**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ**
**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ**
**«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**
Факультет компьютерных наук
Департамент программной инженерии

СОГЛАСОВАНО
Научный руководитель,
Доцент департамента больших данных и информационного поиска, заведующая лабораторией «Научно-учебная лаборатория биоинформатики» факультета компьютерных наук

_____ М.С. Попцова
«___» _____ 2020 г

УТВЕРЖДАЮ
Академический руководитель образовательной программы «Программная инженерия», канд. техн. наук, профессор департамента программной инженерии факультета компьютерных наук

_____ В. В. Шилов
«___» _____ 2020 г.

| | |
|---|---|
| *Подп. и дата* | |
| *Инв. № дубл.* | |
| *Взам. инв. №* | |
| *Подп. и дата* | |
| *Инв. № подл* | |

# ПРОГРАММА ДЛЯ ИССЛЕДОВАНИЯ ПРИМЕНИМОСТИ МЕТОДОВ КЛАСТЕРИЗАЦИИ И СНИЖЕНИЯ РАЗМЕРНОСТИ ДЛЯ АВТОМАТИЧЕСКОГО СРАВНЕНИЯ ЭКСПЕРИМЕНТОВ ОДНОКЛЕТОЧНОГО СЕКВЕНИРОВАНИЯ

**Пояснительная записка**

## ЛИСТ УТВЕРЖДЕНИЯ

**RU.17701729.04.13-01 81 01-1-ЛУ**

Исполнитель
студент группы БПИ182
_____ / И. С. Егоров /
«___»_____ 2020 г.

**Москва 2020**

УТВЕРЖДЕН
RU.17701729.04.13-01 81 01-1-ЛУ

# ПРОГРАММА ДЛЯ ИССЛЕДОВАНИЯ ПРИМЕНИМОСТИ МЕТОДОВ КЛАСТЕРИЗАЦИИ И СНИЖЕНИЯ РАЗМЕРНОСТИ ДЛЯ АВТОМАТИЧЕСКОГО СРАВНЕНИЯ ЭКСПЕРИМЕНТОВ ОДНОКЛЕТОЧНОГО СЕКВЕНИРОВАНИЯ

**Пояснительная записка**

**RU.17701729.04.13-01 81 01-1**

**Листов 20**

| Инв. № подл | Подп. и дата | Взам. инв. № | Инв. № дубл. | Подп. и дата |
|---|---|---|---|---|
| | | | | |

**Москва 2020**

# АННОТАЦИЯ

В данном программном документе приведена пояснительная записка к «Программе для исследования применимости методов кластеризации и снижения размерности для автоматического сравнения экспериментов одноклеточного секвенирования» («The software to study the applicability of clustering and dimensionality reduction methods for automatically comparing unicellular sequencing experiments»), предназначенной для

В данном программном документе, в разделе «Введение», указано наименование программы и документы, на основании которых ведется разработка.

В разделе «Назначение и область применения» указано функциональное и эксплуатационное назначение программы и краткая характеристика области применения программы.

В разделе «Технические характеристики», содержатся следующие подразделы:

- постановка задачи на разработку программы;
- описание алгоритма и функционирования программы;
- описание и обоснование выбора метода организации входных и выходных данных;
- описание и обоснование выбора состава технических и программных средств.

В разделе «Технико-экономические показатели» указана предполагаемая потребность и экономические преимущества разработки по сравнению с отечественными и зарубежными образцами или аналогами.

Перед прочтением данного документа рекомендуется ознакомиться с терминологией, приведенной в списке используемой литературы настоящей Пояснительной записке.

Настоящий документ разработан в соответствии со следующими требованиями:

1) ГОСТ 19.101-77 Виды программ и программных документов [1];
2) ГОСТ 19.102-77 Стадии разработки [2];
3) ГОСТ 19.103-77 Обозначения программ и программных документов [3];
4) ГОСТ 19.104-78 Основные надписи [4];
5) ГОСТ 19.105-78 Общие требования к программным документам [5];
6) ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом [6];
7) ГОСТ 19.404-79 Пояснительная записка. Требования к содержанию и оформлению [7].

## СОДЕРЖАНИЕ

# 1. ВВЕДЕНИЕ

## 1.1.Наименование программы

Наименование программы – «Программа для исследования применимости методов кластеризации и снижения размерности для автоматического сравнения экспериментов одноклеточного секвенирования».

Наименование программы на английском языке – «The software to study the applicability of clustering and dimensionality reduction methods for automatically comparing unicellular sequencing experiments».

## 1.2. Основания для разработки

Разработка выполняется в рамках темы курсовой работы в соответствии с учебным планом подготовки бакалавров по направлению 09.03.04 «Программная инженерия» Национального исследовательского университета «Высшая школа экономики», факультет компьютерных наук, департамент программной инженерии.

Приказ декана факультета компьютерных наук И. В. Аржанцева "Об утверждении тем, руководителей курсовых работ студентов образовательной программы «Программная инженерия» факультета компьютерных наук" № 2.3-02/1112-04 от 11.12.2019.

# 2. НАЗНАЧЕНИЕ И ОБЛАСТЬ ПРИМЕНЕНИЯ

## 2.1. Функциональное назначение

Приложение представляет собой веб-интерфейс для определения сходства входных данных с имеющимися в коллекции.

## 2.2. Эксплуатационное назначение

Приложение предназначено для использования исследователями, ведущими научную работу в области одноклеточного секвенирования и занимающиеся идентификацией клеточных подтипов в когорте клеток.

# 3. ТЕХНИЧЕСКИЕ ХАРАКТЕРИСТИКИ

## 3.1.Постановка задачи

На вход подаётся база RNA-seq, полученная путём одноклеточного секвенирования, в формате 10x_mtx. По данной базе необходимо дать пользователю возможность выполнить одно из следующих действий – либо применить эти данные к уже обученному классификатору и вывести на экран разбиение по кластерам этих данных, либо построить на этих данных свой классификатор и предсказать по нему типы и имена клеток.

Таким образом вся задача делится на две части:

1) Реализации функционала, позволяющего: строить классификаторы по определённому набору данных; строить графики по данным классификатора; предсказывать имена и типы клеток по данным классификатора.
2) Реализация функционала, позволяющего: получать данные от пользователя; выводить пользователю графики в интуитивно-понятном формате; позволять пользователю по запросу получить предсказания имён и типов клеток на основе построенного классификатора.

## 3.2.Описание алгоритма и функционирования программы

Для решения данной задачи мною был выбран путь написания Web-приложения. Во-первых, для пользователя Web-приложение является самым простым методом решения задачи (нет необходимости устанавливать какие-либо сторонние плагины или выполнять сложные действия). Во-вторых, это обусловлено перспективой развития проекта в полноценный ресурс, используемый исследователями по всему миру.

Соответственно для Web-приложения был написан клиент и сервер. Клиент выполняет задачи общения с пользователем, загрузки базы пользователя, визуализации выходных данных в понятном для пользователя виде. Сервер же выполняет задачи построения тематической модели и обработки базы пользователя, в том числе формирования графиков, которые впоследствии будет видеть пользователь.

Далее опишу отдельно алгоритмы, применяемые мной при написании клиента и сервера, а также их функционал.

Работа сервера:

Для начала программа получает на вход набор файлов в формате 10x_mtx. Далее, в зависимости от запроса пользователя по входным данным либо строится новая тематическая модель при помощи алгоритма LDA, либо данные применяются к уже построенной модели. В случае, если было выполнено построение новой тематической модели, то на основе данных из неё делаются предсказания типов и имён клеток. Далее выводится распределение уникальных нуклеотидных идентификаторов по темам в форме графика.

В качестве основы построения тематической модели был выбран алгоритм LDA (Latent Dirichlet allocation). Это алгоритм, позволяющий разбивать некоторые данные по кластерам (топикам). Простым языком этот алгоритм можно объяснить так: пусть у нас есть набор научных статей, их достаточно много. Мы хотим по словам, из этих самых статей, определять тему данной статьи, а затем, статьи с одинаковой тематикой распределить по группам. В данном случае по словам из статей мы будем формировать некоторые кластеры, где кластер — это набор слов, встречаемых в схожем контексте. Далее берутся слова из каждой темы и пересекаются с текстом документов. Так документы распределяются по темам. Проецируя на поставленную задачу: вместо слов, будут гены-маркеры, а вместо научных статей, будут сами клетки, так как клеткам сопоставлен набор некоторых генов-маркеров. Учитывая, что изначально нам неизвестно что за клетка представлена тем или иным набором генов – маркеров, появляется необходимость предположения. Выполняется это аналогично

примеру со статьями. Берётся база клеток человека, в которой каждой клетке сопоставлены гены-маркеры (скачена с интернет-ресурса http://bio-bigdata.hrbmu.edu.cn/CellMarker/download.jsp). Далее для каждого кластера выполняется следующая операция: берутся топ-200 по вероятности вхождения в кластер гены; далее для каждой клетки из скаченной базы находится пересечение соответствующих ей генов-маркеров и генов из топа (конвертируется в численное значение, равное количеству совпавших генов, делённое на общее количество генов-маркеров для клетки); далее все клетки из модели сортируются по значению пересечения, и топ-5 клеток с максимальным значением выводятся пользователю в качестве предсказанных клеток для очередного кластера.

Работа клиента:

Пользователь загружает данные в архиве формата .zip, в котором содержатся файлы в формате 10x_mtx. Эти данные отправляются на сервер. После выполнения сервером всех необходимых операций, динамически создаются элементы, в которых содержатся графики. Далее графики выводятся на экран. Также создаются элементы – ссылки, по нажатию на которые у пользователя есть возможность скачать файл с предсказанными именами и типами клеток и интерактивный график в формате .html.

## 3.3. Описание и обоснование выбора метода организации входных и выходных данных

### 3.3.1. Описание метода организации входных и выходных данных

Входные данные представляют собой базу RNA-seq, полученную путём одноклеточного секвенирования, в формате 10x_mtx. (см. Приложение 3 Технического задания), упакованную в архив с расширением .zip.

Выходные данные представляют собой графики процентного распределения уникальных нуклеотидных идентификаторов по кластерам (количество уникальных нуклеотидных идентификаторов, которые были определены в тот или иной кластер в отношении к общему количеству уникальных нуклеотидных идентификаторов), интерактивный график в формате html (для удобного изучения построенной модели), а также .txt файл с предсказанными именами и типами клеток. *тут будут к каждому пункту картинки м примером*.

### 3.3.2. Обоснование выбора метода организации входных и выходных данных

Входные данные представляют собой базу единичных отсеквенированных клеток в формате 10x_mtx. Данный формат выбран в связи с тем, что он является достаточно распространённым в сфере исследований, связанных с одноклеточным секвенированием.

Выходные данные представлены в виде графиков, так как графики являются самым интуитивно-понятным методом передачи информации.

Файл с предсказаниями имеет формат .txt в силу его универсальности и работоспособности почти в любой операционной системе.

Данные в файле представлены в виде интуитивно – понятного формата для простоты работы с ними.

## 3.4. Описание и обоснование выбора состава технических и программных средств

### 3.4.1. Описание состава технических и программных средств

Для надежной работы программы необходим следующий состав программных средств:

1. любой браузер, способный открывать интернет-ресурсы и отображать картинки в формате .png;
2. программное обеспечение, поддерживающее открытие текстовых файлов в формате .txt.

Для надёжной и бесперебойной работы программы требуется следующий состав технических средств [12]:

1. персональный компьютер, оснащенный 32-разрядным (x86) или 64-разрядным (x64) процессором с тактовой частотой 1 ГГц или выше;
2. 20МБ оперативной памяти или больше;
3. не менее 10 МБ свободного места на жестком диске;
4. видеокарта и монитор, поддерживающие режим SuperVGA с разрешением не менее, чем 1024x768 точек;
5. мышь или совместимое указывающее устройство;
6. клавиатура;
7. подключение к сети Интернет.
8. веб-браузер.

### 3.4.2. Обоснование выбора состава технических и программных средств

Разработка данной программы велась под управлением ОС Linux, и при ее создании использовались языки Python 3, HTML, CSS, JavaScript.

Разработка серверной части велась полностью на языке Python 3, так как именно для данного языка написано огромное количество библиотек, позволяющих выполнить данную задачу.

Разработка клиентской части велась на языках JavaScript, HTML, CSS, так как эти языки являются самыми распространёнными для написания интерфейса и его логики.

20 Мб оперативной памяти требуется для стабильной работы веб-браузера.

10 Мб свободного места на диске требуется для сохранения текстового файла с предсказанными клетками.

# 4. ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ

## 4.1. Предполагаемая потребность

Данный продукт может быть интересен биологам-экспериментаторам и биоинформатикам, работающим в области одноклеточного секвенирования и занимающиеся идентификацией клеточных подтипов в когорте клеток.

## 4.2. Экономические преимущества разработки по сравнению с отечественными и зарубежными аналогами

В рамках данной работы расчет экономической эффективности не предусмотрен.

## СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. ГОСТ 19.101-77 Виды программ и программных документов. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
2. ГОСТ 19.102-77 Стадии разработки. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
3. ГОСТ 19.103-77 Обозначения программ и программных документов. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
4. ГОСТ 19.104-78 Основные надписи. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
5. ГОСТ 19.105-78 Общие требования к программным документам. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
6. ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
7. ГОСТ 19.404-79 Пояснительная записка. Требования к содержанию и оформлению. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
8. ГОСТ 19.603-78 Общие правила внесения изменений. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
9. ГОСТ 19.604-78 Правила внесения изменений в программные документы, выполненные печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
10. ГОСТ 15150-69 Машины, приборы и другие технические изделия. Исполнения для различных климатических районов. Категории, условия эксплуатации, хранения и транспортирования в части воздействия климатических факторов внешней среды. – М.: Изд-во стандартов, 1997.
11. Устинов В. Надежность оптических дисков: как их правильно хранить и использовать. //Журнал «625» №7. М.: Издательство «625», 2005.
12. Системные требования ОС Windows 7. [Электронный ресурс]// URL: http://windows.microsoft.com/systemrequirements?4bcfd458 (Дата обращения: 21.11.2018, режим доступа: свободный).
13. ГОСТ Р 7.02-2006 Консервация документов на компакт-дисках. Общие требования. – М.: ИПК Издательство стандартов, 2006.
14. ГОСТ 19.602-78 Правила дублирования, учета и хранения программных документов, выполненных печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
15. ГОСТ 19.301-79 Программа и методика испытаний. Требования к содержанию и оформлению. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
16. Graph Embeddings – The Summary — URL: https://towardsdatascience.com/graph-embeddings-the-summary-cc6075aba007
17. Cell Types Database: RNA-Seq Data— URL: https://portal.brain-map.org/atlases-and-data/rnaseq
18. SingleCellNet: A Computational Tool to Classify Single Cell RNA-Seq Data Across Platforms and Across Species — URL: https://ars.els-cdn.com/content/image/1-s2.0-S2405471219301991-mmc1.pdf
19. Topic Modeling with Gensim (Python) — URL: https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/
20. Topic Modeling in Python: Latent Dirichlet Allocation (LDA) — URL: https://towardsdatascience.com/graph-embeddings-the-summary-cc6075aba007
21. A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way — URL: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

22. A Gentle Introduction To Math Behind Neural Networks — URL: https://medium.com/datadriveninvestor/a-gentle-introduction-to-math-behind-neural-networks-6c1900bb50e1
23. Capture Cell Heterogeneity in Single Cell RNA-seq by Topic Modeling (Part One) — URL: https://medium.com/@yxu71/capture-cell-heterogeneity-in-single-cell-rna-seq-by-topic-modeling-part-one-f404f71deecf
24. DIMM-SC: A Dirichlet mixture model for clustering droplet-based single cell transcriptomic data — URL: https://arxiv.org/ftp/arxiv/papers/1704/1704.02007.pdf
25. SingleCellNet: A Computational Tool to Classify Single Cell RNA-Seq Data Across Platforms and Across Species — URL: https://www.sciencedirect.com/science/article/pii/S2405471219301991

# ПРИЛОЖЕНИЕ 1
# ОПИСАНИЕ И ФУНКЦИОНАЛЬНОЕ НАЗНАЧЕНИЕ КЛАССОВ И СКРИПТОВ

Таблица 1 — Описание и функциональное назначение классов

| **Класс** | **Назначение** |
|---|---|
| containerImage | Класс, отвечающий за стилизацию графов и подписей к ним. |
| choseFile | Класс, отвечающий за стилизацию кнопок ввода данных. |
| item | Класс, отвечающий за стилизацию элементов. |
| readyImages | Класс, отвечающий за стилизацию изображений графов. |
| textAddition | Класс, отвечающий за стилизацию подписей к графам. |
| textProgress | Класс, отвечающий за стилизацию текстового поля вывода состояния. |
| infoForDownload | Класс, отвечающий за стилизацию ссылок, по которым происходит скачивание файлов. |
| UseBuiltModel | Класс, реализующий применение данных пользователя к уже обученной модели. |
| BuiltModel | Класс, реализующий построение новой модели по пользовательским данным. |

Таблица 2 — Описание и функциональное назначение скриптов

| **Класс** | **Назначение** |
|---|---|
| 10x_mx_cutter.py | Скрипт, реализующий обработку входных данных и готовящий их к дальнейшим операциям с ними. |
| database_prepare.py | Скрипт, реализующий предсказание типов и имён клеток. |
| lda_theme_proportion.py | Скрипт, осуществляющий построение всех графиков. |
| learn_new_model.py | Скрипт, реализующий обучение новой модели. |
| views.py | Скрипт, реализующий логику работы сервера. |
| logic.js | Класс, реализующий логику работы клиента. |

## ПРИЛОЖЕНИЕ 2

## ОПИСАНИЕ И ФУНКЦИОНАЛЬНОЕ НАЗНАЧЕНИЕ ПОЛЕЙ, МЕТОДОВ И СВОЙСТВ

Таблица 3.1 — Описание и функциональное назначение полей, методов и свойств скрипта 10x_mx_cutter.py

| **Поля** | | | | |
|---|---|---|---|---|
| Имя | Модификатор доступа | Тип | | Назначение |
| path_to_files | - | - | | Содержит путь до папки с данными. |
| path_to_data | - | - | | Содержит путь до папки с данными, на которых была обучена модель. |
| input_str | - | - | | Содержит входную строку. |
| arr | - | - | | Массив, содержащий разбиение строки. |
| number_of_barcodes | - | - | | Количество штрихкодов в данных. |
| number_of_genes | - | - | | Количество генов в данных. |
| barcodes | - | - | | Файл с штрихкодами. |
| genes | - | - | | Файл с генами. |
| genes_data | - | - | | Содержит считанный построчно файл с генами. |
| number_of_genes_in_file | - | - | | Содержит количество генов в данных пользователя. |
| matrix | - | - | | Файл с матрицей. |
| data | - | - | | Содержит данные из матрицы. |
| next_line | - | - | | Содердит следующую строчку из матрицы. |
| Методы | | | | |
| Имя | Модификатор доступа | Тип | Аргументы | Назначение |
| cut_line | - | - | string | Разрезает строку и проверяет подходит ли она под параметры |

Таблица 2.2 — Описание и функциональное назначение полей, методов и свойств скрипта database_prepare.py

| **Поля** | | | |
|---|---|---|---|
| Имя | Модификатор доступа | Тип | Назначение |
| path_to_database | - | - | Содержит путь до базы с клетками. |
| path_to_lda_model | - | - | Содержит путь до модели. |
| path_to_data | - | - | Содержит путь до данных, на которых была обучена модель. |
| path_to_output_data | - | - | Содержит путь, куда нужно сохранить выходные данные. |
| file_name | - | - | Имя, которое должны иметь файла на выходе. |
| number_of_genes_in_topics | - | - | Количество генов, которые необходимо достать из каждого кластера. |
| data_base | - | - | База данных с клетками. |
| lines_in_database | - | - | Строки из базы. |
| base_dict | - | - | Словарь, построенный по базе с клетками. |
| lda_model | - | - | Модель. |
| model_data | - | - | Данные, на которых была построена модель. |
| genes_names | - | - | Имена генов. |
| genes_by_theme | - | - | Топ-гены, взятые из кластеров. |
| answer | - | - | Ответ на запрос. |
| genes_intersection | - | - | Массив пересечений клеток из базы с клетками из кластеров. |
| step_answer | - | - | Подобранные клетки для каждого кластера. |
| name_and_type | - | - | Имена и типы клеток. |
| sorted_values | - | - | Массив, отсортированных |

| | | | |
|---|---|---|---|
| | | | значений пересечений для каждого кластера. |
| file | - | - | Файл, содержащий предсказанные клетки. |
| data_for_pyLDA | - | - | Данные, необходимые для построения графика pyLDAvis. |
| vectorizer_n | - | - | Словарь с данными, на которых была обучена модель. |
| count_data | - | - | Индексатор словаря. |
| LDAvis_prepared | - | - | Построенный интерактивный граф pyLDAvis. |

Таблица 2.3 — Описание и функциональное назначение полей, методов и свойств скрипта lda_theme_proportion.py

| **Поля** | | | |
|---|---|---|---|
| Имя | Модификатор доступа | Тип | Назначение |
| path_to_model | - | - | Содержит путь до модели. |
| path_to_model_data | - | - | Содержит путь до данных, на которых была обучена модель. |
| path_to_users_data | - | - | Содержит путь до данных, введённых пользователем. |
| path_to_images | - | - | Содержит путь до папки, в которой должны быть сохранены картинки. |
| file_name | - | - | Содержит имя, которое должны иметь выходные файлы. |
| lda_model | - | - | Модель. |
| model_data | - | - | Данные, на которых была обучена модель. |
| data_out | - | - | Вырезка данных, на которых была обучена модель. |
| lda_output | - | - | Результат работы модели на даных. |

| | | | |
|---|---|---|---|
| df_cell_topic | - | - | Распределение штрихкодов по темам. |
| df_cell_dominant_topic | - | - | Для каждого штрихкода определён доминантный топик. |
| number_of_dominant_topics | - | - | Количество штрихкодов, для которых доминантный топик – i. |
| topics | - | - | Название топиков. |
| values | - | - | Значения количества штрихкодов для каждого топика. |
| dataFrame | - | - | Построенный график. |
| data | - | - | Данные пользователя. |
| data_test | - | - | Вырезка из данных пользователя. |
| topic_probability_scores | - | - | Вероятности попадания в тот или иной топик для каждого штрихкода. |
| df_cell_topic_user | - | - | Распределение штрихкодов по темам. (По пользовательским данным) |
| number_of_dominant_topics_user | - | - | Для каждого штрихкода определён доминантный топик. (По пользовательским данным) |

Таблица 2.4 — Описание и функциональное назначение полей, методов и свойств скрипта learn_new_model.py

| **Поля** | | | |
|---|---|---|---|
| Имя | Модификатор доступа | Тип | Назначение |
| lda_model_name | - | - | Имя модели, которая будет построена. |
| path_to_files | - | - | Путь до данных, на которых должна быть обучена новая модель. |
| number_of_themes | - | - | Количество топиков в модели. |

| | | | |
|---|---|---|---|
| a_Data | - | - | Данные, на которых должна быть построена модель. |
| path_to_model | - | - | Путь, по которому должна быть сохранена обученная модель. |
| data | - | - | Вырезка из данных, на которых должна быть обучена модель. |
| lda_model | - | - | Обученная модель. |
| lda_output | - | - | Выходные данные обученной модели. |

Таблица 2.5 — Описание и функциональное назначение полей, методов и свойств скрипта views.py

| **Методы** | | | | |
|---|---|---|---|---|
| Имя | Модификатор доступа | Тип | Аргументы | Назначение |
| clearData | - | - | string | Полностью очищает папку, путь до которой передан |
| useBloodModel | - | - | string | Применяет данные пользователя к уже обученной модели. |
| buildModelAndUseIt | - | - | string | Обучает новую модель по данным пользоваетля. |
| saveFile | - | - | string, string | Сохраняет переданную строку в файл. |
| formUrlOnServer | - | - | dictionary | Формирует пакет, который будет отправлен на клиент. |

Таблица 2.6 — Описание и функциональное назначение полей, методов и свойств скрипта logic.js

| **Методы** | | | | |
|---|---|---|---|---|
| Имя | Модификатор доступа | Тип | Аргументы | |
| ajaxForm | - | - | string, string, FormData, function | Отправляет ajax запрос с датой Form. |
| createImageField | - | - | string, string | Создаёт элемент с изображением и подписью к нему. |

| | | | | |
|---|---|---|---|---|
| appendImages | - | - | dict{string:string} | Добавляет созданный элемент с изображением и подписью к нему на интерфейс. |
| creteTextField | - | - | string | Создаёт текстовое поле. |
| createLinkToDownload | - | - | string, string | Создаёт ссылку для скачивания данных. |
| appendText | - | - | dict{string:string} | Добавляет текст на интерфейс пользователя. |
| deleteAllInfo | - | - | - | Удаляет элементы с клиента. |
| usingExistingModel | - | - | event | Позволяет обрабатывать событие нажатие кнопки. |
| usingJustBuildModel | - | - | event | Позволяет обрабатывать событие нажатие кнопки. |
| **Поля** | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |
| app | - | - | | Контейнер, содержащий элементы с интерфейса клиента. |
| form | - | - | | Контейнер, содержащий элементы с интерфейса клиента. |
| usingModelEvent | - | - | | Элемент, отвечающий за то, чтобы по нажатию на кнопку выполнялись определённые действия. |
| buildNewModelBtn | - | - | | Элемент, отвечающий за то, чтобы по нажатию на кнопку выполнялись определённые действия. |

Таблица 2.7 — Описание и функциональное назначение полей, методов и свойств класса unary UseBuiltModel

| **Методы** |
|---|

| Имя | Модификатор доступа | Тип | Аргументы | Назначение |
|---|---|---|---|---|
| post | - | - | - | Отвечает за выполнение действий с данными пользователя на уже обученной модели. |

Таблица 2.8 — Описание и функциональное назначение полей, методов и свойств класса unary BuiltModel

| **Методы** | | | | |
|---|---|---|---|---|
| Имя | Модификатор доступа | Тип | Аргументы | Назначение |
| post | - | - | - | Строит новую модель по данным пользователя и выполняет с ней действия |

## ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

| Лист регистрации изменений | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Номера листов (страниц) | | | | | Всего листов (страниц в докум.) | № документа | Входящий № сопроводительного докум. и дата | Подп. | Дата |
| Изм. | Измененных | Заменен ных | Новых | Аннулированх | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

| | | | | |
|---|---|---|---|---|
| Изм. | Лист | № докум. | Подп. | Дата |
| RU.17701729.04.01 —01 81 | | | | |
| Инв. № подл. | Подп. и дата | Взам. инв. № | Инв. № дубл. | Подп. и дата |