MG Korean Federation of Community Credit Cooperatives

KWANGWOON UNIVERSITY

AI Credit Scoring System For University Students

Advisor : Prof . YoungJoo Shin / Team : PassAbility / Members : HanByeol Jang
YoungSang Kwon
YongRak Choi

Overview

- Needs to check credit score automatically
- For university students who can't be checked official credit score because of the loss of details.

SW Architecture

- Machine learning for learning model of financial data with nonlinear characteristics affected by external environment variables.
- Credit scoring with RandomForest, LighGBM and XGBoost that are classification algorithms
  - After that, cross validation for each models and ensemble them to improve the accuracy

IDE & Editor

jupyter ANACONDA — For several machine learning algorithms

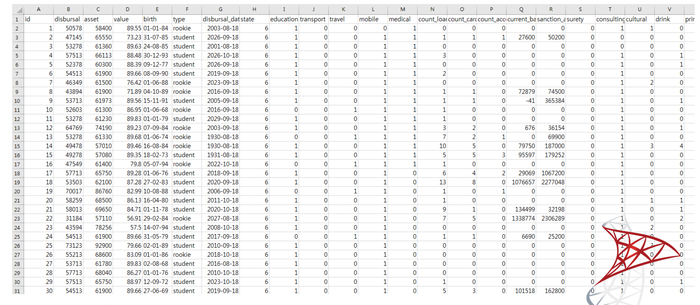
RStudio — Data Preprocessing & Data Augmentation

+ableau — Data Visualization & UI Design

Data Composition

kaggle — To get the datasets of Economics and Finance

kakaobank — For Categorization



After that, Data Preprocessing & Data Augmentation in R

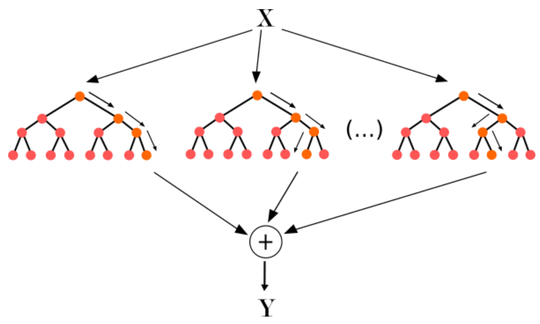
Microsoft SQL Server — Database

Random Forest

- The disadvantage of the decision tree is large variation in performance, which can be overcome by using randomization technologies such as RandomForest Algorithm's Bagging or Randomized node optimization
- Validation methods : Cross validation, Oob(Out of Bag) validation

Parameter	Value
n_estimators	300
max_features	auto
random_state	84
oob_score	True

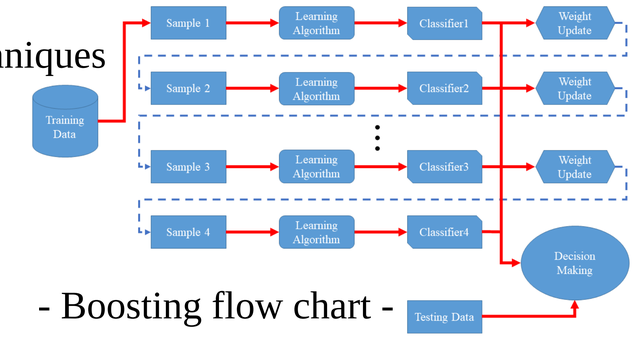
- Parameter values -



- Tree Ensemble process -

Gradient Boosting

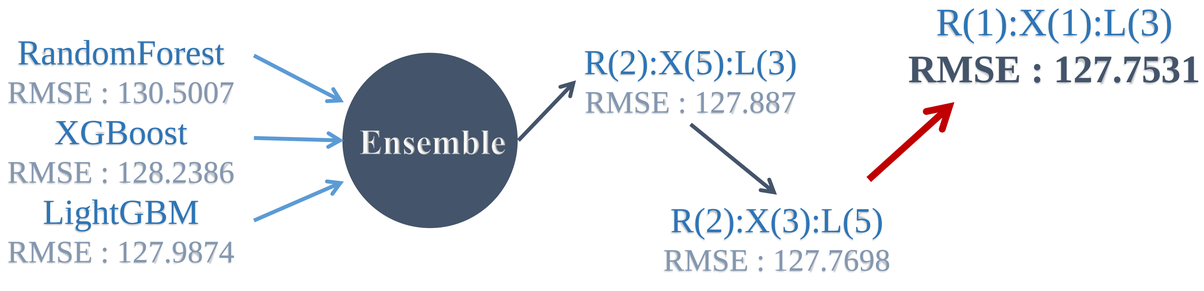
- Unlike Random Forest, which scores as an ensemble among independent trees, the Gradient Boosting complements the residual of the previous classifier to create a strong classifier
- Gradient Boosting used in model implementation
  - LightGBM, XGBoost
- XGBoost
  - Using parallel processing techniques in decision tree configuration
    - ✓ Increase performance
- Performance comparison
  - LightGBM > XGBoost



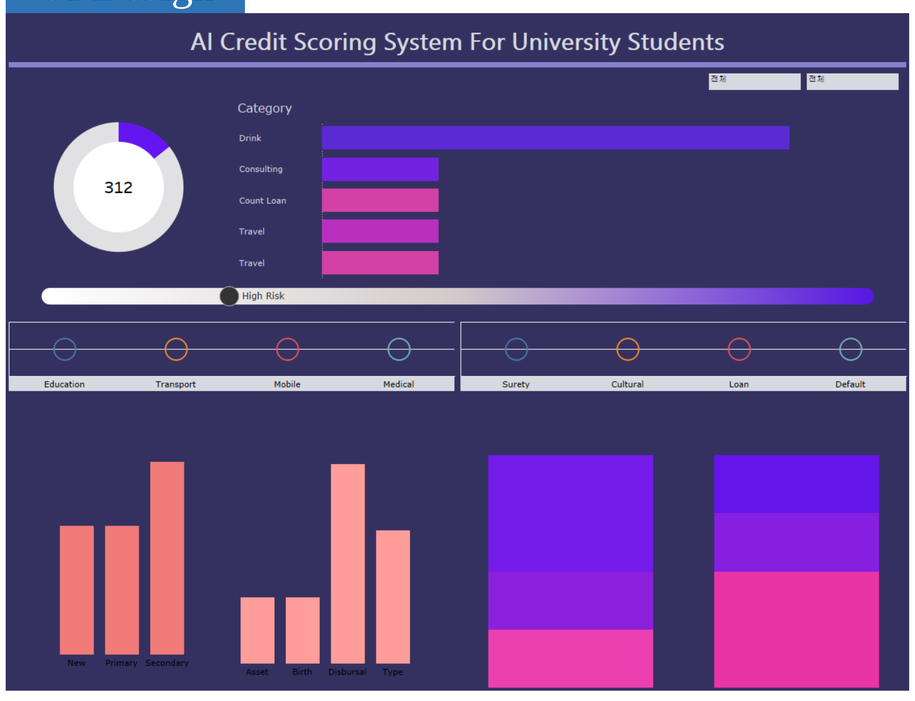
- Boosting flow chart -

Results of Ensemble

- Ensemble Random Forest, XGBoost, LightGBM to derive results
- Accuracy indicator criteria
  RMSE (Root Mean Squared Error)
- Ensemble
  That combine several Machine Learning techniques in order to improve Machine Learning results



UI Design



Conclusion

- LightGBM is faster compared to any other method. Besides, better results. And more exactly after ensemble
- Utilization Plans
  - Using the AI Credit Scoring System, we can simulate the credit evaluation of university students in the bank and classify them to some degrees
  - As a result, It makes more efficient to attract and manage potential customers who can't be checked credit score like university students