

Санкт-Петербургский государственный университет

Кафедра системного программирования

Группа 20Б.11-мм

Федькин Александр Андреевич

Оценка личностных особенностей
пользователя по данным страницы в
Instagram

Отчёт по учебной практике
в форме «Решение»

Научный руководитель:
к.т.н., доцент кафедры информатики СПбГУ Абрамов М. В.

Консультант:
м.н.с. СПб ФИЦ РАН Корепанова А. А.

Санкт-Петербург
2022

Оглавление

Введение	3
1. Постановка задачи	5
2. Обзор	6
2.1. Существующие решения	6
2.1.1. Задача о предсказании личностных особенностей	6
2.1.2. Приложение для сбора данных	10
2.2. Используемые технологии и инструменты	11
3. Приложение для сбора данных	13
3.1. Архитектура приложения для сбора данных	13
3.2. Детали реализации	14
3.2.1. Авторизация аккаунтов	14
3.2.2. Обеспечение безопасности аккаунтов	14
3.2.3. Цепочки запросов	15
4. Предсказание личностных особенностей	16
4.1. Сбор датасета	16
4.2. Анализ датасета	16
4.3. Признаки	20
4.4. Извлечение признаков	21
4.5. Обучение моделей и оценка качества	22
Заключение	26
Список литературы	27

Введение

В настоящее время социальные сети являются важной платформой для самовыражения и обмена мнениями¹. Instagram, насчитывающий более одного миллиарда пользователей по состоянию на декабрь 2021 года², является одной из самых распространенных социальных сетей. Основное применение Instagram — позволить пользователям делиться своими фото и видео, находить людей с помощью хештегов, а также заводить новые знакомства. На данный момент Instagram признан экстремистской организацией в России³.

Согласно модели FFM, [5] личность человека может быть описана пятью независимыми параметрами: экстраверсия, привязанность, самоконтроль, эмоциональная устойчивость, экспрессивность. Информация о личностных характеристиках человека позволяет делать предположения о реакции на различные жизненные ситуации [7]. Оценка степени выраженности личностных особенностей пользователя может облегчить понимание его потребностей, что крайне важно в системах рекомендаций и маркетинга в задачах поиска целевой аудитории для продукта.

Личностная характеристика человека находит отражение в параметрах его профиля социальной сети [3]. Предполагается, что из информации о взаимодействии пользователя Instagram с другими людьми можно извлечь пять параметров FFM, характеризующих личность человека.

Модели оценки в существующих работах по большей части основываются на визуальных признаках, то есть на анализе фотографий. Однако не всегда профиль пользователя содержит достаточное количество фотографий для осуществления оценки. Поэтому видится целесообразным использовать социальные признаки, а именно списки подписок и подписчиков, информацию о публикациях для оценки личностных осо-

¹<https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021> (дата обращения: 2021-12-11).

²<https://www.businessofapps.com/data/instagram-statistics/> (дата обращения: 2021-12-11)

³<https://mos-gorsud.ru/rs/tverskoj/services/cases/civil/details/de7ea6a0-a3ab-11ec-8a7e-51b31fb55b35?participants=meta>

бенностей пользователя.

Целью исследования является сравнение качества моделей машинного обучения, обученных на визуальных, социальных признаках, а также на их комбинации.

Практическая значимость исследования заключается в повышении качества систем рекомендаций и маркетинга при использовании модели оценки степени выраженности личностных особенностей пользователя по его профилю социальной сети.

Степень разработанности темы: на базе лаборатории теоретических и междисциплинарных проблем информатики (ТиМПИ) СПб ФИЦ РАН были сделаны наработки по задачам анализа социальных сетей. По задаче, решаемой в рамках текущей учебной практики, научным руководителем и другими сотрудниками лаборатории ТиМПИ была произведена спецификация программы и формализация задачи, часть из которых нашла отражение в публикациях [8, 9]. Работа велась в рамках более широкого проекта по анализу защищенности пользователей информационных систем от социоинженерных атак (государственное задание СПб ФИЦ РАН № 0073-2019-003; проект РФФИ №20-07-00839).

1. Постановка задачи

Целью данной учебной практики является разработка модели оценки степени выраженности личностных особенностей пользователя по его профилю Instagram.

Для достижения цели были поставлены следующие задачи:

Осенний семестр:

- произвести обзор существующих решений;
- разработать приложение с графическим интерфейсом для сбора данных из Instagram.

Весенний семестр:

- собрать датасет: имя пользователя Instagram — результат психологического теста;
- обучить три модели: основанные на визуальных, социальных признаках и их комбинации;
- произвести сравнение моделей.

2. Обзор

В данном разделе представлен обзор существующих решений задачи сопоставления профиля пользователя Instagram его личностных характеристик. Представлен обзор аналогов приложения для сбора данных. Обоснован выбор используемых при разработке инструментов.

2.1. Существующие решения

2.1.1. Задача о предсказании личностных особенностей

Исследователи проявляют интерес к задаче о предсказании личностных особенностей человека по его профилю социальной сети, что подтверждается увеличением числа статей на эту тему с течением времени. Процесс поиска решений осуществлялся на платформах Web of Science и SCOPUS среди статей, выпущенных за последние пять лет, с числом цитирований большим двадцати. Данные критерии выбора позволяют выявить наиболее актуальные и значимые работы. Среди множества найденных работ были выделены три лучшие по качеству обученной модели:

1. Fusing Social Media Cues: Personality Prediction from Twitter and Instagram [4]
2. Predicting Users' Personality from Instagram Pictures: Using Visual and/or Content Features? [2]
3. Using computer vision techniques on Instagram to link users' personalities and genders to the features of their photos: An exploratory study [10]

Рассмотрим каждую работу отдельно:

1. В работе [4] было произведено сравнение моделей, обученных на визуальных и лингвистических признаках.

К визуальным признакам относятся:

- количество лиц на фото;
- эмоции на лицах;
- доля пикселей, которые попали в определенный диапазон оттенков.

К признакам, извлеченных из текста относятся:

- классификация типа текста;
- количество некоторых ключевых слов в тексте.

Датасет состоит из результатов опроса 193 пользователей Instagram из США с медианным возрастом — 31.

По результатам обучения модель, основанная на визуальных признаках показала себя лучше модели, основанной на лингвистических признаках. Усредненное значение функции ошибки при кросс-валидации модели, основанной на визуальных признаках, представлено в таблице:

Характеристика	RMSE
Открытость	0.51
Добросовестность	0.67
Экстраверсия	0.71
Доброжелательность	0.50
Нейротизм	0.73

Таблица 1: Оценка качества модели, предложенной в [4], при кросс-валидации

2. В работе [2] было произведено сравнение моделей, основанных на визуальных, контентных и комбинированных признаках.

Визуальные признаки такие же, как в работе [4].

Чтобы извлечь контентные признаки, был произведен подсчет количества объектов из различных категорий на фото пользователей с помощью Google Vision API. По итогам этого этапа авторы

получили более 4000 категорий, которые впоследствии были сжаты до 400 с помощью метода ближайших соседей, и затем эти 400 категорий были вручную сжаты в 17. Таким образом, контентными признаками выступали количества объектов из различных категорий, которые были изображены на фотографиях пользователя.

Комбинированные признаки — это совокупность визуальных и контентных.

Датасет состоит из опроса 179 пользователей Instagram из КНДР с медианным возрастом — 22.

По итогам обучения выяснилось, что добавление контентных признаков к визуальным, как и визуальных к контентным не увеличивает качества модели. Был сделан вывод, что добавление признаков не несет какой-либо полезной информации для обучения.

Оценка качества модели, основанной на комбинированных признаках приведена в таблице:

Характеристика	RMSE
Открытость	0.71
Добросовестность	0.62
Экстраверсия	0.98
Доброжелательность	0.61
Нейротизм	0.89

Таблица 2: Оценка качества модели, предложенной в [2], при кросс-валидации

3. В работе [10] были проверены гипотезы о корреляции параметров профиля пользователя с чертами характера, а также обучены три модели.

Первая модель основана на контентных признаках. Алгоритм извлечения признаков аналогичен алгоритму из работы [2], за исключением того, что использовался Microsoft Azure вместо Google Vision API.

Вторая модель основана на пиксельных признаках, то есть на различной информации об оттенках, яркости, контрастности и насыщенности фотографий.

Третья модель основана на признаках, извлекаемых из лиц людей с фотографий: их количества, а также их эмоций.

Датасет состоит из опроса 62 пользователей Instagram, Twitter из США с числом опубликованных фотографий большим тридцати.

По итогам обучения выяснилось, что первая и вторая модель показывают примерно одинаковые результаты при кросс валидации, в свою очередь, вторая модель, основанная на анализе лиц, показывает наилучшее значение функции ошибки при кросс-валидации.

Оценка качества модели, основанной на анализе лиц приведена в таблице:

Характеристика	RMSE
Открытость	0.60
Добросовестность	0.55
Экстраверсия	0.59
Доброжелательность	0.58
Нейротизм	0.66

Таблица 3: Оценка качества модели, предложенной в [10], при кросс-валидации

Помимо этого, существуют работы по другим социальным сетям, например: Twitter, Facebook [1, 6]. В них зачастую используются другие признаки, поскольку ключевыми объектами в этих социальных сетях являются не фотографии, а другие сущности. Например, в Twitter — это текстовые публикации, поэтому используются признаки, связанные с анализом текста.

Подводя итог, большинство существующих работ по Instagram основано исключительно на анализе фотографий пользователя. Наилучшее качество показывают модели, основанные на анализе лиц. Важно отметить, что ни в одной работе не использовались другие параметры

профиля, такие как: подписки, подписчики, при обучении моделей. Поэтому в рамках данной учебной практики планируется обучить модель, основанную на этих признаках, и сравнить с моделью, основанной на визуальных признаках.

2.1.2. Приложение для сбора данных

Поиск аналогов производился по первым двум страницам выдачи по запросам “instagram data parser”, “instagram data scraper”, “instagram data extractor” к поисковой системе Google.

Было выделено четыре наиболее удачных по функциональности и удобству использования решений. Среди них phantombuster, instaloader, botster, apify. Сравнение инструментов производилось по четырем параметрам:

- наличие нужной функциональности: выгрузка подписчиков, подписок, контента;
- поддержка создания цепочек запросов — последовательности вложенных запросов, где результат очередного запроса основывается на результате предыдущего, что является полезным для формирования социального графа;
- наличие графического интерфейса;
- инструмент является бесплатным.

Параметр	phantombuster	instaloader	botster	apify
Функциональность	+	+	+	+
Цепочки запросов	-	-	-	-
Графический интерфейс	+	-	+	+
Бесплатный	-	+	-	-

Таблица 4: Сравнение инструментов для сбора данных из Instagram

Сравнение показало, что ни один из выбранных инструментов не поддерживает создание цепочек запросов. Также все инструменты соответствуют лишь двум поставленным требованиям. Исходя из этого, было принято решение разработать новое приложение для сбора данных.

2.2. Используемые технологии и инструменты

Python был выбран в качестве основного языка программирования. Выбор обоснован наличием богатого набора библиотек для работы с данными и их анализа, а также существованием подходящей библиотеки для взаимодействия с API Instagram.

Instagrapi — легковесная библиотека для сбора данных, является оберткой над API Instagram. Обладает множеством полезных функций, поддержка автоматического ввода кода подтверждения, приходящего на заданную почту при авторизации, выгрузка данных в формате JSON.

Firebase Realtime Database — облачная NoSQL база данных, в которой данные хранятся в формате JSON. Обеспечивает хранение информации о пользователях в древовидном формате. Данный формат хранения позволяет удобно импортировать и экспортировать информацию о конфигурации аккаунтов-исполнителей.

PyQT5 — набор модулей для создания графического интерфейса. Поддерживает интеграцию с Qt Designer — визуальным конструктором графического интерфейса, что быстро создать прототип графического интерфейса на языке Python.

NumPy, pandas — библиотеки для работы с многомерными и табличными данными соответственно. Повсеместно используются при анализе данных, лежат в основе многих других библиотек.

OpenCV — набор алгоритмов компьютерного зрения. Использовалась для перевода фотографий между каналами RGB и HSV, а также извлечения визуальных признаков.

DeepFace — библиотека, содержащая предобученные модели для

выделения лиц на фотографий, а также определения эмоций. Использовалась для извлечения визуальных признаков.

scikit-learn — набор пакетов для машинного обучения, работы с датасетом, оценки качества моделей. Использовался по причине наличия всей необходимой функциональности и также большого количества обучающих материалов.

3. Приложение для сбора данных

В данном разделе представлено описание реализации приложения для сбора данных. Код доступен в репозитории⁴.

3.1. Архитектура приложения для сбора данных

Общая архитектура решения приведена на рисунке 1.

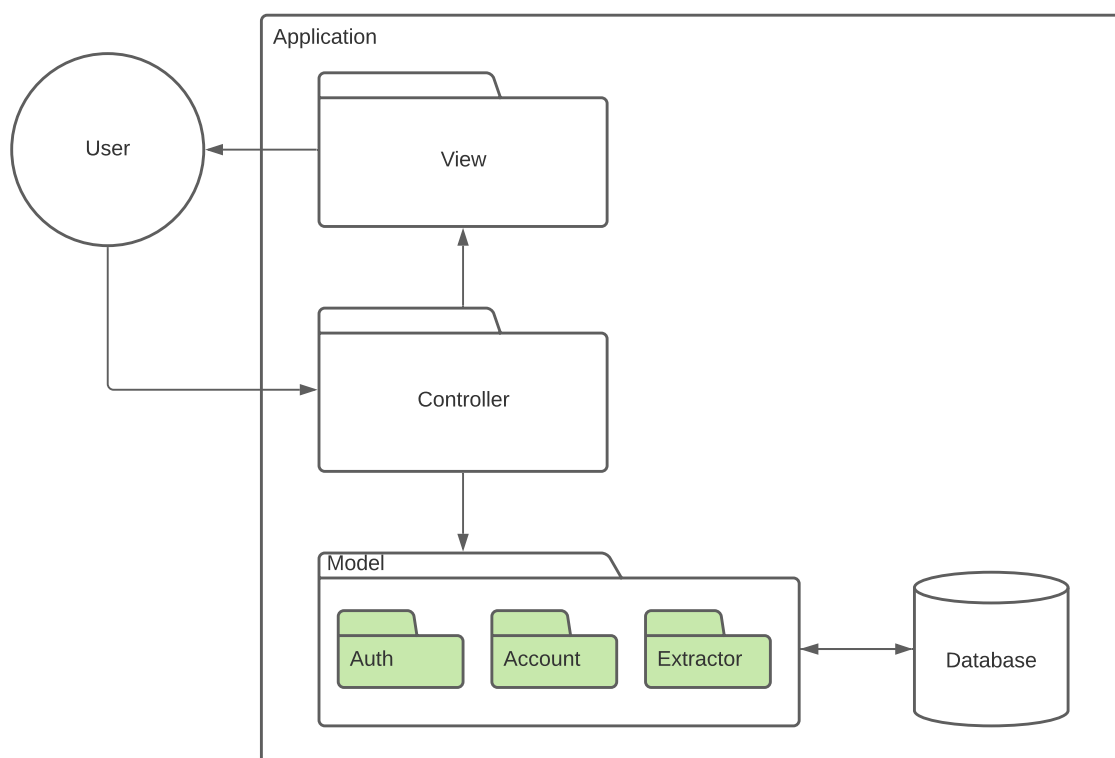


Рис. 1: Общая архитектура предлагаемого решения.

Рассмотрим подробнее компоненты:

- **Model** содержит логику сбора данных, обеспечения безопасности аккаунтов-исполнителей, а также логику авторизации аккаунтов и взаимодействия с базой данных.
- **View** отвечает за представление элементов графического пользовательского интерфейса.

⁴<https://github.com/qsqnk/instagram-data-extractor> (дата обращения: 2021-12-11)

- **Controller** реагирует на взаимодействие пользователя с графическим интерфейсом: обновляет представление, а также уведомляет Model о необходимости совершить некоторое действие.
- **Database** — база данных, хранящая в себе информацию об используемых для выгрузки акканутах-исполнителей.

3.2. Детали реализации

Для выгрузки данных из Instagram была использована библиотека Instagrami, которая работает на основе HTTP-запросов к API Instagram.

3.2.1. Авторизация аккаунтов

Instagram не позволяет получить некоторую информацию неавторизованным пользователям, поэтому возникла необходимость осуществлять выгрузку с использованием авторизованных аккаунтов. Для этого пользователю достаточно передать в приложение файл, где на каждой строке будет логин и пароль аккаунта Instagram.

3.2.2. Обеспечение безопасности аккаунтов

Перед выполнением задачи, например, выгрузки подписчиков конкретного пользователя, с вероятностью 0.1 аккаунт-исполнитель некоторым образом взаимодействует с целевым аккаунтом, например, подписывается на него или оценивает случайную публикацию. После выполнения задачи аккаунт-исполнитель временно замораживается на случайный срок от одной до трех минут. Во время заморозки отправка запросов с помощью этого аккаунта невозможна. Если число запросов к API за последние сутки превышает двести, то аккаунт замораживается на случайный срок от восьми до двенадцати часов. Эти действия нужны для имитации поведения реального пользователя социальной сети и избежания блокировок аккаунта. Сроки заморозок, максимальное количество запросов в день были выбраны исходя из средней активности

пользователя Instagram⁵.

3.2.3. Цепочки запросов

При создании социального графа возникает необходимость создания последовательных запросов на выгрузку некоторых данных. Например, для каждого подписчика пользователя социальной сети получить список его подписчиков. Чтобы позволить пользователю приложения получить желаемое за один запрос, была реализована возможность создания цепочек запросов. На любом шаге цепочки можно сделать отвлечение, чтобы получить некоторые данные по списку пользователей, но не продолжать цепочку с полученными данными. Также реализована возможность сохранения промежуточного результата выполнения одного шага последовательности. В звеньях цепочки могут быть задачи выгрузки подписчиков, подписок, контента, а также информации о пользователе.

Функции обработки последовательности запросов в качестве аргументов передаются начальный список юзернеймов социальной сети и сама последовательность запросов. Не замороженным аккаунтам выдаются задачи и затем загружаются в пул потоков. Если список юзернеймов, которых нужно обработать, больше количества свободных аккаунтов, то раз в несколько секунд осуществляется проверка на наличие свободных аккаунтов. Когда некоторый аккаунт разморозится, исполнение продолжится. По завершении обработки одного звена цепочки аккаунты получают временную заморозку в соответствии с числом запросов к API за последние сутки. Если задача была не последней в цепочке, то начальным списком юзернеймов становится результат выполнения последнего запроса и процесс продолжается итеративно.

⁵<https://www.digitalmarketing.org/blog/how-much-time-does-the-average-person-spend-on-social-media> (дата обращения: 2021-12-11).

4. Предсказание личностных особенностей

4.1. Сбор датасета

Датасет представляет из себя множество пар: имя пользователя Instagram — результаты психологического теста. Для сбора датасета были наняты 279 пользователей Instagram через биржу микрозадач. Респондент должен был пройти психологический тест и оставить ссылку на свой Instagram аккаунт. Фильтрация датасета производилась по следующим условиям:

- малое количество фотографий;
- большое количество фотографий рекламного характера или развлекательных изображений;
- нечестное прохождение опроса.

Честность прохождения опроса проверялась по двум параметрам: время прохождения, результаты опроса. Если время прохождения меньше трех минут, то респондент тратил около одной секунды на каждый вопрос, что с высокой вероятностью говорит о нечестности прохождения. Также если по результатам опроса у человека степень выраженности всех черт характера близка к максимуму, то, скорее всего, человек нажимал на первый вариант ответа, чтобы как можно быстрее пройти опрос.

4.2. Анализ датасета

На таблицах ниже представлен анализ предсказываемых степеней выраженности личностных характеристик в виде гистограмм, где $S.D.$ — стандартное отклонение, $mean$ — среднее, $range$ — отрезок от минимального до максимального значения. Число, означающее степень выраженности любой черты характера находится в отрезке $[1, 5]$.

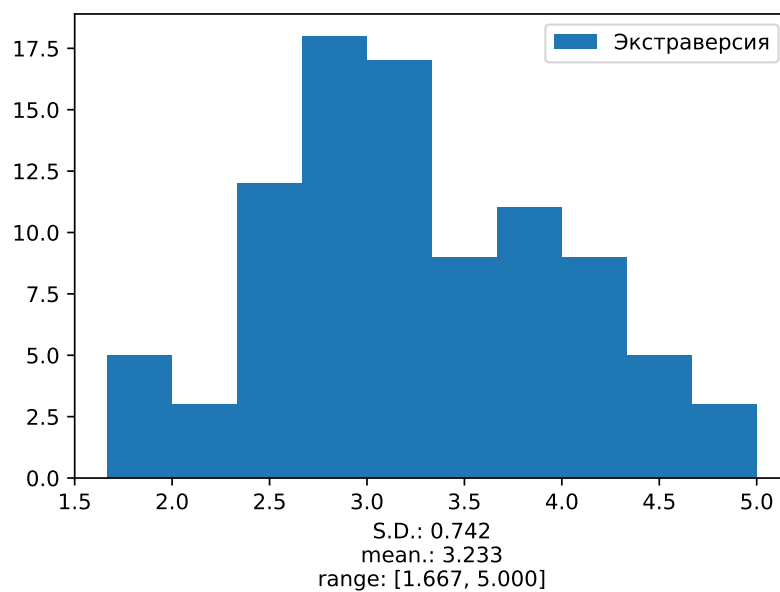


Рис. 2: Распределение степени экстраверсии в датасете.

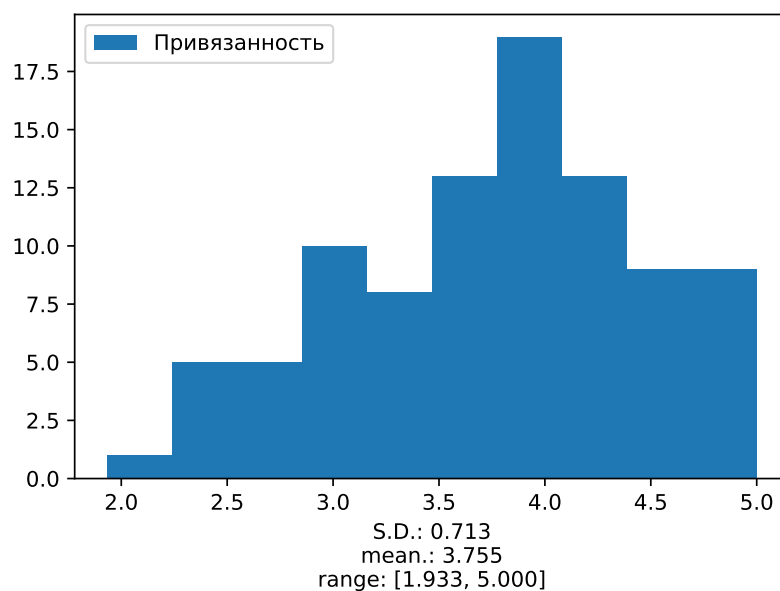


Рис. 3: Распределение степени привязанности в датасете.

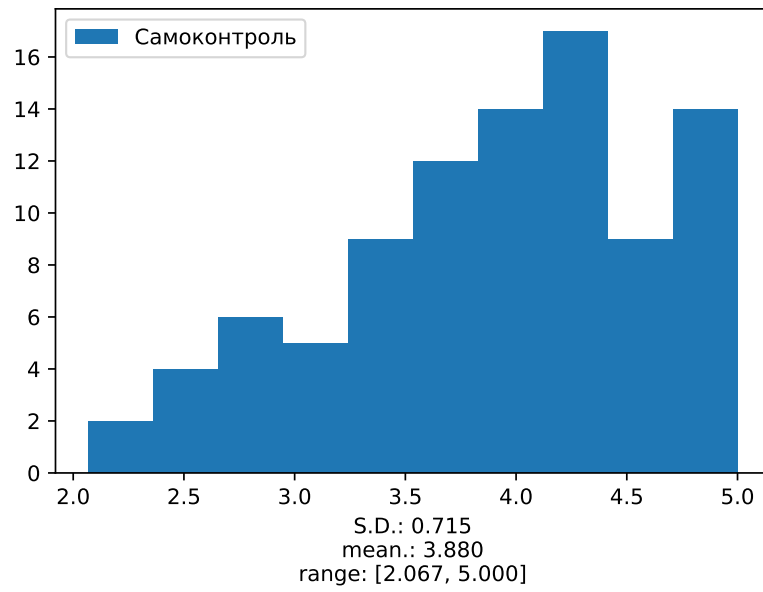


Рис. 4: Распределение степени самоконтроля в датасете.

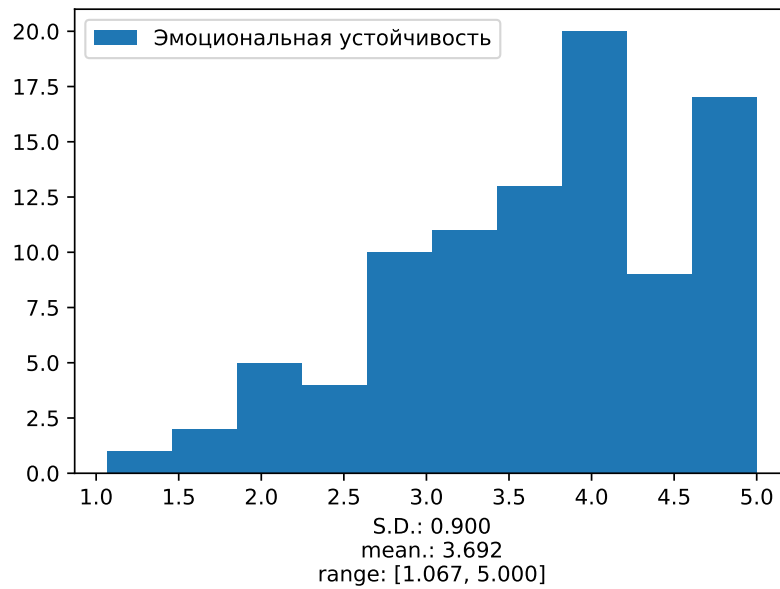


Рис. 5: Распределение степени эмоциональной устойчивости в датасете.

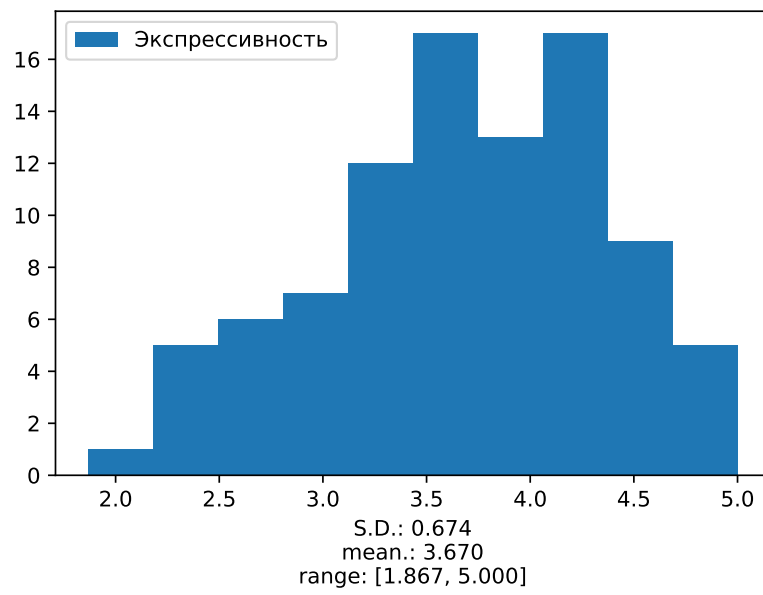


Рис. 6: Распределение степени экспрессивности в датасете.

Степень эмоциональной устойчивости имеет наибольшее стандартное отклонение — 0.9, в отличие от остальных величин, у которых этот показатель приблизительно равен 0.7. Это свидетельствует о сложности предсказания степени выраженности данной характеристики, что подтверждается работами [4, 2].

Стоит отметить, что для каждой черты характера нашелся человек, имеющий ее максимальный показатель степени выраженности, однако нельзя утверждать, что это признак нечестного прохождения опроса, поскольку эти люди имеют приемлемое время прохождения, а также нормальные показатели по другим параметрам личности. Если убрать этих людей из датасета, то значительно понижается среднеквадратическое отклонение у каждой величины, что влечет улучшение значения метрики ошибки при обучении. Но было принято решение этого не делать по описанной выше причине.

4.3. Признаки

Признаки, используемые при обучении модели можно разделить на две категории: визуальные и социальные. К визуальным относятся признаки, которые можно извлечь из фотографий пользователя. К социальным — все остальные, то есть, которые можно извлечь из информации о профиле пользователя.

Признаки, используемые в наших моделях:

Визуальные:

- яркость;
- насыщенность;
- оттенок
- число лиц на фотографиях;
- эмоции на лицах:

- страх;
- отвращение;
- счастье;
- грусть;
- удивление;
- нейтральность;

Социальные:

- число подписчиков;
- число подписок;
- число публикаций;
- число лайков;
- число комментариев.

Степень выраженности каждой эмоции выражается числом от 0 до 100. Все признаки, кроме числа подписчиков, подписок, публикаций, были усреднены.

4.4. Извлечение признаков

Первым этапом извлечения признаков является сбор данных о профиле пользователя с помощью приложения, разработанного в осеннем семестре. Результатом этого этапа являются JSON с информацией о пользователе, а также его фотографии.

После сбора данных для извлечения социальных признаков не нужно совершать каких-либо дополнительных действий, кроме усреднения числа лайков и комментариев на публикациях.

Для извлечения визуальных признаков были использованы библиотеки OpenCV и DeepFace. OpenCV использовалась для перевода фотографий между каналами RGB и HSV, а также для подсчета средних

яркости, насыщенности, оттенка. Для выделения лиц на фотографиях, а также определения эмоций использовалась библиотека DeepFace. Результатом работы библиотек является JSON, содержащий для каждой фотографии пользователя описание лиц и параметров фотографии в формате:

```
1 'emotions': [  
2   {'angry': 87.5247,  
3     'disgust': 0.0097,  
4     'fear': 0.0128,  
5     'happy': 12.0156,  
6     'sad': 0.3954,  
7     'surprise': 2.7406,  
8     'neutral': 0.0414},  
9   {'angry': 44.7129,  
10    'disgust': 0.0010,  
11    'fear': 9.3497,  
12    'happy': 34.2510,  
13    'sad': 0.73262,  
14    'surprise': 0.0015,  
15    'neutral': 10.9509}  
16 ],  
17 'hue': 43.4269,  
18 'saturation': 18.5650,  
19 'brightness': 16.4630
```

4.5. Обучение моделей и оценка качества

Модели были обучены на алгоритмах градиентного бустинга и случайного леса. Подбор гиперпараметров алгоритмов обучения с помощью поиска по сетке проводился независимо для каждой черты личности. В качестве функции ошибки была выбрана функция корня сред-

неквадратичной ошибки $RMSE = \sqrt{\frac{\sum_{i=1}^N (a(x_i) - y_i)^2}{N}}$, где $a(x_i)$ — результат предсказания, y_i — верное значение. Предсказываемые значения являются числами из отрезка $[1, 5]$, поэтому $RMSE \in [0, \sqrt{24}]$. Для оценки качества для каждой модели была произведена кросс-валидация с 10 итерациями, затем полученные 10 значений функции ошибки усреднялись. Оценки качества каждой модели представлены на таблицах, где RF — случайный лес, GB — градиентный бустинг:

Черта характера	Алгоритм	RMSE
Экстраверсия	RF	0.7436
	GB	0.7702
Привязанность	RF	0.7054
	GB	0.7565
Самоконтроль	RF	0.6920
	GB	0.7388
Устойчивость	RF	0.8775
	GB	0.8957
Экспрессивность	RF	0.6767
	GB	0.6796

Таблица 5: Качество модели, основанной на визуальных признаках

Черта характера	Алгоритм	RMSE
Экстраверсия	RF	0.7511
	GB	0.8143
Привязанность	RF	0.7205
	GB	0.8163
Самоконтроль	RF	0.7155
	GB	0.7823
Устойчивость	RF	0.8684
	GB	0.9323
Экспрессивность	RF	0.6686
	GB	0.7263

Таблица 6: Качество модели, основанной на социальных признаках

Черта характера	Алгоритм	RMSE
Экстраверсия	RF	0.7470
	GB	0.7542
Привязанность	RF	0.7093
	GB	0.7336
Самоконтроль	RF	0.6939
	GB	0.7067
Устойчивость	RF	0.8742
	GB	0.8765
Экспрессивность	RF	0.6671
	GB	0.6952

Таблица 7: Качество модели, основанной на комбинированных признаках

Модель, основанная на социальных признаках, позволяет лучше предсказывать устойчивость и экспрессивность, однако по другим психологическим параметрам выигрывает модель, основанная на визуальных признаках. Добавление социальных признаков к визуальным в третьей модели позволяет улучшить качество предсказания экспрессивности и устойчивости, однако качество по остальным параметрам остается практически без изменений. Таким образом, использование модели на социальных признаках имеет место по трем причинам:

- простота извлечения социальных признаков;
- улучшение качества предсказания двух параметров личности;
- несильное ухудшение качества по остальным параметрам.

Заключение

В ходе выполнения данной работы были достигнуты следующие результаты:

- изучена предметная область, произведен обзор аналогов и существующих решений;
- разработано приложение с графическим интерфейсом для выгрузки данных из Instagram.
- собран и отфильтрован датасет, состоящий из 96 пользователей;
- обучены три модели, основанные на социальных, визуальных и комбинированных признаках;
- произведено сравнение моделей.

Код доступен в репозитории⁶, размещённом на веб-сервисе GitHub.

⁶<https://github.com/qsqnk/instagram-data-extractor> (дата обращения: 2021-12-11).

Список литературы

- [1] A. Souril S. Hosseinpour A. Rahmani. Personality classification based on profiles of social networks' users and the five-factor model of personality // [HUMAN-CENTRIC COMPUTING AND INFORMATION SCIENCES](#). — 2018. — Vol. 8.
- [2] B. Ferwerda M. Tkalcic. [Predicting Users' Personality from Instagram Pictures: Using Visual and/or Content Features?](#) // PROCEEDINGS OF THE 26TH CONFERENCE ON USER MODELING, ADAPTATION AND PERSONALIZATION (UMAP'18). — 2018. — P. 157–161.
- [3] G. Vinitzky Y. Amichai-Hamburger. Social network use and personality // [COMPUTERS IN HUMAN BEHAVIOR](#). — 2010. — Vol. 26. — P. 1289–1295.
- [4] M. Skowron M. Tkalcic B. Ferwerda M. Schedl. [Fusing Social Media Cues: Personality Prediction from Twitter and Instagram](#) // PROCEEDINGS OF THE 25TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB (WWW'16 COMPANION). — 2016. — P. 107–108.
- [5] P. Costa R. McCrae. A five-factor theory of personality // The Five-Factor Model of Personality: Theoretical Perspectives. — 1999. — Vol. 2. — P. 51–87.
- [6] T. Tandra D. Suhartono R. Wongso Y. Prasetyo. [Personality Prediction System from Facebook Users](#) // DISCOVERY AND INNOVATION OF COMPUTER SCIENCE TECHNOLOGY IN ARTIFICIAL INTELLIGENCE ERA. — 2017. — Vol. 116. — P. 604–611.
- [7] U. Kumar A. Reganti T. Maheshwari T. Chakraborty. Inducing Personalities and Values from Language Use in Social Network Communities // [INFORMATION SYSTEMS FRONTIERS](#). — 2018. — Vol. 20. — P. 1219–1240.

- [8] V. Oliseenko M. Abramov A. Tulupyev. Identification of user accounts by image comparison: The phash-based approach // [Scientific and Technical Journal of Information Technologies, Mechanics and Optics](#). — 2021. — Vol. 21. — P. 562–570.
- [9] V. Oliseenko T. Tulupyeva M. Abramov. Online Social Network Post Classification: A Multiclass approach // [Lecture Notes in Networks and Systems](#). — 2022. — Vol. 330. — P. 207–215.
- [10] Y. Kim J. Kim. Using computer vision techniques on Instagram to link users' personalities and genders to the features of their photos: An exploratory study // [INFORMATION PROCESSING & MANAGEMENT](#). — 2018. — Vol. 54. — P. 1101–1114.