Report

ในส่วนของ Project นี้ทางทีมได้แบ่งการทำงงานเป็น 2 ส่วน คือ

1. Dataset และ Model
2. Website
3. ) Dataset และ Model

Dataset: ใช้ Dataset จากเว็บไซต์ Kaggle ชื่อ “Loan Approval Prediction” ซึ่งเป็น Dataset ของบริษัท Dream Housing Finance จัดการเรื่องสินเชื่อบ้านทั้งหมด มีอยู่ทั่วทุกพื้นที่ในเมือง กึ่งเมือง และชนบท มี Feature ทั้งหมด 13 Features 614 แถว ได้แก่

1.Loan\_ID= Unique Loan ID

2.Gender= Male/ Female

3.Married= Applicant married (Y/N)

4.Dependents= Number of dependents

5.Education= Applicant Education (Graduate/ Under Graduate)

6.Self\_Employed= Self-employed (Y/N)

7.ApplicantIncome= Applicant income

8.CoapplicantIncome= Coapplicant income

9.LoanAmount= Loan amount in thousands

10.Loan\_Amount\_Term= Term of the loan in months

11.Credit\_History= Credit history meets guidelines

12.Property\_Area= Urban/ Semi-Urban/ Rural

13.Loan\_Status= (Target) Loan approved (Y/N)

แหล่งที่มา: <https://www.kaggle.com/datasets/sonalisingh1411/loan-approval-prediction?select=Training+Dataset.csv>

A picture containing text

Description automatically generatedจำนวน Missing value ของแต่ละ Feature (ทั้งหมด 149 values)

ทางทีมจึงมีสองวิธีในการจัดการกับ Missing values ได้แก่

1. **Regression to predict null value.**
2. **Features Engineering**

โดยการทำ Features Engineering จะเป็นการพิจารณาบทบาทและปริมาณข้อมูงของ Feature นั้น ๆ เพื่อหาวิธีแก้ไข Missing value ที่ดีที่สุด ดังนี้

2.1) Gender (Missing value: 13)

ใช้วิธี KNeighborsClassifier ในการสร้างแบบทำนายเพศ (Gender Predictor) จากชุดข้อมูลสินเชื่อ

|  |  |  |
| --- | --- | --- |
|  | Original Dataset | After preprocess |
| Male | 489 | 112 |
| Female | 489 | 125 |

2.2) Married (Missing value: 3)

จากข้อมูลผู้สมัครที่แต่งงานแล้ว 398 คน ในจำนวนนี้ 309 คนจบการศึกษาแล้ว 326 คนไม่ได้ประกอบอาชีพอิสระ และ 151 คนเป็นชาวเมือง จากนี้ เราสรุปได้ว่าตัวอย่างที่ขาดหายไปทั้งหมดเป็น Married

|  |  |  |
| --- | --- | --- |
|  | Original Dataset | After preprocess |
| Yes | 398 | 401 |
| No | 213 | 213 |

Chart, pie chart

Description automatically generated2.3) Dependents (Missing value: 15)

Chart, pie chart

Description automatically generated

จำนวนคนที่แต่งงานแล้วมีทั้งหมด 401 คน จำนวนคนที่มีลูกมี 227 คน คิดเป็น 56.6 %

จำนวนคนที่ยังไม่ได้แต่งงานมีทั้งหมด 213 คน จำนวนคนที่ไม่มีลูกเลย 175 คน คิดเป็น 82.15 %

ทางทีมจึงพิจารณาว่า Dependents คือ เด็กหรือลูกของผู้ขอสินเชื่อ โดย หากผู้ขอสินเชื่อไม่ได้แต่งงาน เราจะเติมค่าว่างนั้นด้วย 0 และหากผู้ขอสินเชื่อแต่งงานแล้ว เราจะเติมค่าว่างนั้นด้วย 1

|  |  |  |
| --- | --- | --- |
|  | Original Dataset | After preprocess |
| 0 | 345 | 349 |
| 1 | 101 | 113 |
| 2 | 102 | 102 |
| 3+ | 51 | 51 |

2.4) Dependents (Missing value: 32)

จาก Dataset ข้อมูลลูกค้าส่วนใหญ่ (85.9 %) จะไม่มีธุรกิจส่วนตัว ทางทีมเลยพิจารณาว่า ข้อมูลส่วนที่หายไปจะให้เป็น 'No' (ไม่มีธุรกิจส่วนตัว)

|  |  |  |
| --- | --- | --- |
|  | Original Dataset | After preprocess |
| Yes | 82 | 82 |
| No | 500 | 532 |

Graphical user interface

Description automatically generated with low confidence2.5) LoanAmount (Missing value: 22)

จาก Dataset ทางทีมได้พิจารณาจากผลของการอนุมัติสินเชื่อ (Loan\_Status) โดยให้ดูค่าสถิตต่าง ๆ ของ LoanAmount จะเห็นว่า ผลผ่านการอนุมัติจะมีค่าเฉลี่ยของ LoanAmount **151.22** ส่วนผลไม่ผ่านการอนุมัติจะมีค่าเฉลี่ย **144.29** เราจะค่าเฉลี่ยดังกล่าวมาเพิ่มในส่วนของ Missing value

2.6) Loan\_Amount\_Term (Missing value: 14)

จาก Dataset ข้อมูลของ Loan\_Amount\_Term ส่วนใหญ่จะเป็น 360 ทางทีมจึงพิจารณาให้ค่าที่หายไปเป็น 360

Chart, pie chart

Description automatically generated2.7) Credit History (Missing value: 50)

เนื่องจากผู้ขอสินเชื่อ 77.4 % มี Credit\_History ทางทีมจึงเติมค่า Null ด้วย 1.0

|  |  |  |
| --- | --- | --- |
|  | Original Dataset | After preprocess |
| Yes | 475 | 525 |
| No | 89 | 89 |

2.8) Property\_Area

เนื่องจากเป็นข้อมูลที่เป็น string ไม่สามารถนำมาเข้า Model ได้ทางทีมจึงทำ One-Hot Encoding หรือ Dummy Encoding เพื่อให้โมเดลสามารถจำแนกลักษณะของเมืองที่อยู่ได้

|  |  |
| --- | --- |
| Property\_Area | Count |
| Rural (ชนบท) | 179 |
| Semiurban | 233 |
| urban | 202 |

Graphical user interface, application

Description automatically generated

**Percentage Importance of each feature with respect to Loan\_Status**

Text

Description automatically generatedExtraTreesClassifier คืออัลกอริทึมที่ใช้ในการสร้างและเรียนรู้ข้อมูลในรูปแบบของต้นไม้ตัดสินใจ (Decision Trees) หลาย ๆ ต้น และประเภทของเส้นทางตัดสินใจที่สุ่มขึ้นมาจากข้อมูล (Extremely Randomized Trees) ซึ่งเป็นวิธีการปรับปรุงที่คล้ายกับ Random Forest แต่ ExtraTreesClassifier สุ่มข้อมูลอย่างมากขึ้นในขั้นตอนของการแยกแยะข้อมูลที่แต่ละโหนด

ในส่วนของ Model ทางทีมได้ทำการทดสอบ Decision Tree โดยเพิ่มปัจจัย มา 2 อย่างคือ Standard Scaling และ PCA โดยจะทำการแบ่ง Trian Test เป็นอัตราส่วน 80:20

1. Decision Tree Classifier Model without Standard Scaling and without PCA

Training Accuracy: 85.53 %

Testing Accuracy: 85.36 %

1. Decision Tree Classifier Model with Standard Scaling and without PCA

Training Accuracy: 85.74 %

Testing Accuracy: 85.36 %

1. Decision Tree Classifier Model with Standard Scaling and with PCA

Training Accuracy: 84.92 %

Testing Accuracy: 83.73 %

จากผลการทดลองพบว่าทั้ง 3 วิธีได้ค่า Accuracy ที่ใกล้เคียงกันมาก ทางทีมจึงเลือกใช้วิธีที่ 1 เพราะไม่ต้องทำ scaling และ PCA เพื่อลดเวลาการทำงาน

เมื่อนำวิธีการจัดการ Missing value ทั้งสองวิธีมาใช้กับ Model Decision Tree ได้ผลลัพธ์ดังนี้

|  |  |
| --- | --- |
|  | Accuracy |
| Regression to predict null value | 0.81 |
| Features Engineering | 0.85 |

จากผล Accuracy พบว่าวิธี Features Engineering ได้ผลลัพธ์ที่ดีกว่าวิธี Regression to predict null value เราจึงเลือกใช้วิธี Features Engineering ในการ preprocess Data และใช้ Model Decision Tree ในการ classification การอนุมัติสินเชื่อ

1. ) Website

รูปแบบหน้าทั้งหมดมีดังต่อไปนี้

• หน้าแรก

• หน้าสมัครสมาชิกของพนักงาน

• หน้าเข้าสู่ระบบของสมาชิก

• หน้ากรอกข้อมูลของลูกค้า

• หน้าตัดสินใจในการให้สินเชื่อลูกค้า

**หน้าแรก**

Graphical user interface, application

Description automatically generated

หน้าแรกจะแบ่งออกเป็น 4 แถบตามด้านบนโดยมีการทำงานดังต่อไปนี้

1. CPE: แถบสำหรับลิงค์ไปยัง <https://www.facebook.com/cpe.eng.kps> หน้า Facebook ภาควิชาวิศวกรรมคอมพิวเตอร์ กำแพงแสน
2. Home: แถบสำหรับการกลับมาหน้าหลักของตัวเว็บไซต์
3. About Us: แถบสำหรับลิงค์ไปหน้าสมาชิกภาพในกลุ่มดังภาพด้านล่างนี้

A picture containing text, screenshot, indoor, different

Description automatically generated

1. Register: แถบสำหรับลิงค์ไปหน้าสมัครสมาชิกในกลุ่มดังภาพด้านล่างนี้

Graphical user interface, application, email

Description automatically generated

1. Graphical user interface, website

   Description automatically generatedLogin: แถบสำหรับลิงค์ไปหน้าเข้าสู่ระบบของสมาชิกในกลุ่มดังภาพด้านล่างนี้

**หน้าสมัครสมาชิกของพนักงาน**

Graphical user interface, application, email

Description automatically generated

มีรายละเอียดดังต่อไปนี้คือ

* กรอกชื่อที่ใช้ในการแสดงกับหน้าเว็บ
* กรอกอีเมลล์ที่ใช้สำหรับการเข้าสู่ระบบ
* กรอกรหัสที่ใช้สำหรับการเข้าสู่ระบบ
* กรอกรหัสยืนยันที่ใช้สำหรับการเข้าสู่ระบบ
* เลือกวันเกิดตามรูปแบบปฏิทิน

หน้าเข้าสู่ระบบของสมาชิก

Graphical user interface, website

Description automatically generated

มีรายละเอียดดังต่อไปนี้

* กรอกอีเมลล์ที่สำหรับการเข้าสู่ระบบ [ต้องมีอยู่ฐานข้อมูล]
  + หากไม่มี จะแสดงผลดังภาพต่อไปนี้

Graphical user interface

Description automatically generated

* กรอกรหัสที่สำหรับการเข้าสู่ระบบ [ต้องตรงกับฐานข้อมูล]
  + Graphical user interface, application

    Description automatically generatedหากไม่ตรงกับฐานข้อมูล จะแสดงผลดังภาพต่อไปนี้

หน้ากรอกข้อมูลของลูกค้า

Graphical user interface, application, email

Description automatically generated

มีรายละเอียดดังต่อไปนี้

* กรอกข้อมูลของลูกค้า ชื่อและนามสกุล
* กรอกข้อมูลของลูกค้า อายุ เพศและเบอร์โทรศัพท์
* เลือกข้อมูลการแต่งงานของลูกค้าว่า แต่งงานแล้ว หรือ ยังไม่แต่งงาน
* เลือกข้อมูลประวัติเครดิตของลูกค้าว่า มี หรือ ไม่มี (ใช้การพิจารณาของโมเดล)
* กรอกข้อมูลของลูกค้าจำนวนเงินที่ขอกู้ยืม (ใช้การพิจารณาของโมเดล)
* กรอกข้อมูลของลูกค้ารายได้ของผู้กู้ยืม (ใช้การพิจารณาของโมเดล)
* กรอกข้อมูลของลูกค้ารายได้จากผู้สมรสหรือผู้ร่วมขอสินเชื่อที่ใช้ร่วมกันกับผู้กู้ยืม (ใช้การพิจารณาของโมเดล)
* เลือกข้อมูลของลูกค้าจำนวนคนที่อยู่ในด้านการเลี้ยงดูของผู้ยืม (0, 1, 2, +3) (ใช้การพิจารณาของโมเดล)
* เลือกข้อมูลของลูกค้าระดับการศึกษาของผู้กู้ยืม (Graduate, Not Graduate)
* เลือกข้อมูลสถานะการเป็นเจ้าของธุรกิจของผู้กู้ยืมของลูกค้าว่า มี หรือ ไม่มี
* กรอกข้อมูลของลูกค้าระยะเวลาในการกู้ยืม
* เลือกข้อมูลของลูกค้าที่อยู่อาศัยของผู้กู้ยืม ('Urban', 'Rural', 'Semiurban')

**หน้าตัดสินใจในการให้สินเชื่อลูกค้า**

Graphical user interface, application, Teams

Description automatically generated

มีรายละเอียด คือ การแสดงผลของลูกค้าว่าระบบจะแนะนำให้ปล่อยกู้หรือไม่โดยพนักงานสามารถกดได้ 2 Actions

* Denied: ไม่อนุมัติ
* Approve: อนุมัติ