# Beta2 18 Feature Model:

Training model with top 18 features

Data downloaded from Monthly\_Income\_Estimation/Target\_Encoded\_Artifacts/companytrgencode.joblib

Index(['global\_mean', 'smoothing\_factor', 'loan\_company\_name',

'freq\_encodedcompanyName', 'category\_mean\_companyName',

'category\_size\_companyName', 'target\_encodedcompanyName',

'encoded\_company\_name\_group'],

dtype='object')

Preparing data...

The shape of dfcn before duplicate drop is: (64774, 4)

The shape of dfcn is: (64774, 4)

The shape of df\_train is: (61584, 22)

['digitalLoanAccountId', 'freq\_encodedcompanyName', 'target\_encodedcompanyName', 'loan\_company\_name', 'loan\_type']

Dropped columns due to multicollinearity: []

Training model...

Making predictions...

Calculating metrics...

Training accuracy: 61.20%

Test accuracy: 71.08%

Model Performance Metrics:

Mean Absolute Error (MAE): 4247.56

Mean Squared Error (MSE): 157514277.77

Root Mean Squared Error (RMSE): 12550.47

R-squared (R2): 0.7108

Mean Absolute Percentage Error (MAPE): 0.1153

Generating plots...

A graph with a red line and blue dots

Description automatically generated

A graph with numbers and a line

Description automatically generated

A graph with a line and a line

Description automatically generated

A graph with an orange line

Description automatically generated

Top 10 Most Important Features:

feature importance

0 encoded\_company\_name\_group 76.92

1 ln\_loan\_type 5.53

2 ln\_Industry\_desc 3.32

3 ln\_age 2.95

4 ln\_education\_level 2.64

5 ln\_Employment\_type 1.81

6 ln\_city 1.03

7 ln\_osversion\_type 0.78

8 ln\_docType 0.77

9 ln\_source\_funds 0.74

10 ln\_brand 0.67

11 ln\_purpose\_desc 0.55

12 ln\_postalcode 0.54

13 ln\_nature\_of\_work 0.53

14 ln\_dependent\_count 0.39

15 ln\_gender 0.37

16 ln\_email\_domain 0.24

17 loan\_barangay 0.22

A graph with blue and red lines

Description automatically generated

A graph with red and white text

Description automatically generated

Model successfully saved to: /home/jupyter/Models/Income\_Estimation\_Models/Income\_Estimation\_Model/model\_Good\_Customer\_Top18SHAPFeatures\_Beta2WithOutApp\_Step2\_1\_0.joblib

Metrics structure: (4247.563617548964, 157514277.76759616, 12550.469224996974, 0.7108496482101371, 0.11532908180185107)

# Alpha (Beta2 + CIC)

Training model with top 16 features

Data downloaded from Monthly\_Income\_Estimation/Target\_Encoded\_Artifacts/companytrgencode.joblib

Index(['global\_mean', 'smoothing\_factor', 'loan\_company\_name',

'freq\_encodedcompanyName', 'category\_mean\_companyName',

'category\_size\_companyName', 'target\_encodedcompanyName',

'encoded\_company\_name\_group'],

dtype='object')

Preparing data...

The shape of dfcn before duplicate drop is: (61616, 4)

The shape of dfcn is: (61616, 4)

The shape of df\_train is: (61584, 21)

['digitalLoanAccountId', 'freq\_encodedcompanyName', 'target\_encodedcompanyName', 'loan\_company\_name', 'loan\_type']

Dropped columns due to multicollinearity: []

Training model...

The cat feature indices are: [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]

Model fitting started

Making predictions...

Calculating metrics...

Training accuracy: 61.48%

Test accuracy: 73.54%

Model Performance Metrics:

Mean Absolute Error (MAE): 4035.36

Mean Squared Error (MSE): 144156524.47

Root Mean Squared Error (RMSE): 12006.52

R-squared (R2): 0.7354

Mean Absolute Percentage Error (MAPE): 0.1124

Generating plots...

A graph showing a line and a red line

Description automatically generated with medium confidence

A graph with numbers and lines

Description automatically generated

A graph with a line and a line

Description automatically generated

A graph with an orange line

Description automatically generated

Top 10 Most Important Features:

feature importance

0 encoded\_company\_name\_group 68.231943

1 loan\_product\_type 9.206202

2 loan\_education\_level 5.595243

3 employment\_type 2.525495

4 max\_amt\_active\_contracts 2.516286

5 industry\_description 2.259833

6 age 2.222626

7 osversionType 1.369698

8 loan\_city 1.350848

9 CreditAvgCreditLimit 1.315839

10 loan\_brand 0.730407

11 loan\_purpose 0.725097

12 Gender 0.582358

13 source\_funds 0.494160

14 dependentsCount 0.461656

15 loan\_docType 0.412309

A graph of a company

Description automatically generated with medium confidence

A graph with red and white text

Description automatically generated

Model successfully saved to: /home/jupyter/Models/Income\_Estimation\_Alpha/model\_Good\_Customer\_Top16SHAPFeatures\_Alpha\_Step\_1\_0.joblib

Metrics structure: (4035.3578824320603, 144156524.471493, 12006.520081667835, 0.7353705939899805, 0.11242919061541877)