
УДК 004.021

СРАВНЕНИЕ ОЦЕНОК КАЧЕСТВА МОДЕЛЕЙ, ОСНОВАННЫХ НА АНСАМБЛЕВОМ МЕТААЛГОРИТМЕ МАШИННОГО ОБУЧЕНИЯ – БУСТИНГ

Безрукова Д.Е., студентка 5 курса факультета безопасности
Научный руководитель – Катаева Е.С., старший преподаватель
ФГБОУ ВО ТУСУР

Ключевые слова: машинное обучение, прогнозирование, ансамблевый метод, бустинг, лесные пожары, оценка эффективности моделей.

Работа посвящена сравнительному анализу качества моделей, основанных на ансамблевом метаалгоритме машинного обучения – Бустинг: XGBoost, LightGBM, CatBoost, AdaBoost на базовом оценщике «Дерева решений».

Введение. На сегодняшний день основной из главных научных задач является получение достоверных предсказаний о поведении сложной системы в будущем на основе предыдущего опыта системы [1]. В этом и выражается сущность применения методов машинного обучения (МО). Также особенно остро обостряется экологическая проблема в виде лесных пожаров. Достоверное предсказание с помощью МО минимизирует угрозу окружающей среде, экономике, людям, позволит противопожарным структурам оперативно принять меры по предотвращению очагов горения и спланировать эффективно работу.

Цель работы. Целью данной работы является проведение сравнительного анализа оценок качества моделей, основанных на ансамблевом метаалгоритме машинного обучения – Бустинг, для выявления наилучшего алгоритма для задачи прогнозирования.

Результаты исследования. В ходе исследования был проанализирован набор данных [2] о лесных пожарах в природном парке Монтесиньо. Ранее было подтверждено, что ансамблевые методы МО, действительно, являются более мощным инструментом для

прогнозирования по сравнению с классическими методами МО. По анализируемому набору данных с применением МО решались две задачи: задача классификации, задача регрессии.

Реализация методов МО была выполнена на языке Python. Для реализации МО набор данных был разделён с помощью модуля `train_test_split` в следующем соотношении: обучение – 80 %, тестирование – 20%, также был реализован метод перекрестной проверки со значением $N = 5$.

Бустинг — это ансамблевый метаалгоритм машинного обучения, идея которого заключается в том, что несколько однородных моделей объединяются для создания более сильной модели. Причем обучение модели происходит последовательно, исправляя ошибки друг друга. В работе были рассмотрены такие алгоритмы как: XGBoost, LightGBM, CatBoost, AdaBoost на базовом оценщике «Деревья решений».

В ходе исследования были выявлены следующие особенности реализации алгоритмов Бустинга:

1. Высокая точность предсказания, простота в реализации.
2. Эффективное использование памяти и использование распределённых расчетов для улучшения скорости обучения.
3. Использование для решения задач и классификации, и регрессии.
4. Гибкость, так как позволяет использовать различные алгоритмы машинного обучения.
5. Хорошее использование на несбалансированных данных.

На рисунках 1 и 2 представлены графики сравнения оценок качества моделей Бустинг для решения задач классификации и регрессии.

При решении задачи регрессии (Рис. 2) у модели AdaBoost были выявлены достаточно высокие показатели ошибок и низкий показатель коэффициента детерминации, что свидетельствует о наименьшей точности прогнозирования площадей лесных пожаров в непрерывных значениях.



Рис. 1 – Сравнение оценок качества моделей Бустинг для задачи классификации

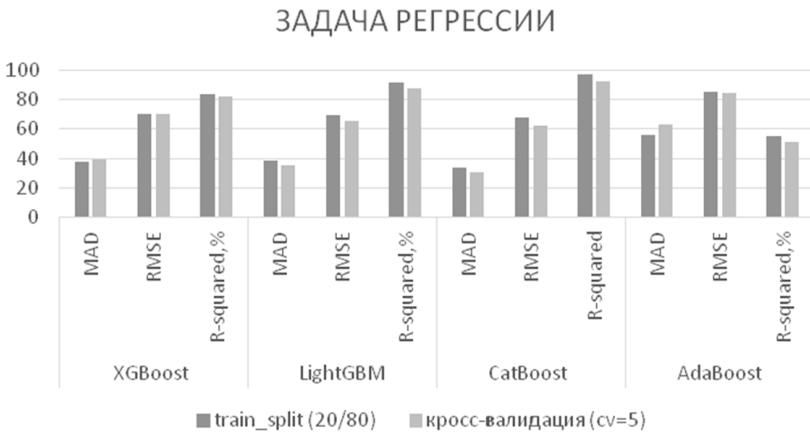


Рис. 2 – Сравнение оценок качества моделей Бустинг для задачи регрессии

Закключение. При решении задачи классификации алгоритм LightGBM показал наилучшие результаты предсказания лесных пожаров по метеорологическим параметрам, а при решении задачи регрессии наилучшие показатели были выявлены у алгоритма CatBoost. Стоит отметить также, что LightGBM имеет улучшенные средства для

обработки категориальных признаков. А CatBoost является более устойчивым по сравнению с LightGBM, что делает его более предпочтительным в данных, содержащих шум.

Таким образом, использование моделей, основанных на ансамблевом метаалгоритме машинного обучения – Бустинг, гарантирует создание более точных и производительных моделей. Результаты исследования в дальнейшем будут использованы при разработки автоматизированной системы предотвращения лесных пожаров на основе метеорологических показателей с использованием ансамблевых методов машинного обучения и глубоких нейронных сетей.

Библиографический список:

1. Математическая модель оптимизации прибытия пожарного подразделения с использованием информационных систем мониторинга транспортной логистики города Воронежа. [Электронный ресурс]. – URL: <https://cyberleninka.ru/article/n/matematicheskaya-model-optimizatsii-pribytiya-pozharnogo-podrazdeleniya-s-ispolzovaniem-informatsionnyh-sistem-monitoringa> (дата обращения: 09.01.2023). – Текст: электронный.
2. Forest Fires Data Set. [Электронный ресурс]. – URL: <https://www.kaggle.com/datasets/elikplim/forest-fires-data-set> (дата обращения: 05.12.2022). – Текст: электронный.

COMPARISON OF QUALITY ESTIMATES OF MODELS BASED ON THE ENSEMBLE META-ALGORITHM OF MACHINE LEARNING – BOOSTING

Bezrukova D.E

Keywords: *machine learning, forecasting, ensemble method, boosting, forest fires, evaluation of the effectiveness of models.*

The work is devoted to a comparative analysis of the quality of models based on the ensemble meta-algorithm of machine learning – Boosting: XGBoost, LightGBM, CatBoost, AdaBoost on the basic estimator «Decision Trees».