Содержание

1 Введение

1.1 Расстройство личности

Расстройство личности с 1960-х достигло высокого уровня изученности медицинским сообществом, и в настоящее время признано психиатрией как серьёзное заболевание [5]. Поэтому теперь оно рассматривается наравне с другими психическими расстройствами. В качестве критериев наличия личностного расстройства в одиннадцатом издании международной классификации болезней (МКБ-11) предлагаются наличие неадекватных способов познания, поведения, эмоциональных переживаний и реакций. Также выделяются проблемы в психосоциальном функционировании, что сильнее всего проявляется в межличностных отношениях. При этом проявление нарушений характерно для различных межличностных и социальных ситуаций, а не ограничивается конкретными случаями. Нарушение относительно стабильно во времени и продолжительно, наиболее часто расстройство личности впервые обнаруживается в детстве и выраженно проявляется в подростковом возрасте [16]. Далее при постановке диагноза определяют тяжесть течения и уже потом особенности личности у конкретного пациента.

В мире распространненость расстройства личности растёт, и по оценке ВОЗ к 2018 году оно затрагивало около 7.8% населения [3]. Некоторые из этих людей переносят тяжелые формы расстройства, что вносит нарушения в функционирование общества [5]. Это подчёркивает важность проведения терапии. В ходе ряда исследований было обнаружено, что выявление личностных расстройств оказывает сильное влияние на успешность лечения прочих психических заболеваний при совместном протекании с ними [15] [27]. Так, например, вероятность безрезультатного лечения текущей депрессии при накладывающемся расстройстве личности снижается в среднем в 2-2.3 раза [23]. Встречается и резистентность к некоторым разновидностям медикаментозного лечения [18]. Межличностная психотерапия также отрицательно сказывается на процессе выздоровления пациентов с расстройством личности [22]. Всё это необходимо учитывать при лечении этих людей, чтобы вернуть их к полнофункциональной жизни наиболее быстро и эффективно. Поэтому так важно своевременно выявлять личностные расстройства и начинать профессиональное вмешательство на их ранних стадиях.

В настоящее время удаётся достаточно надёжно диагностировать расстройство личности, но охватить с этой целью всех людей не представляется возможным. Низкая доступность квалифицированных специалистов в некоторых регионах особенно заметна [17]. Многие случаи нарушения остаются невыявленными, что вредит как самому больному, так и его окружению. Расстройство вовлекает несколько сфер личности

и влияет на социализацию и коммуникацию. Это отражается на общительности, устной и письменной речи человека [2]. Последняя является объектом изучения психолингвистики - науки, которая исследует связи процессов синтеза речи с психологическими особенностями её автора и использует для этого обширную методологию. Обработка естественного языка с применением машинного обучения предоставляет возможность проводить автоматический психолингвистический анализ в больших объёмах и нивелировать нехватку специалистов для выявления ментальных расстройств [26]. В эпоху коммуникаций через социальные сети стали реализовываться исследования текстов, размещаемых их пользователями [6]. Однако, выявление психического статуса человека по его текстам в социальных сетях ещё недостаточно изучено для русской речи.

1.2 Цель и данные

Основная цель данной работы состоит в использовании методов машинного обучения для обработки естественных языков с целью решения проблемы выявления расстройств личности при помощи текстов, написанных пользователями социальных сетей. Для решения этой задачи была использована коллекция из 702 текстов пользователей Twitter, поделенная на обучающую и тестовую выборки. Обучающая выборка состоит из 57 текстов, написанных здоровыми людьми, и 391 текста, написанного людьми с подтвержденным специалистом расстройством личности. Распределение текстов по целевому признаку *Healthy/Unhealthy* в тестовой выборке условно считается неизвестным. Деление расстройств личности на виды в данной работе не учитывается. С точки зрения машинного обучения, решается задача бинарной классификации. В качестве признаков используются текстовые, психолингвистические, словарные признаки, признаки, характеризующие активность коммуникаций пользователя в социальной сети.

2 Обзор литературы

2.1 Психометрический анализ

Существующие исследования по психометрии употребления слов показывают, что выбор слов людьми достаточно стабилен во времени и последователен, чтобы использовать язык в качестве меры различий личностей [21] [9]. При этом для анализа используются как простые лингвистические показатели, так и более психологически обоснованные особенности речи.

Местоимения и служебные части речи составляют более половины слов, которые мы используем в разговорной речи [10]. В этом исследовании испытуемые с расстройствами личности использовали значительно меньше таких слов по сравнению со здоровыми контрольными группами. В частности, люди с пограничным расстройством личности использовали меньше местоимений от первого лица множественного числа, от второго лица и больше местоимений от третьего лица единственного числа.

Предыдущие труды показали, что меньшее использование множественного местоимения от первого лица связано с меньшим эмоциональным дистанцированием [21]. Участники с пограничным расстройством также использовали больше слов, связанных с негативными эмоциями (страх, беспокойство и т.д.) и слов гнева (ненависть, убийство и др.). В то время как участники контрольной группы использовали значительно больше слов относящихся к положительным эмоциям (любовь, хороший, дорогой и др.). Интересно, что одно из недавних исследований также показало, что тяжесть симптомов пограничного расстройства выражалась в меньшем проявлении печали и более частой враждебностью по отношению к другим [31].

Прошлые психолингвистические исследования показывают, что использование глаголов, слов эмоций и слов мыслительной деятельности значительно варьируется в ответ на травмирующий опыт [21]. Анализ показал, что в группе с расстройствами личности используют больше глаголов прошедшего времени и ругательств. Напротив, участники контрольной группы использовали значительно больше глаголов будущего времени. Здоровая контрольная группа в целом больше использовала лексики когнитивных процессов, а участники с личностным расстройством реже оперировали словами, относящимися к категориям "итуиция"(думать, знать, считать и др.), "несоответствие"(абсурд, ошибка, чушь и др.), и "попытка"(можем, надеюсь, гипотетически и др.) в программе для анализа текста LIWC – Linguistic Inquiry and World Count, и в дополнение к этому они больше употребляли слов, выражающих причинно-следственные связи (потому что, следовательно и др.).

Недавно было проведено исследование абсолютистского мышления,

которое большинством когнитивных методов лечения тревоги и депрессии считается когнитивным искажением. При таком мышлении человек убеждён, что какое-то действие обязательно повлечёт только один конкретный результат, и никакой другой. Выполненное исследование показало, что интернет-форумы, посвящённые тревоге, депрессии и суицидальным наклонностям, содержали больше абсолютистских слов, чем контрольные форумы, касающиеся других тем для общения [13]. Эксперимент представлял собой анализ текстов 63 интернет-форумов с более чем 6400 участниками, в результате был сформирован словарь абсолютистских слов. В него вошли такие слова, как всё, каждый, никогда, полностью и т.д.

Другая работа, посвященная разбору нарушений речи при пограничном расстройстве личности, выявила иную особенность в использовании слов больными [2]. Эти люди склонны чаще обращаться к своим воспоминаниям, сравнивать своё состояние до и после некоторого события в их жизни, например переезда, смены школы или травмирующего опыта. Для этого человек использует в своей речи различные временные маркеры (раньше, прежде, до того как, потом, после, теперь и др.), и сказуемыми в его предложениях чаще выступают глаголы в прошедшем времени. Эти результаты согласуются с более ранними исследованиями других научных групп [21].

2.2 Современные тенденции исследований

Большинство последних исследований, использующих методы обработки естественнего языка для прогнозирования психических заболеваний, были сосредоточены на текстовых образцах, полученных из социальных сетей испытуемых. Как правило, использовались тексты, которые пересекаются по времени с протеканием психического заболевания. Например, в работе [11] обрабатывались тексты сообщений людей в Facebook, полученные от тех людей, кто во время исследования время находился в состоянии депрессии, что отражено в их оценках нейротизма, полученных в опроснике Big 5 - модели личности человека, демонстрирующей взаимное восприятие людей [12]. В результате был достигнут коэффицент корреляции Пирсона r = 0,386 между предсказанными и действительными ментальными статусами с помощью линейной регрессии. В качестве признаков использовались N-grams, topic model и признаки, полученные с использованием LIWC.

В другой работе использовали тематическую модель латентного размещения Дирихле с учителем (sLDA) для прогнозирования депрессии пользователей Twitter [4]. Классификация основывалась на самоидентифицирующих утверждения в сообщениях, таких как "У меня диагностировали депрессию". При помощи Supervised Anchor Algorithm [30] иссле-

дователям удалось достичь precision = 74 % при полноте recall = 0.5 и precision = 62% при полноте recall = 0.75.

На соревновании CLPsych 2015[6] его участники использовали посты в Twitter для прогнозирования депрессии и Посттравматическое стрессовое расстройство (ПТСР), в которых принадлежность к группе с определенным состоянием психического здоровья так же была получена на основе заявлений в Twitter, указывающих на то, что у пользователя было определенное психическое заболевание [20]. Лучший результат по метрике accuracy составил 87%, он был достигнут с использованием лексических признаков и tf-idf [1]. Другая команда, использовавшая N-grams получила сопоставимый результат precicision = 86%.

В недавней работе, использующей глубокое обучение для обнаружения ментальных расстройств, результат показал, что использование чувств, эмоций и негативных слов в высказывании очень влияет на определение уровня депрессии [32]. Человек в депрессии чаще использует негативные слова, которые указывают на его отчаяние, длительную печаль, даже мысли о самоубийстве (например, "грустно", "страшно", "умереть", "самоубийство"). В задаче классификации LSTM-сеть обеспечила следующие показатели метрик: accuracy = 70.89%; precision = 50.24%; recall = 70.89%.

Другая группа исследователей в своей работе сравнила метод градиентного бустинга и свёрточные нейронные сети в задаче классификации различных ментальных заболеваний, в том числе пограничного расстройства личности [7]. Классификация проводилась по текстам пользователей социальной сети Reddit. В результате при выявлении пограничного личностного расстройства градиентный бустинг дал accuracy = 85.14%, а свёрточная нейронная сеть обеспечила accuracy = 90.49% при полноте recall = 99.54% для класса здоровых пользователей.

Из исследования американского сегмента социальных сетей в 2019 году, посвящённого предсказанию будущих психических заболеваний, были выделены особенности текстов, характерные для различных диагнозов [33]. Пользователи с тревожным синдромом как правило употребляли слова для выражения беспокойства и опасения (беспокойство, нервозность, дискомфорт и др.), а также часто говорили о соматических проблемах организма (дрожь, дыхание, живот и др.). В сообщениях о биполярном расстройстве часто упоминались различные временные периоды (порой, иногда и др.) и сторонняя помощь (больница, отец, бог). Этот контент-анализ помогает прояснить, как психическое заболевание может отражаться на людях до развития тяжёлых стадий болезни. При депрессии происходят сильные изменения в эмоциях, при тревоге - влияние на настроение, а при биполярном расстройстве - неусточивость во времени эмоций человека и его отношения к происходящему вокруг. В результате удалось построить классификатор, на котором была достиг-

нута F1-мера равная 0.77 для предсказания по текстам по тематике ментального здоровья и F1-мера равная 0.38 для текстов без ограничений по теме.

2.3 Вывод

Анализ смежных исследований показывает, что в силу разнородности данных и различий в постановках задач, объективно сравнивать эти работы с данной затруднительно. Подавляющее большинство исследований проведено только над англоязычным сегментом интернета, поэтому не все психолингвистические выводы могут быть перенесены на изучение русского языка в социальных сетях, необходимо перепроверять работоспособность описанных методов на новых данных. Резонно провести эксперименты с моделями, использующими лингвистические и психологические признаки, провести работу с N-граммами, применять TF-IDF модели. Хорошо себя показывает применение словарных признаков - словари часто встречающихся слов, словари тональностей и сборники из программы Linguistic Inquiry and Word Count. Работа с социальными сетями позволяет рассматривать как признаки различные показатели активности пользователя. Они так же обосновывают состояние человека с точки зрения психологии и показывают свою эффективность в классификации пользователей.

3 Постановка задачи и используемые методы

Рассмотрим математическую постановку задачи и ее методы решения.

Пусть X – множество описаний объектов, Y – множество классов, в нашем случае 2 класса – Unhealthy, Healthy. Существует неизвестная целевая зависимость – отображение $y^* : X \longrightarrow Y$, значения которой известны только на объектах обучающей выборки.
$X^m = \{(x_1, y_1), ..., (x_m, y_m)\}$. Требуется построить алгоритм $a : X \longrightarrow Y$, способный классифицировать произвольный объект $x \in X$.

В данной работе используются 3 классификатора для решения данной задачи: Random Forest, SVM, Gradient Boosting.

3.1 Random Forest

В основе алгоритма случайного леса лежит дерево решений. Алгоритм построения случайного леса, состоящего из N деревьев, выглядит следующим образом:
Для каждого $n = 1, ..., N$:.

- Сгенерировать выборку X_n
- Построить решающее дерево b_n по выборке X_n:
  - по заданному критерию выбирается лучший признак, делается разбиение в дереве по нему и так до исчерпания выборки
  - дерево строится, пока в каждом листе не более n_{min} объектов или пока не достигнем определенной высоты дерева
  - при каждом разбиении сначала выбирается m случайных признаков из n исходных, и оптимальное разделение выборки ищется только среди них

Итоговый классификатор $a(x) = \frac{1}{N}\sum_{i=1}^{N} b_i(x)$, т.е. в задаче классификации выбирается решение голосованием по большинству.

3.2 SVM – Support Vector Machine

Каждая точка x_i принадлежит какому-то из классов. x_i – это p-мерный вещественный вектор. Чтобы классифицировать объекты, мы хотим построить разделяющую гиперплоскость, которая имеет вид

$$w \times x - b = 0$$

Вектор w - перпендикуляр к разделяющей гиперплоскости. Параметр $\frac{b}{||w||}$ равен по модулю расстоянию от гиперплоскости до начала координат. Обозначим *Unhealthy, Healthy* как 1, -1. $c_i = +1$ или $c_i = -1$

Далее получим следующее условие:

$$c_i(w \times x_i - b) \geq 1,$$

Задача построения оптимальной разделяющей гиперплоскости сводится к следующей оптимизационной задаче:

$$\begin{cases} ||w||^2 \longrightarrow min \\ c_i(w \times x_i - b) \geq 1 \end{cases}$$

Решая эту задачу, получим, что алгоритм классификации может быть записан в виде:

$$a(x) = sign\left(\sum_{i=1}^{n} \lambda_i c_i x_i \times x - b\right),$$

где $\lambda = (\lambda_1, ..., \lambda_n)$ – вектор двойственных переменных.

3.3 Gradient Boosting

Обозначения: M – число итераций, набор данных $\{(x_i, y_i)\}$, восстанавливаемая зависимость $y = f(x)$, функция потерь $L(y, f)$, которую мы будем минимизировать.
Алгоритм градиентного бустинга:

- Инициализировать начальное приближение параметров $\hat{\theta} = \hat{\theta_0}$
- Для каждой итерации $t = 1, ..., M$ повторить:
  1. Посчитать градиент функции потерь $\nabla L_\theta(\hat{\theta})$ при текущем приближении $\hat{\theta}$
     $\nabla L_\theta(\hat{\theta}) = \left[\frac{\partial L(y, f(x,\theta))}{\partial \theta}\right]_{\theta=\hat{\theta}}$
  2. Задать текущее итеративное приближение $\hat{\theta}_t$ на основе посчитанного градиента
     $\hat{\theta}_t \leftarrow -\nabla L_\theta(\hat{\theta})$
  3. Обновить приближение параметров $\hat{\theta}$:
     $\hat{\theta} \leftarrow \hat{\theta} + \hat{\theta}_t = \sum_{i=0}^{t} \hat{\theta}_i$
- Сохранить итоговое приближение $\hat{\theta}$
- $\hat{f}(x) = f(x, \hat{\theta})$

В данной работе используется реализация всех этих алгоритмов из библиотеки sklearn.

3.4 Метрики качества

В задачах классификации в основном используют следующие метрики качества: accuracy, precision, recall, F1-мера, площадь под ROC-кривой – ROC AUC.

Наиболее очевидной из них является accuracy – доля правильных ответов. В основном используется, когда классы сбалансированы. Для несбалансированных классов используются остальные метрики.

	$y = 1$	$y = 0$
$\hat{y} = 1$	True Positive (TP)	False Positive (FP)
$\hat{y} = 0$	False Negative (FN)	True Negative (TN)

Таблица 1: Матрица ошибок.

Здесь $\hat{y}$ – это ответ алгоритма на объекте, а y – истинная метка класса на этом объекте. В терминах этой матрицы выше перечисленные метрики будут выглядеть следующим образом:

$$\begin{aligned} accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\ recall &= \frac{TP}{TP + FN} \\ precision &= \frac{TP}{TP + FP} \\ F_1 &= 2 \cdot \frac{precision \cdot recall}{precision + recall} \end{aligned} \tag{1}$$

Precision – доля людей, которых классификатор определил как Unhealthy и при этом реально имеющие расстройство личности.
Recall – какую долю людей с личностным расстройством из всех объектов с меткой Unhealthy нашел алгоритм.
ROC-AUC score – площадь под кривой ROC. Данная кривая представляет из себя линию от (0;0) до (1;1) в координатах True Positive Rate –TPR и False Positive Rate – FRP:

$$\begin{aligned} TPR &= \frac{TP}{TP + FN} \\ FPR &= \frac{FP}{FP + TN} \end{aligned} \tag{2}$$

В качестве основной метрики в этой работе была выбрана ROC-AUC.

4 Признаки для обучения

Признаки, которые используются для обучения моделей в этой работе, можно стратифицировать на 3 класса:

1. Текстовые признаки;
2. Признаки, базирующиеся на словарях;
3. Признаки, характеризующие вовлеченность пользователя в процесс общения в социальной сети;

4.1 Текстовые признаки

В силу того, что работа идет с текстами, написанными людьми, в словах встречаются ошибки и опечатки. Чтобы они не препятствовали анализу, тексты были преобразованы при помощи YandexSpeller для Python [24]. Признаки для обучения были получены из текстов посредством библиотек Mystem и NLTK. С их помощью была проведена токенизация, лемматизация и морфологический анализ слов. Пример их работы проиллюстрирован на рисунке 1.

```
import pyaspeller
from pyaspeller import YandexSpeller
speller = YandexSpeller()
example = 'Унас жывёт кот Семён, он дбрй и очен пушыстый.'
spelled_example = speller.spelled(example)
spelled_example
```

```
'У нас живет кот Семён, он добрый и очень пушистый.'
```

```
import pymystem3
mystem=pymystem3.Mystem()
for word in mystem.analyze(spelled_example):
    if 'analysis' in word:
        print(word['analysis'][0])
```

```
{'lex': 'у', 'wt': 0.9993940324, 'gr': 'PR='}
{'lex': 'мы', 'wt': 1, 'gr': 'SPRO,мн,1-л=(пр|вин|род)'}
{'lex': 'жить', 'wt': 1, 'gr': 'V,несов,нп=непрош,ед,изъяв,3-л'}
{'lex': 'кот', 'wt': 1, 'gr': 'S,муж,од=им,ед'}
{'lex': 'семен', 'wt': 1, 'gr': 'S,имя,муж,од=им,ед'}
{'lex': 'он', 'wt': 1, 'gr': 'SPRO,ед,3-л,муж=им'}
{'lex': 'добрый', 'wt': 1, 'gr': 'A=(вин,ед,полн,муж,неод|им,ед,полн,муж)'}
{'lex': 'и', 'wt': 0.9999770357, 'gr': 'CONJ='}
{'lex': 'очень', 'wt': 1, 'gr': 'ADV='}
{'lex': 'пушистый', 'wt': 1, 'gr': 'A=(вин,ед,полн,муж,неод|им,ед,полн,муж)'}
```

Рис. 1: Пример работы YandexSpeller и Mystem

Пример показывает, что стеммер может определить для каждого слова его часть речи, род, лицо, число и другие свойства. На основе этого морфологического разбора рассчитаны "базовые"признаки. Их статистика представлена в таблице 2.

Наименование признака	Группа без РЛ	Группа с РЛ
№ слов во всех постах	127.07 ± 80.5	350.75 ± 366.27
№ постов от пользователя	9.2 ± 6.01	23.5 ± 24.85
№ слов в ед. числе	65.2 ± 41.43	169.42 ± 175.83
№ местоимений	10.4 ± 7.54	37.92 ± 42.31
№ местоимений 1-ом лице	4.53 ± 3.48	17.71 ± 19.22
№ местоимений во 2-ом 3-ем лице	2.8 ± 3.17	8.33 ± 9.5
№ местоимений в ед. числе	8.73 ± 6.15	29.75 ± 33.07
№ местоимений во мн. числе	1.6 ± 1.84	7.0 ± 9.37
№ союзов	10.93 ± 7.98	30.0 ± 32.63
№ частиц	8.13 ± 5.55	22.96 ± 25.15
№ глаголов	21.13 ± 12.79	58.67 ± 60.44
№ глаг. в прошедшем времени	6.53 ± 6.53	17.96 ± 17.19
№ глаг. не в прошедшем времени	9.27 ± 5.99	26.08 ± 28.8
№ глаг. в инфинитиве	4.4 ± 3.14	12.33 ± 13.56
№ глаг. в ед. числе	12.53 ± 7.23	34.25 ± 34.47
№ глаг. во мн. числе	3.6 ± 3.64	11.12 ± 14.39
№ глаг. мужского рода	2.47 ± 4.21	6.04 ± 9.22
№ глаг. женского рода	2.4 ± 2.44	6.21 ± 6.39
№ прилагательных	15.4 ± 11.41	36.29 ± 43.55
№ прил. в женском роде	4.8 ± 4.43	10.67 ± 12.54
№ существительных	35.93 ± 24.14	92.96 ± 94.0
№ предлогов	12.47 ± 8.81	35.33 ± 36.17
№ наречий	9.13 ± 6.33	26.62 ± 26.8

Таблица 2: Среднее значение и среднее отклонение для «базовых» признаков. № – количество.

Видно, что за один и тот же период времени, люди с расстройствами личности в среднем публикуют текстов в совокупном объёме почти в 3 раза больше, чем здоровые пользователи. И относительное стандартное отклонение у страдающих расстройствами в 1,67 раза превышает тот же самый показатель у контрольной группы людей. На графике распределения постов пользователей по длине видно, что у здоровых пользователей длина одного текста редко превышает 25 символов, тогда как у людей с личностными расстройствами доля таких текстов больше, как и их максимальная длина. Можно сделать предположение, что здоровые пользователи пишут меньше, и размеры их текстов стабильнее во времени.

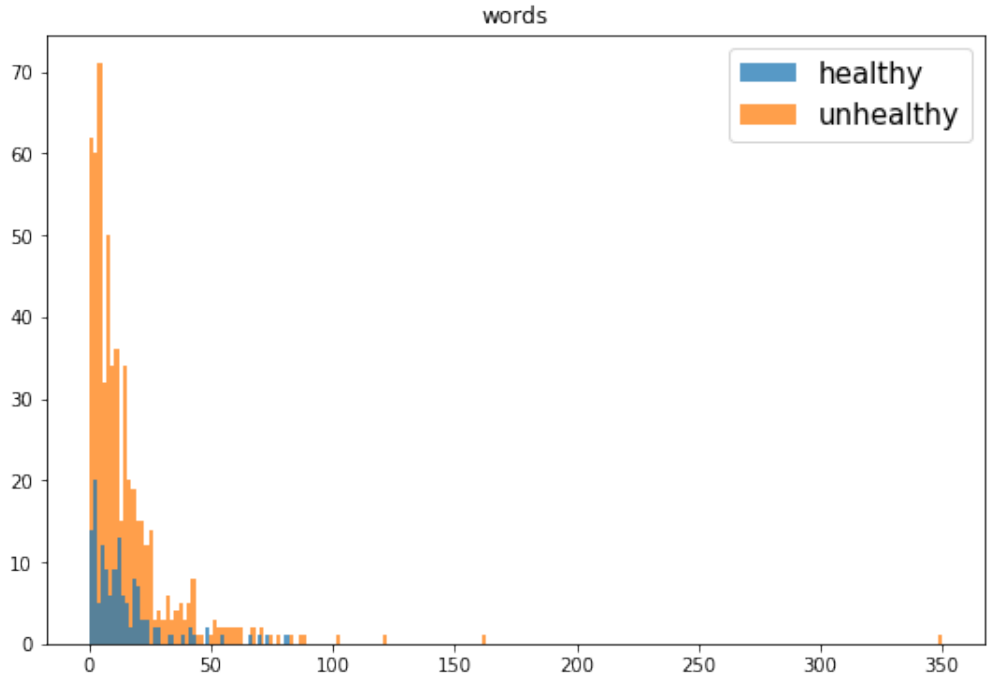

Рис. 2: Количество слов в посте

Далее, на основе «базовых» признаков, были рассчитаны следующие «смешанные» признаки:

- Количество слов на пост
- Доля предлогов [34]
- Доля союзов [34]
- Доля частиц
- Сложность поста – количество глаголов, разделенное на количество постов [34].
- Singularity index – доля слов в единственном числе [34].
- Coherence index – сумма количества частиц, союзов и предлогов, деленная на утроенное количество предложений [34].
- Pronominalisation index – отношение количества местоимений к количеству существительных [34].
- Formality metric – вычисляется по формуле: (сущ. + прил. + предлоги - местоим. - глаг. - наречия + 100)/2 [19].
- Trager index – количество глаголов на одно прилагательное [29].
- Readiness – число глаголов на одно существительное [29].

- Aggressiveness – количество глаголов, деленное на общее число слов [29].
- Activity – вычисляется по формуле: глаг/(глаг. + прил. + наречия)
- Autosemantic index – вычисляется по формуле (сущ. + прил. + местоим. + глаг. + наречия)/общее количество слов [25].
- Доля слов в ед. числе 1-го лица.
- Доля местоимений в первом лице.
- Доля обсценной лексики.
- Plural_first_second_pronoun – доля местоимений 1 и 2 лица множественного числа.
- Single_third_pronoun – доля местоимений 3 лица единственного числа.
- Доля глаголов в прошедшем времени.
- Mark3p5 – (сущ. + глаг.)/(нареч. + прил.) [14]
- Connection – число предлогов, приходящееся на один пост [14].
- Dinamo – (сущ. + прил.)/(глаг. + частиц) [14]

Исходя из результатов исследований [21] [10], нас будут наиболее сильно интересовать такие признаки как доля местоимений 1 и 2 лица множественного числа и доля местоимений 3 лица единственного числа (Рис 3, 4). Так же большое значение имеют процент глаголов прошедшего времени и процент обсценной лексики.

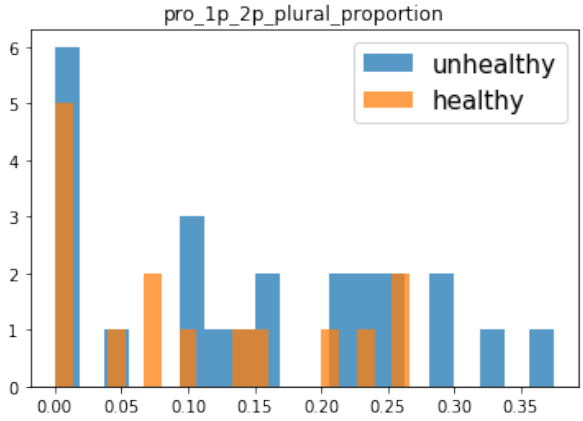

Рис. 3: Мест. 1 и 2 лица мн.ч.

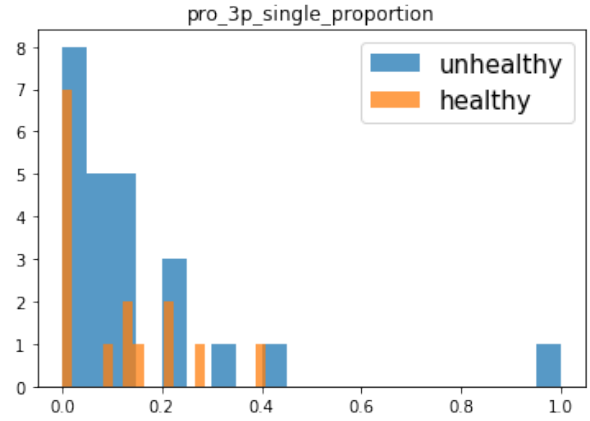

Рис. 4: Мест. 3 лица ед.ч.

Название признака	Группа без РЛ	Группа с РЛ
Число слов на пост	13.842 ± 7.378	16.569 ± 6.794
Сложность поста	2.329 ± 1.012	2.864 ± 1.197
Singularity index	0.535 ± 0.078	0.5 ± 0.059
Coherence index	1.136 ± 0.755	1.369 ± 0.665
Pronominalisation index	0.289 ± 0.19	0.407 ± 0.22
Formality metric	0.082 ± 0.059	0.059 ± 0.059
Trager index	1.804 ± 0.722	1.996 ± 0.874
Readiness	0.754 ± 0.149	0.889 ± 0.0
Aggressiveness	0.185 ± 0.054	0.175 ± 0.029
Activity	0.498 ± 0.112	0.494 ± 0.08
Autosemantic index	0.736 ± 0.059	0.73 ± 0.043
Доля предлогов	0.092 ± 0.034	0.105 ± 0.025
Доля союзов	0.079 ± 0.027	0.079 ± 0.027
Доля частиц	0.063 ± 0.036	0.059 ± 0.026
Доля местоимений 1 лица	0.037 ± 0.03	0.051 ± 0.029
Plural_first_second_pronoun	0.01 ± 0.012	0.016 ± 0.013
Single_third_pronoun	0.01 ± 0.014	0.014 ± 0.018
Доля глаголов прошедшего времени	0.275 ± 0.178	0.383 ± 0.197
Доля обсценной лексики	0.017 ± 0.041	0.013 ± 0.016
Mark3p5	2.784 ± 1.08	2.605 ± 0.808
Connection	1.36 ± 0.966	1.769 ± 0.85
Dinamo	1.714 ± 0.588	1.643 ± 0.474

Таблица 3: Среднее значение и среднее отклонение для «смешанных» признаков.

Предположение о большей доле местоимений 3 лица единственного числа в текстах пользователей социальных сетей с расстройством личноси подтверждают и наши данные (Рис. 4). Но доля местоимений 1 и 2 лица множественного числа тоже выросла (Рис. 3), вопреки выводам исследования [21] о зависимости социального дистанцирования человека и употребления им различных местоимений. Вероятнее всего это объясняется различием в данных на английском и русском языках и тем, как местоимения в этой форме употребляются в них обоих. Графики для других «смешанных» и «базовых» признаков можно найти в приложении к работе.

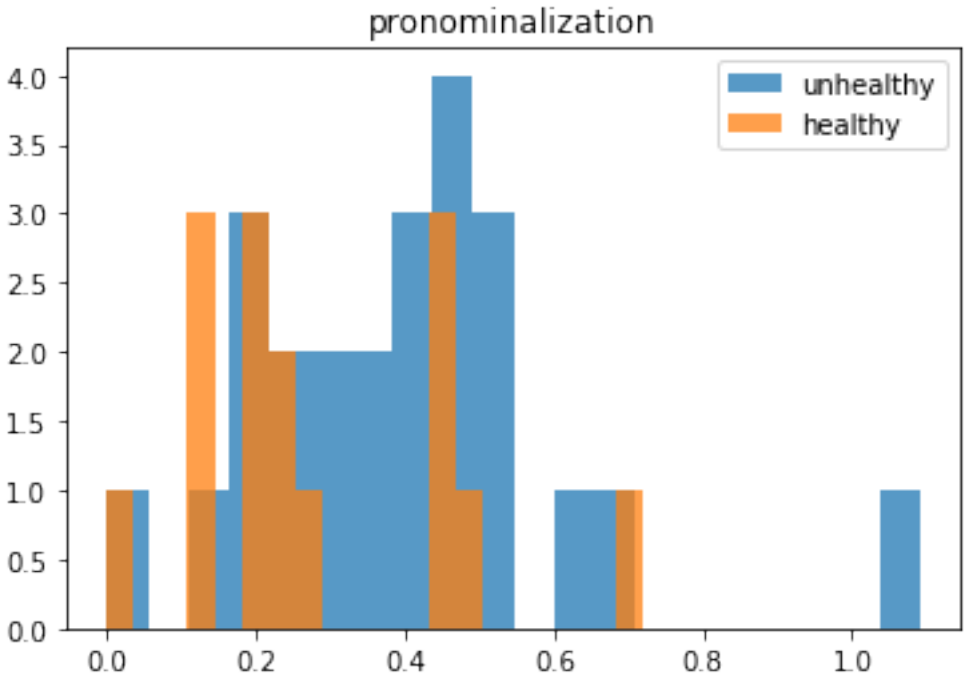

Рис. 5: Pronominalisation index

Из графика на рисунке 5 и значений в таблице 3 можно сделать вывод, что Pronominalisation index в группе пользователей с расстройством личности превышает показатели контрольной группы здоровых людей. Это означает, что процент местоимений в их текстах действительно больше.

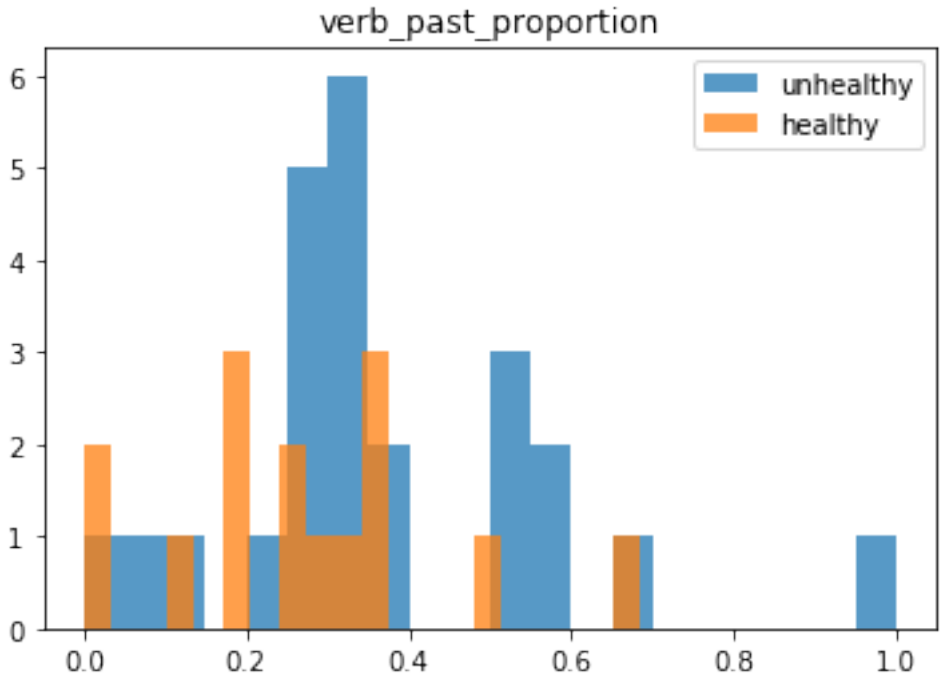

Рис. 6: Глаголы прошедшего времени

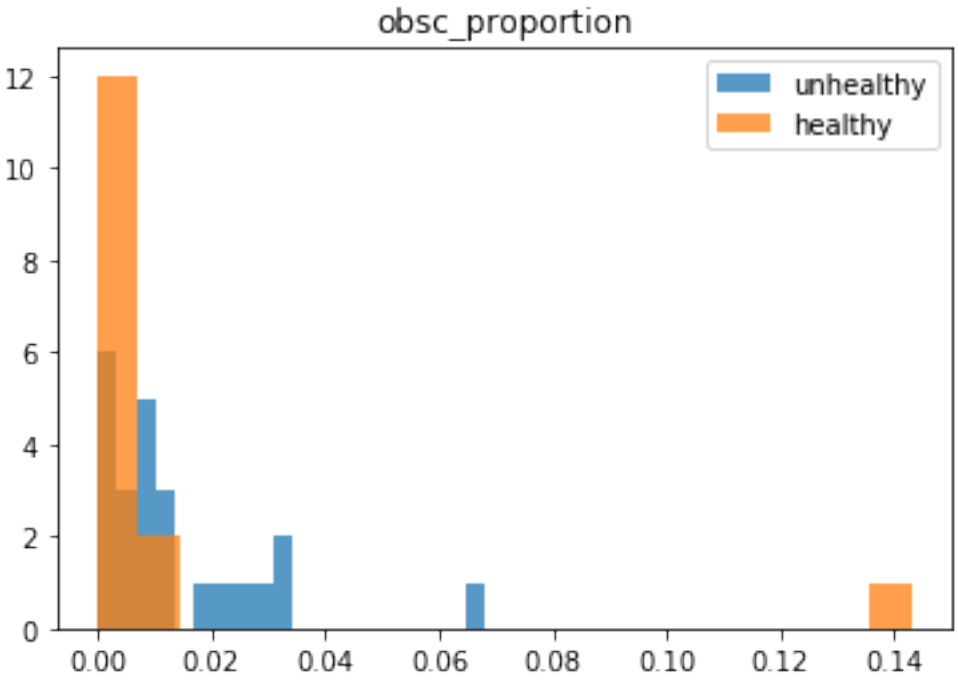

Рис. 7: Обсценная лексика

Из таблицы 3 и рисунков 6, 7 видно, что наши данные подтверждают вывод рассмотренных работ [21] и [10] о увеличении у лиц с личностными расстройствами процента обсценной лексики и глаголов прошедшего времени в речи. Последнее объясняется частой рефлексией своего личного прошлого людьми с расстройствами личности. Они чаще обращаются к своим воспоминаниям о важных моментах своей жизни, которые зачастую оказывались травмирующими и в некоторой степени явились причинами возникновения заболевания у человека. Это подтверждается в работе [2].

4.2 Словарные признаки

В работах с англоязычными текстами использующих Linguistic Inquiry and Word Count [2] отмечается, что в письменной речи людей с расстройством личности меньшее чем у здоровых использование слов из встроенных словарей когнитивной деятельности: Интуиция, Несоответсвтие, Попытка (рисунки 8, 9, 10). Кроме того, эти тексты хорошо классифицируются по вхождениям слов в словари эмоциональной лексики: Негатив, Позитив, Гнев, Беспокойство, Грусть. Воспользуемся этими словарями для русского языка и для каждого текста, подсчитаем долю слововхождений в этих сборники.

Рис. 8: Слова из словаря "Интуиция"

Рис. 9: Слова из словаря "Попытка"

Рис. 10: Слова из словаря "Несоответствие"

Слова-абсолютисты
абсолютно, безусловно, конечно, совершенно, безусловно, стопроцентно, полностью, точно, несомненно, бесспорно, очевидно, верно, именно, правильно, неверно, четко, несомненный, весь, вечно, вечный, то и дело, определенно, ясно, однозначно, понятно, всегда, только, каждый, любой, наверняка, просто, обязательно, вовсе, непременно, необязательно, нужно, должен, необходимо, надо, обязан, никогда, ничего

Таблица 4: Слова-абсолютисты

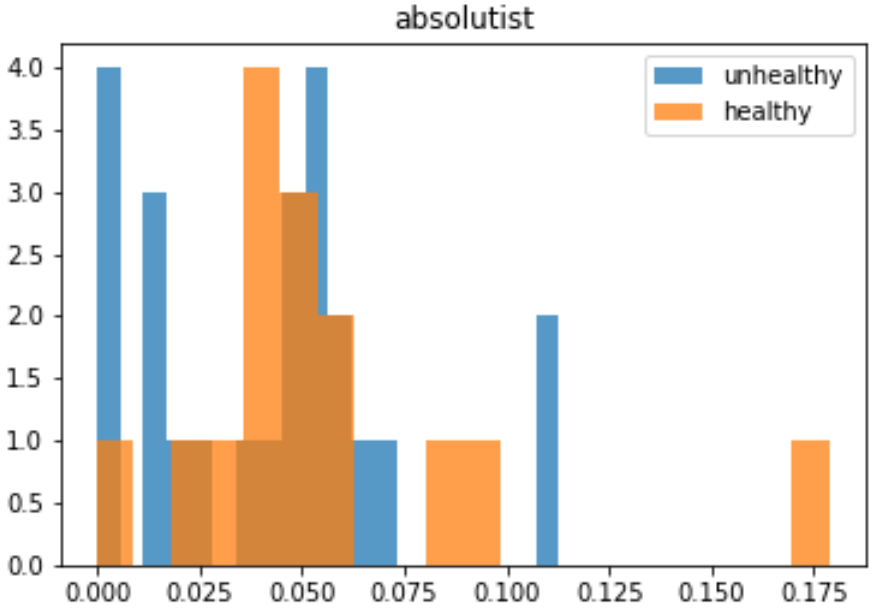

Рис. 11: Слова-абсолютисты

В исследовании [13] было замечено, что люди с проблемами ментального здоровья склонны употреблять в сообщениях «слова-абсолютисты» (таблица 4). В связи с этим был создан словарь этих слов. Словарь был расширен при помощи функции, использующей Word2Vec модель, обученную на корпусе из интернет-страниц на русском языке [28] для подбора наиболее близких слов для данного слова. Для каждого текста было подсчитано, какую его часть составляют слова из этого словаря.

Также было отобрано топ-10, топ-25, топ-50 наиболее используемых слов текстов каждого из классов Healthy, Unhealthy. Они были дополнены синонимами, подобранными с помощью той же модели Word2Vec. Доля вхождений слов из текста в эти корпуса была использована в качестве признака.

Рис. 12: Слова из класс Unhealthy

Среди 50 самых частоупотребляемых слов пользователей с расстройством личности встретилось несколько слов обсценной лексики, тогда как в контрольной группе здоровых людей ни одно нецензурное слово в топ-50 по популярности не вошло. Это объясняет результаты, полученные при рассчете текстовых признаков, где обсценная лексика у здоровых пользователей имела сравнимую среднюю долю, но её относительное отклонение было в 2 раза больше: она встречалась крайне редко, но в большем количестве. Такой результат согласуется с выводами о нецензурной лексике из работ [21] и [10]

Рис. 13: Слова из класса Healthy

Двумя разными по данным и методике исследованиями [32], [21] были получены сходные результаты о большем употреблении людьми с расстройствами личности слов, связанных с негативными эмоциями. В этой работе словарь негативной лексики Negative также использовался в качестве признака. Его значения в таблице 5 и график на рисунке 14 позволяют сделать вывод о том, что разница в использовани этих слов между классами Healthy и Unhealthy действительно есть. Пользователям с расстройством личности лексика негативных эмоций свойственна больше чем здоровым.

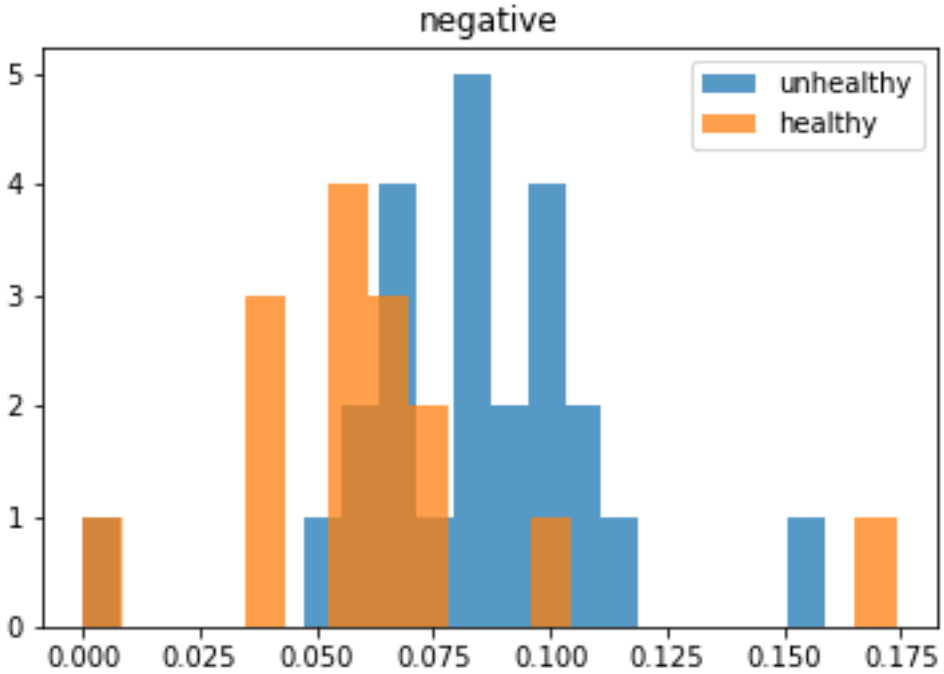

Рис. 14: Негативные слова

Название признака	Группа без РЛ	Группа с РЛ
top50_healthy	0.046 ± 0.036	0.034 ± 0.022
top25_healthy	0.027 ± 0.031	0.024 ± 0.019
top10_healthy	0.008 ± 0.012	0.01 ± 0.009
top50_unhealthy	0.062 ± 0.036	0.072 ± 0.041
top25_unhealthy	0.033 ± 0.022	0.032 ± 0.026
top10_unhealthy	0.016 ± 0.016	0.009 ± 0.01
absolutist	0.055 ± 0.041	0.041 ± 0.031
negative	0.064 ± 0.038	0.083 ± 0.028
positive	0.101 ± 0.029	0.12 ± 0.045
anger	0.035 ± 0.025	0.039 ± 0.023
anxiety	0.001 ± 0.003	0.007 ± 0.008
sadness	0.011 ± 0.013	0.022 ± 0.025
insight	0.054 ± 0.043	0.067 ± 0.033
discrepancy	0.028 ± 0.029	0.024 ± 0.025
tentative	0.026 ± 0.026	0.034 ± 0.021

Таблица 5: Среднее значение и среднее отклонение для «словарных» признаков.

4.3 Признаки активности в социальной сети

В работе [8], посвященной выявлению депрессии через Twitter, были предложены признаки, характеризующие активность и вовлеченность пользователя:

- posts_per_day_normalised - нормированное число постов в день
- replies_proportion - доля твитов, в которых упоминаются другие пользователи
- retweets_proportion - доля ретвитов
- questions_proportion - доля твитов, в которых пользователь задает вопросы
- shares_proportion - доля твитов, в которых пользователь делится сторонними материалами (изображения, видео, ссылки)

Название признака	Группа без РЛ	Группа с РЛ
posts_per_day_normalised	1.192 ± 0.78	2.932 ± 3.111
replies_proportion	0.587 ± 0.343	0.747 ± 0.439
retweets_proportion	0.092 ± 0.127	0.221 ± 0.412
questions_proportion	0.162 ± 0.196	0.129 ± 0.104
shares_proportion	0.578 ± 0.47	0.827 ± 0.792

Таблица 6: Среднее значение и среднее отклонение для признаков активности в социальной сети.

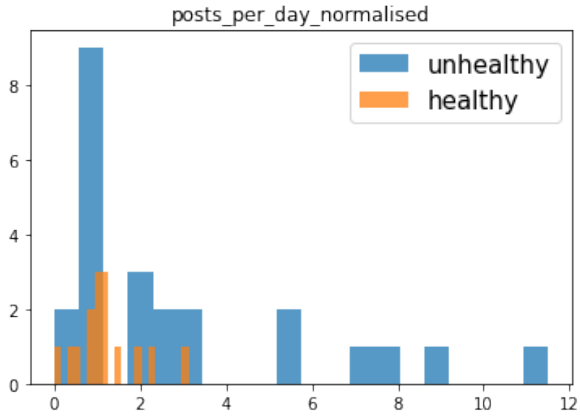

Рис. 15: Posts_per_day_normalised

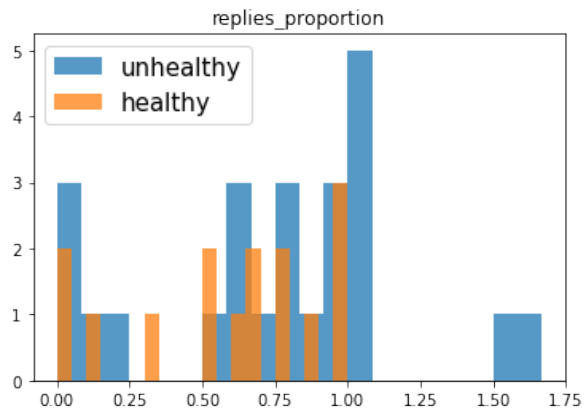

Рис. 16: Replies

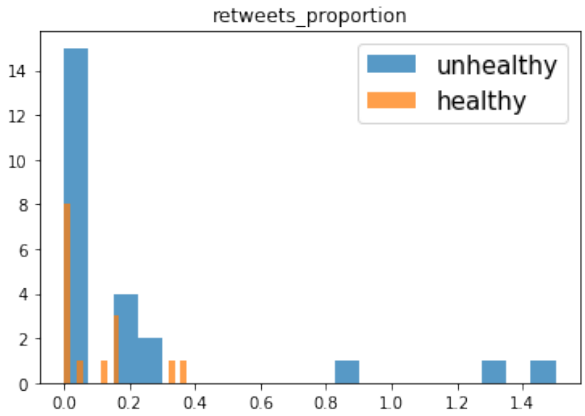

Рис. 17: Retweets

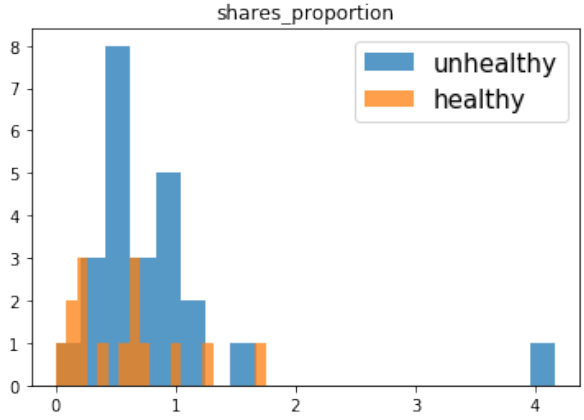

Рис. 18: Shares

Графики на рисунках 15, 16, 17, 18 и таблица средних значений и отклонений 6 указывают на большую социальную активность пользователей с расстройством личности. Они чаще публикуют новые тексты, обращаются к другим пользователям и делятся сторонними материалами с остальными.

5 Обучение и отбор признаков

Для каждого набора признаков были обучены классификаторы Random Forest, SVM и Gradient Boosting. Подбор параметров осуществлялся с помощью GridSearchCV с n_folds = 3.

Признаки	Классификатор	ROC-AUC
Текстовые	Random Forest	**0.649**
	SVM	0.517
	Gradient Boosting	0.514
Словарные	Random Forest	0.830
	SVM	0.740
	Gradient Boosting	**0.851**
Активности	Random Forest	0.688
	SVM	0.611
	Gradient Boosting	**0.689**

Таблица 7: Значение метрики ROC-AUC

После этого при помощи матрицы корреляции признаков и встроенного метода feature_importance Random Forest Classifier были выбраны лучшие признаки.

- best_text_features - лучшие текстовые признаки и признаки активности
- best_dict_features - лучшие словарные признаки

Содержащиеся в этих наборах признаки описаные в приложении. Далее представлены результаты работы классификаторов с этими признаками.

Классификатор	ROC-AUC
Random Forest	**0.781**
SVM	0.635
Gradient Boosting	0.618

Таблица 8: ROC-AUC для набора best_text_features

Классификатор	ROC-AUC
Random Forest	**0.847**
SVM	0.701
Gradient Boosting	0.819

Таблица 9: ROC-AUC для набора best_dict_features

Классификатор	ROC-AUC
Random Forest	**0.885**
SVM	0.674
Gradient Boosting	0.755

Таблица 10: ROC-AUC для совокупности наборов best_text_features и best_dict_features

6 Результаты

Здесь представлены результаты работы топ-3 моделей на тестовой выборке.

	Precision	Recall	F1-score
Healthy	0.92	0.93	0.91
Unhealthy	0.82	0.78	0.83
Avg	0.87	0.86	0.87

Таблица 11: ТОП-1 модель, ROC-AUC – 0.861

ТОП-1 модель это классификатор Random Forest Classifier на совокупном наборе признаков best_text_features и best_dict_features (таблицы 14, 15).

ТОП-2 модель это классификатор Random Forest Classifier только на наборе признаков best_dict_features(таблица 15).

ТОП-3 модель это классификатор Gradient Boosting на наборе признаков best_dict_features(таблица 15).

	Precision	Recall	F1-score
Healthy	0.92	0.90	0.93
Unhealthy	0.83	0.79	0.80
Avg	0.88	0.85	0.86

Таблица 12: ТОП-2 модель, ROC-AUC – 0.847

	Precision	Recall	F1-score
Healthy	0.90	0.88	0.86
Unhealthy	0.71	0.67	0.70
Avg	0.80	0.77	0.78

Таблица 13: ТОП-3 модель, ROC-AUC – 0.767

7 Заключение

В представленной работе оценивается потенциал использования машинного обучения для классификации текстов, написанных людьми с расстройством личности и без в социальных сетях на русском языке.

Лучший результат был достигнут классификатором Random Forest Classifier на наборе, включающем в себя лучшие признаки активности в социальной сети, текстовые и словарные признаки – ROC-AUC 0.861 (таблиц 11).

Из признаков стоит отметить словарь слов-абсолютистов, словари LIWC и признаки активности в социальной сети. Последние два ещё не получили широкого распространения в работах, посвященных выявлению психического статуса автора по его текстам на русском языке. Некоторые из них оказались особенно эффективны при классификации пользователей (таблицы 14, 15).

Результаты, полученные в данной работе могут найти применение в быстроразвивающейся цифровой медицине. Распознавание ментального статуса у пользователей социальных сетей может позволить своевременно оказывать адресную помощь людям для предотвращения развития серьёзных форм заболеваний.

8 Приложение

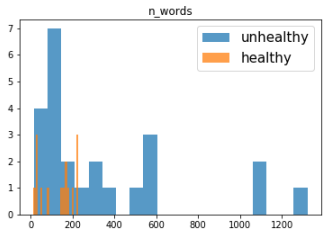

Рис. 19: Количество слов

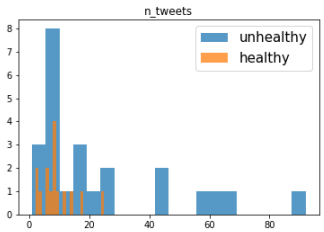

Рис. 20: Количество постов

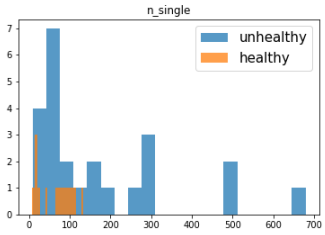

Рис. 21: Слов в ед.ч.

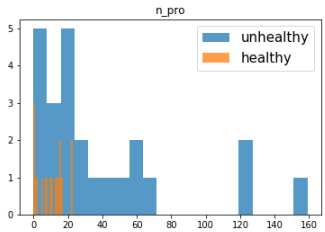

Рис. 22: Количество местоимений

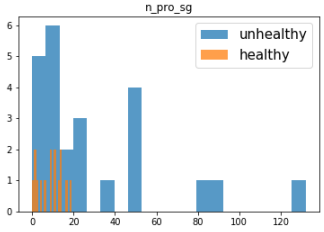

Рис. 23: Местоимений в ед.ч.

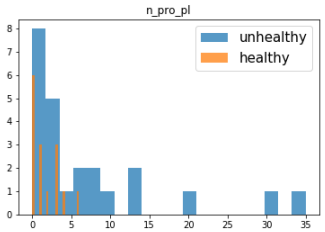

Рис. 24: Местоимений во мн.ч.

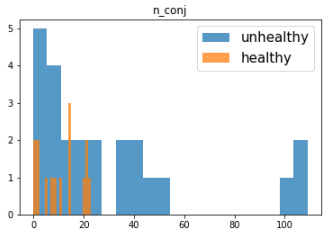

Рис. 25: Количество союзов

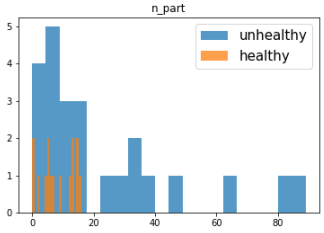

Рис. 26: Количество частиц

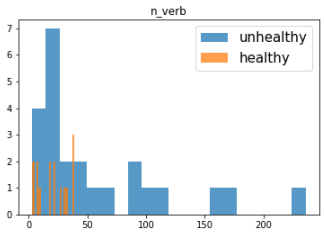

Рис. 27: Количество глаголов

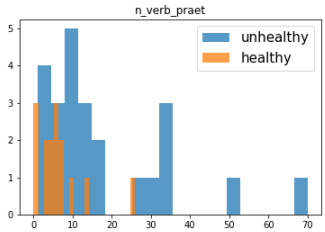

Рис. 28: Глаголов в прош.вр.

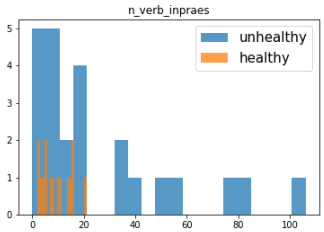

Рис. 29: Глаголов **не** в прош.вр.

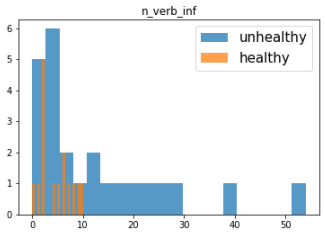

Рис. 30: Глаголов в инфинитиве

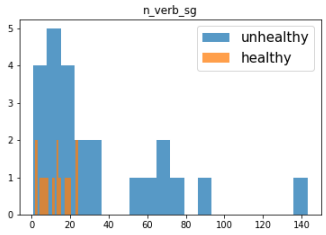

Рис. 31: Глаголов в ед.ч.

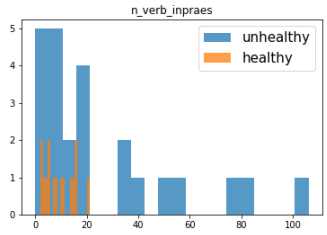

Рис. 32: Глаголов **не** в прош.вр.

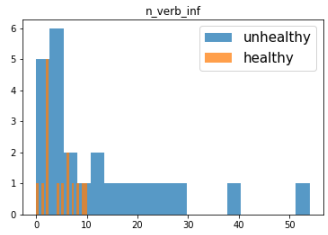

Рис. 33: Глаголов в инфинитиве

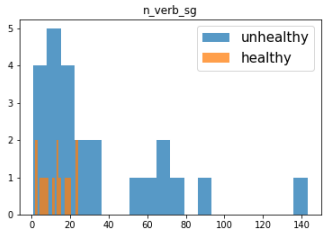

Рис. 34: Глаголов в ед.ч.

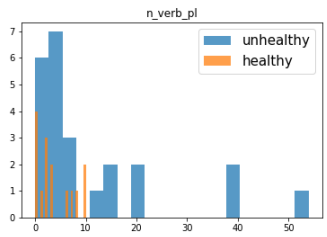

Рис. 35: Глаголов во мн.ч.

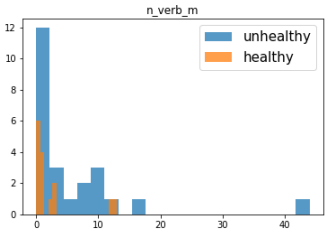

Рис. 36: Глаголов мужского рода

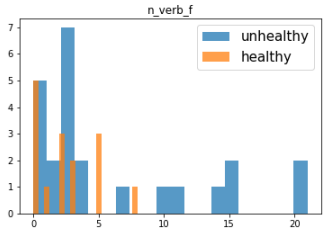

Рис. 37: Глаголов женского рода

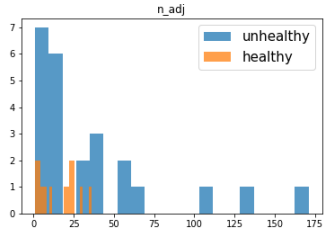

Рис. 38: Количество прилагательных

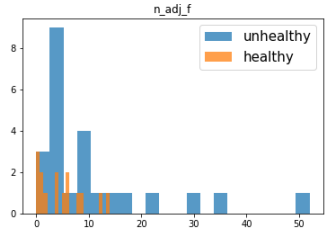

Рис. 39: Прил. женского рода

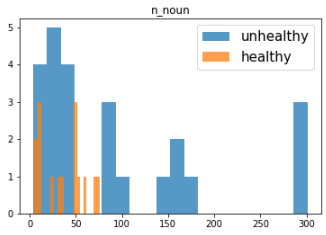

Рис. 40: Количество сущ.

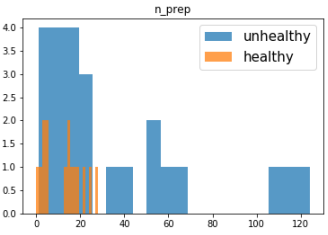

Рис. 41: Количество предлогов

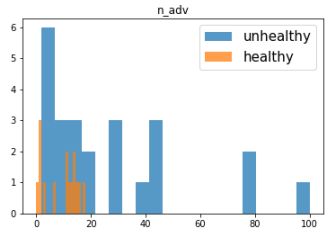

Рис. 42: Количество наречий

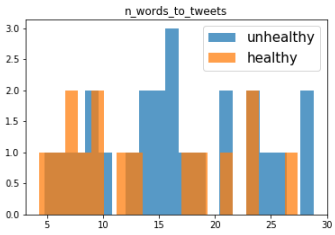

Рис. 43: Число слов на пост

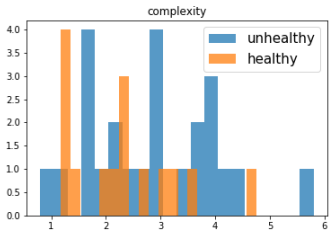

Рис. 44: Сложность

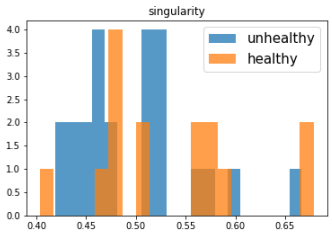

Рис. 45: Singularity

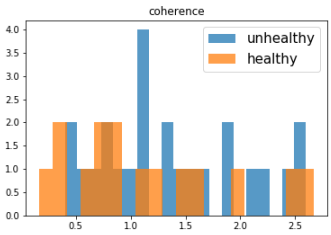

Рис. 46: Coherence

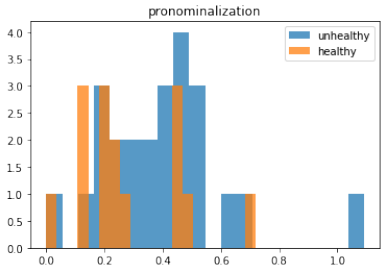

Рис. 47: Pronominalisation

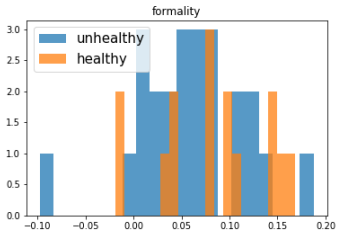

Рис. 48: Formality

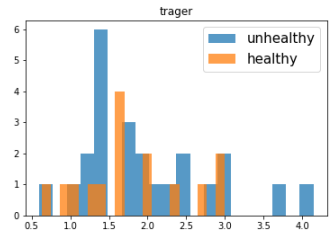

Рис. 49: Trager

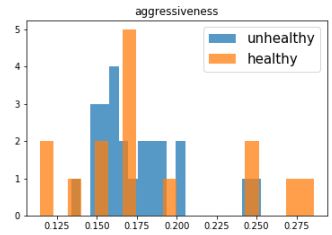

Рис. 50: Aggressiveness

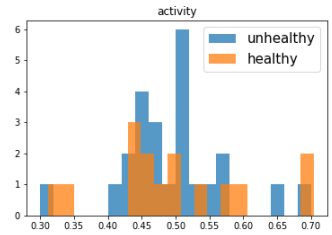

Рис. 51: Activity

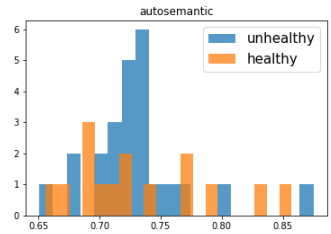

Рис. 52: Trager

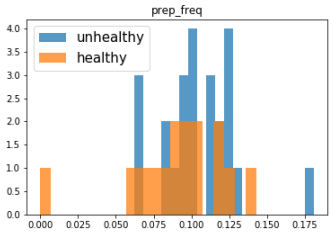

Рис. 53: Доля предлогов

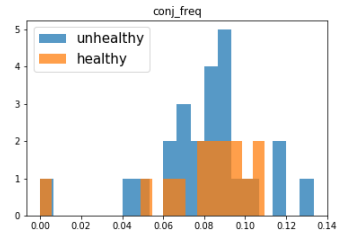

Рис. 54: Доля союзов

n_words, n_tweets, complexity, singularity, pronominalisation, readiness, pro_1p_2p_plural_proportion, verb_past_proportion, posts_per_day_normalised, replies_proportion, retweets_proportion, connection, prep_freq

Таблица 14: Набор признаков best_text_features

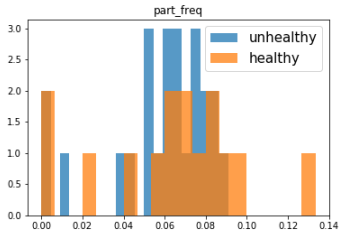

Рис. 55: Trager

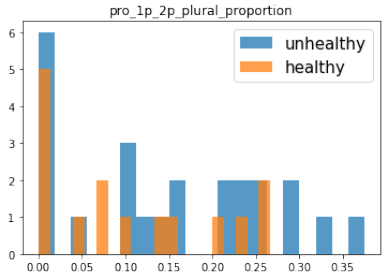

Рис. 56: Plural first second pronoun

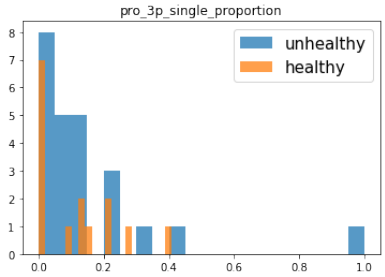

Рис. 57: Single third pronoun

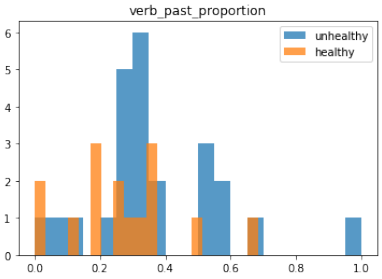

Рис. 58: Доля глаголов прош.вр.

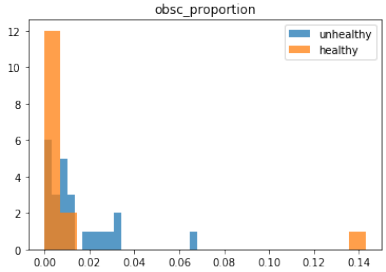

Рис. 59: Доля обсценной лексики

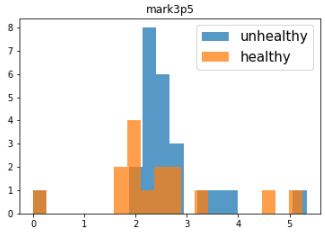

Рис. 60: Mark3p5

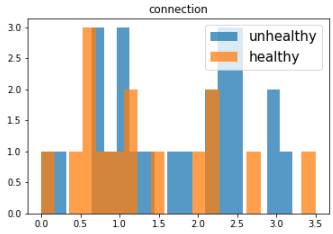

Рис. 61: Connection

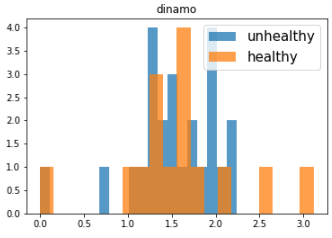

Рис. 62: Dinamo

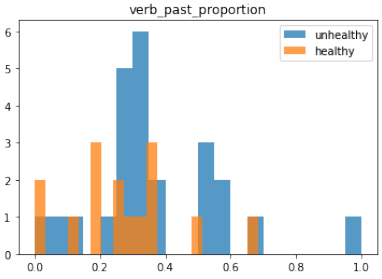

Рис. 63: Доля глаголов прош.вр.

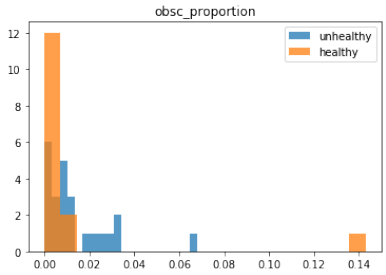

Рис. 64: Доля обсценной лексики

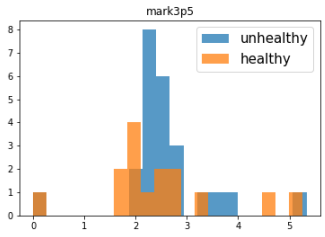

Рис. 65: Mark3p5

top_50_healthy, top10_unhealthy, absolutist, negative, positive, anxiety, sadness

Таблица 15: Набор признаков best_dict_features

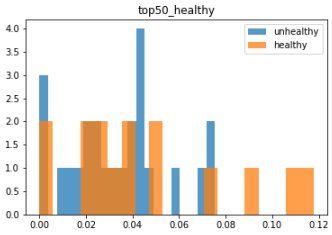

Рис. 66: top50_healthy

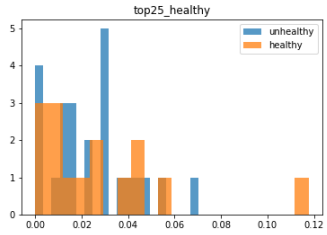

Рис. 67: top25_healthy

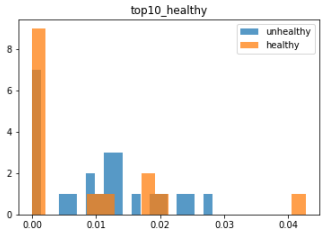

Рис. 68: top10_healthy

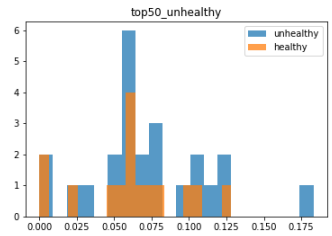

Рис. 69: top50_unhealthy

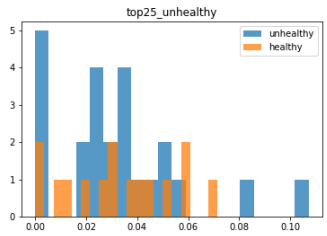

Рис. 70: top25_unhealthy

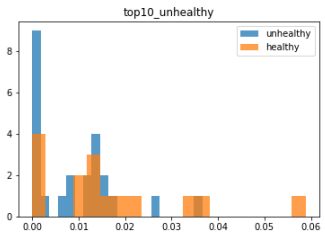

Рис. 71: top10_unhealthy

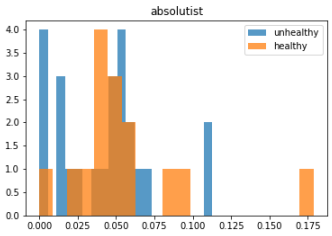

Рис. 72: Absolutist

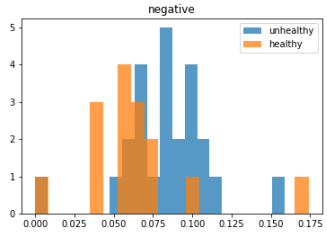

Рис. 73: Negative

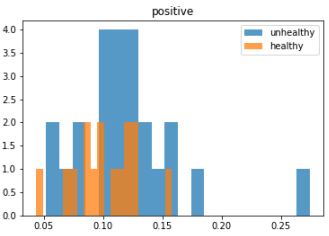

Рис. 74: Positive

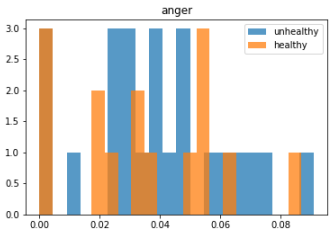

Рис. 75: Anger

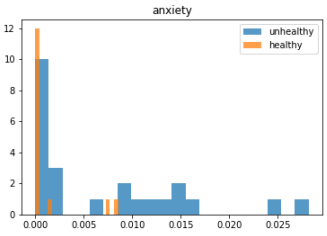

Рис. 76: Anxiety

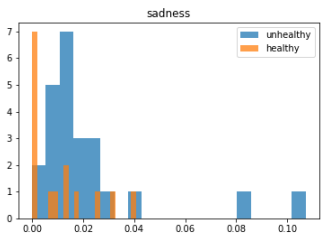

Рис. 77: Sadness

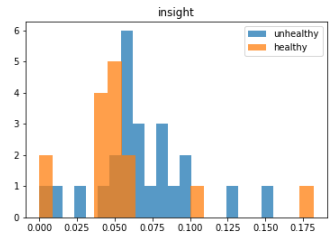

Рис. 78: Insight

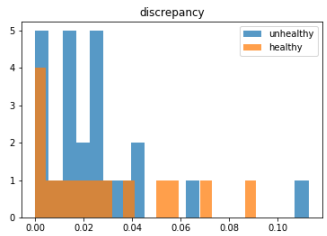

Рис. 79: Discrepancy

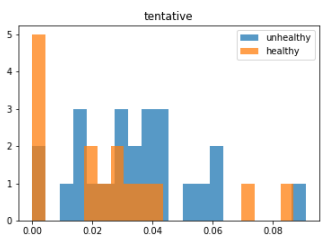

Рис. 80: Tentative

Список литературы

[1] William Armstrong. *Using Topic Models to Investigate Depression on Social Media*. University of Maryland, 2015.

[2] Phoebe Carter. *Understanding expressive language disturbance in borderline personality disorder*. University of Wollongong, 2011.

[3] Andrew Thompson Catherine Whisper Ayten Bilgin. *The prevalence of personality disorders in the community: a global systematic review and meta-analysis*. The British Journal of Psychiatry, 2019.

[4] Philip Resnik William Armstrong Leonardo Claudino. *Beyond LDA: Exploring Supervised Topic Modeling for Depression-Related Language in Twitter*. Proceedings of the 2nd Workshop on Computational Linguistics и Clinical Psychology: From Linguistic Signal to Clinical Reality, 2015.

[5] Peter Tyrer Roger Mulder Mike Crawford. *Personality disorder: a new global perspective*. World Psychiatry, 2010.

[6] Margaret Mitchell Glen Coppersmith Mark Dredze Craig Harman Kristy Hollingshead. *CLPsych 2015 Shared Task: Depression and PTSD on Twitter*. Conference: Proceedings of the 2nd Workshop on Computational Linguistics и Clinical Psychology: From Linguistic Signal to Clinical Reality, 2015.

[7] Jina Kim Jieon Lee Eunil Park Jinyoung Han. *A deep learning model for detecting mental illness from user content on social media*. Scientific Reports, 2020.

[8] Munmun De Choudhury Michael Gamon Scott Counts Eric Horvitz. *Predicting Depression via Social Media*. Seventh International AAAI Conference on Weblogs и Social Media, 2013.

[9] Schnurr P. P. Rosenberg S. D. Oxman T. E. Tucker G. J. *A methodological note on content analysis: Estimates of reliability*. Journal of Personality Assessment, 1986.

[10] Kate G. Niederhoffer James W. Pennebaker Matthias R. Mehl. *Psychological Aspects of Natural Language Use: Our Words, Our Selves*. Annual Review of Psychology, 2003.

[11] H. Andrew Schwartz Johannes Eichstaedt Margaret L. Kern. *Towards Assessing Changes in Degree of Depression through Facebook*. Proceedings of Conference of the Association for Computational Linguistics (ACL), 2014.

[12] Goldberg LR. *Language and individual differences: The search for universals in personality lexicons*. Review of Personality и social psychology, 1981.

[13] Tom Johnstone Mohammed Al-Mosaiwi. *In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation.* Clinical Psychological Science, 2018.

[14] Alexander Sboev D. Gudovskikh Roman Rybka Ivan Moloshnikov. *A Quantitative Method of Text Emotiveness Evaluation on Base of the Psycholinguistic Markers Founded on Morphological Features.* Procedia Computer Science, 2015.

[15] Anagnostakis K Newton-Howes G Tyrer P. *The prevalence of personality disorder, its comorbidity with mental state disorders, and its clinical significance in community mental health teams.* Soc Psychiatry Psychiatr Epidemiol, 2010.

[16] World Health Organization. *International statistical classification of diseases and related health problems (11th ed.)* 2019.

[17] World Health Organization. *Mental Health Atlas 2017.* World Health Organization, 2018.

[18] F Rouillon P Gorwood. *Treatment response in major depression: effects of personality dysfunction and prior depression.* British Journal of Psychiatry, 2010.

[19] S.H.Lovibond P.F.Lovibond. *The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories.* Behavior Research и Therapy, 1995.

[20] Daniel Preoţiuc-Pietro Johannes Eichstaedt Gregory Park. *The role of personality, age, and gender in tweeting about mental illness.* Association for Computational Linguistics, 2015.

[21] Lay T. C. Pennebaker J. W. *Language Use and Personality during Crises: Analyses of Mayor Rudolph Giuliani's Press Conferences.* Journal of Research in Personality, 2002.

[22] Janice M McKenzie Peter R Joyce. *Temperament, childhood environment and psychopathology as risk factors for avoidant and borderline personality disorders.* Australian & New Zealand Journal of Psychiatry, 2003.

[23] Peter Tyrer Priya Bajaj. *Managing mood disorders and comorbid personality disorders.* Current Opinion in Psychiatry, 2005.

[24] *Python Yandex Speller is a search tool typos in the text, files and websites.* 2020.

[25] Gabriel Altmann Radek Čech Ioan-Iovitz Popescu. *Metody kvantitativní analýzy (nejen) básnických textů.* 2014.

[26] David N. Milne Rafael A. Calvo. *Natural language processing in mental health applications using non-clinical texts.* Natural Language Engineering, 2017.

[27] Milošeska K Ranger M Tyrer P. *Cost-effectiveness of nidotherapy for comorbid personality disorder and severe mental illness: randomized controlled trial.* Epidemiol Psichiatria Soc, 2009.

[28] *RusVector corpus.* 2020.

[29] Alexander Sboev Tatiana Litvinova Dmitry Gudovskikh Roman Rybka. *Machine Learning Models of Text Categorization by Author Gender Using Topic-independent Features.* Procedia Computer Science, 2016.

[30] Thang Nguyen Jordan Boyd-Graber Jeffrey Lund Kevin D. Seppi. *Is Your Anchor Going Up or Down? Fast and Accurate Supervised Topic Models.* Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015.

[31] Marc-Andre Bouchard Serge Lecours. *Verbal elaboration of distinct affect categories and BPD symptoms.* Psychology, Psychotherapy: Theory, Research и Practice, 2011.

[32] Kholifah B. Syarif I. Badriyah T. *Mental Disorder Detection via Social Media Mining using Deep Learning.* Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, и Control, 2020.

[33] Robert Thorstad Phillip Wolff. *Predicting future mental illness from social media: A big-data approach.* The Psychonomic Society, 2019.

[34] Tatiana Litvinova P.V. Seredin Olga Litvinova Olga Zagorovskaya. *Profiling a set of personality traits of text author: What our words reveal about us.* Research in Language, 2016.