อทำนายความเสี่ยงการขาดทุนในผู้ป่วยเด็ก [5]

จากบริบทดังกล่าวการศึกษานี้จึงมุ่งเปรียบเทียบประสิทธิภาพของแบบจำลอง RANDOM FOREST และ XGBOOST เพื่อระบุปัจจัยที่สัมพันธ์กับการขาดทุนของโรงพยาบาล โดยผลการศึกษาจะถูกนำมาวิเคราะห์เพื่อเสนอแนะแนวทางเชิงนโยบายและการบริหารจัดการที่สามารถลดความเสี่ยงในการขาดทุน และเพิ่มประสิทธิภาพการให้บริการผู้ป่วยในโรงพยาบาลได้อย่างเหมาะสม

วัตถุประสงค์ในการศึกษา

1. เพื่อวิเคราะห์ปัจจัยสำคัญที่สัมพันธ์กับผลลัพธ์ทางการเงิน (กำไร/ขาดทุน) ของการชดเชยค่าบริการรักษาผู้ป่วยในภายใต้ระบบหลักประกันสุขภาพแห่งชาติ
2. เพื่อพัฒนาและเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ (Random Forest และ XGBoost) ในการทำนายผลลัพธ์ดังกล่าว
3. เพื่อนำข้อมูลเชิงลึกที่ได้จากแบบจำลองมาเสนอแนะแนวทางเชิงกลยุทธ์ที่ขับเคลื่อนด้วยข้อมูล สำหรับการเพิ่มรายได้และปรับปรุงประสิทธิภาพการดำเนินงานของโรงพยาบาล

วิธีการศึกษา

1. รูปแบบการศึกษาและแหล่งข้อมูล

การศึกษานี้เป็นการวิเคราะห์ข้อมูลเชิงย้อนหลัง (Retrospective analysis) โดยใช้ข้อมูลจากระบบสารสนเทศของโรงพยาบาล (Hospital Information System: HIS) และใบแจ้งยอดการชดเชยค่าบริการ จาก สปสช. (Statement) ของโรงพยาบาล นครปฐม ข้อมูลเป็นข้อมูลผู้ป่วยในสิทธิหลักประกันสุขภาพแห่งชาติ ครอบคลุมช่วงปีงบประมาณ 2566 ถึง 2567 ภายหลังการทำความสะอาดข้อมูล มีข้อมูลที่ใช้ในการศึกษาจำนวน 67,115 รายการ

2. การเตรียมข้อมูลและนิยามตัวแปร

1. ปัจจัยนำเข้า (Predictors): ตัวแปรที่ใช้ในการสร้างแบบจำลองประกอบด้วย:
 - ระยะเวลาวันนอน (Length of Stay: LOS): จำนวนวันนอนพักรักษาตัวในโรงพยาบาลทั้งหมด
 - รหัสโรคหลัก (Main ICD): รหัสการวินิจฉัยโรคหลักตามบัญชีจำแนกทางสถิติระหว่างประเทศของโรคและปัญหาสุขภาพที่เกี่ยวข้อง ฉบับที่ 10 (ICD-10)
 - หมวดหมู่ค่าใช้จ่าย: ค่าใช้จ่าย 6 หมวด ได้แก่ ค่ายาและสารอาหารในเส้นเลือด เวชภัณฑ์ที่มีขาย ค่าตรวจวินิจฉัยทางเทคนิคการแพทย์และพยาธิวิทยา ค่าทำหัตถการและวิสัญญี

ค่าอวัยวะเทียมและอุปกรณ์ในการบำบัดรักษาโรค และค่าตรวจวินิจฉัยและรักษาทางรังสีวิทยา

2. ตัวแปรตาม (Target): ตัวแปรตามที่ต้องการทำนายคือ กำไรหรือขาดทุน ซึ่งเป็นตัวแปรทวิภาค (Binary variable) ที่นิยามดังนี้:

- 1 (กำไร - Profit): กรณีที่ค่าชดเชยที่ได้รับจาก สปสช. มากกว่าค่าใช้จ่ายในการรักษาพยาบาลของโรงพยาบาล
- 0 (ขาดทุน - Loss): กรณีที่ค่าชดเชยที่ได้รับจาก สปสช. น้อยกว่าหรือเท่ากับค่าใช้จ่ายในการรักษาพยาบาลของโรงพยาบาล

3. การพัฒนาและประเมินแบบจำลอง

- การเลือกแบบจำลอง: เลือกใช้อัลกอริทึม Random Forest และ XGBoost ในการเปรียบเทียบ เนื่องจากเป็นเทคนิคการเรียนรู้แบบรวมกลุ่ม (Ensemble methods) ที่มีประสิทธิภาพสูงและได้รับการยอมรับอย่างกว้างขวางในบริบทการเงินสาธารณสุข [6]
- กลยุทธ์การตรวจสอบความถูกต้อง: ใช้เทคนิคการตรวจสอบไขว้แบบ 5 ส่วน (5-fold cross-validation) เพื่อป้องกันปัญหาการเรียนรู้เกินพอดี (Overfitting)
- ตัวชี้วัดประสิทธิภาพ: ประเมินแบบจำลองโดยใช้ตัวชี้วัดมาตรฐาน ได้แก่ Accuracy, Precision, Recall (Sensitivity), F1-Score และ ROC AUC

4. การวิเคราะห์ปัจจัย

- การวิเคราะห์สหสัมพันธ์: สร้างเมทริกซ์สหสัมพันธ์แบบเพียร์สัน (Pearson correlation matrix) เพื่อตรวจสอบความสัมพันธ์เชิงเส้นระหว่างปัจจัยนำเข้าและตัวแปรตาม
- การวิเคราะห์ความสำคัญของปัจจัย: สกัดค่าความสำคัญของปัจจัย (Feature importance score) จากแบบจำลอง Random Forest ที่ผ่านการฝึกฝนแล้ว เพื่อประเมินอิทธิพลของแต่ละปัจจัย

ผลการศึกษา

1. การวิเคราะห์เชิงพรรณนาของปัจจัยนำเข้า การวิเคราะห์ข้อมูลเบื้องต้นทำให้เห็นภาพรวมและลักษณะของปัจจัยที่สำคัญ
 - 1.1 การวิเคราะห์กลุ่มโรคหลัก (Main ICD) การวิเคราะห์เชิงพรรณนา

จากข้อมูลการชดเชยค่าบริการและค่าใช้จ่ายในการรักษา ได้ระบุกลุ่มโรคที่ส่งผลกระทบต่อทางการเงินต่อโรงพยาบาลอย่างชัดเจนทั้งในด้านบวกและลบ (ตารางที่ 1 และ 2)

ตารางที่ 1 แสดง 5 อันดับแรกกลุ่มโรคหลักที่มียอดชดเชยรวมน้อยกว่าค่าใช้จ่ายในการรักษา (ขาดทุนรวม)

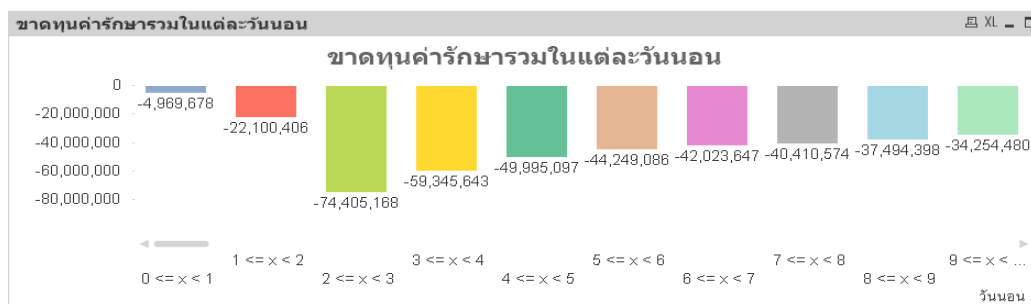
Main ICD (กลุ่มโรคหลัก)	ยอดชดเชยรวมน้อยกว่าค่าใช้จ่ายในการรักษา (บาท)
Disorders related to short gestation and low birth weight, not elsewhere classified (P07)	-56,164,567.91
Intracranial injury (S06)	-29,746,320.85
Acute myocardial infarction (I21)	-29,335,842.29
Viral pneumonia, not elsewhere classified (J12)	-27,934,514.29
Bacterial pneumonia, not elsewhere classified (J15)	-27,551,167.14

ตารางที่ 2 5 อันดับแรกกลุ่มโรคหลักที่มียอดชดเชยรวมมากกว่าค่าใช้จ่ายในการรักษา(กำไรรวม)

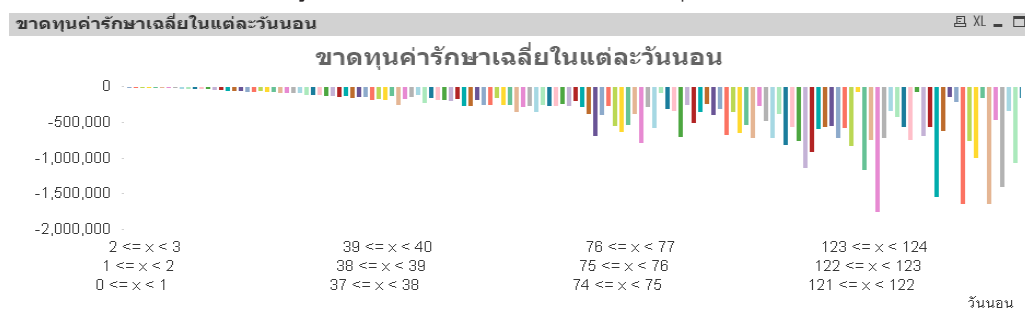
Main ICD (กลุ่มโรคหลัก)	ยอดชดเชยรวมมากกว่าค่าใช้จ่ายในการรักษา (บาท)
Malignant neoplasm of ovary (C56)	1,148,569.23
Other diseases of intestine (K63)	910,156.56
Other coagulation defects (D68)	729,113.02
Schizophrenia (F20)	205,860.43
Other disorders of veins (I87)	164,642.56

1.2 การวิเคราะห์ปัจจัยระยะเวลาวันนอน (LOS) ค่าฐานนิยม(Mode) ของระยะเวลาวันนอนในชุดข้อมูลนี้คือ 3 วัน โดยยอดขาดทุนรวมสูงสุด เกิดขึ้นในช่วงวันนอน 2-4 วัน ซึ่งเป็นช่วงที่มีจำนวนผู้ป่วยมากที่สุด (รูปที่ 1)

แต่เมื่อพิจารณาค่าเฉลี่ยต่อราย พบว่ายิ่งระยะเวลาวันนอนนานขึ้น ยอดขาดทุนเฉลี่ยต่อรายจะยิ่งสูงขึ้นอย่างมีนัยสำคัญ (รูปที่ 2)



รูปที่ 1 ระยะเวลาวันนอน (LOS) ที่ขาดทุนรวม



รูปที่ 2 ระยะเวลาวันนอน (LOS) ที่ขาดทุนเฉลี่ยต่อราย

1.3 การวิเคราะห์ปัจจัยด้านหมวดค่ารักษา การวิเคราะห์โครงสร้างต้นทุน (ตารางที่ 3) พบว่าหมวดที่มีมูลค่าสูงสุดคือ ค่าอวัยวะเทียมและอุปกรณ์ในการบำบัดรักษาโรค ตามมาด้วย

ค่าทำหัตถการและวัสดุยา และค่ายาและสารอาหารในเส้นเลือดตามลำดับ

ตารางที่ 3 แสดง 5 อันดับแรกหมวดค่ารักษาที่มียอดรวมค่ารักษาสูงสุด

หมวดค่ารักษา	ยอดรวมค่ารักษา (บาท)
ค่าอวัยวะเทียมและอุปกรณ์ในการบำบัดรักษาโรค	111,921,406
ค่าทำหัตถการและวัสดุยา	80,953,420
ค่ายาและสารอาหารในเส้นเลือด	53,314,655
ค่าตรวจวินิจฉัยทางเทคนิคการแพทย์และพยาธิวิทยา	36,844,049
ค่าตรวจวินิจฉัยและรักษาทางรังสีวิทยา	32,593,922

2. ผลการวิเคราะห์แบบจำลองเชิงพยากรณ์

2.1 ประสิทธิภาพเปรียบเทียบของแบบจำลอง (ตารางที่ 4) แสดงผลการเปรียบเทียบประสิทธิภาพระหว่างแบบจำลอง

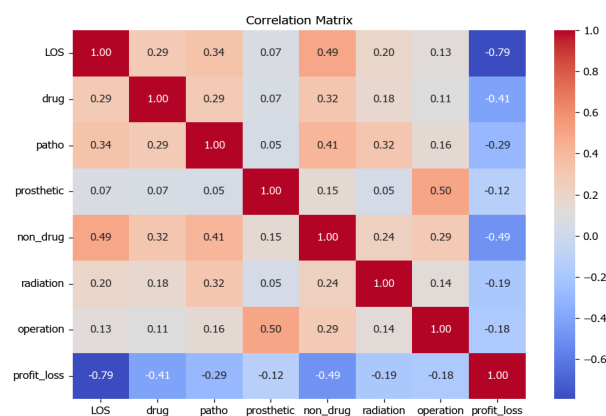
Random Forest และ XGBoost พบว่าแบบจำลอง Random Forest มีประสิทธิภาพสูงกว่าในทุกตัวชี้วัด ดังนั้นจึงได้เลือกแบบจำลอง Random Forest เพื่อทำการวิเคราะห์ในเชิงลึกต่อไป

ตารางที่ 4 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง

แบบจำลอง	Accuracy	Precision	Recall	F1-Score	ROC
Random Forest	0.8448	0.6602	0.3549	0.4616	0.7875
XGBoost	0.8270	0.5778	0.2892	0.3853	0.7695

2.2 การวิเคราะห์ความสัมพันธ์ของปัจจัย (Correlation Analysis) เมทริกซ์สหสัมพันธ์ (รูปที่ 3) พบความสัมพันธ์เชิงลบที่แข็งแกร่งที่สุดระหว่างกำไร/ขาดทุน และ LOS (-0.79) ซึ่งยืนยันว่าเมื่อวันนอนเพิ่มขึ้น โอกาสที่จะเกิดกำไรจะลดลงอย่างชัดเจน นอกจากนี้ยังพบความสัมพันธ์เชิงลบปานกลางกับค่าเวชภัณฑ์ที่มีใช้ยา (-0.49) และค่ายา และสารอาหารในเส้นเลือด (-0.41)

2.3 ปัจจัยเชิงพยากรณ์ที่สำคัญที่สุด (Feature Importance) การวิเคราะห์ความสำคัญของปัจจัยจากแบบจำลอง Random Forest (ตารางที่ 5) ยืนยันว่าระยะเวลาวันนอน (LOS) เป็นปัจจัยที่มีอิทธิพลสูงสุดอย่างท่วมท้น (ค่าความสำคัญ = 0.69) ตามมาด้วยปัจจัยด้านค่าใช้จ่ายต่างๆ



รูปที่ 3 Correlation Matrix

ตารางที่ 5 ปัจจัยเชิงพยากรณ์ที่สำคัญที่สุดจากแบบจำลอง Random Forest

อันดับ	ปัจจัย (Feature)	ค่าความสำคัญ (Importance Score)
1	LOS	0.690560547
2	ค่ายาและสารอาหารในเส้นเลือด	0.107000116
3	ค่าเวชภัณฑ์ที่มีใช้	0.040581660
4	ค่าอวัยวะเทียมและอุปกรณ์ในการบำบัดรักษาโรค	0.029140602
5	ค่าตรวจวินิจฉัยทางเทคนิคการแพทย์และพยาธิวิทยา	0.022242866
6	ค่าตรวจวินิจฉัยและรักษาทางรังสีวิทยา	0.017631163
7	ค่าทำหัตถการและวิสัญญี	0.015227342
8	Chronic sinusitis (J32)	0.004186493
9	Viral pneumonia, not elsewhere classified (J12)	0.003355300
10	Intracranial injury (S06)	0.002928088
11	Disorders related to short gestation and low birth weight (P07)	0.002746770
12	Bacterial pneumonia, not elsewhere classified (J15)	0.002384701
13	Fracture of skull and facial bones (S02)	0.002095591
14	Acute myocardial infarction (I21)	0.002094513

4.2.4 การประเมินประสิทธิภาพของแบบจำลอง Random Forest Confusion Matrix (รูปที่ 4) พบว่าแบบจำลองมีความสามารถสูงในการระบุกรณีที่จะขาดทุนได้อย่างถูกต้อง (True Negative = 52,232) ส่งผลให้มีค่า True Negative Rate สูงถึง 95.8% ($52,232 / (52,232 + 2,299)$) อย่างไรก็ตามแบบจำลองยังมีความสามารถจำกัดในการตรวจจับกรณีที่จะมีกำไรทั้งหมด โดยมีค่า Recall อยู่ที่ 35.5% ($4,466 / (8,118 + 4,466)$)

Confusion Matrix -Random Forest		
Reference	Predict	
	0	1
0	52,232	2,299
1	8,118	4,466

รูปที่ 4 Confusion Matrix

บทวิจารณ์

ผลการศึกษานี้พบว่า ระยะเวลาวันนอน (LOS) เป็นปัจจัยที่สำคัญที่สุดต่อผลลัพธ์การชดเชยค่ารักษาพยาบาล เมื่อพิจารณาปัจจัยด้านค่าใช้จ่าย เปรียบเทียบระหว่างยอดรวมค่าใช้จ่าย (ตารางที่ 3) กับค่าความสำคัญของปัจจัยจากแบบจำลอง (ตารางที่ 5) แม้ว่าค่าอวัยวะเทียมและอุปกรณ์ในการบำบัดรักษาโรคจะเป็นหมวดที่มีค่าใช้จ่ายรวมสูงสุด แต่แบบจำลองกลับให้น้ำหนักความสำคัญกับค่าที่เป็นอันดับสองรองจาก LOS อาจเป็นเพราะระบบ DRG เป็นการจ่ายแบบเหมาจ่ายซึ่งได้คำนวณรวมต้นทุนมาตรฐานของหัตถการและอุปกรณ์ไว้แล้ว ในขณะที่ค่ายา โดยเฉพาะยาที่ใช้ในภาวะแทรกซ้อนหรือยานอกบัญชี อาจเป็นส่วนที่ระบบ DRG ไม่ได้ครอบคลุมทั้งหมด ทำให้ความแปรปรวนของค่าใช้จ่ายในหมวดนี้ส่งผลโดยตรงต่อการขาดทุนของโรงพยาบาลอย่างมีนัยสำคัญ

นอกจากนี้ผลการศึกษาพบความสัมพันธ์ระหว่างการวิเคราะห์เชิงพรรณนาและการพยากรณ์ของแบบจำลอง กลุ่มโรคหลักที่ก่อให้เกิดผลขาดทุนรวมสูงสุด 5 อันดับแรก (ตารางที่ 1) ได้แก่ P07, S06, I21, J12, และ J15 ซึ่งล้วนเป็นกลุ่มโรคหลักที่แบบจำลองให้ความสำคัญในระดับสูงในการทำนาย (ตารางที่ 5) การค้นพบนี้เป็นการยืนยันความถูกต้องของแบบจำลอง (Model validation) ว่าไม่เพียงแต่จะทำนายผลลัพธ์โดยรวมได้ดี แต่ยังสามารถระบุกลุ่มโรคหลักที่มีผลต่อการขาดทุนได้อย่างแม่นยำ

ในเชิงปฏิบัติ แม้ว่าแบบจำลองจะมีค่า Recall สำหรับการทำนายกำไรที่ค่อนข้างต่ำ (35.5%) ซึ่งเป็นข้อจำกัดจากความไม่สมดุลของข้อมูล แต่ค่า True Negative Rate ที่สูง (95.8%) กลับเป็นจุดแข็งที่สำคัญที่สุด เพราะหมายความว่าแบบจำลองสามารถทำนายกรณีที่จะขาดทุนได้อย่างแม่นยำ ที่น่าสนใจคือกรณีที่แบบจำลองทำนายขาดทุน แต่ในความเป็นจริงกลับมีกำไร อาจสะท้อนถึงแนวปฏิบัติที่ดีของโรงพยาบาลซึ่งสามารถจัดการผู้ป่วยความเสี่ยงสูงได้ดีกว่าที่โมเดลคาดการณ์ การศึกษาเชิงลึกในกลุ่มนี้อาจช่วยค้นหา best practices ที่สามารถนำไปขยายผลได้

ผลการศึกษาพบว่าระยะเวลาวันนอน (LOS) และค่าใช้จ่ายในการรักษาแต่ละหมวดเป็นปัจจัยสำคัญในการทำนายผลลัพธ์ทางการเงินนั้นสอดคล้องกับงานวิจัยของต่างประเทศ [7,8] ซึ่งช่วยยืนยันว่าผลการศึกษานี้มีความน่าเชื่อถือและสอดคล้องกับองค์ความรู้ในปัจจุบัน

ข้อจำกัดของการศึกษา

1. แบบจำลองถูกพัฒนาขึ้นจากข้อมูลของโรงพยาบาลนครปฐมเพียงแห่งเดียว ซึ่งเป็นโรงพยาบาลตติยภูมิ ดังนั้นความสามารถในการนำไปใช้กับโรงพยาบาลระดับอื่นอาจมีข้อจำกัด
2. ชุดปัจจัยนำเข้าที่จำกัด: แบบจำลองนี้ยังไม่ได้รวมปัจจัยอื่นๆ ที่อาจมีความสำคัญ เช่น โรคร่วม (Comorbidities), ภาวะแทรกซ้อน และข้อมูลประชากรของผู้ป่วย ซึ่งงานวิจัยสนับสนุนว่าปัจจัยเหล่านี้มีผลต่อค่าใช้จ่ายและวันนอน [9,10]
3. ความไม่สมดุลของข้อมูล: ความไม่สมดุลของจำนวนข้อมูลระหว่างกรณีที่ขาดทุนและมีกำไรส่งผลกระทบต่อค่า Recall ของแบบจำลอง การศึกษาในอนาคตอาจพิจารณาใช้เทคนิคการจัดการข้อมูลที่ไม่สมดุล เช่น SMOTE (Synthetic Minority Over-sampling Technique) เพื่อปรับปรุงประสิทธิภาพความแม่นยำของแบบจำลอง [11]

สรุปและข้อเสนอแนะ

การศึกษานี้แสดงให้เห็นว่า Machine Learning โดยเฉพาะ Random Forest สามารถใช้เป็นเครื่องมือสนับสนุนการตัดสินใจทางการเงินในโรงพยาบาลได้อย่างมีประสิทธิภาพ โดยระยะเวลาวันนอน (LOS) และค่าใช้จ่ายค่ายาและสารอาหารในเส้นเลือด/ค่าเวชภัณฑ์ที่มีใช้ยา เป็นปัจจัยหลักที่มีผลต่อผลลัพธ์กำไรและขาดทุน

ข้อเสนอเชิงกลยุทธ์ ได้แก่:

1. บริหารจัดการระยะเวลาวันนอน(LOS) เชิงรุก: ใช้แบบจำลองเพื่อคัดกรองผู้ป่วยที่มีความเสี่ยงนอนโรงพยาบาลนานตั้งแต่ระยะเริ่มต้น
2. เน้นกลุ่ม DRG ที่มีความเสี่ยงสูง เช่น P07, S06, I21, J12 โดยกำหนด clinical pathway และมาตรการควบคุมต้นทุนเฉพาะโรค
3. เพิ่มประสิทธิภาพการจัดการเวชภัณฑ์และยา โดยใช้การพยากรณ์ความต้องการเพื่อควบคุมต้นทุนและวางแผนการจัดซื้อจัดหา

การผนวก predictive analytics เข้ากับการจัดการโรงพยาบาลจะช่วยลดความไม่แน่นอนทางการเงิน และเพิ่มประสิทธิภาพการใช้ทรัพยากร

กิตติกรรมประกาศ

ผู้ศึกษาขอขอบคุณเจ้าหน้าที่ฝ่ายสารสนเทศของโรงพยาบาลนครปฐมทุกท่านที่มีส่วนร่วมในการเก็บรวบรวมข้อมูล และให้การสนับสนุนการศึกษานี้จนสำเร็จลุล่วงด้วยดี

เอกสารอ้างอิง

- [1] O. Khiaocharoen et al., "Cost by Diagnosis Related Group: Outputs from the First Phase Cost per Disease Project," *HISPA Compendium*, vol. 1, no. 8, pp. 113–129, 2023. doi: 10.14456/hispa.2023.8.
- [2] งานประกันสุขภาพ โรงพยาบาลโชคชัย, "ศูนย์จัดเก็บรายได้คุณภาพ," โรงพยาบาลโชคชัย, นครราชสีมา, ประเทศไทย, 2565. [ออนไลน์]. เข้าถึงได้จาก: <https://www.uckkpho.com/wpcontent/uploads/2022/05/07.ศูนย์จัดเก็บรายได้มีคุณภาพ.pdf> [เข้าถึงเมื่อ: 28 ก.ย. 2568].
- [3] S. Kulkarni, S. S. Ambekar, and M. Hudnurkar, "Predicting the inpatient hospital cost using a machine learning approach," *International Journal of Innovation Science*, vol. 13, no. 1, pp. 87–104, 2021, doi: 10.1108/IIS-09-2020-0175.
- [4] W. Thongpeth, A. Lim, A. Wongpairin, T. Thongpeth, and S. Chaimontree, "Comparison of linear, penalized linear and machine learning models predicting hospital visit costs from chronic disease in Thailand," *Informatics in Medicine Unlocked*, vol. 26, Art. no. 100769, 2021, doi: 10.1016/j.imu.2021.100769.
- [5] W. Siwaprapapakorn, "Data analytic and predictive model for analysis of causes for a loss from diagnosis related groups (DRGs) payment of pediatric inpatient at Sunpasitthiprasong hospital," *Journal of the Thai Medical Informatics Association*, vol. 1, pp. 12–20, 2021.
- [6] B. Langenberger, T. Schulte, and O. Groene, "The application of machine learning to predict high-cost patients: A performance-comparison of different models using healthcare claims data," *PLoS ONE*, vol. 18, no. 1, p. e0279540, Jan. 2023, doi: 10.1371/journal.pone.0279540.
- [7] K. Stone, R. Zwiggelaar, P. Jones, and N. Mac Parthalluin, "A systematic review of the prediction of hospital length of stay: Towards a unified framework," *PLOS Digital Health*, vol. 1, no. 4, p. e0000017, Apr. 2022, doi: 10.1371/journal.pdig.0000017.
- [8] R. Jain, M. Singh, A. R. Rao, and R. Garg, "Predicting hospital length of stay using machine learning on a large open health dataset," *BMC Health Services Research*, vol. 24, Art. no. 860, 2024, doi: 10.1186/s12913-024-11238-y.
- [9] S. Deniz, A. Şengül, Y. Aydemir, J. ʌ. Emre, and M. H. Özhan, "Clinical factors and comorbidities affecting the cost of hospital-treated COPD," *International Journal of COPD*, vol. 11, pp. 3023–3030, Dec. 2016, doi: 10.2147/COPD.S121637.
- [10] S. Haidar, R. Vazquez, and G. Medic, "Impact of surgical complications on hospital costs and revenues: retrospective database study of Medicare claims," *Journal of Comparative Effectiveness Research*, Art. no. e230080, Jun. 2023, doi: 10.57264/cer-2023-0080.
- [11] A. M. Sowjanya and O. Mrudula, "Effective treatment of imbalanced datasets in health care using modified SMOTE coupled with stacked deep learning algorithms," *Applied Nanoscience*, vol. 13, pp. 1829–1840, 2023, doi: 10.1007/s13204-021-02063-4.