УДК 004.89.

Гиперпараметры градиентного бустинга с использованием байесовской оптимизации для оценки кредитного риска заемщиков

Яковенко Я.В.

КФ МГТУ им. Н.Э. Баумана

г.Калуга, Российская Федерация

Борсук Н.А.

КФ МГТУ им. Н.Э. Баумана

г.Калуга, Российская Федерация

Аннотация

Оценка кредитного риска относится к задачам, в которых ошибка при классификации заемщика может приводить к финансовым потерям. Одним из применяемых подходов является использование моделей градиентного бустинга, однако качество таких моделей существенно зависит от выбранных гиперпараметров. В работе рассматривается настройка модели XGBoost с использованием байесовской оптимизации.

Ключевые слова

кредитный скоринг; оценка риска заемщика; машинное обучение; градиентный бустинг; байесовская оптимизация; настройка гиперпараметров.

Введение

Современный финансовый сектор работает с большими объемами данных, которые используются при принятии решений о кредитовании[1]. Банки обрабатывают информацию о заемщиках, и качество прогнозирования их платежеспособности напрямую связано с возможными финансовыми потерями. Поэтому при рассмотрении кредитных заявок используются различные методы оценки риска, позволяющие заранее выявлять заемщиков с высокой вероятностью невозврата кредита.

Логистическая регрессия долгое время применялась в кредитном скоринге благодаря своей интерпретируемости, однако ее линейность ограничивает возможность выявления нелинейных зависимостей между признаками[2].

Поэтому для решения подобных задач также используются ансамблевые методы градиентного бустинга, в том числе XGBoost, LightGBM и CatBoost, которые позволяют учитывать более сложные зависимости в исходных данных

В отличие от оценки модели только по Accuracy и ROC-AUC[3-4], дополнительно используются Recall, F1-score и MCC. Это позволяет оценить качество определения заемщиков с повышенным риском при несбалансированных данных.

Материалы и методы

Работа решает задачу бинарной классификации кредитного риска заемщиков. Использовано открытое хранилище данных, содержащее сведения о 1000 заемщиках и 20 признаках (кредитная история, доход, размер и срок кредита, занятость, возраст и др.). Целевой переменной является кредитный статус заемщика, принимающий два значения: добросовестный заемщик и дефолт.

В качестве модели машинного обучения использован алгоритм градиентного бустинга XGBoost.

Результаты

Был проведен эксперимент по оценке влияния байесовской оптимизации гиперпараметров на качество модели градиентного бустинга в задаче кредитного риска. Сравнивались базовый XGBoost со стандартными параметрами и модель после байесовской оптимизации (50 итераций; целевая функция – максимум AUC ROC по пятиблочной стратифицированной кросс-валидации).

Для оценки эффективности предложенного подхода было выполнено сравнение базовой модели и модели после оптимизации.

Таблица 1 – Сравнение качества моделей

Метрика	Базовая модель	После оптимизации	Изменение, %

После оптимизации значения всех рассмотренных метрик оказались выше, чем у базовой модели XGBoost.

Обсуждение результатов

После подбора параметров модель показала лучшие результаты по сравнению с исходным XGBoost. Больше всего изменились значения Recall, F1-score, ROC-AUC и коэффициента Мэттьюса.

Заключение

В работе рассмотрена настройка модели XGBoost с использованием байесовской оптимизации гиперпараметров для оценки кредитного риска заемщиков. Были сравнены результаты базовой модели и модели после подбора параметров. В результате после оптимизации увеличились значения Accuracy, Precision, Recall, F1-score, ROC-AUC и MCC. Наиболее заметное изменение получено для MCC и Recall. Это показывает, что настройка гиперпараметров в проведенном эксперименте повлияла не только на общую точность модели, но и на определение заемщиков с повышенным риском.

Список использованной литературы

брагимов Б. Л., Гусев Г. Г. Оптимизация стохастического градиентного бустинга с помощью out-of-sample оценок качества // Труды Московского физико-технического института. – 2024. – Т. 16, № 3. – С. 49–60. – URL: https://mipt.ru/upload/Труды%20мфти/63/05.pdf (дата обращения: 27.07.2026).

ценка исправности блока усилителя мощности радиолокационной аппаратуры методами машинного обучения // КиберЛенинка : научная электронная библиотека. – URL: clck.ru/3VqtjC (дата обращения: 27.07.2026).

опоставление библиотек для создания моделей машинного обучения на основе методов градиентного бустинга // КиберЛенинка : научная электронная библиотека. – URL: https://lix.li/xRYj6p (дата обращения: 27.07.2026).

нсамблирование безэталонных мер качества изображений для повышения

корреляции с экспертными оценками // КиберЛенинка : научная электронная библиотека. – URL: https://lix.li/EOVKVU (дата обращения: 27.07.2026).