

# ОБРАБОТКА РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ МИКРОКЛИМАТИЧЕСКИХ ПАРАМЕТРОВ ПРИ ВЫРАЩИВАНИИ РАСТЕНИЙ

**Ю. В. Черепанова**

*Белорусский государственный университет, г. Минск;*

*j.cher.30.11@gmail.com;*

*науч. рук. – А. В. Микулович, ст. преп.*

В данной работе рассмотрены методы интеллектуального анализа данных, которые применяются в сельском хозяйстве. Проведено сравнение популярных алгоритмов машинного обучения и на основе самого точного метода реализовано прогнозирование по выращиванию наилучшей сельскохозяйственной культуры.

**Ключевые слова:** интеллектуальный анализ данных, случайный лес, классификация, машинное обучение.

## ВВЕДЕНИЕ

На сегодняшний день набирает популярность точное земледелие, эта современная технология земледелия использует данные о характеристиках почвы, данные об урожайности, погодных условиях и предлагает фермерам наиболее оптимальную культуру для выращивания в своих хозяйствах для получения максимального урожая и прибыли [1].

Данный подход к выращиванию может быть реализован посредством использования методов интеллектуального анализа данных, поэтому их изучение и использование актуально на сегодняшний день.

Целью данной работы является изучение и применение методов анализа данных в области земледелия.

Методы интеллектуального анализа данных (ИАД) позволяют решить многие задачи, из них основными являются: классификация, кластеризация, регрессия, ассоциация, визуализация [2].

## МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ

В работе была рассмотрена задача классификации, которая заключается в получении категориального ответа на основе набора признаков [3]. Задача рассматривалась на основе набора данных, который был получен путем дополнения и объединения различных общедоступных наборов данных Индии [4]. В наборе рассматриваются следующие параметры: соотношение содержания азота, фосфора, калия в почве в кг/га, температура в градусах Цельсия, относительная

влажность воздуха в %, значение pH почвы и количество осадков в мм. Всего набор данных содержит 2200 записей параметров. Данное исследование выполнено с помощью языка программирования Python и с использованием следующих библиотек: pandas, numpy, scikit-learn, seaborn, matplotlib.

В исследовании были рассмотрены следующие методы:

### 1. К-ближайших соседей

Одним из самых популярных алгоритмов машинного обучения является метод ближайших соседей. В случае использования метода для классификации объект присваивается тому классу, который является наиболее распространённым среди k соседей данного элемента, классы которых уже известны [5]. Реализация представлена на рисунке 1.

### 2. Логистическая регрессия

Ещё одним методом, который используется в данном исследовании, является логистическая регрессия. Данный метод полезен в ситуациях, когда необходимо предсказать итог на основании значений набора переменных [6]. Реализация представлена на рисунке 1.

K-Nearest Neighbors	Logistic Regression
<pre>from sklearn.neighbors import KNeighborsClassifier  knn = KNeighborsClassifier() knn.fit(X_train,y_train)  pred_knn = knn.predict(X_test)  x = metrics.accuracy_score(y_test, pred_knn) acc.append(x) model.append('K-NN') print(x)</pre>	<pre>from sklearn.linear_model import LogisticRegression  lr = LogisticRegression() lr.fit(X_train, y_train)  pred_lr = lr.predict(X_test)  x = metrics.accuracy_score(y_test, pred_lr) acc.append(x) model.append('LR') print(x)</pre>
0.9886363636363636	0.9522727272727273

Рис. 1. Реализация и показатель точности для методов k - ближайших соседей и логистической регрессии

### 3. Дерево решений

Дерево решений является одним из методов автоматического анализа данных, основанным на нахождении правил или решений, где каждому объекту соответствует единственный узел, дающий решение. За счет обучающего множества правила генерируются автоматически в процессе обучения [7]. Реализация метода представлена на рисунке 2. Для обучения был выбран критерий энтропии, а также было наложено ограничение на количество разбиений в ветках.

### 4. Случайный лес

Случайный лес — это наиболее гибкий и простой в использовании алгоритм. Лес является набором из множества деревьев решений. Он создаёт деревья для случайно выбранных семплов данных, получает

прогноз от каждого дерева и выбирает наилучшее решение посредством голосования. Алгоритм предотвращает переобучение, создавая деревья на случайных выборках. В то же время метод схож с методом ближайших соседей. Лес осуществляет предсказания для объектов на основе меток похожих объектов из обучения. Схожесть при этом тем выше, чем чаще эти объекты оказываются в одном и том же листе дерева [8]. Реализация представлена на рисунке 2.

Random Forest	Decision Tree
<pre>from sklearn.ensemble import RandomForestClassifier  rf_model = RandomForestClassifier(n_estimators = 20) rf_model.fit(X_train, y_train)  pred_rf = rf_model.predict(X_test)  x = metrics.accuracy_score(y_test, pred_rf) acc.append(x) model.append('RF') print(x)  0.9977272727272727</pre>	<pre>from sklearn.tree import DecisionTreeClassifier  decision_tree = DecisionTreeClassifier(criterion = 'entropy',  max_depth = 5) decision_tree.fit(X_train, y_train)  pred_tree = decision_tree.predict(X_test)  x = metrics.accuracy_score(y_test, pred_tree) acc.append(x) model.append('DTree') print(x)  0.9386363636363636</pre>

*Рис. 2.* Реализация и показатель точности для методов случайного леса и дерева решений

Сравнивая все полученные показатели точности, можно сделать вывод, что самую лучшую точность показал метод «Случайного леса». Точность составила 0,998. На основе этого алгоритма было реализовано предсказание к выращиванию наилучшей сельскохозяйственной культуры в определенной местности. Реализация представлена на рисунке 3.

## Predict

```
data = np.array([[100, 82, 50, 27, 80, 5,105]])
prediction = rf_model.predict(data)
print(prediction)
```

```
['banana']
```

*Рис. 3.* Результат прогнозирования наилучшей к выращиванию культуры

## ЗАКЛЮЧЕНИЕ

Подводя итоги, можно отметить, что методы интеллектуального анализа данных, а также и те, что были представлены выше, набирают популярность в различных областях жизни, в том числе, в сельском хозяйстве.

Использование данных алгоритмов помогает строить более эффективные стратегии ведения сельского хозяйства, а также помогает экономить хозяйственные и природные ресурсы. При этом открываются реальные возможности производства качественной продукции и сохранения окружающей среды. Ведь точное земледелие позволяет сократить затраты на внесение удобрений и семян.

#### Библиографические ссылки

1. *Rositsa Beluhova-Uzunova, Dobri Dunchev* Precision farming – concepts and perspectives – 2019. DOI:10.30858/zer/112132
2. *Яцков Н. Н.* Интеллектуальный анализ данных. Минск: БГУ, 2014, 4 - 11 с.
3. *Дьяконов А. Г.* Методы решения задач классификации с категориальными признаками// Прикладная информатика и математика. Труды факультета Вычислительной математики и кибернетики МГУ имени М. В. Ломоносова, № 46, 2014, с 103 – 127
4. Интернет – адрес: <https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset>
5. Интернет – адрес: <https://realpython.com/knn-python/>
6. *Paklin N.B.* Logistic regression and ROC- analysis – a mathematical tool//Official site of BaseGroup Labs URL: <http://www.basegroup.ru/library/analysis/regression/logistic/>
7. Nagesh Singh Chauhan, **Decision Tree Algorithm**, Data Science Enthusiast on February 9, 2022 in Machine Learning
8. Интернет–адрес: <https://www.javatpoint.com/machine-learning-random-forest-algorithm>