

УДК:517.534

## ПРОГНОЗИРОВАНИЕ ЗАДАЧ СЕЛЬСКОГО ХОЗЯЙСТВА С ПРИМЕНЕНИЕМ МАШИННОГО ОБУЧЕНИЯ

Сабитов<sup>1</sup> Б.Р., Картанова<sup>2</sup> А., Орозобекова<sup>2</sup> А.К., Шеримбекова<sup>2</sup> Э.Б., Турдубаева<sup>3</sup> А.Б., Тенирбеди<sup>2</sup> у И, Алмабекова<sup>1</sup> Н.А., Джунушалиева<sup>1</sup> А.Д., Цю Ифен<sup>1</sup>, Чжан Цайцзюй<sup>1</sup>

<sup>1</sup>КНУ им. Ж.Баласагына, <sup>2</sup>КТУ им. И.Раззакова, <sup>3</sup>КРСУ им.Б.Ельцина

В данной статье визуализируются результаты прогнозирования с применением продвинутых алгоритмов машинного обучения. Получены оценки точности моделей с помощью методов машинного обучения. Мы изучили процесс прогнозирования для расширения возможностей метода RF для прогнозирования урожайности кукурузы с помощью других наиболее продвинутых алгоритмов машинного обучения.

**Ключевые слова:** методы машинного обучения, моделирование урожайности, база данных, случайный лес, оценка точности модели.

## МАШИНА МЕНЕН ҮЙРӨНҮҮ МЕНЕН АЙЫЛ ЧАРБА МИЛДЕТТЕРИН БИЛҮҮ

Сабитов<sup>1</sup> Б.Р., Картанова<sup>2</sup> А., Орозобекова<sup>2</sup> А.К., Шеримбекова<sup>2</sup> Э.Б., Турдубаева<sup>3</sup> А.Б., Тенирбеди<sup>2</sup> у И, Алмабекова<sup>1</sup> Н.А., Джунушалиева<sup>1</sup> А.Д., Цю Ифен<sup>1</sup>, Чжан Цайцзюй<sup>1</sup>

Ж. Баласагына ат. <sup>2</sup>КМУ, И.Раззаков ат. <sup>1</sup>КМТУ, Б.Ельцин ат. <sup>3</sup>КРСУ

Бул макалада өнүккөн машина үйрөнүү алгоритмдерин колдонуу менен болжолдоо натыйжалары визуализацияланат. Моделдин тактыгына баа берүү машиналарды үйрөнүү ыкмаларын колдонуу менен алынат. Биз башка эң алдыңкы машина үйрөнүү алгоритмдери менен жүгөрүнүн түшүмдүүлүгүн болжолдоо үчүн RF методунун күчүн кеңейтүү үчүн болжолдоо процессин изилдедик.

**Баштапкы сөздөр:** машинаны үйрөнүү ыкмалары, түшүмдү моделдөө, маалымат базасы, кокус токой, моделдин тактыгын баалоо.

## FORECASTING AGRICULTURAL TASKS USING MACHINE LEARNING

Sabitov<sup>1</sup> B.R., Kartanova<sup>2</sup> A., Orozobekova<sup>2</sup> A.K., Sherimbekova<sup>2</sup> E.B., Turdubayeva<sup>3</sup> A.B., Tenirbedi<sup>2</sup> u I, Almabekova<sup>1</sup> N.A., Dzhunushalieva<sup>1</sup> A.D., Zu Ifen<sup>1</sup>, Chjan Zayzshuy<sup>1</sup>

<sup>1</sup>KNU named of J. Balasagyn, <sup>2</sup>KTU named of I. Razzakov,

<sup>3</sup>KRSU named of B.Yeltsin

This article visualizes the results of forecasting using advanced machine learning algorithms. Model accuracy estimates are obtained using machine learning methods. We have studied the forecasting process to extend the power of the RF method for corn yield forecasting with other most advanced machine learning algorithms.

**Keywords:** machine learning methods, yield modeling, database, random forest, model accuracy estimation.

В последние годы сельскохозяйственный сектор, в аграрных странах, включая Иссык-Кульский регион, сильно пострадал от недавних климатических изменений. Основной категорией служит урожайность культур. Известно, что погодные условия играют существенную роль в формировании урожайности. Например, ожидаемый дождь может не пойти в ожидаемое время из-за недавних изменений в характере осадков. Это исследование в основном направлено на решение проблем, возникших из-за сложности выявления несоответствия между текущим количеством осадков, почвы, влажности и температуры и традиционным графиком сезонного выращивания сельскохозяйственных культур. Даже с передовыми технологиями прогнозирование погоды стало сложной задачей.

В этой статье представлен подход машинного обучения, основанный на методе случайный лес, для прогнозирования урожайности точного земледелия. Для прогнозирования урожайности были выбраны коррели-

рующие характеристик урожайности как с расположением региона, минимальное и максимальное температура, влажность, характеристики почвы, которые были улучшены с помощью надлежащей предварительной обработки данных. 80-20 данных с перекрестной проверкой были первоначально обучены и протестированы с помощью алгоритма Random Forest, что достигла уровня точности 90. Результаты демонстрируют значительную точность подхода, основанного на методах ансамбля, для прогнозирования погоды в Иссык-Кульской области. Сельскохозяйственное производство в основном зависит от погодных условий. Оба аспекта внутренне связаны друг с другом в нескольких аспектах, поскольку изменение климата является ключевым фактором биотических и абиотических стрессов растений, оказывающих неблагоприятное влияние на мировое сельскохозяйственное производство. Сельскохозяйственные угодья подвержены влиянию Проблемной ситуацией является получение информации из необработанных наборов данных, что привело к расширению новых методов, таких как машинное обучение (МО), которые можно значительно использовать для интеграции информации с оценкой урожайности. Эта глава задумана как попытка представить моделирование урожайности с использованием методов машинного обучения в многомерных наборах данных.

Проблемной ситуацией является получение информации из необработанных наборов данных, что привело к расширению новых методов, таких как машинное обучение (МО), которые можно значительно использовать для интеграции информации с оценкой урожайности. Эта статья задумана как попытка представить моделирование урожайности с использованием методов машинного обучения в многомерных наборах данных.

Для построения моделей, использованы алгоритмы ансамблевого обучения для построения различных моделей. Перед построением и обучением модели нам нужно отделить зависимую переменную и разделить набор данных на набор для обучения и тестирования. В качестве базы

данных использовано реальная база данных с учетом временной особенности, региональный учет почвенных характеристик Иссык-кульского региона и погодные условия. Вот структура базы данных:

	Year	Location	Soil Type	Max_Temperature	Min_Temperature	Rainfall	Humidity	Yield Kg/ha
7083	2014	Ak_Suuiskey_peg	Black Soil	25.5	13.0	559.8	68.0	345.0
1891	2018	Tupsky_peg	Black Soil	21.5	13.0	543.8	50.0	2134.0
5644	2019	Issyk_Kulsky_peg	Black Soil	24.5	14.0	811.2	66.0	240.0
6918	2014	Tupsky_peg	Black Soil	26.5	13.0	559.8	74.0	2725.0
2173	2018	Jety_Ogyz_peg	Light Soil	31.0	11.0	543.8	73.0	2985.0
6300	2014	Ak_Suuiskey_peg	Black Soil	22.5	13.0	559.8	67.0	135.0
3643	2017	Jety_Ogyz_peg	Light Soil	31.0	14.0	612.0	63.0	1560.0
4854	2019	Issyk_Kulsky_peg	Black Soil	24.5	12.5	811.2	36.0	650.0
3490	2017	Tupsky_peg	Black Soil	30.5	14.5	612.0	43.0	713.0
6998	2014	Tupsky_peg	Black Soil	29.0	11.5	559.8	41.0	2025.0

Для построения модели с учетом особенностей регионов Иссык-кульской области исходная база данных без учета временных особенностей была преобразовано в следующую структуру.

	Location	Soil Type	Max_Temperature	Min_Temperature	Rainfall	Humidity	Yield Kg/ha
0	Issyk_Kulsky_peg	Black Soil	31.0	13.0	811.2	68.0	980.0
1	Issyk_Kulsky_peg	Black Soil	30.5	16.0	811.2	72.0	670.0
2	Issyk_Kulsky_peg	Black Soil	29.5	14.0	811.2	68.0	730.0
3	Issyk_Kulsky_peg	Black Soil	30.0	13.0	811.2	54.0	710.0
4	Issyk_Kulsky_peg	Black Soil	31.0	12.0	811.2	70.0	780.0

Для обучения модели выделим теперь прогнозируемую и основную обучающую часть базы данных. Вот ее структура. Обучающая множество базы данных.

	Location	Soil Type	Max_Temperature	Min_Temperature	Rainfall	Humidity
0	Issyk_Kulsky_peg	Black Soil	31.0	13.0	811.2	68.0
1	Issyk_Kulsky_peg	Black Soil	30.5	16.0	811.2	72.0
2	Issyk_Kulsky_peg	Black Soil	29.5	14.0	811.2	68.0
3	Issyk_Kulsky_peg	Black Soil	30.0	13.0	811.2	54.0
4	Issyk_Kulsky_peg	Black Soil	31.0	12.0	811.2	70.0

Ниже представлено

```
0      980.0
1      670.0
2      730.0
3      710.0
4      780.0
...
8072   1710.0
8073   2030.0
8074   1325.0
8075   1743.0
8076   1765.0
Name: Yield Kg/ha, Length: 8077, dtype: float64
```

---

Для дальнейшего изучения базы данных необходимо, создать фиктивные переменные для обучающего множества в следующем виде

	Max_Temperature	Min_Temperature	Rainfall	Humidity	Location_Issyk_Kulsky_peg	Location_Jety_Ogyz_peg	Location_Tupsky_peg	Soil Type_Light Soil
612	27.0	14.0	811.2	60.0	0	0	1	0
845	29.5	12.5	811.2	66.0	0	0	1	0
624	26.5	16.0	811.2	51.0	0	0	1	0
1640	30.0	14.5	543.8	52.0	0	1	0	1

Далее мы должны стандартизовать исходную базу и разделит ее на данные обучения и тестирования.

```
# Вывод формы нашего обучающего и тестового набора
print(X_train.shape)
print(X_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(6461, 8)
(1616, 8)
(6461,)
(1616,)
```

Как видно из данных всего обучающих данных, включая урожайность 6451 с 8 признаками. Для тестирования выделено 1616 тестовых данных.

Для анализа ниже приведен `y_test-y_pred` прогнозирование урожайности с помощью алгоритма случайный лес.

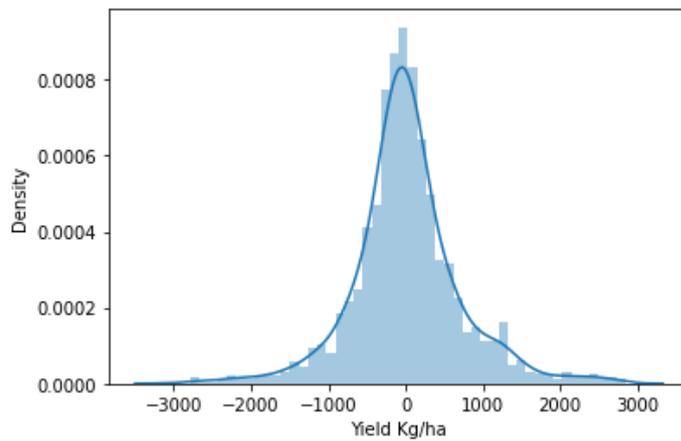


Рис.1. Расположения данных разности прогнозирования  $Y_{test} - Y_{Pred}$   
 Расположения данных тестирования  $Y_{test}$  и прогнозирования  $Y_{Pred}$  выглядит таким образом.

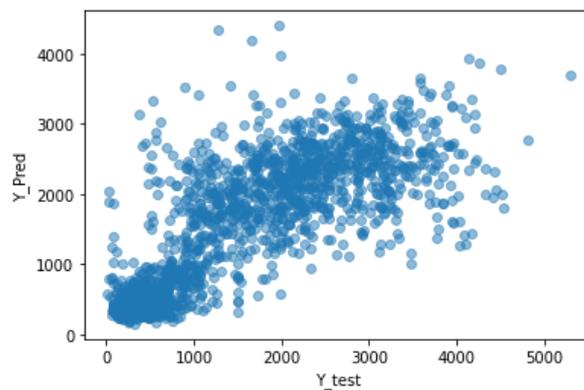


Рис.2. Расположения данных прогнозирования  $Y_{test}, Y_{Pred}$  на основе случайного леса

Вот характеристика  $R^2$ , являющиеся наиболее важным свойством модели.

```
metrics.r2_score(y_test,y_pred)
0.592475911991121
```

Из листингов выше мы видим, данные разности  $y_{test} - y_{pred}$  расположены по нормальному распределению и что тестовые и обучающие

данные густо расположены и лидером прогноза является алгоритм градиентного бустинга –xgboost с прогнозными значениями на тестовых и обучающих данных.

```
model.score(X_test,y_test)
```

```
0.6111430552452045
```

```
model.score(X_train,y_train)
```

```
0.8135386918155929
```

Изменения разности  $y_{test}-y_{pred}$  при этом выглядит следующим образом.

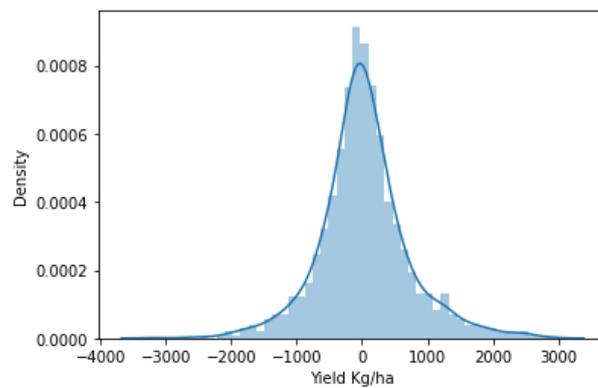


Рис.3. Расположения данных разности прогнозирования  $Y_{test}-Y_{Pred}$  для градиентного бустинга.

Расположения данных  $Y_{test}$  и  $Y_{Pred}$

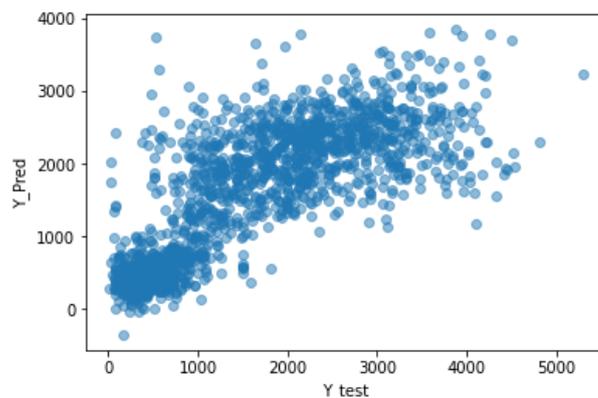


Рис.4. Расположения данных прогнозирования  $Y_{test}, Y_{Pred}$  с помощью алгоритма градиентного бустинга

Построение регрессионной модели на основе Адабусты –Ada Boost дали следующие результаты

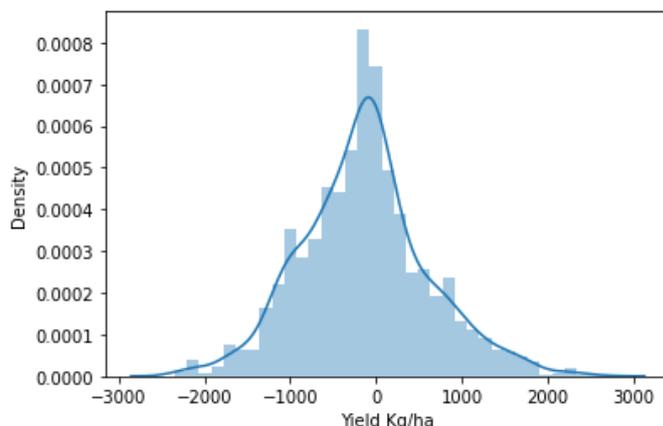


Рис.5. Расположения данных разности прогнозирования  $Y_{test}-Y_{Pred}$  с помощью алгоритма ада бустинга

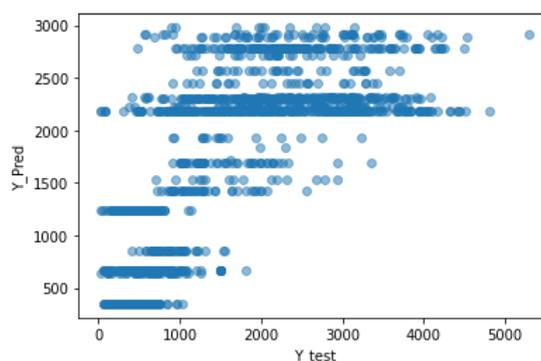


Рис.6. Расположения данных прогнозирования  $Y_{test}, Y_{Pred}$  с помощью алгоритма ада бустинга

**Заключение.** Случайный лес (RF) — это метод машинного обучения, который доказал свою высокую точность во многих сельскохозяйственных приложениях. Мы изучили процесс прогнозирования для расширения возможностей метода RF для прогнозирования урожайности кукурузы с помощью других наиболее продвинутых алгоритмов машинного обучения. В данной статье мы, также визуализировали результаты прогнозирования с применением продвинутых алгоритмов машинного обучения. Получены оценки точности моделей с помощью методов машинного обучения.

## ЛИТЕРАТУРА

1. Ali, A. M., Abouelghar, M. A., Belal, A.-A., Saleh, N., Younes, M., Selim, A., et al. (2022). Crop yield prediction using multi sensors remote sensing (review article). *Egypt. J. Remote Sens. Space Sci.* 25, 711–716. doi:10.1016/j.ejrs.2022.04.006
2. Archontoulis, S. V., Castellano, M. J., Licht, M. A., Nichols, V., Baum, M., Huber, I., et al. (2020). Predicting crop yields and soil-plant nitrogen dynamics in the US Corn Belt. *Crop Sci.* 60 (2), 721–738. doi:10.1002/csc2.20039
3. Arnault, J., Rummeler, T., Baur, F., Lerch, S., Wagner, S., Fersch, B., et al. (2018). Precipitation sensitivity to the uncertainty of terrestrial water flow in WRF-Hydro: An ensemble analysis for central Europe. *J. Hydrometeorol.* 19 (6), 1007–1025. doi:10.1175/jhm-d-17-0042.1