



**САБИТОВ Б.Р., ОРОЗОБЕКОВА А.К., ЖАДИЛОВ Б.М., СЕЙТБЕКОВ А.,
ШАМЫРОВА Д.Р., ШЕРИМБЕКОВА Э.Б.**

¹ Кыргызский национальный университет им. Ж.Баласагына, Бишкек,
Кыргызская Республика

² КГУСТА им. Н.Исанова, Бишкек, Кыргызская Республика

**SABITOV B.R., OROZOBKOVA A.K., ZHADILOV B.M., A. SEITBEKOV A.,
SHAMYROVA D.R., SHERIMBEKOVA E.B.**

¹ Kyrgyz National University named after J. Balasagyn, Bishkek, Kyrgyz Republic

² KSUCTA n.a. N. Isanov, Bishkek, Kyrgyz Republic

ВИЗУАЛИЗАЦИЯ ДАННЫХ УРОЖАЙНОСТИ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР С ПРИМЕНЕНИЕМ PYTHON-ТЕХНОЛОГИЙ

VISUALIZATION OF CROP YIELD DATA USING PYTHON TECHNOLOGIES

Изучается анализ и визуализация данных урожайности сельскохозяйственных культур с применением технологий машинного обучения. Рассматривается применение новейшей системы FbProphet для анализа временных рядов.

Ключевые слова: машинное обучение, алгоритмы, анализ данных, Python, АПК, прогнозирование, тестирование.

Айыл чарба өсүмдүктөрүнүн түшүмү боюнча маалыматтарды машина үйрөнүү технологияларын колдонуу менен талдоо жана визуалдаштыруу каралган. Убакыт катарлар аркылуу талдоо үчүн жаңы FbProphet системасын колдонуу каралган.

Өзөк сөздөр: машина үйрөнүү, алгоритмдер, маалыматтарды талдоо, Python, агроөнөр жай комплекси, болжолдоо, тестирлөө.

The analysis and visualization of data on the yield of agricultural crops using machine learning technologies is being studied. The application of the latest FbProphet system for time series analysis is considered.

Key words. Machine learning, algorithms, data analysis, Python, agro-industrial complex, forecasting, testing

В данной работе изучен анализ данных с применением Python технологий. Изучены общие вопросы и технологии работы с временными рядами в Python. Построены модели конкретных задач для агропромышленного комплекса (АПК) с применением временных рядов. Изучены процессы анализа данных, с использованием интеллектуальной системы FbProphet. Система прогнозирования FbProphet, считается современной и универсальной технологией сегодняшнего дня, наряду с технологией прогнозирования машинного обучения –XGBoost. В применении данной технологии удобно управлять данные временных рядов, а также визуализировать их с помощью библиотек Python.

В статье, данную технологию применим к задачам АПК с реальными данными. Изучим процесс построения сложных нелинейных моделей и рассмотрим задачу прогнозирования, используя данную технологию, в зависимости от требуемого качества прогнозирования, продолжительности прогнозируемого периода и, конечно же, времени, в течение которого, мы должны выбирать функции и параметры настройки для достижения желаемых результатов.



Программный пакет, подключение необходимого программного обеспечения библиотек Python для ее работы с графикой и данными осуществляется следующими системами и библиотеками Python.

```
import numpy as np          # Работа с векторами и матрицами
import pandas as pd        # Работа с таблицами и их манипуляции
import matplotlib.pyplot as plt  # Графика
import seaborn as sns      # Работа с большими визуализациями
sns.set()
from dateutil.relativedelta import relativedelta # Работа с данными и их стилями
from scipy.optimize import minimize # Минимизация функций
import statsmodels.formula.api as smf # Статистика и эконометрика
import statsmodels.tsa.api as smt
import statsmodels.api as sm
import scipy.stats as scs
from itertools import product # Некоторые полезные функции
from tqdm import tqdm_notebook
import warnings # `Игнорирование предупреждений Anaconda
warnings.filterwarnings('ignore')
%matplotlib inline
```

В качестве примера рассмотрим реальные данные урожайности картофеля и ее стоимость на отдельно взятом регионе, например по Иссык-Кульской области Кыргызстана, в котором отражены реальные данные отдельных айыл окмоту. В качестве региона выберем айыл окмоту Иссык-Кульской области. Ниже приведены данные по урожайности и стоимости картофеля. Рассмотрим визуализацию стоимости картофеля по годам.

```
st_currency = pd.read_csv('k_currency.csv', index_col=['Time'], parse_dates=['Time'])
```

Графически это выглядит следующим образом. По оси y отложим стоимость, а по оси x годы по месяцам



Рис.1. График изменения стоимости картофеля по годам и месяцам

Теперь посмотрим базу данных урожайности картофеля. По оси y отложим урожайность в центнерах с гектара, а по оси x месяцы и годы.

```
st_ads_kart = pd.read_csv('ads_kart.csv', index_col=['Time'], parse_dates=['Time'])
```

Поле признаков ads_kart отражает урожайность картофеля в центнерах на га, сначала по дням, месяцам и годам. Данные получены из разных айыл окмоту соответствующего региона.

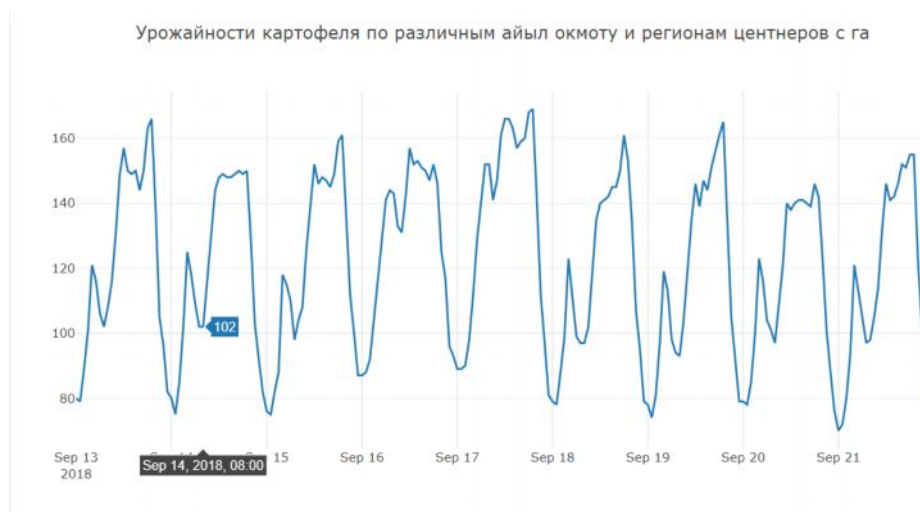


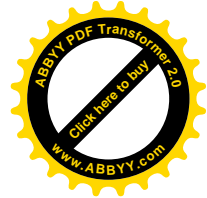
Рис. 2. График изменения урожайности картофеля по определенному отрезку времени или в целом по годам



Рис. 3. Колебания изменения урожайности картофеля.

Использование интеллектуальной системы FbProphet для визуализации данных. Для реализации визуализации данных данной задачи будем использовать недавно разработанную новейшую систему Core Data Science из Facebook (2017 год). В данной технологии предложен новый метод прогнозирования временных рядов – система FbProphet (Пророк), который позволяет аналитикам и разработчикам выполнять прогнозирование в Python 3. Для работы с временными рядами в данную систему включены множества различных библиотек Python. Интеллектуальная система Fbprophet по результатам применения ко многим прикладным задачам показала себя как мощную и универсальную технологию прогнозирования. В работе данный метод применено к задачам прогнозирования урожайности в сельском хозяйстве. Изучаемая, интеллектуальная система, кроме языка программирования Python также, использует дополнительно язык программирования Ruytan. Для работы понадобятся библиотеки pandas, matplotlib, numpy, cython и fbprophet. Как и большинство других пакетов Python, pandas, numpy, cython и matplotlib можно установить через pip в Python. Библиотека fbprophet еще зависит от языка программирования STAN, поэтому сначала нужно установить ruytan – как оболочку Python для STAN.

```
%matplotlib inline
```



```
import pandas as pd
from fbprophet import Prophet
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight') #Кроме того, нужно определить стиль matplotlib # (в
данном случае используется fivethirtyeight)
```

Для работы с данной системой прогнозирования рассмотрим набор данных об урожайности картофеля в разные годы и разных айыл окмоту, который содержит временные ряды о ежемесячном количестве картофеля за период с 2007 по 2018 годы. После этого нужно загрузить временные ряды. Загрузим ее.

```
df = pd.read_csv('Urojai_kart.csv')
df.head(20)
```

Вот фрагмент данных (урожайность картофеля указано в центнерах с га)

	Month	Urojai_kart
0	2007-01	112
1	2007-02	118
2	2007-03	132
3	2007-04	129
4	2007-05	121
5	2007-06	135
6	2007-07	148
7	2007-08	148
8	2007-09	136
9	2007-10	119
10	2007-11	104
11	2007-12	118
12	2008-01	115
13	2008-02	126
14	2008-03	141
15	2008-04	135
16	2008-05	125
17	2008-06	129
18	2008-07	120
19	2008-08	130

Всего получены данные урожайности 144 фермерских хозяйств в разные годы.

```
df.shape
(144, 2)
```

Укажем тип данных

```
df.dtypes
Month      object
Urojai_kart int64
dtype: object
```

Согласно требованиям библиотеки Prophet, поступающий в качестве ввода DataFrame должен содержать столбец с информацией о времени, и столбец, содержащий метрику, которую нужно спрогнозировать. Важно отметить, что столбец времени должен содержать данные типа datetime. Как видно, столбец Month-месяц содержит данные другого типа, не datetime, поэтому его нужно конвертировать. Далее, для работы с системой, преобразуем эти данные- месяцы в Datetime.

```
df['Month'] = pd.DatetimeIndex(df['Month'])
df.dtypes
```

Вот результат типа данных.

```
Month      datetime64[ns]
Urojai_kart int64
dtype: object
```

Теперь все в порядке с данными. Также Prophet требует, чтобы столбцы входных данных назывались ds (столбец времени) и y (столбец метрики). Учитывая, эти требования системы FbProphet столбцы датафрейма, преобразуем в переменные ds и y.

```
df = df.rename(columns={'Month': 'ds',
```



```
'Urojai_kart': 'y'})  
df.head(10)
```

Получили новый фрейм. Преобразованная таблица выглядит следующим образом.

	ds	y
0	2007-01-01	112
1	2007-02-01	118
2	2007-03-01	132
3	2007-04-01	129
4	2007-05-01	121
5	2007-06-01	135
6	2007-07-01	148
7	2007-08-01	148
8	2007-09-01	136
9	2007-10-01	119

Давайте посмотрим на визуализацию данных.

```
ax = df.set_index('ds').plot(figsize=(12, 8))  
ax.set_ylabel('Урожайность картофеля центнеров с га')  
ax.set_xlabel('Data') plt.show()
```

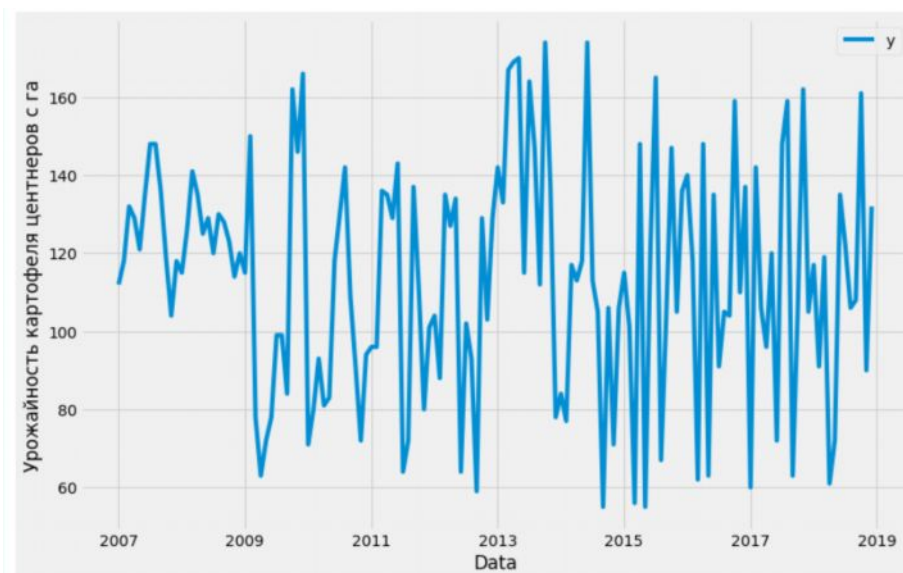


Рис. 4. Колебания изменения урожайности картофеля.

Заключение. В данной работе изучены различные методы анализа данных временных рядов с применением технологий FbProphet. На примере урожайности, изучили технологии визуализации данных в системе FbProphet. Ознакомились с технологией современной визуализации, не прибегая к большим трудностям, как другие статистические технологии прогнозирования. Система прогнозирования FbProphet, считается современной и универсальной технологией сегодняшнего дня, наряду с XGBoost –градиентным бустингом.

Список литературы

1. Орельен Ж. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow [Текст] / Ж.орельен. - М.: Диалектика-Вильямс, 2020.
2. Официальный сайт scikit-learn: <https://scikit-learn.org>
3. Официальный сайт FbProphet: <https://prophet.com>