

Классификация агрокультур по NDVI

Решение задачи на чемпионате в Иннополисе
в рамках «Цифрового прорыва — 2022».

Автор — Павел Сёмин

Задача

- Что есть: около 5 тысяч полей, для каждого из них — медианные значения NDVI (вегетационный индекс) за 70 дат с апреля по август, площадь и контур поля, и культура, растущая на нём.
- Что нужно: научиться предсказывать культуру, опираясь на NDVI и доступные геоданные.

Решение

- **LightGBM Classifier** с параметрами, подобранными с помощью *grid search*.
- Набор признаков:
 - площадь поля;
 - все значения NDVI;
 - широта, долгота поля, число углов в контуре поля;
 - описательные статистики временного ряда NDVI (сумма, среднее, число пропусков и некоторые другие);
 - Canonical Time-series Characteristics (Catch22) временного ряда NDVI (по методу из библиотеки sktime).
- Значение *recall* на тренировочных данных по итогам кроссвалидации — более **0.97**.

Альтернативы

- Decision tree, logistic regression, ridge classifier, SVM, KNN — хуже (иногда значительно).
- Random Forest, CatBoost, XGBoost, Multi-layer perceptron — незначительно хуже (полнота меньше на 0.01–0.02).

Итог

- Обученный классификатор LightGBM распознает культуры по временному ряду NDVI и геоданным (контуру и площади поля). Полнота классификации — более 97%.
- Классификаторы CatBoost, XGBoost, Random Forest могут быть использованы в качестве альтернативы.