

Симпозиум «Технологии анализа естественного языка в медицине»  
в рамках 15-го Международного форума MedSoft-2019

# Анализ депрессивных состояний в социальных сетях с помощью технологий NLP

Иван Валентинович Смирнов

Институт проблем искусственного интеллекта ФИЦ ИУ РАН

# Признаки депрессии в сообщениях социальных медиа

Депрессивные пользователи социальных сетей:

- Употребляют специфичную лексику (исключительную, негативную лексику и т.п.)
- Имеют измененный стиль письма
- Публикуют меньше сообщений
- Имеют меньше социальных связей (подписки и подписчики)
- Чаще употребляют личные местоимения
- Чаще публикуют изображения в других оттенках, используют особенные фильтры
- И т.д.

De Choudhury M., Counts S., Horvitz E. Social media as a measurement tool of depression in populations //Proceedings of the 5th Annual ACM Web Science Conference. – ACM, 2013. – С. 47-56.

Al-Mosaiwi M., Johnstone T. In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation //Clinical Psychological Science. – 2017. – С. 2167702617747074.

Reece A. G., Danforth C. M. Instagram photos reveal predictive markers of depression //EPJ Data Science. – 2017. – Т. 6. – №. 1. – С. 15.

# Предсказание депрессии на соревновании Clef/eRisk 2017 Early Risk Prediction on the Internet

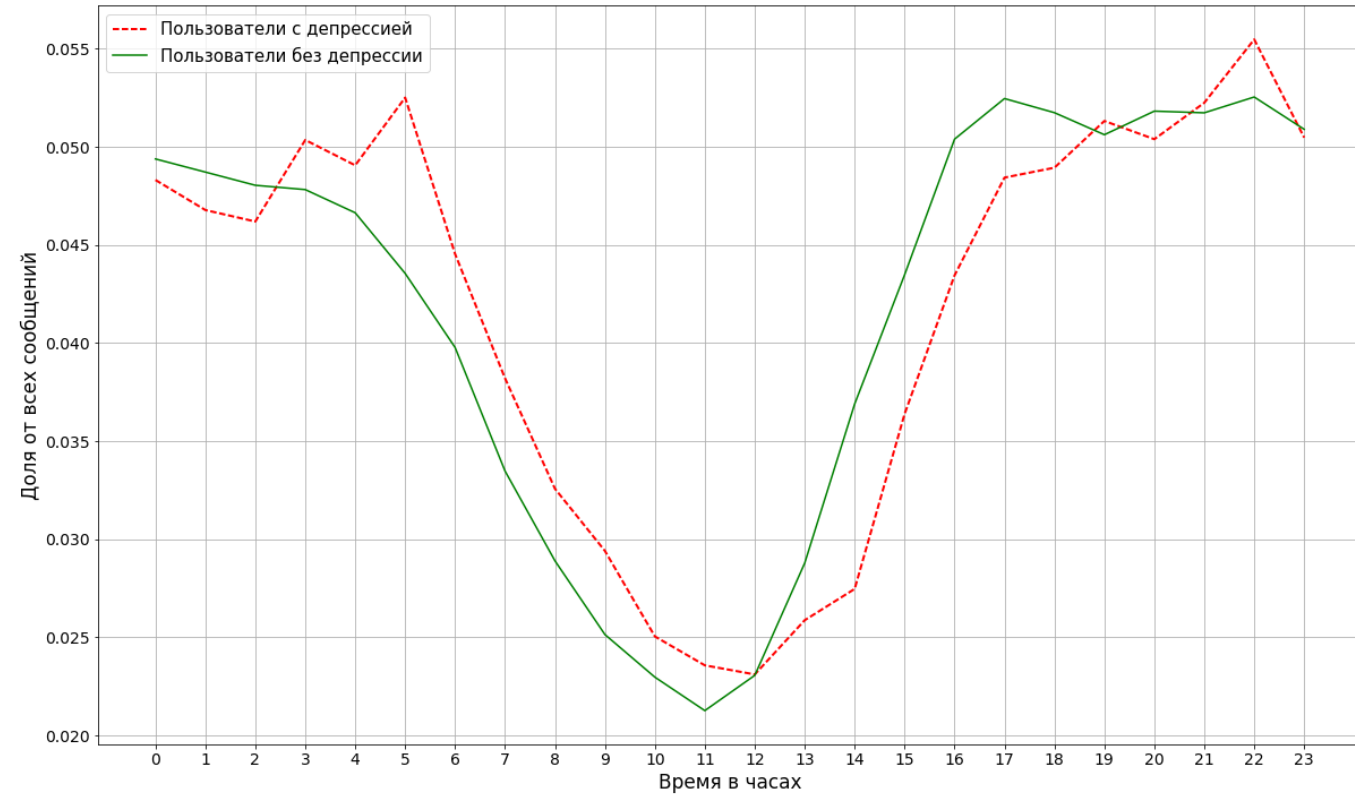
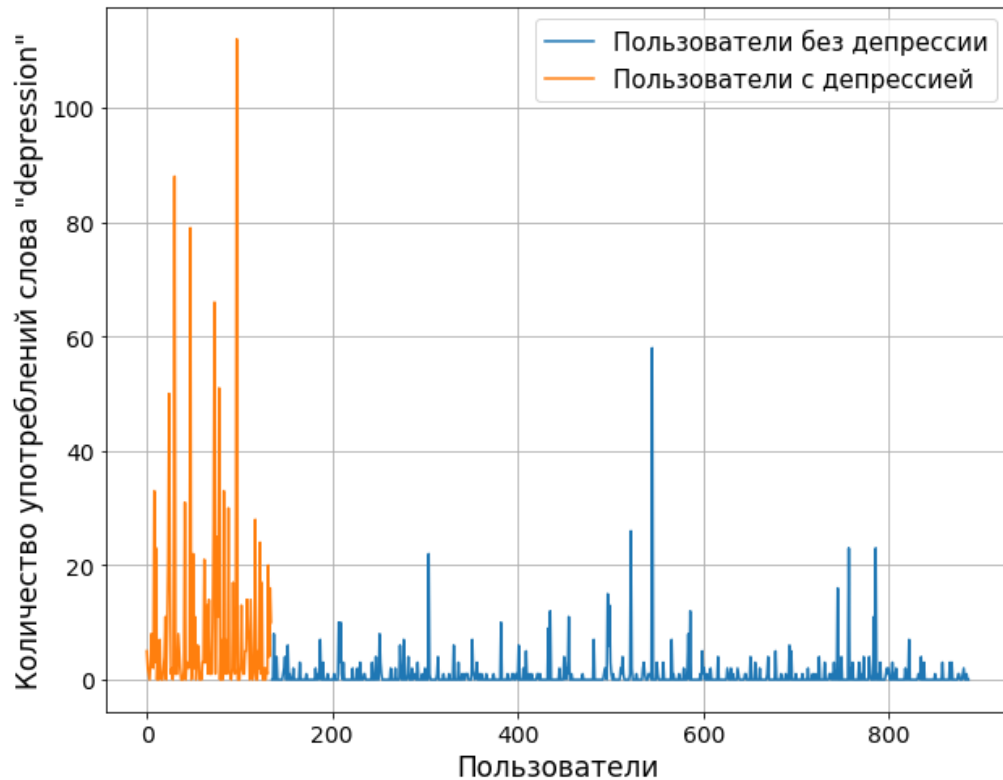
Задача:

Предсказать наличие признаков депрессии у пользователей социальной сети  
Reddit на основе анализа их текстовых сообщений (на английском языке)

## *Статистика*

Количество пользователей	887
Пользователей с депрессией	135 (15%)
Пользователей без депрессии	752 (85%)
Обучающая часть	486 (83 / 403)
Тестовая часть	401 (52 / 349)

# Особенности в активности пользователей



- Депрессивные и не депрессивные пользователи пишут сообщения с разным количеством слов и стилем
- Депрессивные и не депрессивные пользователи пишут разное количество сообщений в некоторые периоды времени

# Лингвистические маркеры

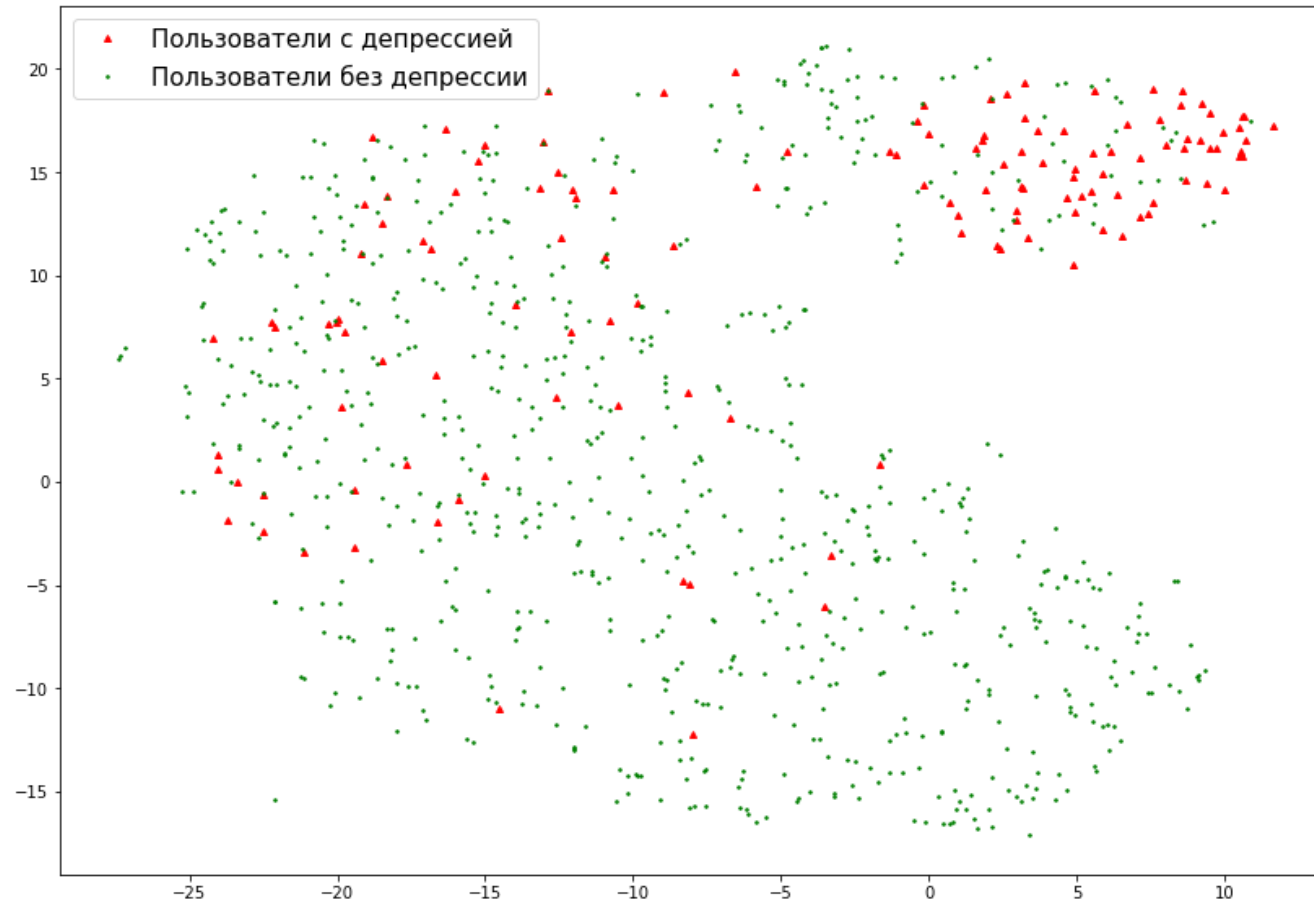
## Средние значения стилистических признаков

Средние значения	Пользователи с депрессией	Пользователи без депрессии
Слов в одном сообщении	41.39 ± 19.94	27.77 ± 13.98
Предложений в одном сообщении	2.89 ± 1.20	2.01 ± 0.68
Слов в одном предложении	14.77 ± 3.82	13.09 ± 3.36

## Пропорции использования различных частей речи

<i>Часть речи</i>	<i>В текстах с депрессией, %</i>	<i>В текстах без депрессии, %</i>
Существительные	19.36	25.18
Глаголы	22.04	20.26
Местоимения	13.55	10.00
Прилагательные	6.70	7.81
Артикли	7.82	8.80
Наречия	13.69	12.12
Союзы	3.60	2.93

# Результаты экспериментов



Метод определяет наличие признаков депрессии в текстах пользователей социальной сети с эффективностью Accuracy score = 92%, F1 score = 62.79

Stankevich M., Isakov V., Devyatkin D. and Smirnov I. (2018). Feature Engineering for Depression Detection in Social Media. In Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods, pages 426-431. DOI: 10.5220/0006598604260431

<i>Model</i>	<i>F1-score</i>	<i>Precision</i>	<i>Recall</i>
SVM: tf-idf+stl+mrph	<b>0.63</b>	0.65	0.61
SVM: emb+stl	0.61	0.48	<b>0.84</b>
RF: tf-idf+mrph	0.62	<b>0.79</b>	0.51

# Что для русского языка?

- Эссе на тему “Я, другие, мир”:
  - написанные людьми с клинической депрессией (группа депрессивных)
  - написанные здоровыми людьми (группа здоровых)

	<b>Группа депрессивных</b>	<b>Группа здоровых</b>
Количество эссе	93	223
Среднее кол. символов	1883 ± 895	1994 ± 207
Среднее кол. слов	294.83 ± 145.89	317.75 ± 35.6
Среднее кол. предложений	22.6 ± 11.28	23.11 ± 6.93

# Лингвистические маркеры депрессии

*Среднее количество слов в предложении*

*(кол-во инфинитивов) / (кол-во глаголов)*

*Среднее количество символов в слове*

*(кол-во глаголов единственного числа первого*

*Лексикон: (кол-во уникальных слов) / (кол-во слов)*

*лица прошедшего времени) / (кол-во глаголов)*

*Средняя глубина синтаксического дерева*

*(кол-во глаголов первого лица) / (кол-во глаголов)*

**+ N-граммы слов**

*(кол-во глаголов) / (кол-во прилагательных)*

*(кол-во глаголов третьего лица) / (кол-во*

*(кол-во союзов + кол-во предлогов) / (кол-во*

*глаголов)*

*предложений sentences)*

*(кол-во местоимений первого лица) / (кол-во*

*местоимений)*

Признаки	Полнота, %	Точность, %	F1, %	Аккуратность, %
ЛМ	65.53 ± 8.31	<b>77.52</b> ± 4.91	70.65 ± 5.39	<b>84.16</b> ± 2.15
Униграммы	69.83 ± 9.36	70.87 ± 7.31	69.69 ± 4.51	82.27 ± 2.4
Униграммы + ЛМ	72.01 ± 10.03	70.7 ± 10.18	70.48 ± 5.36	82.28 ± 3.4
Униграммы+биграмммы	<b>76.26</b> ± 7.37	70.42 ± 5.69	72.72 ± 2.44	83.21 ± 1.77
Униграммы+биграмммы+ЛМ	74.18 ± 3.11	72.12 ± 4.16	<b>73.01</b> ± 2.11	83.85 ± 1.46



# Что мы измеряем в социальных сетях?

- Депрессия
  - Высокая
  - Средняя
- Личностные черты (по 3-х балльной шкале: высокий, средний, низкий)
  - Нейротизм
  - Экстраверсия
  - Сознательность
  - Открытость опыту
  - Готовность к согласию

# Собираемые данные

- Результаты прохождения опросников: Шкала депрессии Бека, Большая пятерка личностных черт (NEO-FFI)
- Собирается следующая информация из Вконтакте:
  - Посты
  - Репосты
  - Комментарии
  - Фото
  - Лайки
  - Друзья
  - Сообщества
  - И т.п.
- Более 200 испытуемых

# Ключевые признаки

- ❖ Пол и возраст
- ❖ Количество лайков и репостов
- ❖ Количество друзей и подписчиков
- ❖ Количество групп
- ❖ Количество фото в основном профиле
- ❖ Активность в ночное время
- ❖ Общее количество постов с января 2017
- ❖ Отношение количества личных сообщений и репостов
- ❖ И т.д.

# Результаты определения личностных черт

Процент испытуемых	Нейротизм, %	Сознательность, %	Экстраверсия, %	Открытость опыту, %	Готовность к согласию, %
<b>Высокий</b>	16.3	24.8	13.9	23.1	11.5
<b>Средний</b>	49.6	55.7	58.1	66.3	73.3
<b>Низкий</b>	33.9	19.3	27.8	10.6	15.1

	Random Forest			SVM		
	Полнота, %	Точность, %	F1-score, %	Полнота, %	Точность, %	F1-score, %
<b>Нейротизм</b>	<b>49.07</b>	<b>53.01</b>	<b>49.51</b>	33.88	49.17	33.38
<b>Экстраверсия</b>	35.19	37.12	35.46	<b>37.54</b>	<b>41.69</b>	<b>36.17</b>
<b>Открытость опыту</b>	<b>46.41</b>	46.79	<b>46.38</b>	40.02	<b>47.04</b>	41.78
<b>Согласие</b>	<b>44.65</b>	47.50	<b>45.46</b>	32.28	<b>52.47</b>	35.07
<b>Сознательность</b>	<b>51.15</b>	56.04	<b>53.15</b>	38.10	<b>57.59</b>	43.26

# Отчет об автоматической оценке психологических характеристик

## Большая пятерка личностных черт

### У вас низкий уровень **Нейротизма**

Низкие показатели говорят об эмоциональной стабильности, самодостаточности, уверенности в себе. Такие люди спокойны, уравновешены. Их тяжело смутить. Для этих людей характерна высокая стрессоустойчивость. О таких людях говорят, как об эмоционально устойчивых.

### У вас низкий уровень **Экстраверсии**

Низкие значения говорят о замкнутости, сдержанности, сниженном эмоциональном фоне. Таких людей называют одиночками. Они часто серьезно относятся к жизни, избегают сильных эмоций. Предпочитают проводить время за такими занятиями, как, например, чтение книг. В больших компаниях испытывают тревогу.

### У вас средний уровень **Открытости опыту**

Средние значения говорят, что любознательность и воображение проявляются у такого человека умеренно. В зависимости от ситуации, такие люди могут как быть консервативны, так и заинтересоваться чем-то новым, неизведанным. Такие люди склонны действовать привычным образом, однако, когда видят пользу, могут согласиться на новаторские методы и даже разработать их.

### У вас средний уровень **Готовности к согласию**

Средние значения говорят о том, что такие люди склонны к сотрудничеству и кооперации, однако решающим для них будет их собственное мнение. Они умеренно независимы, способны проявлять дружелюбие и отзывчивость, если того требует ситуация. Находясь в команде, они склонны искать компромисс, однако, и соревновательный дух им не чужд.

### У вас средний уровень **Сознательности**

Средние значения говорят о том, что решения такого человека могут быть как сознательными и взвешенными, так и ситуативными. Эти люди умеют делать размеренную работу, быть усидчивыми. Однако и нестандартные ситуации их привлекают.

## Депрессия

**Выраженная депрессия (средней тяжести).** Важно помнить, что депрессия хорошо поддается лечению. На сегодняшний день психологи и клинические психотерапевты работают с этим заболеванием как чисто психологическими методами, так и с использованием медикаментозного сопровождения.



**Институт  
проблем  
искусственного  
интеллекта**

Иван Валентинович Смирнов

ИПИИ ФИЦ ИУ РАН

117312, Москва, пр-т. 60-летия Октября, 9

Телефон/факс: +7 (499) 135-90-20

e-mail: [ivs@isa.ru](mailto:ivs@isa.ru)

Исследования проводились при финансовой поддержке Министерства образования и науки РФ

Уникальный идентификатор соглашения RFMEFI60417X0194