



Столыпинский  
вестник

Научная статья

Original article

УДК 659.118

**БИБЛИОТЕКА STREAMLIT КАК ИНСТРУМЕНТ ОБРАБОТКИ И  
ВИЗУАЛИЗАЦИИ БОЛЬШИХ ДАННЫХ**  
STREAMLIT LIBRARY AS A TOOL FOR PROCESSING AND  
VISUALIZATION OF BIG DATA

**Конова Полина Сергеевна**, Студент 2 курс, факультет Экономики и управления АПК, Российский государственный аграрный университет – МСХА имени К.А.Тимирязева, Россия, г. Москва

**Konova Polina Sergeevna**, 2nd year student, Faculty of Economics and Agribusiness Management, Russian State Agrarian University – Moscow Agricultural Academy named after K.A.Timiryazev, Russia, Moscow

**Аннотация.** В статье анализируется библиотека Streamlit как инструмент обработки и визуализации больших данных. Рассматриваются понятие, характеристики и особенности обработки больших данных и проблемы их анализа, связанные с размерностью выборок. Приводятся основные методы и технологии обработки больших данных, характерные черты и преимущества метода визуализации. Выявляются принципы использования, особенности и преимущества библиотеки Streamlit и её возможности для обработки и визуализации больших данных.

**Annotation.** The article analyzes the Streamlit library as a tool for processing large big data. The conclusions, characteristics and features of big data processing and the problems of their analysis related to the dimension of the samples are considered. When using the main methods and technologies for processing big data, the features and benefits are useful. The rationale for the use, features and benefits of the Streamlit library and its capabilities for processing and detecting big data are revealed.

**Ключевые слова:** большие данные, Streamlit, Python, обработка больших данных, визуализация.

**Keywords:** big data, Streamlit, Python, big data processing, visualization.

Основной тенденцией, определяющей архитектуру современных информационных систем и приложений, является наличие значительного объёма данных, часто имеющих слабоструктурированный либо неструктурированный вид и требующих оперативной обработки [1]. Анализ больших данных требует особых подходов и может осуществляться посредством специализированных программ и приложений, способных своевременно предоставлять качественную информацию, в последующем используемую для принятия значимых управленческих решений. Одной из сфер, в которой требуется внедрение подобных программ, является сельское хозяйство, от развития которого зависит глобальная общественная стабильность [2]. Использование информационных технологий для обработки больших данных повышает производительность и эффективность деятельности различных отраслей сельского хозяйства, в частности фермерских хозяйств, что делает актуальным исследование инструментов для разработки приложений для анализа данных этой сферы.

Целью работы является изучение библиотеки Streamlit как инструмента обработки и визуализации больших данных. Для её достижения были использованы методы анализа и синтеза научных публикаций и литературных источников по рассматриваемой теме. Научная новизна исследования

заключается в применении нового подхода к обработке и визуализации данных о фермерских хозяйствах с использованием библиотеки Streamlit.

Большие данные представляют собой технологию обработки структурированных и неструктурированных данных, имеющих постоянный прирост [3]. Основными характеристиками больших данных являются объём, скорость, вариативность и ценность. Скорость обработки сведений в больших данных приближена к реальному времени. Благодаря большому объёму данных при обработке их массива повышается точность обнаружения различных связей, используемых для последующего представления аналитики в агрегированном виде. Вариативность данных даёт возможность выявлять неочевидные зависимости.

Значимым вопросом при работе с большими данными является их анализ, который не может быть осуществлён посредством методов и критериев из арсенала классической математической статистики в силу размерности выборок [4]. Рост размерности осложняет анализ информации в части вычисления оценок параметров, поскольку требует кардинального увеличения вычислительных затрат и сопровождается ухудшением сходимости используемых при нахождении оценок итерационных алгоритмов. Основная сложность анализа больших данных при помощи критериев проверки статистических гипотез заключается в зависимости между распределениями статистик этих критериев и объёмами выборок. При этом наличие предельных распределений статистик критериев не может гарантировать корректности статистических выводов при анализе больших выборок. Данные обстоятельства обуславливают актуальность разработки новых методов обработки больших данных.

Можно выделить следующие особенности обработки больших данных по сравнению с обработкой обычных данных [5]:

- хранение информации в распределённых источниках;
- параллельная распределённая архитектура, применение кластерных платформ с несколькими узлами;

- разнородность и динамичность данных, собираемых из различных источников;
- разнородность формата данных, наличие взаимосвязи сведений, требующей их извлечения из различных узлов;
- неизвестность источников данных и этапов предварительной обработки, большое количество неточных, некорректных и неразмеченных сведений;
- потоковое поступление данных, сосуществование ценных и бесполезных сведений;
- отсутствие статистической значимости массивных наборов данных.

В настоящее время можно выделить следующие основные методы и технологии обработки больших данных [6]:

- краудсорсинг – ручной анализ, осуществляемый с привлечением большого количества людей;
- смешение и интеграция данных – приведение полученной из разных источников информации к одному виду, её дополнение и уточнение;
- нейронные сети и машинное обучение – создание программ, умеющих анализировать и принимать решения посредством выстраивания логических связей;
- предиктивная аналитика – прогнозирование на базе собранных сведений;
- имитационное моделирование – использование больших данных для построения моделей, позволяющих проводить эксперименты в компьютерной реальности;
- статистический анализ – подсчёт сведений по формулам и выявление имеющихся сходств, тенденций и закономерностей;
- data mining – получение новой значимой информации из значительного объёма данных;

- визуализация – представление самих больших данных и результатов их обработки в виде понятных человеку графиков и схем.

Визуализация является инструментом сокращения сложности данных, в основе которого лежит создание зрительных образов визуальными элементами, такими как диаграммы, графы и карты [7]. Эти элементы дают возможность упрощать сложные явления и объекты для их понимания аналитиками или исследователями, а также передавать информацию.

Посредством визуализации обеспечивается наглядное представление геометрической метафоры сведений, лаконичное описание существующих в наборе данных внутренних закономерностей, сжатие заключённой в данных информации, восстановление имеющихся в данных пробелов, решение прогностических задач и задач построения регрессионных зависимостей, имеющихся между признаками [8]. Визуализация данных направлена на помощь пользователю в решении прикладных задач в процессе предварительной обработки информации и фактического анализа данных. Благодаря ей повышается качество восприятия свойств данных.

В настоящее время для обработки и визуализации больших данных чаще всего используется язык программирования Python [9]. Во многом это обусловлено возможностью точного построения с его помощью самообучающихся систем и нейронных сетей. На практике при создании программ и приложений используются различные библиотеки данных, позволяющие значительно сократить время и трудозатраты на написание кода для конкретных целей.

Одной из таких библиотек является Streamlit, представляющая собой фреймворк Python с открытым исходным кодом [10]. С её помощью можно визуализировать модели обработки больших данных и создавать базовые программы в форме интерактивных веб-приложений с использованием нескольких строк кода. Основными особенностями фреймворка Streamlit являются:

1. Использование сценариев Python. Приложения Streamlit представляют собой выполняющиеся сверху вниз скрипты без скрытого состояния. Для создания приложений требуется написание нескольких строк кода с использованием предоставляемых фреймворком API. По мере внесения изменений в код приложение обновляется автоматически.
2. Аналогичность виджетов переменным. При использовании Streamlit добавление виджетов в приложение аналогично объявлению переменных в Python, при этом каждое взаимодействие перезапускает сценарий сверху вниз, что позволяет получить чистый код.
3. Мгновенное развёртывание. Streamlit позволяет легко и быстро создавать приложения и управлять ими.
4. Возможность повторного использования данных и вычислений. Streamlit является примитивом кэша, использующимся как постоянное, по умолчанию неизменное хранилище данных, которое позволяет приложениям Streamlit легко и безопасно повторно использовать информацию.

При работе Streamlit весь сценарий запускается заново для каждого взаимодействия с пользователем, при этом каждой переменной присваивается актуальное значение с учётом состояния виджета [11]. Благодаря кэшированию Streamlit пропускает избыточные вычисления и выборки данных, что оптимизирует работу приложений. Можно выделить следующие ключевые преимущества Streamlit:

- приложения Streamlit являются чистыми файлами Python, что позволяет использовать любой редактор и отладчик;
- скрипты Pure Python совместимы с Git и другим программным обеспечением для управления версиями и совместной работы, которое может быть использовано при работе с фреймворком;
- Streamlit предоставляет среду прямого кодирования в реальном времени;

- фреймворк разработан для графических процессоров, поэтому обеспечивает прямой доступ к примитивам машинного уровня и дополняет библиотеки;
- Streamlit является бесплатной библиотекой с открытым исходным кодом, поэтому обслуживание созданных с её помощью приложений может осуществляться локально.

Виджеты в библиотеке Streamlit задаются при помощи API (программного интерфейса приложения), представляющего собой набор подпрограмм, протоколов, функций и/либо команд, используемых для облегчения взаимодействия между различными программными службами [12]. API-интерфейсы позволяют одному приложению получать доступ к данным из другого приложения или программы без необходимости знания о механизме его работы. Streamlit содержит справочник по API, организованный по типу действия, к примеру по отображению данных или оптимизации производительности [13]. Каждый раздел включает связанные с типом действия методы и содержит примеры, что значительно облегчает создание приложений при помощи Streamlit.

Таким образом, Streamlit представляет собой фреймворк Python, позволяющий создавать удобные приложения с применением технологий машинного обучения без необходимости написания большого объёма кода и значительных временных затрат. Готовые приложения способны обрабатывать значительные массивы данных и выстраивать статистические модели для прогнозирования бизнес-процессов, что делает перспективным использование библиотеки Streamlit в областях, требующих анализа больших данных, включая и сельское хозяйство.

### Список литературы

1. Пролетарский А.В. Методы ситуационного анализа и графической визуализации потоков больших данных / А.В. Пролетарский, Д.В. Березкин, Ю.Е. Гапанюк, И.А. Козлов, А.Ю. Попов, Р.С. Самарев, В.И.

- Терехов // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. – 2018. – № 2. – С. 98-123.
2. Махмудова Н.Р. Информационные технологии в сельскохозяйственной экономике // Проблемы современной науки и образования. – 2020. – № 6-2 (151). – С. 31-34.
  3. Тихонов А.А. Большие данные и глубокое машинное обучение в искусственных нейронных сетях // Наука и образование сегодня. – 2018. – № 6 (29). – С. 35-38.
  4. Лемешко Б.Ю., Лемешко С.Б., Семенова М.А. К вопросу статистического анализа больших данных // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2018. – № 44. – С. 40-49.
  5. Ромичева Е.В. Методы обработки и визуализации больших данных // Аллея науки. – 2017. – Т. 3, № 16. – С. 976-982.
  6. Шпрингер Е. Технологии big data: как анализируют большие данные, чтобы получить максимум прибыли [Электронный ресурс] // VK Cloud Solutions. – 2020. – URL: <https://mcs.mail.ru/blog/tekhnologii-big-data-kak-analiziruyut-bolshie-dannye> (дата обращения: 22.07.2022).
  7. Платонов В.В. Визуализация больших данных в экономических науках в условиях информационного общества // Вопросы инновационной экономики. – 2020. – Т. 10, № 4. – С. 1831-1848.
  8. Лежебоков А.А., Кулиев Э.В. Технологии визуализации для прикладных задач интеллектуального анализа данных // Известия Кабардино-Балкарского научного центра РАН. – 2019. – № 4 (90). – С. 14-23.
  9. Джангаров А.И., Сулейманова М.А. Анализ больших данных // Colloquium-journal. – 2019. – № 28 (52). – С. 116-117.
  10. Streamlit [Электронный ресурс] // Streamlit. – URL: <https://streamlit.io/> (дата обращения: 22.07.2022).
  11. Treuille A. Turn Python Scripts into Beautiful ML Tools [Электронный ресурс] // Medium. – 2019. – URL: <https://towardsdatascience.com/coding->



ml-tools-like-you-code-ml-models-ddba3357eace (дата обращения: 22.07.2022).

12. Солёный С.В., Воропаев И.А., Давиденко М.Н. Об API и архитектурных методах защиты при его разработке // Волновая электроника и инфокоммуникационные системы: материалы XXV Международной научной конференции. – СПб., 2022. – С. 251-256.
13. API reference. [Электронный ресурс] // Streamlit. – URL: <https://docs.streamlit.io/library/api-reference> (дата обращения: 22.07.2022).

#### **List of literature**

1. Proletarsky A.V. Methods of situational analysis and graphical visualization of big data flows / A.V. Proletarsky, D.V. Berezkin, Y.E. Gapanyuk, I.A. Kozlov, A.Y. Popov, R.S. Samarev, V.I. Terekhov // Bulletin of Bauman Moscow State Technical University. Ser. Instrumentation. - 2018. – No. 2. – С. 98-123.
2. Makhmudova N.R. Information technologies in agricultural economics // Problems of modern science and education. – 2020. – № 6-2 (151). – Pp. 31-34.
3. Tikhonov A.A. Big data and deep machine learning in artificial neural networks // Science and Education today. – 2018. – № 6 (29). – Pp. 35-38.
4. Lemeshko B.Yu., Lemeshko S.B., Semenova M.A. On the issue of statistical analysis of big data // Bulletin of Tomsk State University. Management, computer engineering and computer science. - 2018. – No. 44. – pp. 40-49.
5. Romicheva E.V. Methods of processing and visualization of big data // Alley of Science. – 2017. – Vol. 3, No. 16. – pp. 976-982.
6. Springer E. Big data technologies: how big data is analyzed to get maximum profit [Electronic resource] // VK Cloud Solutions. – 2020. – URL: <https://mcs.mail.ru/blog/tekhnologii-big-data-kak-analiziruyut-bolshie-dannye> (accessed: 07/22/2022).
7. Platonov V.V. Visualization of big data in economic sciences in the Information society // Issues of innovative economy. – 2020. – Vol. 10, No. 4. – pp. 1831-1848.

8. Lezhebokov A.A., Kuliyeв E.V. Visualization technologies for applied problems of data mining // Izvestiya Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences. – 2019. – № 4 (90). – Pp. 14-23.
9. Dzhangarov A.I., Suleymanova M.A. Big data analysis // Colloquium-journal. – 2019. – № 28 (52). – Pp. 116-117.
10. Streamlit [Electronic resource] // Streamlit. – URL: <https://streamlit.io/> (accessed: 07/22/2022).
11. Treuille A. Turn Python Scripts into Beautiful ML Tools [Electronic resource] // Medium. – 2019. – URL: <https://towardsdatascience.com/coding-ml-tools-like-you-code-ml-models-ddba3357eace> (accessed: 22.07.2022).
12. Salty S.V., Voropaev I.A., Davidenko M.N. About API and architectural methods of protection during its development // Wave Electronics and infocommunication systems: proceedings of the XXV International Scientific Conference. – St. Petersburg, 2022. – pp. 251-256.
13. API reference. [Electronic resource] // Streamlit. – URL: <https://docs.streamlit.io/library/api-reference> (accessed: 07/22/2022).

© Конова П.С., 2022 Научный сетевой журнал «СтолЫПИНСКИЙ вестник»  
№3/2022

**Для цитирования:** Конова П.С. БИБЛИОТЕКА STREAMLIT КАК ИНСТРУМЕНТ ОБРАБОТКИ И ВИЗУАЛИЗАЦИИ БОЛЬШИХ ДАННЫХ// Научный сетевой журнал «СтолЫПИНСКИЙ вестник» №3/2022