

Прогнозирование урожайности подсолнечника

по гидрометеорологической информации

Олег Воропаев



Цель проекта

Разработка предиктивной модели для прогнозирования урожая подсолнечника с помощью ML

Целевая переменная

Средняя урожайность подсолнечника по муниципальным районам.

Метрика качества

Root mean square error (RMSE).

(целевое значение $RMSE \leq 10\%$)

Типы данных

1

Гидрометеорологическая информация

[Сайт расписания погоды gr5.ru](http://gr5.ru)

Набор данных содержит 30 параметров с гидрометеорологической информацией.

2

Данные об урожайности подсолнечника

[Официальный сайт Росстат](http://rosstat.gov.ru)

Набор данных содержит информацию об урожайности сельскохозяйственных культур по муниципальным районам с 2007 года, всего 11 параметров.

3

Информация о населенных пунктах

[Сайт платформы ИНИД](http://nnp.ru)

Набор данных содержит информацию о населенных пунктах РФ (гео-координаты, административно-территориальная классификация, численность и др.), всего 18 параметров.

Этапы работы над проектом

1

Разработка web-скраперов для автоматизации процесса сбора входных данных.

2

Создание базы данных для хранения собранной информации на платформе PostgreSQL.

3

Разведочный анализ, предобработка, очистка и подготовка данных для обучения и тестирования модели

4

Разработка предиктивной модели для прогнозирования урожайности подсолнечника

Библиотеки

Модели градиентного бустинга

CatBoost Regressor
XGB Regressor

Модель линейной регрессии

Sklearn Linear Regression

Нейронные сети

LSTM Bidirectional
Полносвязная нейронная сеть

Результаты прогнозирования урожайности подсолнечника



Параметр	Значение	В % к средней урожайности
Средняя урожайность, ц/га	16,2	—
$RMSE_{wg}$	1,458	9,0%
CatBoost	1,498	9,3%
XGBoost	1,601	9,9%

