

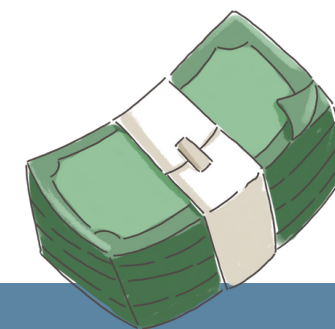
# ระบบคาดการณ์การอนุมัติสินเชื่อจากธนาคารที่ใช้ Deep Q-learning

An ensemble Deep Q-learning based bank loan approval predictions system

640315 สิทธิกร เฉลิมกิตติชัย

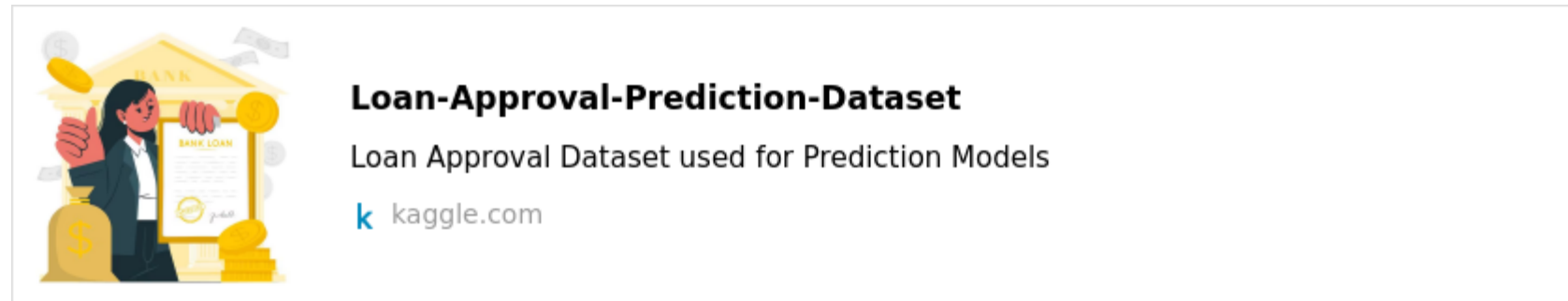
640549 ชัยณุชา อัครกุลพิชา

640542 ชัยกฤษ พุ่มเทียน

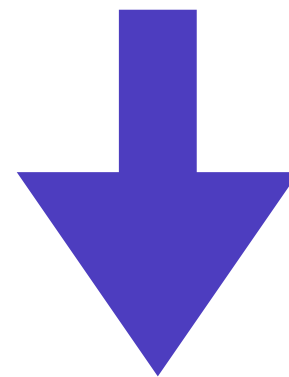


# บทที่ 1 ความเป็นมาและความสำคัญ

ในยุคดิจิทัลที่การเข้าถึงบริการทางการเงินต้องรวดเร็วและแม่นยำ การพัฒนากระบวนการอนุมัติสินเชื่อจึงมีความสำคัญมากขึ้น การใช้เทคโนโลยีที่เรียกว่า Deep Q-learning ในระบบอนุมัติสินเชื่อของธนาคาร ได้ช่วยให้กระบวนการนี้เป็นไปอย่างอัตโนมัติและมีประสิทธิภาพมากขึ้น ระบบนี้ใช้หลักการของการเรียนรู้แบบเสริมแรงเพื่อปรับปรุงการตัดสินใจในการอนุมัติสินเชื่อ ช่วยให้ธนาคารสามารถลดความผิดพลาด ความเสี่ยง

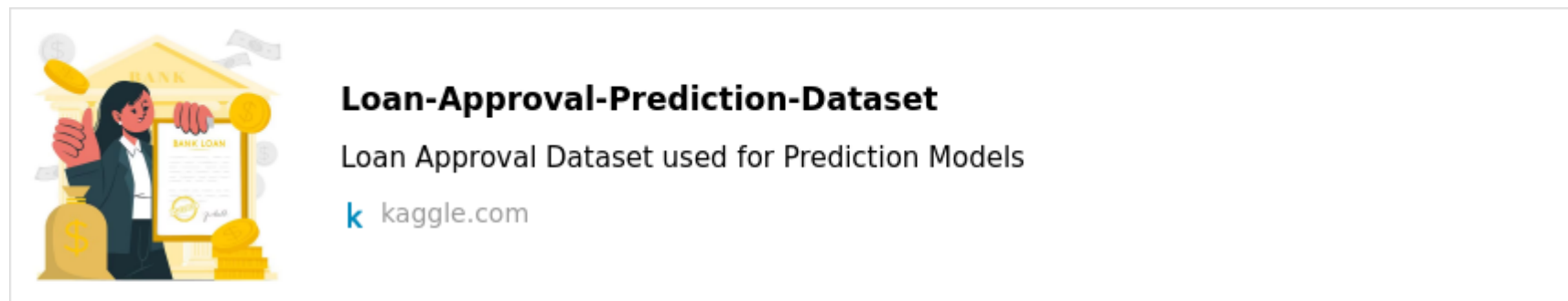


พวกเราอยากศึกษาวิธีการทางปัญญาประดิษฐ์ด้านการปล่อยสินเชื่อด้วยข้อมูลจาก kaggle.com



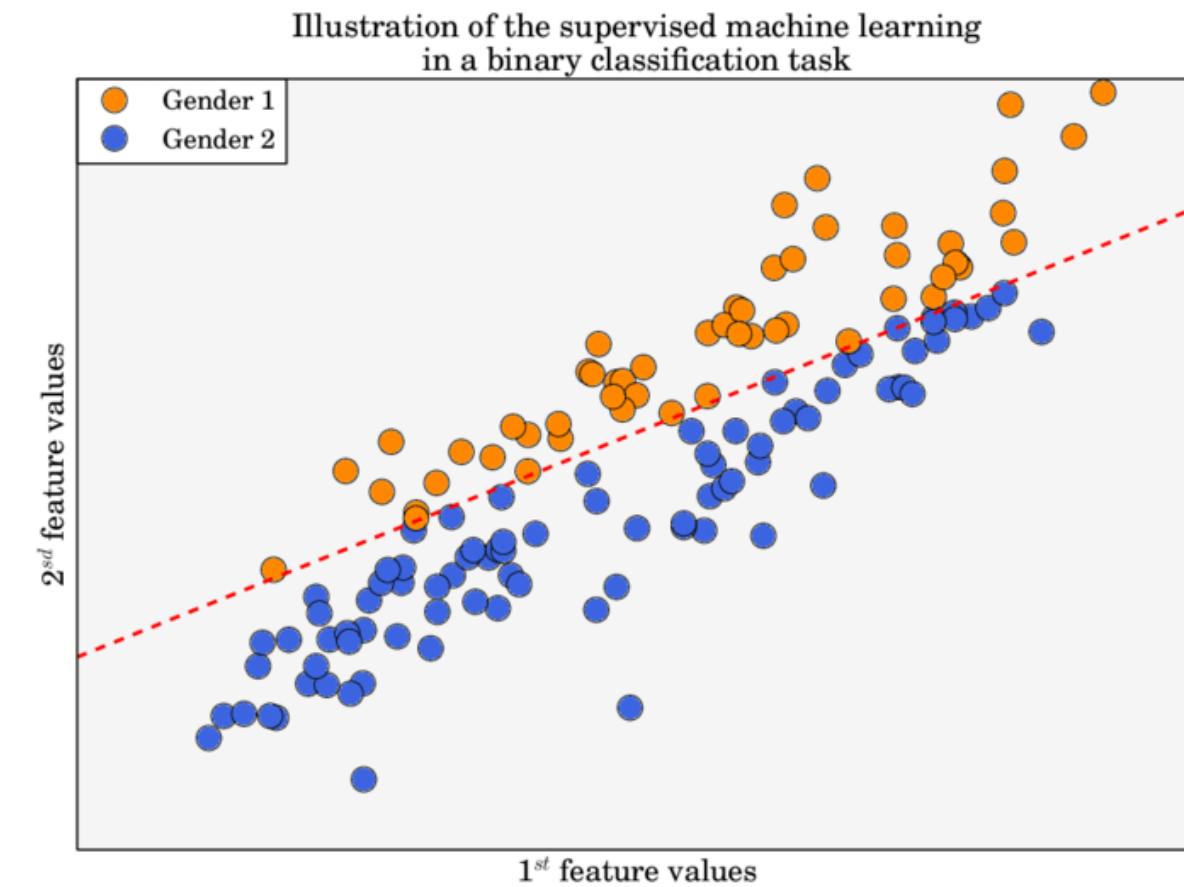
มีคนทำ model อะไรแล้วไปบ้าง  
แล้วมี model อะไรที่ยังไม่เคยใช้งานในด้านการปล่อยสินเชื่อ

# บทที่ 1 ขอบเขตของโครงการ



ศึกษา model ที่ยังไม่เคยลองกับข้อมูลการปล่อยสินเชื่อ  
แล้วนำมาทดลองกับข้อมูลจาก kaggle.com

ข้อมูล 07/2023



ใช้เทคนิคจำแนกประเภท (Classification) ด้วยวิธีการต่างๆ

# บทที่ 1

## วัตถุประสงค์

1. เพื่อเปรียบเทียบรายได้เฉลี่ยของผู้ขอสินเชื่อระหว่างผู้ที่ได้รับสินเชื่อและไม่ได้รับสินเชื่อ
2. เพื่อทำนายการปล่อยสินเชื่อด้วยวิธีDeep Q-learning
3. เปรียบเทียบประสิทธิภาพของเทคนิคทางปัญญาประดิษฐ์

## ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อสนับสนุนพนักงานฝ่ายการปล่อยสินเชื่อ
2. เพื่อลดระยะเวลาในการปล่อยสินเชื่อ
3. ลดความเสี่ยงในการอนุมัติกับผู้ที่ไม่เหมาะสม

# บทที่ 2 งานวิจัยที่เกี่ยวข้องกับงานที่จะทำ

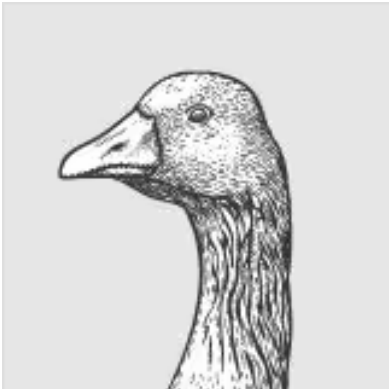


**Loan Status Prediction with 97%**  
Explore and run machine learning code with Kaggle Notebooks | Using data from Loan-Approval-Prediction-Dataset  
k Kaggle / Apr 10

- LabelEncoder
  - education
  - self\_employed
  - loan\_status
- split
  - train 80
  - test 20


| Name                | Accuracy          |
|---------------------|-------------------|
| Logistic Regression | 59.83606557377049 |
| Decision Tree       | 97.54098360655738 |
| Random Forest       | 97.54098360655738 |

# บทที่ 2 งานวิจัยที่เกี่ยวข้องกับงานที่จะทำ



**LoanApprove\_98%acc\_easyExplained**

Explore and run machine learning code with Kaggle Notebooks | Using data from Loan-Approval-Prediction-Dataset

 Kaggle / Feb 12

LabelEncoder

- education
- self\_employed
- loan\_status

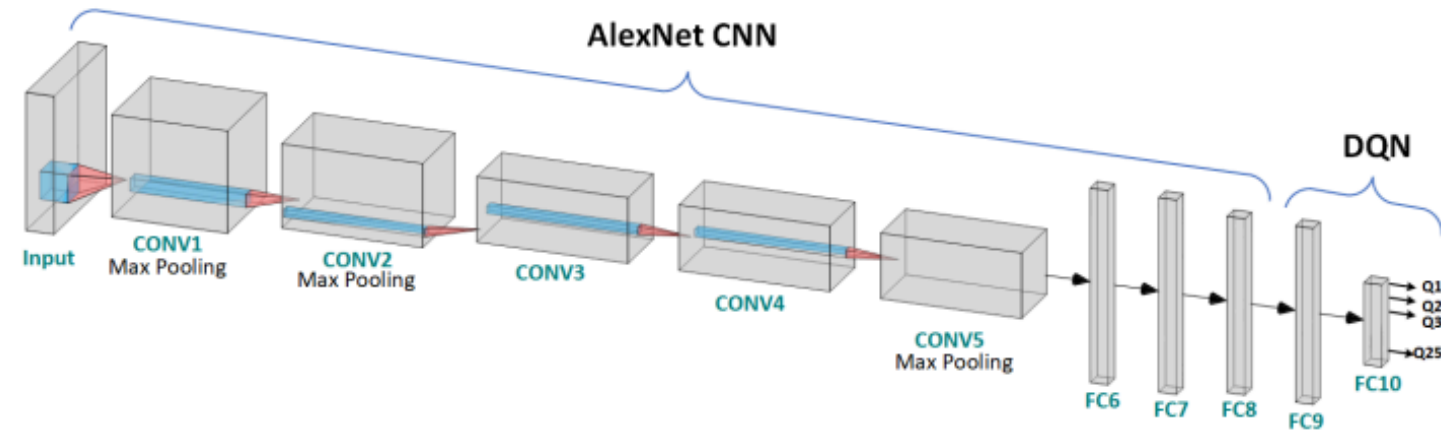
split

- train 80
- test 20

| Name                | Accuracy          |
|---------------------|-------------------|
| Logistic Regression | 63.23185011709602 |
| Random Forest       | 98.36065573770492 |

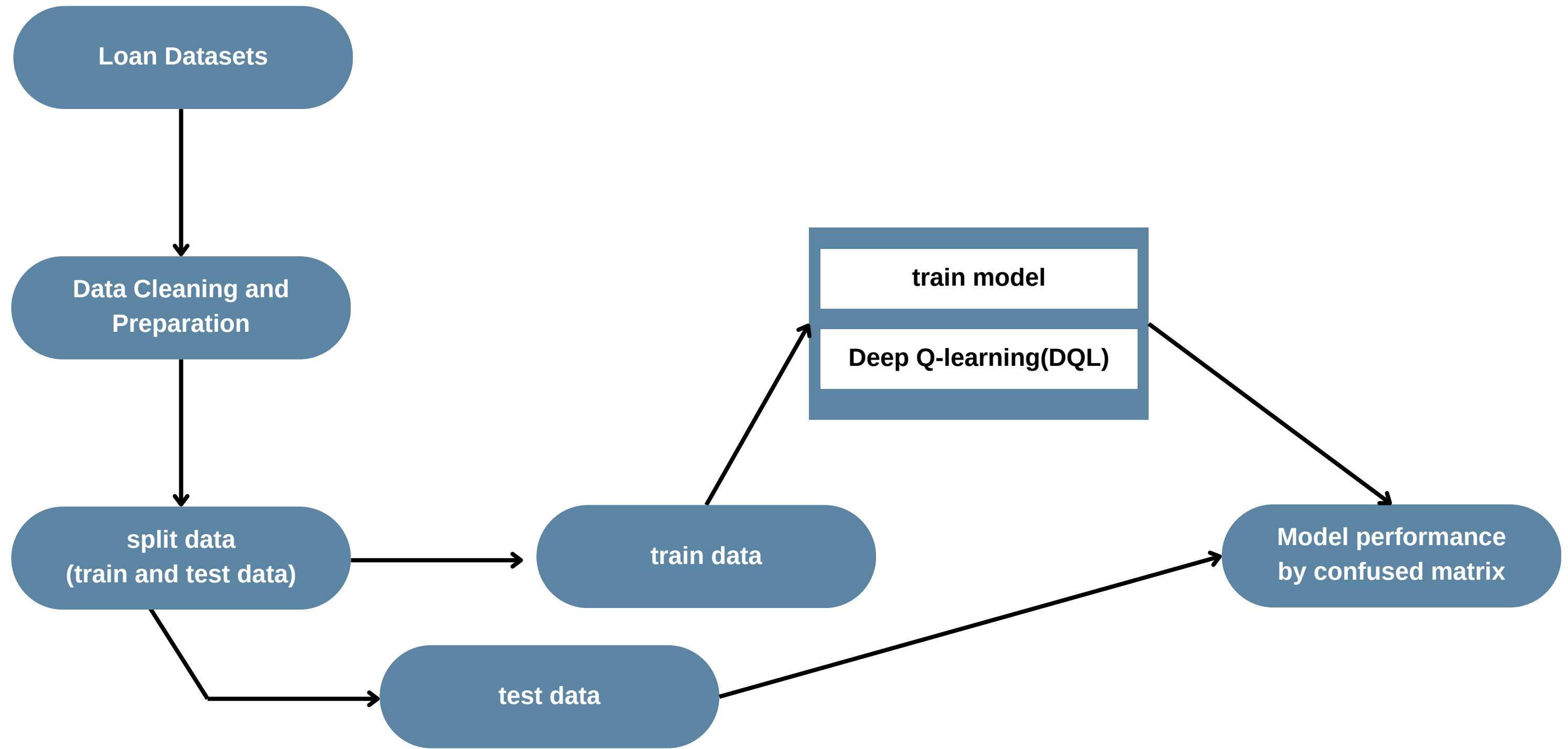
## บทที่ 2 งานวิจัยที่เกี่ยวข้องกับงานที่จะทำ

### 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 3: วิธีดำเนินการ





# บทที่ 3: วิธีดำเนินการ

## 3.1 ข้อมูลและตัวแปร

### รายละเอียดข้อมูล:

1. loan\_id (type int64) เลขสีนเชื่ออ้างอิง
2. no\_of\_dependents (type int64) คนในครอบครัวที่อยู่ในการปกครอง
3. education (object) จบการศึกษาปริญญาตรีหรือไม่
4. self\_employed (object) ทำงานอิสระหรือไม่
5. income\_annum (type int64) รายได้ต่อปี(รูป )
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:

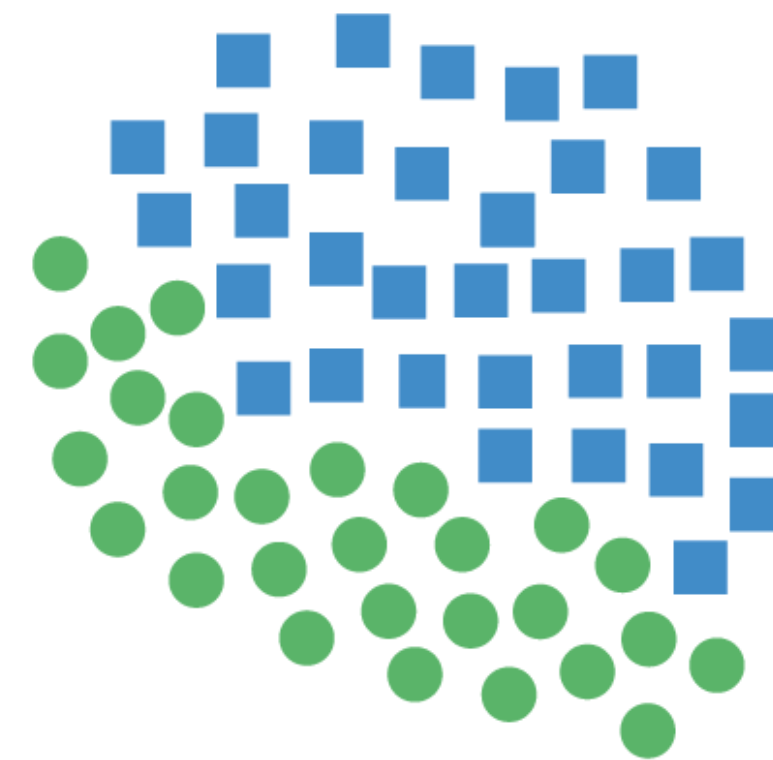
1. no\_of\_dependents (type int64) คนในครอบครัวที่อยู่ในการปกครอง
2. education (object) จบการศึกษาปริญญาตรีหรือไม่
3. self\_employed (object) ทำงานอิสระหรือไม่
4. income\_annum (type int64) รายได้ต่อปี(รูป )
5. loan\_amount (type int64) จำนวนเงินในการกู้
6. loan\_term (type int64) ระยะเวลาการผ่อนชำระ
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

## บทที่ 3: วิธีดำเนินการ



Original Dataset

**data imbalance**



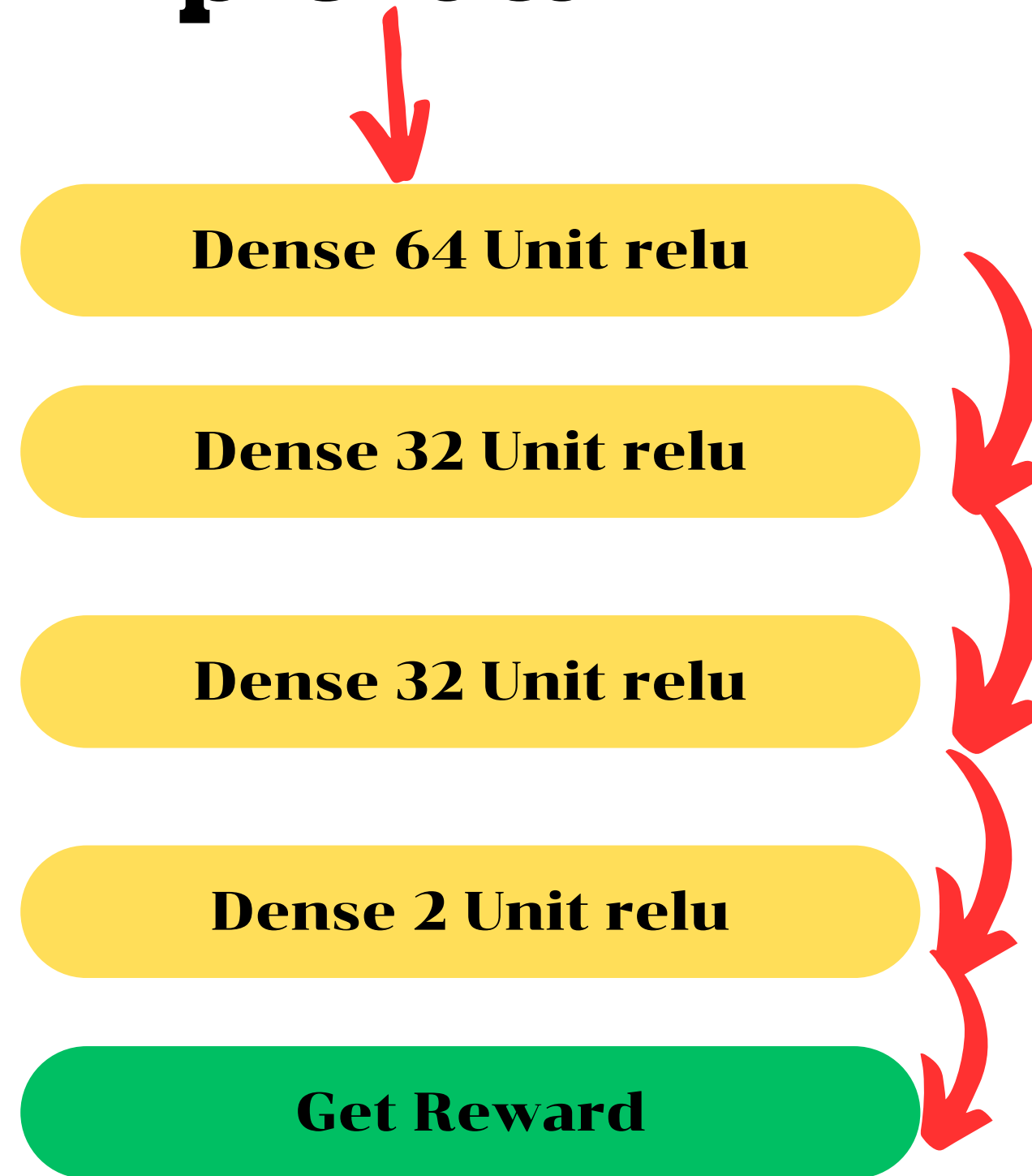
Resampled Dataset

**data balance**

**eliminate data redundancy**

### บทที่ 3: วิธีดำเนินการ

**preData**



## บทที่ 3: วิธีดำเนินการ

### 3.3 สมมติฐานการวิจัย

สมมติฐานหลัก: ค่าเฉลี่ยของรายได้ผู้ที่ได้สินเชื่อ  $\geq$  ค่าเฉลี่ยของรายได้ผู้ที่ไม่ได้สินเชื่อ

สมมติฐานรอง: ค่าเฉลี่ยของรายได้ผู้ที่ได้สินเชื่อ  $<$  ค่าเฉลี่ยของรายได้ผู้ที่ไม่ได้สินเชื่อ

### 3.4 การวิเคราะห์ข้อมูล

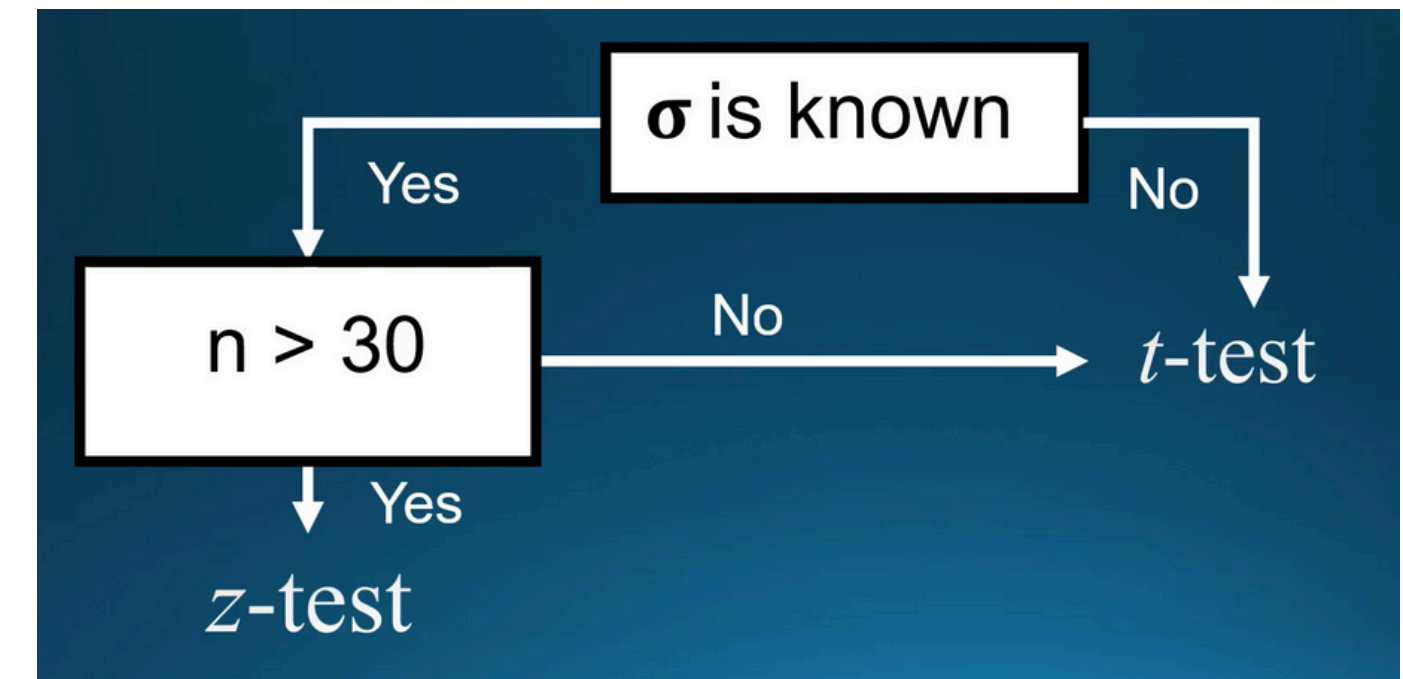
- ระดับนัยสำคัญ = 0.01
- สถิติ  $t$  ( $n=4,269$ )
- training Deep Q-learning model

### 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) |
| Predicted Negative (0) | False Negatives (FNs) | True Negatives (TNs)  |



## บทที่ 4: ผลการวิจัยและการอภิปรายผล

### 1. ตั้งสมมติฐานทางสถิติ

สมมติฐานหลัก: ค่าเฉลี่ยของรายได้ผู้ที่ได้สินเชื่อ  $\geq$  ค่าเฉลี่ยของรายได้ผู้ที่ไม่ได้สินเชื่อ

สมมติฐานรอง: ค่าเฉลี่ยของรายได้ผู้ที่ได้สินเชื่อ  $<$  ค่าเฉลี่ยของรายได้ผู้ที่ไม่ได้สินเชื่อ

### 2. กำหนดระดับนัยยะสำคัญทางสถิติ

ระดับนัยยะสำคัญ = 0.01

| F-Test Two-Sample for Variances |                     |                    |
|---------------------------------|---------------------|--------------------|
|                                 | <i>income_annum</i> | <i>loan_status</i> |
| Mean                            | 5059123.917         | 0.622159756        |
| Variance                        | 7.87835E+12         | 0.235132073        |
| Observations                    | 4269                | 4269               |
| df                              | 4268                | 4268               |
| F                               | 3.35061E+13         |                    |
| P(F<=f) one-tail                | 0                   |                    |
| F Critical one-tail             | 1.073828325         |                    |

## บทที่ 4: ผลการวิจัยและการอภิปรายผล

|   |                     |                     |
|---|---------------------|---------------------|
| t-Test: Two-Sample Assuming Unequal Variances |                     |                     |
|   |                     |                     |
|   | <i>income_annum</i> | <i>dummy_status</i> |
| Mean  | 5059123.917         | 0.622159756         |
| Variance                                      | 7.87835E+12         | 0.235132073         |
| Observations                                  | 4269                | 4269                |
| Hypothesized Mean Difference                  | 0                   |                     |
| df  | 4268                |                     |
| t Stat  | 117.7662131         |                     |
| P(T<=t) one-tail                              | 0                   |                     |
| t Critical one-tail                           | 2.327221917         |                     |
| P(T<=t) two-tail                              | 0                   |                     |
| t Critical two-tail                           | 2.576981745         |                     |

### 3.สรุปผล

p-value น้อยกว่า ระดับนัยสำคัญ แปลว่า ปฏิเสธสมมุติฐานหลัก แปลว่าค่าเฉลี่ยของรายได้ผู้ที่ได้สินเชื่อมีค่าน้อยกว่าผู้ที่ไม่ได้สินเชื่อ

## บทที่ 4: ผลการวิจัยและการอภิปรายผล

### DQN 15 EPOCH

```
1/1 [=====] - 0s 20ms/step
1/1 [=====] - 0s 19ms/step
1/1 [=====] - 0s 21ms/step
1/1 [=====] - 0s 19ms/step
1/1 [=====] - 0s 18ms/step
1/1 [=====] - 0s 19ms/step
1/1 [=====] - 0s 19ms/step
1/1 [=====] - 0s 18ms/step
1/1 [=====] - 0s 20ms/step
1/1 [=====] - 0s 25ms/step
1/1 [=====] - 0s 22ms/step
1/1 [=====] - 0s 21ms/step
1/1 [=====] - 0s 20ms/step
1/1 [=====] - 0s 25ms/step
1/1 [=====] - 0s 20ms/step
1/1 [=====] - 0s 21ms/step
1/1 [=====] - 0s 20ms/step
1/1 [=====] - 0s 21ms/step
1/1 [=====] - 0s 24ms/step
1/1 [=====] - 0s 23ms/step
1/1 [=====] - 0s 22ms/step
1/1 [=====] - 0s 23ms/step
Accuracy on the test set: 0.82
```

# บทที่ 5: สรุปผลการวิจัยและข้อเสนอแนะ

## ผลการวิจัย

พวกเราได้ทดลองใช้ Deep Q-learning + SMOTE แล้วได้ค่า Accuracy = 82% และ epoch 15 รอบเท่านั้น

| Name                    | Accuracy          |
|-------------------------|-------------------|
| Logistic Regression     | 63.23185011709602 |
| Random Forest           | 98.36065573770492 |
| Deep Q-learning + SMOTE | 82.00             |

## ข้อเสนอแนะ

แนะนำให้ใช้ method อื่น เพราะว่า method นี้ได้ Accuray น้อยกว่าคนอื่นที่ทำงานด้านนี้



# 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>]

**THANK YOU**