

УДК:517.534

## **ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОПТИМИЗАЦИИ ЗАДАЧ УРОЖАЙНОСТИ СЕЛЬСКОГО ХОЗЯЙСТВА В УСЛОВИЯХ ИСПОЛЬЗОВАНИЯ ПЕСТИЦИДОВ**

**<sup>1</sup>Сабитов Б.Р., <sup>2</sup>Кожокматов Т.М., <sup>2</sup>Шамырова Д.Р., <sup>3</sup>Турдубаева А.Б., <sup>1</sup>Джакипов А.Р., <sup>1</sup>Исабек уулу Б., <sup>1</sup>Кенжебек уулу А., <sup>1</sup>Ма-Тянь –Ю. Р.**

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

В данной работе изучается важная задача сельского хозяйства повышение качества урожайности с умеренным использованием пестицидов. Особенно важно учитывать массовое производство фермерами и сельхозпроизводителями продуктов сельского хозяйства. Одной из важных задач в данном случае является управления качеством производимой продукции. Изучается возможность применения алгоритмов машинного обучения к широкому кругу задач сельского хозяйства.

**Ключевые слова:** прогнозирование урожайности, искусственные нейронные сети, модель многослойного персептрона, линейная регрессия, метод опорных векторов.

## **ПЕСТИЦИДДЕРДИ КОЛДОНУУ АРТЫНДА АЙЫЛ ЧАРБА ТҮШҮМҮ МАКСАТТАРЫН ОПТИМАЛАШТЫРУУ ҮЧҮН МАШИНА ҮЙРӨНҮҮНҮ КОЛДОНУУ**

**<sup>1</sup>Сабитов Б.Р., <sup>2</sup>Кожокматов Т.М., <sup>2</sup>Шамырова Д.Р., <sup>3</sup>Турдубаева А.Б., <sup>1</sup>Джакипов А.Р., <sup>1</sup>Исабек уулу Б., <sup>1</sup>Кенжебек уулу А., <sup>1</sup>Ма-Тян-Ю. Р.**

<sup>1</sup>КНУ им. Ж.Баласагына, <sup>2</sup>КМТУ им. И.Раззакова, <sup>3</sup> Б.Ельцин атындагы КРСУ

Бул макалада айыл чарбасынын маанилуу милдети — пестициддерди нормалдуу пайдалануу менен тушумдун сапатын жогорулатуу болуп саналат. Айрыкча дыйкандардын жана айыл чарба продук-цияларынын вндуруштук масса-сын эсепке алуу маанилуу. Мында маанилуу милдеттердин бири продукциянын сапатын башкаруу болуп

саналат. Машина менен уйренуунун алгоритмдерин айыл чарба проблемаларынын кенири чейресуне колдонуу мумкунчулугу изилденип жатат.

**Баштапкы сөздөр:** түшүмдү болжолдоо, жасалма нейрон түйүндөрүү, көп катмарлуу перцептрон модели, сызыктуу регрессия, колдоо вектордук ыкма.

## APPLICATION OF MACHINE LEARNING TO OPTIMIZE AGRICULTURAL YIELD OBJECTIVES UNDER PESTICIDE USE

<sup>1</sup>Sabitov B.R., <sup>2</sup>Kozhokmatov T.M., <sup>2</sup>Shamyrova D.R., <sup>3</sup>Turdubaeva A.B., <sup>1</sup>Dzhakipov A.R., <sup>1</sup>Isabek uulu B., <sup>1</sup>Kenzhebek uulu A., <sup>1</sup>Ma-Tian-Yu. R.

<sup>1</sup>KNU im. Zh. Balasagyna, <sup>2</sup> KSTU im. I. Razzakova, <sup>3</sup> B. Yeltsin KRSU

In this paper, the important task of agriculture is to improve the quality of yields with moderate use of pesticides. It is especially important to take into account the mass production of agricultural products by farmers and agricultural producers. One of the important tasks in this case is product quality management. The possibility of applying machine learning algorithms to a wide range of agricultural problems is being studied.

**Keywords:** yield prediction, artificial neural networks, multilayer perceptron model, linear regression, support vector machine.

### Введение

Прогнозирование урожайности является важнейшей задачей исследований по оптимизации и интенсификации сельскохозяйственного производства. Точное предвидение природных условий на год вперед может оказать существенное влияние на управленческие решения, касающиеся выбора культур, размещения в севооборотах, применяемых агротехнических приемов и долгосрочного планирования землепользования. Одним из наиболее важных аспектов точного земледелия является устойчивость. Новизна этого исследования заключается в том, чтобы доказать влияние параметров окружающей среды температуры, пестицидов и осадков на устойчивое сельское хозяйство и экономическую эффективность на уровне посевных площадей регионов республики. Кроме того, прогнозирование будущих

значений урожайности основных культур для 7 регионов. Исследованию рассматривается использование метода опорных векторов машинного обучения и искусственного интеллекта (ИИ) для оценки влияния факторов окружающей среды и агротехнических параметров на урожайность сельскохозяйственных культур и прогнозирования урожайности. С помощью искусственных нейронных сетей (ИНС) была построена высокоэффективная модель многослойного персептрона (MLP) для точного прогнозирования урожайности, температуры, инсектицидов и осадков на основе данных об окружающей среде. Набор данных собран из разных регионов республики, включая температуру, пестициды, количество осадков и урожайность пшеницы. В этом исследовании мы использовали линейную регрессию, метод опорных векторов и нейронные сети для точного прогнозирования урожайности, температуры, инсектицидов и осадков на основе данных об окружающей среде.

Пестициды незаменимы в сельскохозяйственном производстве. Они использовались фермерами для борьбы с сорняками и насекомыми, и сообщалось о значительном увеличении их количества в сельскохозяйственной продукции. Около трети сельскохозяйственной продукции производится в зависимости от применения пестицидов. Без использования пестицидов потери урожая фруктов составили бы 78 %, урожая овощей — 54 %, а урожая зерновых — 32 %.

Таким образом, пестициды играют решающую роль в снижении заболеваемости и повышении урожайности во всем мире. Таким образом, необходимо обсудить процесс развития сельского хозяйства; историческая перспектива, типы и особенности использования пестицидов; и поведение пестицидов, его загрязнение и неблагоприятное воздействие на природную среду. Обзорное исследование показывает, что развитие сельского хозяйства во многих местах мира имеет долгую

историю. Историю использования пестицидов можно разделить на три периода времени. Пестициды классифицируются по различным классификационным терминам, таким как химические классы, функциональные группы, способы действия и токсичность. Пестициды используются для уничтожения вредителей и борьбы с сорняками с использованием химических ингредиентов; следовательно, они также могут быть токсичными для других организмов, включая птиц, рыб, полезных насекомых и нецелевые растения, а также для воздуха, воды, почвы и сельскохозяйственных культур. Кроме того, загрязнение пестицидами удаляется от целевых растений, что приводит к загрязнению окружающей среды. Такие химические остатки воздействуют на здоровье человека через загрязнение окружающей среды и пищевых продуктов. Кроме того, факторы, связанные с изменением климата, также влияют на применение пестицидов и приводят к увеличению использования пестицидов и загрязнению пестицидами. Таким образом, этот обзор предоставит научную информацию, необходимую для применения пестицидов и управления ими в будущем.

### **Методы исследования.**

Целевая переменная: урожайность/площадь.

Модель прогнозирования: линейная регрессия, SVR, ANN

Понимание данных: отсутствующие данные, выбросы, корреляция, выбор признаков, конечная нелинейная функция выглядит следующим образом:

$$\frac{\text{Урожай}}{\text{Площадь}}=F$$

(вода, ультрафиолетовая излучение, удобрения, пестициды, регионы, категории), где  $F$  сигмоидальная функция нейронной сети.

В случае линейной модели, коэффициент линейной модели указывает, какие функции оказывают наибольшее влияние на результат

прогнозирования. Например, пестициды в квадрате имеют самый сильный эффект, а область пересечения 2 (фиктивная переменная) имеет самый слабый эффект.

Эта работа организована следующим образом: Сначала данные считываются и исследуются. Затем выполняется очистка данных. Затем исследуются и сравниваются различные методы моделирования. Для работы с базами данных устанавливаем следующий необходимый пакет программ. Здесь мы включили обычный пакет работы с графикой, представление элементов базы данных в массивы данных и работы с базами данных. Установили, также необходимый пакет sklearn с соответствующими библиотеками работы с базами данных, алгоритмами линейной регрессии и метода опорных векторов. Отдельно установлено пакеты keras и tensorflow для работы с нейронными сетями. Вот пакет установленных библиотек.

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import KFold
from sklearn.svm import SVR
from sklearn.feature_selection import RFECV
from keras.models import Sequential
from keras.layers.core import Dense, Activation
from keras.callbacks import EarlyStopping
from keras.callbacks import ModelCheckpoint
from keras.wrappers.scikit_learn import KerasRegressor
import tensorflow as tf
```

Рис.1. Список, установленных библиотек.

Для работы с базами данных и ее предварительной обработки и загружаем данные.

	id	water	uv	area	fertilizer_usage	yield	pesticides	region	categories
0	169	5.615	65.281	3.230	0	7.977	8.969	0	b,a,c
1	476	7.044	73.319	9.081	0	23.009	7.197	0	c,a,d
2	152	5.607	60.038	2.864	2	23.019	7.424	0	d,a
3	293	9.346	64.719	2.797	2	28.066	1.256	0	d
4	10	7.969	NaN	5.407	1	29.140	0.274	0	c,d

Рис .2. Данные

Основные признаки базы данных следующие.

1. Вода: У воды был выброс данных, который был удален. Среднее количество воды, полученное гектаром (используется в исходной форме, имеет выброс в id 586 или строку 38)
2. Ультрафиолет -uv: Среднее количество света, получаемого гектаром (используется в исходной форме). Кажется, что у них нет никакой корреляции между урожайностью.
3. Удобрения использование: уровень удобрения (используется в исходной форме, мы предполагаем, что эти уровни кратны постоянному значению)
4. Пестициды: Количество пестицидов, использованных на гектар (используется в исходной форме)
5. Регион: код региона (6 фиктивных переменных (существует 7 регионов) определены для представления каждого региона)
6. Категории: список используемых пестицидов через запятую.
7. Зависимая прогнозируемая переменная определяется следующим образом:

Урожайность на гектар=общий урожай, разделенный на площадь фермы в гектарах.

Некоторые корреляционные связи между функцией и целевыми переменными:

На основе структуры базы данных, возможные отношения между каждой функцией и выходными данными следующие:

1. Вода: линейная
2. УФ-ультрафиолет: Нет видимой корреляции
3. Использование удобрения: линейное
3. Пестициды: Квадрат
4. Регион: рассматривается как фиктивная переменная
5. Категории: Считаются фиктивными переменными

Наш общий модель теперь будет иметь следующий вид:

Урожайность/площадь= $f$ (вода, ультрафиолет, удобрения\_использование, пестициды, регион, категории).

**Результаты исследования.** Приведем результаты прогнозирования алгоритмов линейной регрессии, метода опорных векторов и нейронной сети.

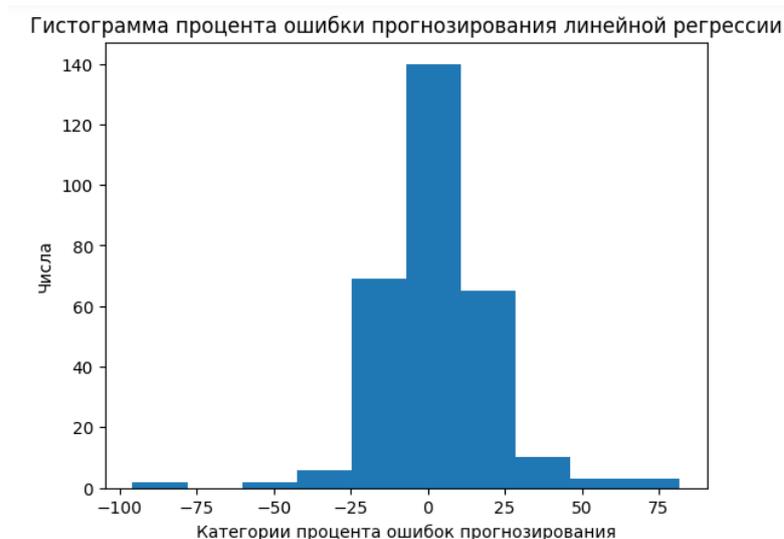


Рис.3. Гистограмма прогнозирования ошибки при линейной регрессии.

Теперь приведем сравнение линейной регрессии с точными данными.

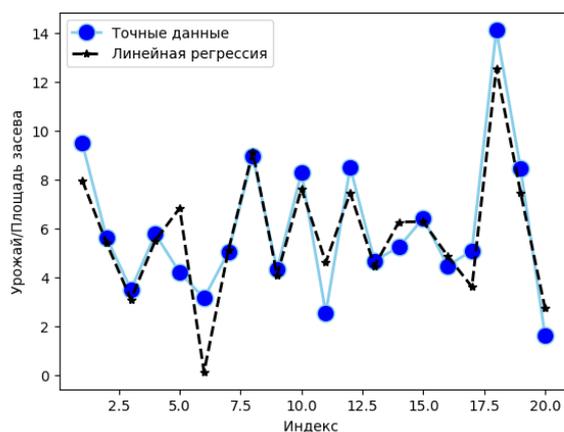


Рис.4. Прогнозирование при линейной регрессии.

Ошибка алгоритма MSE -со всеми исходными данными следующая:

$$0.0109 (+/- 0.0017)$$

$$R^2\text{- модели: } 0.7020 (+/- 0.0618)$$

Значение г-квадрата, равное 0,7, означает, что линейная зависимость недействительна. В следующем методе ультрафиолет uv заменяется категориями в квадрате. В последнем случае для ввода используются все исходные данные с возведенными в квадрат категориями.

Применение метода опорных векторов дает следующий процент ошибки результата.

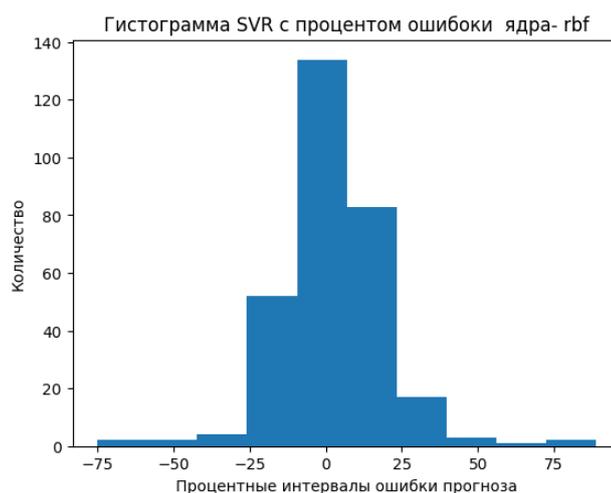


Рис.5. Гистограмма прогнозирования ошибки при использовании алгоритма метода опорных векторов.

Сравним прогнозируемые данные с точными значениями методом опорных векторов.

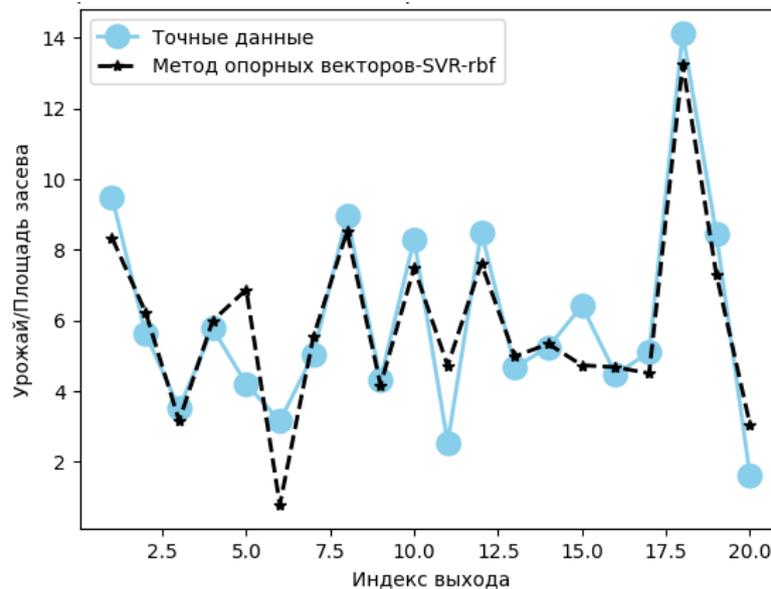


Рис.6. Прогнозирование методом опорных векторов.

Из сравнительного графика видно, что данный метод лучше подходит по сравнению с методом линейной регрессии. В данном случае ошибки варьируются следующим образом

MSE для svr\_rbf: 0.0046 (+/- 0.0003)

MSE для svr\_lin: 0.0054 (+/- 0.0007)

MSE для svr\_poly: 0.0050 (+/- 0.0006)

Результаты, также показывают, что SVR-метод опорных векторов с ядром радиальной базисной функции дает наилучшие результаты, наименьшую ошибку MSE и самый надежный метод (минимум std). Поэтому в качестве представителя выберем SVR с ядром rbf из этой группы.

Теперь перейдем более совершенному методу искусственным нейронным сетям. Ниже приведены результаты реализации ИНС

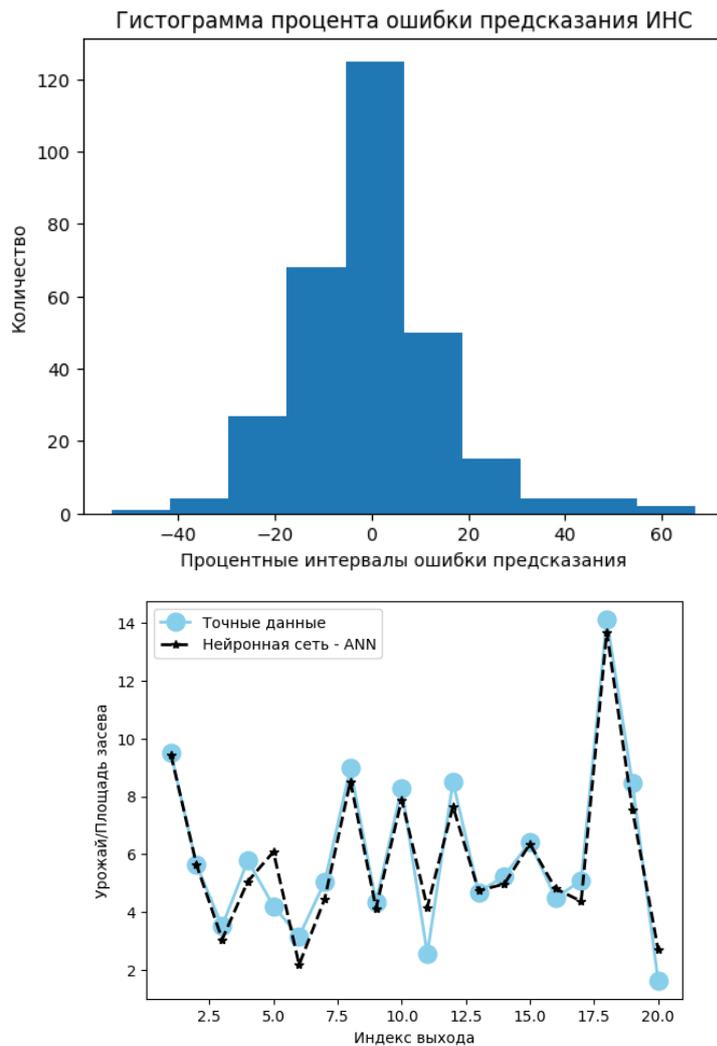


Рис.7. Гистограммы интервалов ошибки и прогнозирование искусственными нейронными сетями (ИНС) .

В данном случае ИНС обучается моделировать систему. Сети имеют три средних скрытых слоя с 20 нейронами. Функции активации выбраны как сигмоидальные, и используется пакет Keras. Ошибка MSE составляет:

Ошибка MSE для ИНС составляет 0.0051 (+/- 0.0016)

### Сравнение моделей

Теперь сравним все методы. Здесь сравниваются результаты лучшей модели каждого раздела. Как видно из таблицы ниже, ANN дает лучший прогноз, а SVR\_rbf — самый надежный метод.

В данном случае можно использовать и другие методы, такие как случайный лес или градиентный бустинг. Они могут дать более результативные результаты прогноза и моделирования. Ниже приведены таблица результатов сравнения и графическое представление полученных результатов:

	Методы	MSE	Среднеквадратичное отклонение
0	Линейная регрессия	0.005316	0.000345
1	Метод опорных векторов	0.004604	0.000132
2	Искусственный нейронный сеть	0.004804	0.000220

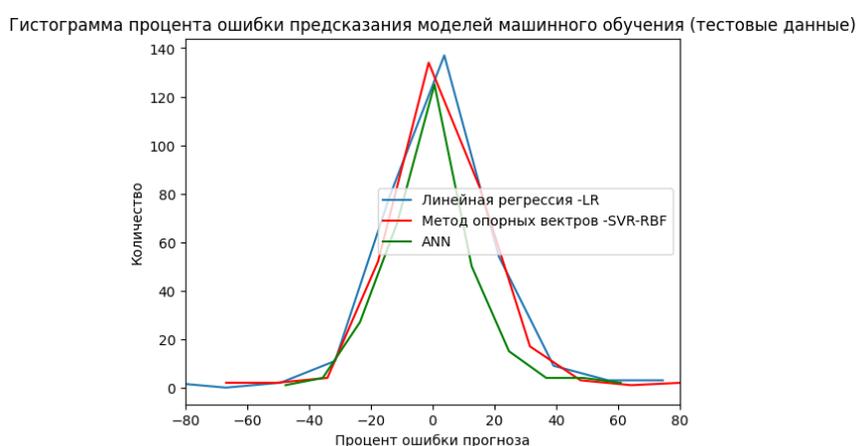


Рис.8. Сравнение результатов прогнозирования машинным обучением и искусственными нейронными сетями (ИНС) .

## ЗАКЛЮЧЕНИЕ

В этой работе использовались различные методы машинного обучения для моделирования урожайности на основе таких характеристик, как вода, пестициды, солнечная радиация, естественные удобрения, пестициды. Были реализованы и сравнено несколько методов. Модели отсортированы от самых простых алгоритмов машинного обучения до самых сложных использованных нейронных сетей, и все они были хорошо обучены. Коэффициент линейной модели указывает, какие функции оказывают наибольшее влияние на результат. Например, пестициды в квадрате имеют самый сильный эффект

коэффициента корреляции. Некоторые фиктивные переменные как ультрафиолет, а регионы имеют самый слабый эффект корреляции.

## ЛИТЕРАТУРА

1. Polder, G.; Blok, P.M.; de Villiers, H.A.C.; van der Wolf, J.M.; Kamp, J. Potato Virus Y Detection in Seed Potatoes Using Deep Learning on Hyperspectral Images. *Front. Plant Sci.* 2019, 10. [Google Scholar] [CrossRef] [PubMed]
2. Zhang, X.; Han, L.; Dong, Y.; Shi, Y.; Huang, W.; Han, L.; González-Moreno, P.; Ma, H.; Ye, H.; Sobeih, T. A Deep Learning-Based Approach for Automated Yellow Rust Disease Detection from High-Resolution Hyperspectral UAV Images. *Remote Sens.* 2019, 11, 1554. [Google Scholar] [CrossRef]
3. Golhani, K.; Balasundram, S.K.; Vadamalai, G.; Pradhan, B. A review of neural networks in plant disease detection using hyperspectral data. *Inf. Process. Agric.* 2018, 5, 354–371. [Google Scholar] [CrossRef]