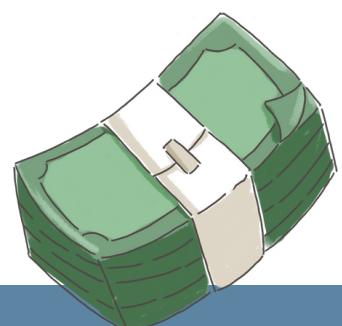
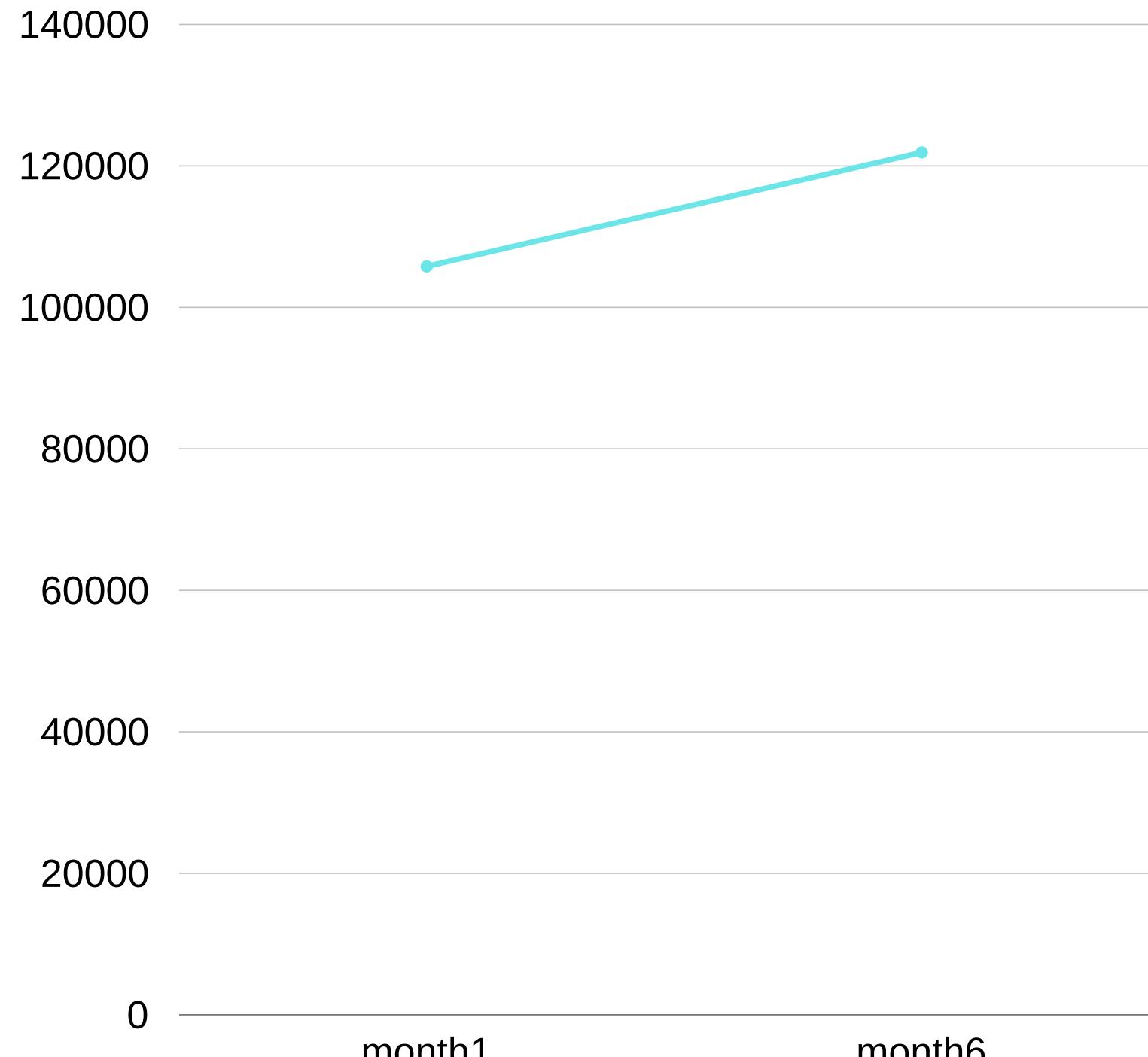


ระบบคาดการณ์การอนุมัติสินเชื่อจากธนาคารที่ใช้ Transformers & Reinforcement

An ensemble Transformers & Reinforcement based bank loan approval predictions system



บทที่ 1 ความเป็นมาและความสำคัญ



มีกำไรสุทธิรวมกัน **121,917.19** ล้านบาท

ปรับตัวเพิ่มขึ้น **16,121.56** ล้านบาท

ธนาคารพาณิชย์ 10 แห่ง แจ้งผลประกอบการ 6 เดือนแรกของ 2566

บทที่ 1 ความเป็นมาและความสำคัญ



กำไรเพิ่มขึ้นจากการได้ดอกเบี้ย ค่าธรรมเนียม และการเข้มงวดในการปล่อยสินเชื่อ เพื่อบริหารจัดการหนี้ที่ไม่เกิดรายได้

บทที่ 1 ความเป็นมาและความสำคัญ

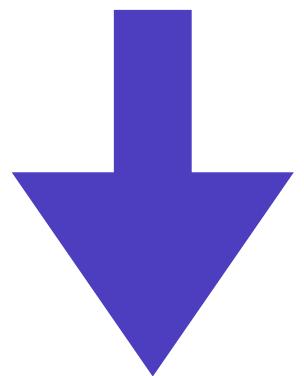


Loan-Approval-Prediction-Dataset

Loan Approval Dataset used for Prediction Models

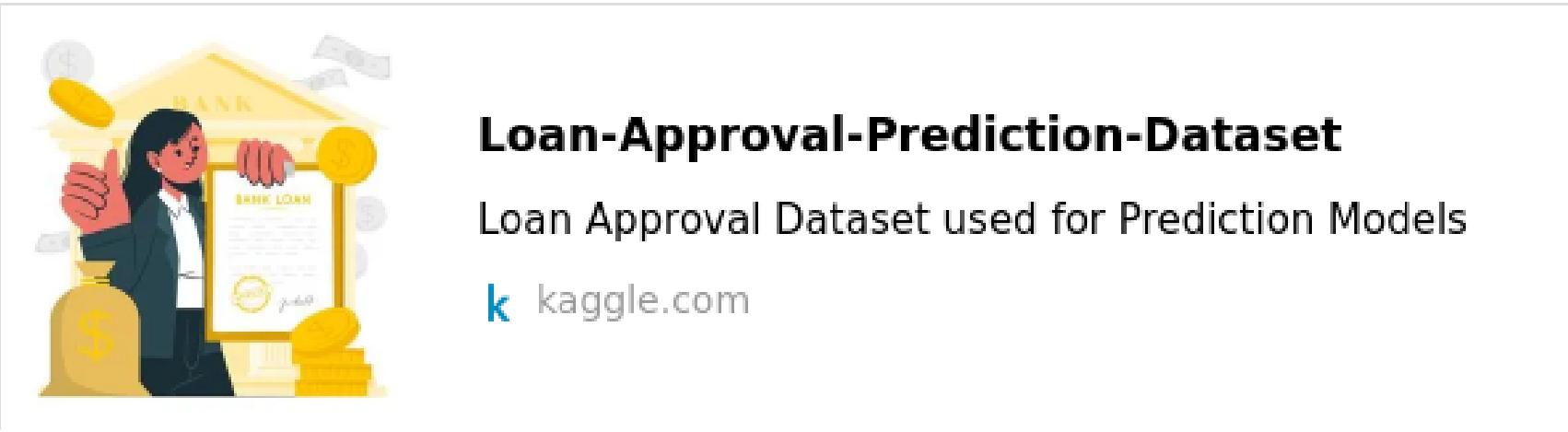
[kaggle.com](https://www.kaggle.com)

พวกราอยากศึกษาวิธีการทำการปัญญาประดิษฐ์ด้านการปล่อยสินเชื่อด้วยข้อมูลจาก kaggle.com

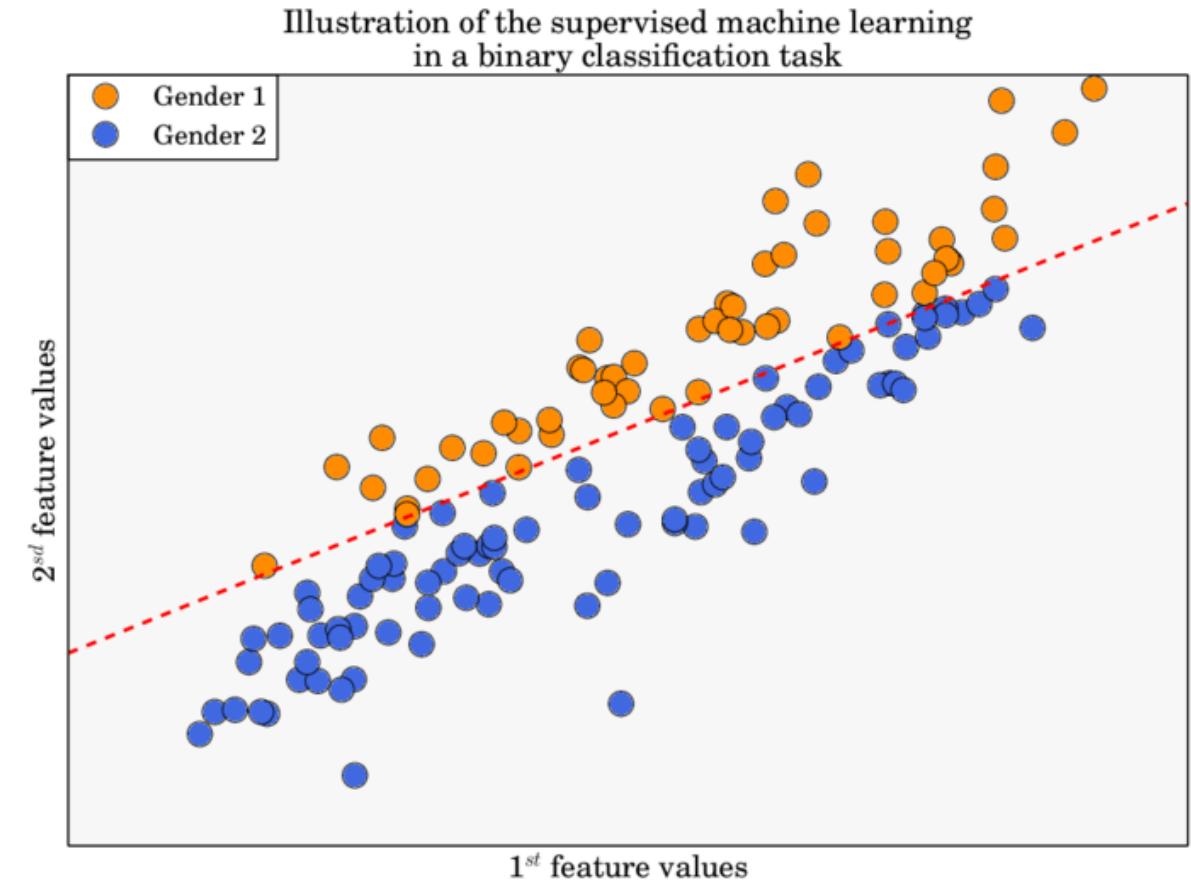


มีคนทำ model อะไรแล้วไปบ้าง
แล้วมี model อะไรที่ยังไม่เคยใช้งานในด้านการปล่อยสินเชื่อ

บทที่ 1 ขอบเขตของโครงการ



ศึกษา model ที่ยังไม่เคยลองกับข้อมูลการปล่อยสินเชื่อ
แล้วนำมาทดลองกับข้อมูลจาก kaggle.com



ឧបត័រ 1

វត្ថុប្រសិទ្ធភាព

1. เพื่อศึกษาพัฒนาระบบกลุ่มผู้ขอสินเชื่อร่วงผู้ที่ได้รับสินเชื่อและไม่ได้รับสินเชื่อ
2. สร้างโมเดลคัดกรองกลุ่มของผู้ขอสินเชื่อด้วยวิธีการใหม่

ប្រយោជន៍កំណត់តម្លៃ

1. ลดการทำงานของผู้ให้สินเชื่อ เพื่อให้บุคลากรได้ใช้เวลาไปทำงานในส่วนอื่นๆ
2. เพิ่มประสิทธิภาพของ model เพื่อให้สามารถจำแนกการให้อនุมัติอย่างแม่นยำยิ่งขึ้น

บทที่ 2 งานวิจัยที่เกี่ยวข้องกับงานที่จะทำ

An ensemble machine learning based bank loan approval predictions system with a smart application



Uddin et al. (2023) **The dataset contains 806 rows and 13 columns from kaggle**

- data balancing using SMOTE (Chatterjee, 2021)
- Logistic Regression (LR)
- Decision Tree (DT)
- Random Forest (RF)
- Extra Trees (ET)
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)
- Gaussian Naive Bayes (GB)
- AdaBoost (AdB)
- Gradient Boosting (GB)
- Dense Neural Network (DNN)
- recurrent neural networks (RNN)
- long short-term memory (LSTM)

บทที่ 2 งานวิจัยที่เกี่ยวข้องกับงานที่จะทำ

An ensemble machine learning based bank loan approval predictions system with a smart application



Parameter	Precision		Recall		F1-Score		Accuracy (%)
	Yes	No	Yes	No	Yes	No	
Logistic Regression	0.63	0.94	0.97	0.42	0.77	0.58	70.06
Decision Tree	0.81	0.77	0.76	0.82	0.78	0.8	78.98
Random Forest	0.83	0.85	0.86	0.82	0.84	0.84	84.71
Extra Trees	0.83	0.88	0.89	0.82	0.86	0.85	86.64
SVM	0.62	0.94	0.97	0.38	0.75	0.55	68.15
KNeighbors	0.87	0.74	0.7	0.9	0.77	0.81	79.61
GaussianNB	0.62	0.87	0.94	0.42	0.75	0.57	68.15
AdaBoost	0.68	0.76	0.81	0.62	0.74	0.68	71.33
Gradient Boosting	0.75	0.86	0.89	0.71	0.81	0.77	79.61
Dense Neural Network	0.65	0.90	0.94	0.51	0.78	0.65	73.24
Long Short-Term Memory	0.64	0.88	0.93	0.48	0.76	0.62	71.33
Recurrent Neural Network	0.69	0.79	0.83	0.62	0.75	0.70	73.24

บทที่ 2 งานวิจัยที่เกี่ยวข้องกับงานที่จะทำ

An ensemble machine learning based bank loan approval predictions system with a smart application



	Model	Accuracy (%)	Paper
Existing Work	Extra Tree	86.2	(Anand et al., 2022)
	Decision Tree with AdaBoost	84	(Kumar et al., 2022)
	Logistic Regression	78.5	(Dosalwar et al., 2021)
	Multilayer Perceptron	80	(Alsaleem & Hasoon, 2020)
	Naïve Bayesian	80.4	(Blessie & Rekha, 2019)

บทที่ 2 งานวิจัยที่เกี่ยวข้องกับงานที่จะทำ

Bank Loan Classification of Imbalanced Dataset Using Machine Learning Approach

The dataset contains 887k rows and 17 columns from kaggle
(M Yasser H, 2024)

Classifiers	Accuracy (%)	AUC (%)	Precision (%)	Recall (%)	F1-Score (%)
LR + SMOTE	82	83	97	83	89
RF + SMOTE	96	89	96	99	98
Bagging + SMOTE	96	89	96	99	98
Adaboost + SMOTE	90	81	96	94	95

The dataset contains 14,8671 rows and 34 columns from kaggle
(M Yasser H, 2024)

Babo and Beyene (2023)

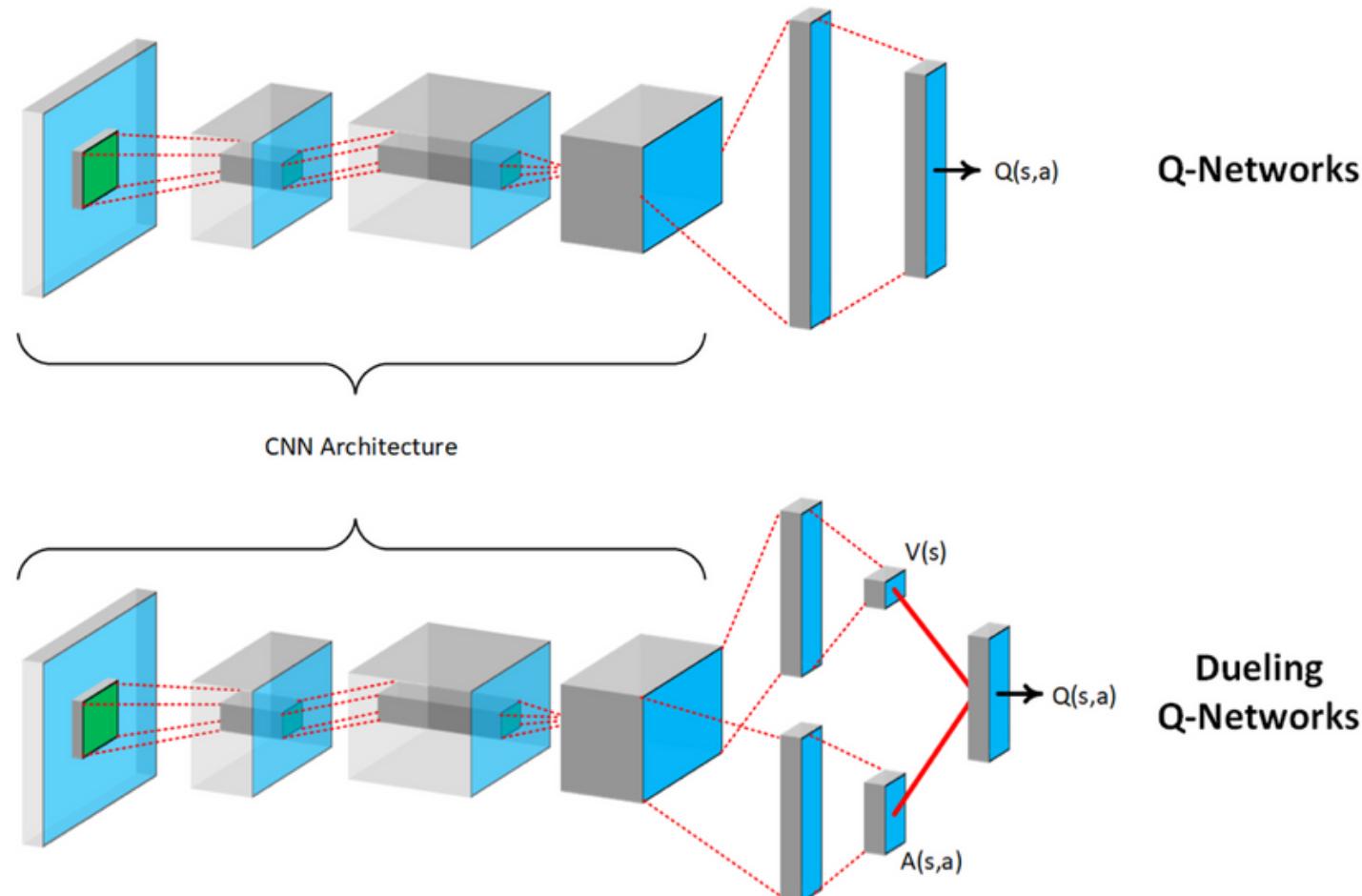
- Resampling the Imbalanced
 - SMOTE
 - MSMOTE
- random forest algorithm
- Logistic regression
- Classifier Ensemble Techniques
 - Bagging Based
 - Boosting Based

Bank Loan Classification of Imbalanced Dataset Using Machine Learning...

Before giving loans to borrowers, banks decide whether the borrower is bad (defaulter) or good...

บทที่ 2 งานวิจัยที่เกี่ยวข้องกับงานที่จะทำ

Reinforcement Learning Algorithms: An Overview and Classification



Deep Q-Learning (DQN) is a key algorithm for solving problems with vast state spaces and discrete actions. It utilizes convolutional and fully connected layers, like those found in AlexNet CNN, to produce Q-values representing the best actions to take in a given state. By iteratively learning from environment feedback, DQN enables agents to make informed decisions and optimize rewards in complex scenarios.

บทที่ 2 งานวิจัยที่เกี่ยวข้องกับงานที่จะทำ

Sparse Binary Transformers for Multivariate Time Series Modeling



Sparse Binary Transformers for Multivariate Time Series Modeling

Compressed Neural Networks have the potential to enable deep learning across new applications and smaller computational environments. However, understanding the range of learning tasks in which...

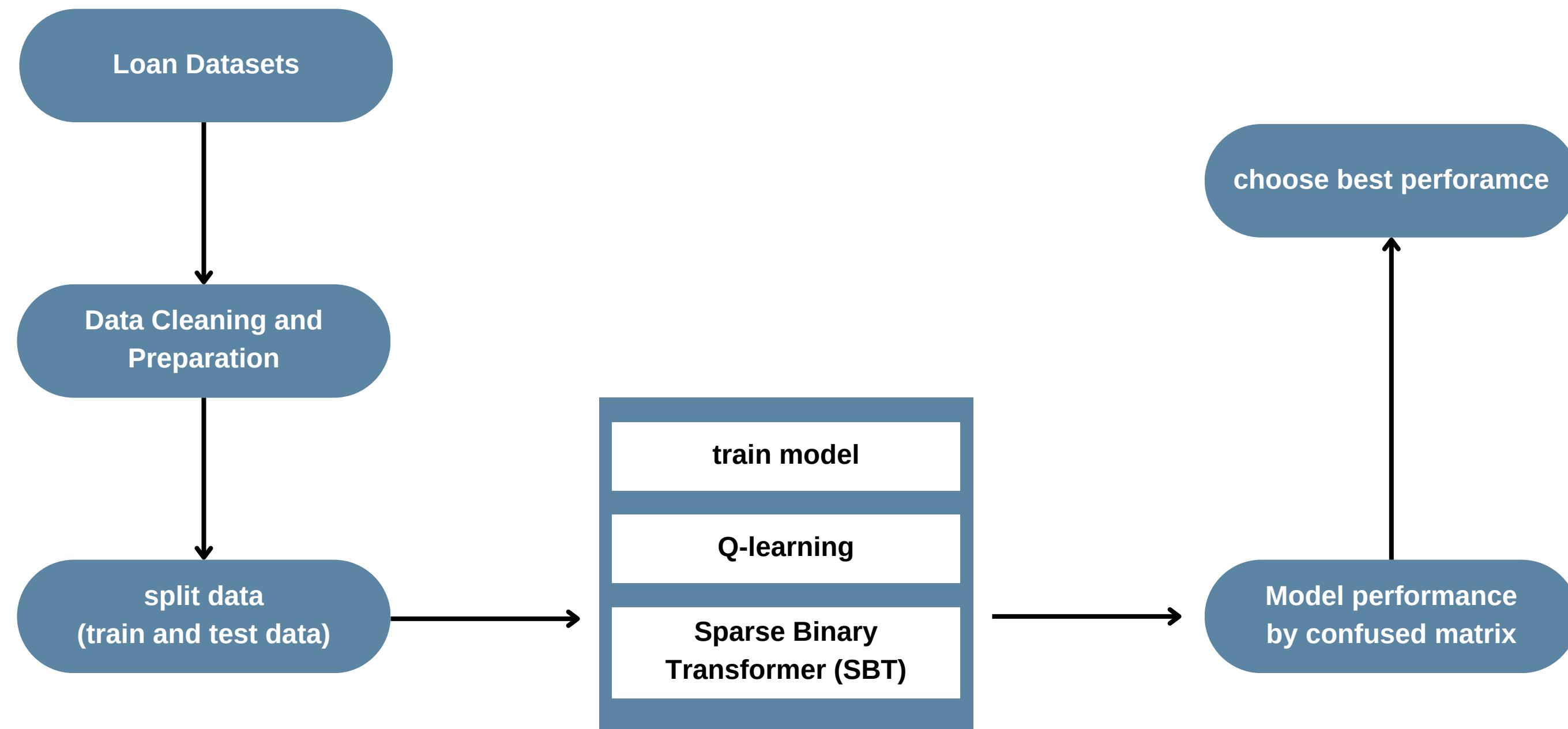
 arXiv.org

Matt Gorbett , Hossein Shirazi and Indrakshi Ray (2023)

Dataset	Metric	Manual Threshold		POT Threshold	
		Dense	$SBT_{p=0.75}$	Dense	$SBT_{p=0.75}$
MSL	P	92.7	96.8	85.5	82.9
	R	100	100	100	100
	F1	96.2	96.8	92.1	91.1
SMD	P	85.4	85.3	99.9	100
	R	100	100	100	100
	F1	92.1	92.1	100	99.9
SMAP	P	93.9	93.7	85.9	84.9
	R	100	100	100	100
	F1	96.9	96.8	92.4	91.8

Table 2: Anomaly detection results with benign sample windows. We evaluate Precision (P), Recall (R), and the F1 score using both manual threshold and POT threshold technique. We find that the single time step prediction window achieves high accuracy when each past time-step in w is benign. $w = 200$ for SMD and $w = 50$ for SMAP and MSL. These results indicate that when given time to stabilize after an anomalous event, our SBT framework can detect new anomalies with high accuracy. We evaluate our results using a manual threshold ($\tau=0.5\%$ for SMD, 1% for others) and the POT automatic threshold selector.

ឧបត្ថម្ភ ៣: វិវាទការ



บทที่ 3: วิธีดำเนินการ

3.1 ข้อมูลและตัวแปร

รายละเอียดข้อมูล:

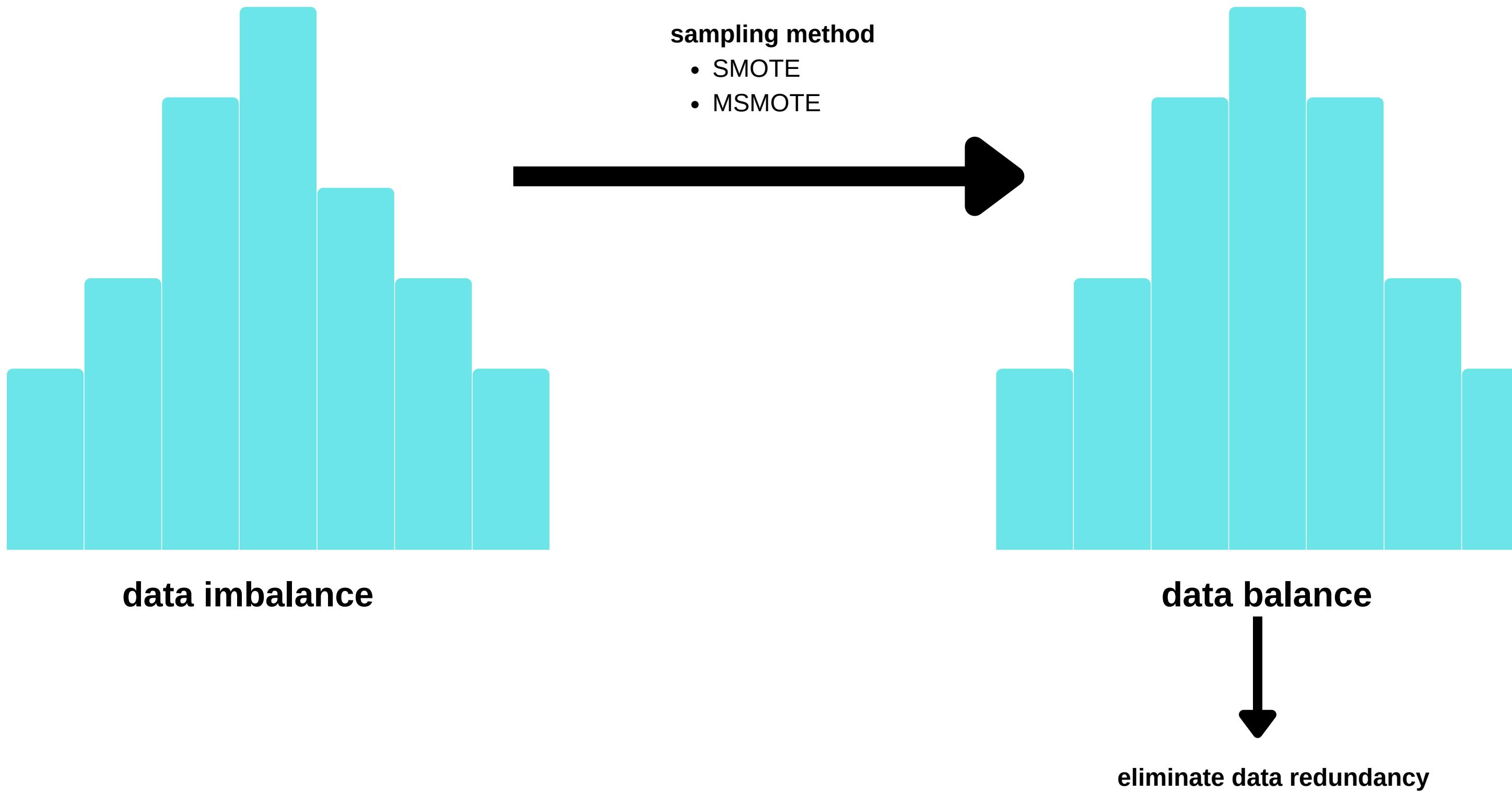
1. **loan_id** (type int64) เลขสินเชื่ออ้างอิง
2. **no_of_dependents** (type int64) คนในครอบครัวที่อยู่ในการประกอง
3. **education** (object) จบการศึกษาปริญญาตรีหรือไม่
4. **self_employed** (object) ทำงานอิสระหรือไม่
5. **income_annum** (type int64) รายได้ต่อปี(รูปี)
6. **loan_amount** (type int64) จำนวนเงินในการกู้
7. **loan_term** (type int64) ระยะเวลาการผ่อนชำระ
8. **cibil_score** (type int64) เครดิต
9. **residential_assets_value** (type int64) มูลค่าของที่อยู่อาศัย
10. **commercial_assets_value** (type int64) มูลค่าทรัพย์สินเชิงพาณิชย์
11. **luxury_assets_value** (type int64) มูลค่าทรัพย์สินที่มีค่า (สร้อยทอง, เพชร, เครื่องประดับ)
12. **bank_asset_value** (type int64) เงินที่ฝากกับธนาคาร
13. **loan_status** (object) อนุมัติหรือไม่

3.2 data cleaning

รายละเอียดข้อมูล:

1. **no_of_dependents** (type int64) คนในครอบครัวที่อยู่ในการประกอง
2. **education** (object) จบการศึกษาปริญญาตรีหรือไม่
3. **self_employed** (object) ทำงานอิสระหรือไม่
4. **income_annum** (type int64) รายได้ต่อปี(รูปี)
5. **loan_amount** (type int64) จำนวนเงินในการกู้
6. **loan_term** (type int64) ระยะเวลาการผ่อนชำระ
7. **cibil_score** (type int64) เครดิต
8. **residential_assets_value** (type int64) มูลค่าของที่อยู่อาศัย
9. **commercial_assets_value** (type int64) มูลค่าทรัพย์สินเชิงพาณิชย์
10. **luxury_assets_value** (type int64) มูลค่าทรัพย์สินที่มีค่า (สร้อยทอง, เพชร, เครื่องประดับ)
11. **bank_asset_value** (type int64) เงินที่ฝากกับธนาคาร
12. **loan_status** (object) อนุมัติหรือไม่

ឧបត្ថម្ភ ៣: វិវត្តការ



บทที่ 3: วิธีดำเนินการ

3.3 สมมติฐานการวิจัย

สมมติฐานหลัก: ค่าเฉลี่ยของรายได้ผู้ที่ได้สินเชื่อ \geq ค่าเฉลี่ยของรายได้ผู้ที่ไม่ได้สินเชื่อ

สมมติฐานรอง: ค่าเฉลี่ยของรายได้ผู้ที่ได้สินเชื่อ $<$ ค่าเฉลี่ยของรายได้ผู้ที่ไม่ได้สินเชื่อ

3.4 การวิเคราะห์ข้อมูล

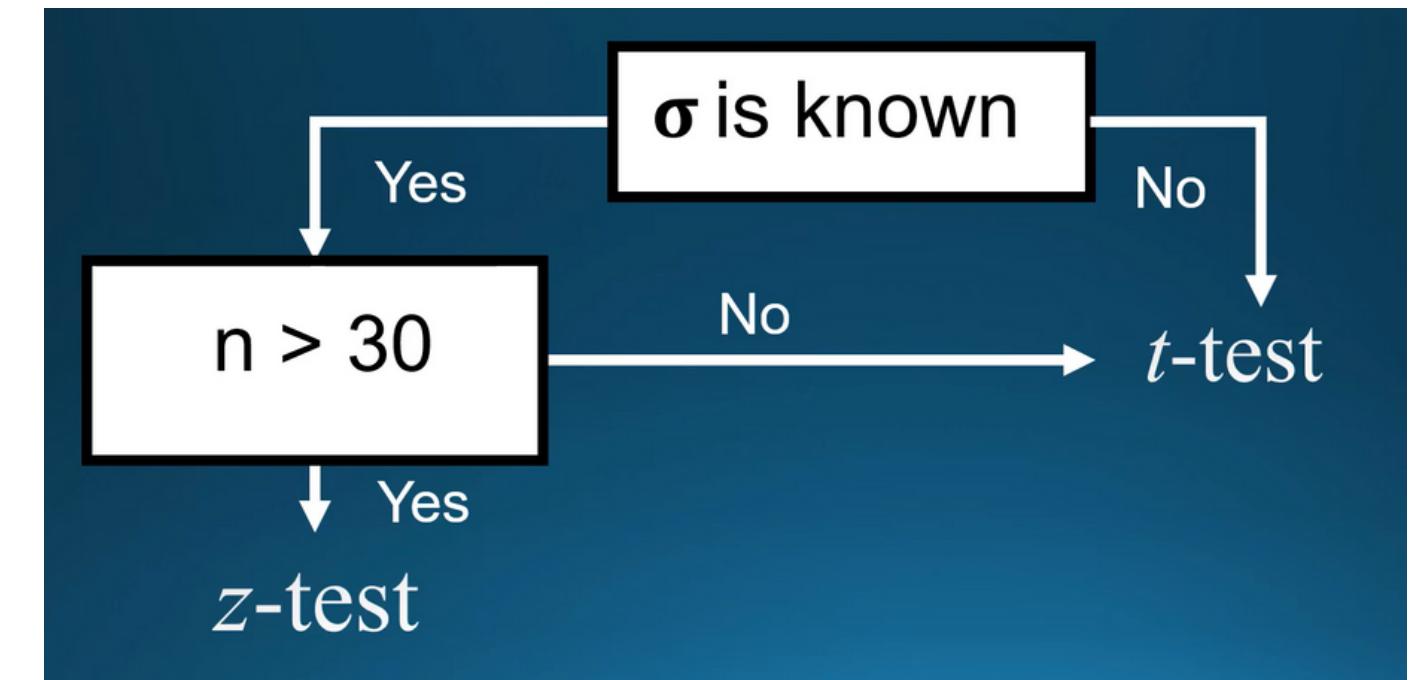
- ค่าความเชื่อมั่น = 0.01
- สถิติ t ($n=4,269$)
- model
 - Q-learning
 - sparse binary Transformers (SBT)

3.6 การประเมินผล

Confusion Matrix (such accuracy recall precision)

Confusion Matrix

		Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)	
	False Negatives (FNs)	True Negatives (TNs)	
Predicted Negative (0)			



Reference

- Uddin, N., Ahamed, M. K. U., Uddin, M. A., Islam, M., Talukder, M. A., & Aryal, S. (2023). An ensemble machine learning based bank loan approval predictions system with a smart application. *International Journal of Cognitive Computing in Engineering*, 4, 327–339.
<https://doi.org/10.1016/j.ijcce.2023.09.001>
- Soreti Bekele Babo, Asrat Mulatu Beyene. **Bank Loan Classification of Imbalanced Dataset Using Machine Learning Approach**, 15 March 2023, PREPRINT (Version 1) available at Research Square [<https://doi.org/10.21203/rs.3.rs-2667057/v1>]
- <https://arxiv.org/pdf/2308.04637.pdf>