

# ОНЛАЙН-ПЛАТФОРМА ДЛЯ АНАЛИЗА И ОБРАБОТКИ ДАННЫХ

*Климец А.А.*

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Куприянова Д.В. – ассистент*

Представлен прототип разрабатываемой системы в рамках концепции Data Mining as a Service, т.е. предоставление сервиса по анализу и обработке данных, доступ к которому можно получить через браузер. Преимуществами системы является: ориентированность на клиент-серверную архитектуру, взаимодействие с любой библиотекой алгоритмов обработки данных, предоставляющей доступ с помощью языка программирования Python.

Объём генерируемой каждый день информации существенно увеличивается каждый день (например, по результатам исследований, проводимых в 2019 году в компании Domo [1], в Instagram создается 277 777 текстовых сообщений, а Google обрабатывает 4 497 420 поисковых запросов ежеминутно). Все создаваемые информационные потоки анализируются, обрабатываются и учитываются в различных сферах деятельности (реклама, прогноз успеха той или иной кампании). Так как «вручную» обработать весь поток информации невозможно, огромными темпами развивается такое направление, как Data Mining.

Data Mining – исследование и обнаружение в сырых данных скрытых знаний, которые ранее не были известны, нетривиальны, практически полезны, доступны для интерпретации человеком [2]. В данной области широкое развитие получили как коммерческие (например, MatLab, Statistica), так и свободные (например, Weka, R и др.) специализированные инструменты. Однако они обладают как преимуществами («заточка» на решение той или иной задачи), так и недостатками (написание кода, усложненный интерфейс со множеством параметров, работа только на персональном компьютере).

Разрабатываемая система предоставляет простой и удобный интерфейс пользователя, не

требующий какого-либо навыка в области написания программного кода или предварительного изучения того, как следует взаимодействовать с системой. В ней реализуются обширные возможности для расширения за счёт использования принципа модульности: можно подключить любую библиотеку на языке Python, который является главным языком программирования, используемый в проектах, связанных с математической статистикой, машинным обучением, нейронными сетями, анализом данных и т.д.

Система состоит из двух частей: сервера, предоставляющего API для взаимодействия с библиотеками языка Python и формирования конечной программы, которая будет осуществлять работу с данными, и веб-интерфейса, доступ к которому можно получить посредством сети Интернет. Интерфейс состоит из двух областей: списка доступных для использования функций, представляемого в виде многоуровневого списка, и области, в которой выстраивается порядок выполнения операций над данными, представленного в виде отдельных вкладок для каждой функции. Макет интерфейса пользователя приведён на рисунке 1.

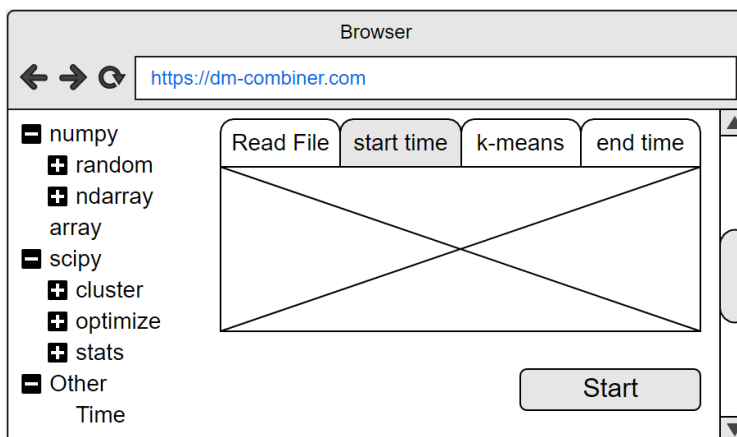


Рисунок 1 – Макет интерфейса пользователя системы

**Список использованных источников:**

1. Data Never Sleeps 7.0 [Электронный ресурс].– Режим доступа: <https://www.domo.com/learn/data-never-sleeps-7>.– Дата доступа: 25.03.2020
2. Барсегян, А.А. Анализ данных и процессов: учеб. пособие / А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. – 3-е изд., перераб. и доп.– СПб.: БХВ-Петербург, 2009. – 512 с.