

# **Home Credit - Credit Risk Model Stability with XGBoost**

**20211531 나호영**

# Introduction

데이터를 사용하여 대출 상환 가능성을 예측하는 모델을 개발하고,  
이를 통해 금융 기관이 보다 정확한 신용 평가를 할 수 있도록 돕는 것을 목적으로 한다

# Process



# Data PreProcessing

- 1. Data type conversion**
- 2. Date handling**
- 3. Remove unnecessary heat**
- 4. Feature engineering**
- 5. Measure and optimize memory usage**

# Model Training

## XGBoost with StratifiedGroupKFold

## Model Parameter Setting

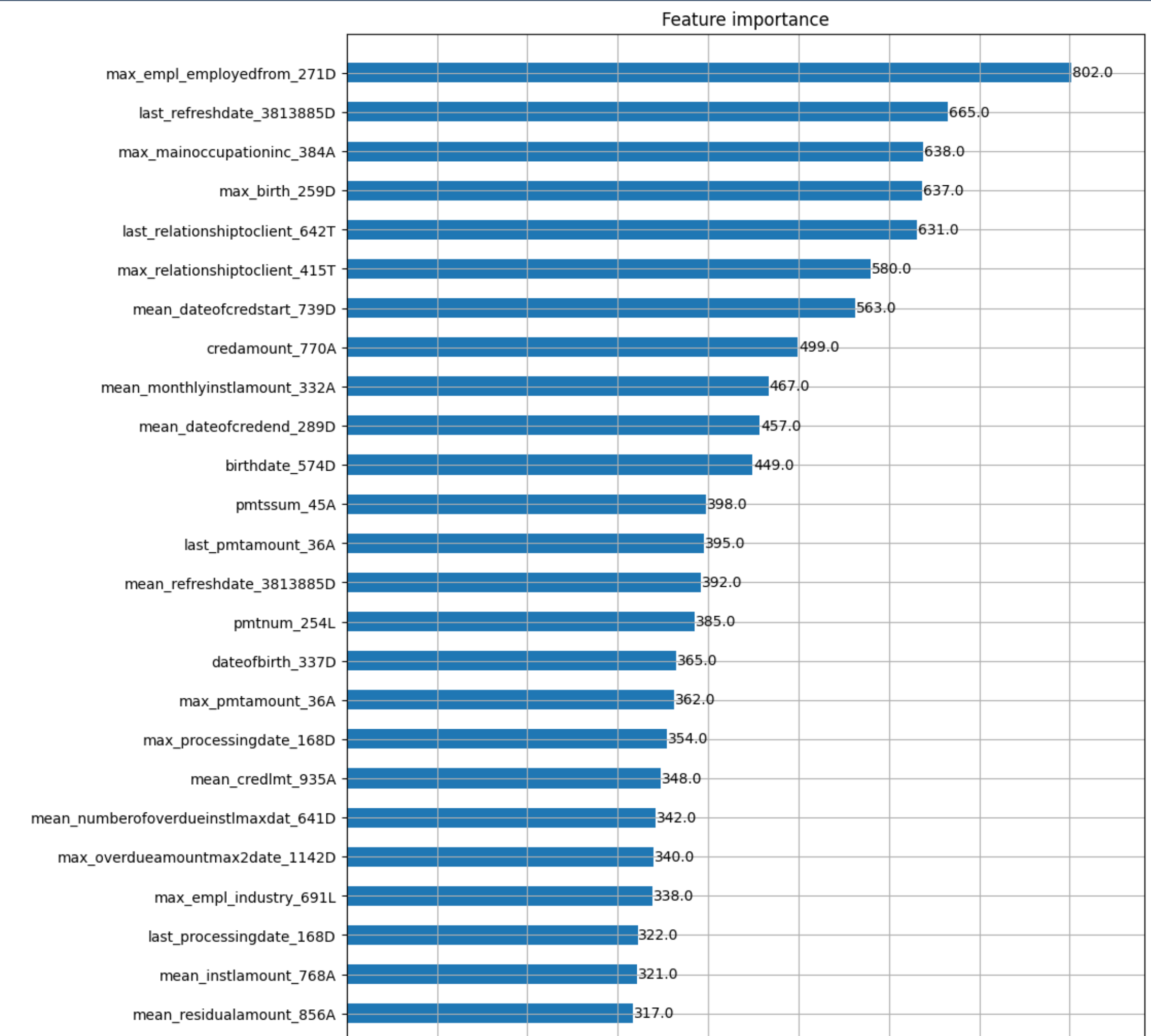
## 5-fold Cross-Validation

[0]	train-auc:0.70237	valid-auc:0.64751
[181]	train-auc:0.98820	valid-auc:0.73343
[0]	train-auc:0.70782	valid-auc:0.67067
[200]	train-auc:0.98912	valid-auc:0.74143
[213]	train-auc:0.98998	valid-auc:0.74031
[0]	train-auc:0.67743	valid-auc:0.67698
[197]	train-auc:0.98654	valid-auc:0.77845
[0]	train-auc:0.68986	valid-auc:0.62665
[187]	train-auc:0.98509	valid-auc:0.74116
[0]	train-auc:0.72283	valid-auc:0.63371
[200]	train-auc:0.98908	valid-auc:0.72326
[202]	train-auc:0.98922	valid-auc:0.72340

CV AUC scores: [0.7397231066556114, 0.7456849007133438, 0.7854618401478609, 0.7417501725327813, 0.7369001802359127]  
Maximum CV AUC score: 0.7854618401478609

# Model Training - Feature Importance

## Feature Importance visualization with plot\_importance



	features	importance
0	max_empl_employedfrom_271D	802.0
1	last_refreshdate_3813885D	665.0
2	max_mainoccupationinc_384A	638.0
3	max_birth_259D	637.0
4	last_relationshiptoclient_642T	631.0
..	...	...
292	pctinstlsallpaidlate4d_3546849L	1.0
293	last_sex_738L	1.0
294	clientscnt_493L	1.0
295	last_empls_economicalst_849M	1.0
296	last_empls_employer_name_740M	1.0
[297 rows x 2 columns]		
Number of features which are not important: 173		

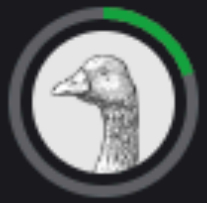
# Result

case_id	score
57543	0.010575
57549	0.026809
57551	0.009973
57552	0.017614
57569	0.046458
57630	0.018087
57631	0.034438
57632	0.023390
57633	0.024913
57634	0.039544

2645

▲ 96

nhy dev



0.46087

# Discussion

**XGBoost vs LightGBM**

**why LightGBM is better ?**



**Thank you**