

GRADIENT BOOSTING (XGBOOST) YORDAMIDA KREDIT RISKINI BAHOLASH

Farg'ona davlat texnika universiteti (FDTU)

Pozilov Abdulaziz Pahlavonjon o'g'li

e-mail: abdulazizpozilov06@gmail.com

telefon nomer: +998 94 937 38 11

Annotatsiya: Ushbu maqolada kredit riskini baholash uchun Gradient Boosting va uning samaradorligi oshirilgan turi — XGBoost algoritmi qo'llanilishi tahlil qilinadi. Maqolada kredit riskining mohiyati, ma'lumotlarni tayyorlash bosqichlari, modelni qurish va baholash jarayoni bosqichma-bosqich ko'rsatildi. Bu yondashuv bank tizimlarida qarz berish qarorlarini avtomatlashtirish, moliyaviy xavflarni kamaytirish va kredit portfellarini optimallashtirishda muhim ahamiyat kasb etadi. Eksperimental tavsif va baholash metrikalari orqali XGBoost modelining an'anaviy usullarga nisbatan afzalliklari yoritiladi.

Kalit so'zlar: XGBoost, Prognozlash, Mashinali o'qitish, AUC-ROC, Matrix

Abstract: This article analyzes the application of Gradient Boosting and its enhanced version, the XGBoost algorithm, for credit risk assessment. It discusses the essence of credit risk, data preparation stages, and the step-by-step process of model building and evaluation. The proposed approach is crucial for automating lending decisions in banking systems, reducing financial risks, and optimizing credit portfolios. Experimental results and evaluation metrics highlight the advantages of the XGBoost model compared to traditional methods.

Keywords: XGBoost , Prediction / Forecasting , Machine Learning , AUC-ROC , Confusion Matrix

Аннотация: В данной статье анализируется применение алгоритма Gradient Boosting и его усовершенствованной версии — XGBoost для оценки кредитного риска. Рассматриваются суть кредитного риска, этапы подготовки данных, а также пошаговый процесс построения и оценки модели. Предлагаемый подход имеет важное значение для автоматизации решений по кредитованию в банковских системах, снижения финансовых рисков и оптимизации кредитных портфелей. Экспериментальные результаты и метрики оценки демонстрируют преимущества модели XGBoost по сравнению с традиционными методами.

Ключевые слова: XGBoost , Прогнозирование , Машинное обучение, AUC-ROC (Площадь под ROC-кривой) , Матрица ошибок

Kirish: Banklar va moliyaviy tashkilotlar kreditlashda qaror qabul qilish uchun mijozlar haqidagi katta miqdordagi ma'lumotlarni qayta ishlaydi. Kredit riskini noto'g'ri baholash banklarga katta moliyaviy zarar keltirishi mumkin. An'anaviy ball tizimlari yoki ekspert bahosi ko'pincha sub'ektiv bo'lib, murakkab chiziqsiz munosabatlarni ushlab qolmaydi. Shu sababli mashinaviy o'rganish metodlari, xususan Gradient Boosting va XGBoost, kredit riskini bashoratlashda keng foydalanilmoqda.

Metodologiya

Mijozning demografik ma'lumotlari, moliyaviy ko'rsatkichlar, to'lov tarixlari va kredit tarixiga oid atributlar.

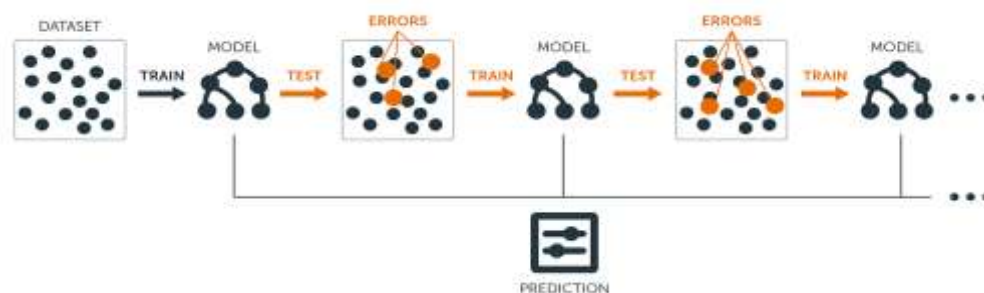
Yo'qolgan qiymatlarni to'ldirish, kategorik o'zgaruvchilarni kodlash, shkalalash, outlierlarni tekshirish.

Baholash metrikalari AUC-ROC, Precision, Recall, F1-score, Accuracy, Confusion Matrix.

Model

Asosiy model — XGBoost (Gradient Boosted Decision Trees).

XGBoost (Extreme Gradient Boosting) — bu Gradient Boosting algoritmining tezkor va optimallashtirilgan versiyasidir. U qaror daraxtlarini ketma-ket qurib, har bir keyingi daraxt orqali avvalgi xatoliklarni kamaytiradi va model aniqligini oshiradi. XGBoost katta hajmdagi ma'lumotlar bilan samarali ishlaydi, parallel hisoblashni qo'llab-quvvatlaydi va xotiradan oqilona foydalanadi. Model overfitting xavfini kamaytirish uchun L1 va L2 regulyarizatsiyaga ega. Shuningdek, yo'q (missing) qiymatlarni avtomatik qayta ishlaydi hamda atributlarning prognozga ta'sir darajasini aniqlash imkonini beradi, bu kredit riskini tahlil qilishda muhim.



1-Rasm “XGBoost. Ekstremal Gradient Kuchaytirish Algoritmi”

1-qadam: Ma'lumotlar to'plami trening va test to'plamlariga bo'linganidan so'ng, trening to'plamidan tasodifiy tanlangan ma'lumotlar bilan (**model1**) quriladi.

2-qadam: Keyin, model butun trening to'plamida sinovdan o'tkaziladi.

3-qadam: Model sinovdan o'tgach, ba'zi trening ma'lumotlari nuqtalari yaxshi bashorat qilinmaganini aniqlaymiz va bu nuqtalar sezilarli xatolikka olib kelishi mumkin.

4-qadam: Shuning uchun, yana bir model (model2) quriladi, bunda trening to‘plamidan tasodifiy tanlangan nuqtalar ishlatiladi, lekin oldingi modelda katta xatoga sabab bo‘lgan nuqtalarning tanlanish ehtimoli yuqori bo‘ladi.

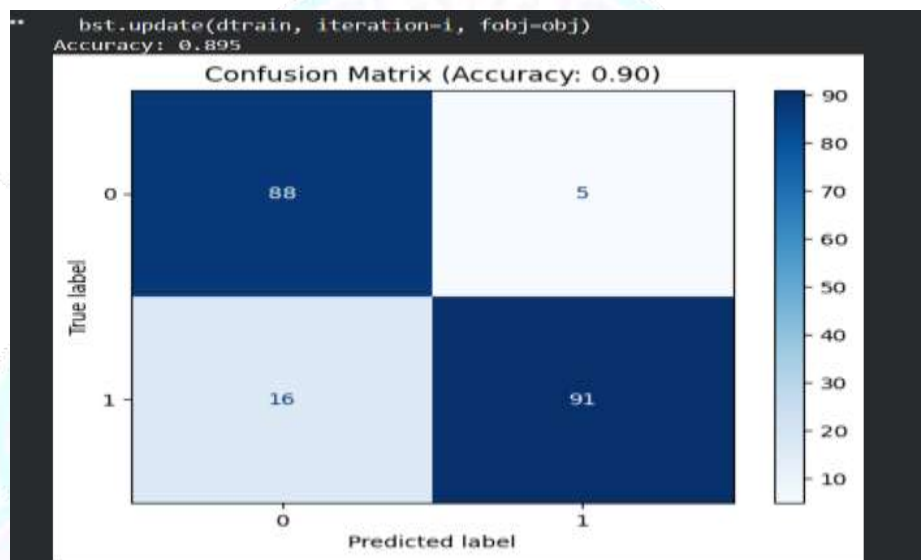
5-qadam: Keyinchalik, ikkala model (model1 va model2) butun trening to‘plamida sinovdan o‘tkaziladi va ularning natijalari birlashtiriladi. Butun ma’lumotlar bo‘yicha xato o‘lchanadi. Model2 ning o‘rtacha kvadrat xatosi (Mean Squared Error) model1 ga nisbatan pastroq bo‘ladi. Keyingi modelni qurishda xatoga sabab bo‘lgan nuqtalar tanlanadi va jarayon davom etadi, to‘xtashga hech qanday yaxshilanish qolmaguncha. Quydagi dastur misol qilib ko‘rib chiqsak. Bu dastur XGBoost modelini yaratadi, uni o‘qitadi va test datasida bashorat qilish orqali modelning aniqligini (accuracy) va Confusion Matrix’ini chiqaradi.

Dastur kodi:

```
from xgboost import XGBClassifier
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay,
accuracy_score
import matplotlib.pyplot as plt
# 1. Ma'lumotlar yaratish va bo'lish
X, y = make_classification(n_samples=1000, n_features=20, n_classes=2,
random_state=42)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
# 2. XGBoost modeli
model = XGBClassifier(
    max_depth=3,
    learning_rate=0.1,
    n_estimators=100,
    objective='binary:logistic',
    eval_metric='logloss',
    use_label_encoder=False )
# 3. Modelni o'qitish va bashorat qilish
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
# 4. Accuracy
acc = accuracy_score(y_test, y_pred)
print("Accuracy:", acc)
# 5. Confusion Matrix chizish
cm = confusion_matrix(y_test, y_pred)
```

```
disp = ConfusionMatrixDisplay(confusion_matrix=cm,
display_labels=model.classes_)
disp.plot(cmap=plt.cm.Blues)
plt.title(f"Confusion Matrix (Accuracy: {acc:.2f})") plt.show()
```

NATIJA



XGBoost modelini ishlash prinsipi va natijasi

XGBoost yordamida aniqlikni baholashda turli metrikalardan foydalaniladi, jumladan: AUC-ROC, Precision.

AUC-ROC (Area Under the Curve — Receiver Operating Characteristic) — bu klassifikatsiya modelining sezgirlik va xatolikni balanslash qobiliyatini baholovchi ko'rsatkich. Kredit riskini baholashda AUC-ROC qanchalik yuqori bo'lsa, model riskli va risksiz mijozlarni to'g'ri ajrata oladi.

Natijalar va baholash

Natijalar ROC curve, Confusion Matrix, va Feature Importance grafiklari yordamida baholandi. Model AUC-ROC natijasida yuqori aniqlikka erishgan. Shuningdek, SHAP tahlili yordamida o'zgaruvchilarning modelga ta'siri aniqlangan.

Xulosa

XGBoost algoritmi yuqori aniqlik **yordamida kredit riskini** va modelning samaradorligini baholash uchun ishlatiladigan asosiy metrikalar tahlil qilindi. Kelajakda SHAP tahlili va fairness metrikalarini qo'shish orqali yanada adolatli tizimlar yaratish mumkin.

Foydalanilgan adabiyotlar

1. **XGBoost: A Scalable Tree Boosting System.** *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*
2. **Master Machine Learning Algorithms.** Machine Learning Mastery.
3. Hastie, T. **The Elements of Statistical Learning.** Springer.