

User feedback analysis

Анализ тональности отзывов пользователей

Гумеров Ильмаз  
Пашкин Артем  
Залялутдинов Муса

# Text Corpus. Corpus 1.

## Pharmacy reviews (2 Russian, 1 English)

"Всем здравствуйте. Сегодня хочу рассказать вам про один из видов леденцов для детей - это Мультивитаминные леденцы Vitatone. Были как-то на днях в аптеке с ребенком и он увидел на витрине данный леденец на палочке. Никогда не думал, что в аптеках могут продавать сладости, а оказалось, что это леденец-витаминка. Решили купить попробовать, тем более что цена была относительно невысокая. Хотя если подумать, что за эту сумму можно было купить несколько обычных леденцов в магазине - но этот же - витаминный). Сам леденец неплохой - содержит витамин С и другие микроэлементы. К тому же имеет приятный вкус. Для деток это очень важно, так как они не всегда хотят принимать обычные витамины, которые безвкусные или имеют неприятный вкус. Также следует отметить и очень привлекательную и яркую упаковку данного витамина, что немаловажно для деток. Если развернуть упаковку - то сам леденец имеет интересную форму - в виде сердечка и красный цвет. К тому же он еще и на палочке - нашему ребенку такое нравится. Еще одним немаловажным моментом является то, что данный витамин производитель не рекомендует принимать деткам до 3-х лет. В общем - теперь каждый наш последующий поход в аптеку с ребенком наверное будет сопровождаться этой покупкой). Год выпуска/покупки: 2016  
Общее впечатление : Хорошее решение для деток",positive

"Во время путешествия по Вьетнаму в одной из аптек я наткнулась вот на такой бальзам. Было несколько слов о его свойствах, стоил сущие копейки и решила взять. Теперь жалею, что не взяла побольше. Вот в такой коробочке он находится, яркая красная упаковка. Слоник небольшой нарисован. А внутри масло-бальзам в небольшой бутылочке объемом 3 мл. Применять бальзам можно для борьбы с гриппом, головной болью, насморком, заложенностью носа, мазать места укусов насекомых. Я применяю его когда болит голова. Очень легко избавиться от головной боли, если помазать виски этим средством и углубление между шеей и головой (не знаю как это место называется), где углубление. Проходит через 5 минут. У меня все на

# Corpus 2

```
{"title":"Невозможно дозвониться в регистратуру","text":"Второй день не могу дозвониться до регистратуры. Явно сняли трубку и положили рядом. Посему отзыв отрицательный.", "url":"https://www.spr.ru/forum_vyvod.php?id_tema=2806550", "date":"2014-01-16 15:37", "author":"Максим", "path":"Главная > Мещанский > Городская поликлиника № 137 > Отзывы", "positive":false} {"title":"Благодарность медсестре","text":"Выражаю благодарность медсестре физитерапевтического кабинета Царевой Галине Валерьевне. Всегда приветливая, улыбчивая, находит подход даже к самому капризному пациенту. Поздравляю с приближающимся днем медицинского работника и прошу главного врача отметить работу такой прекрасной медсестры!", "url":"https://www.spr.ru/forum_vyvod.php?id_tema=3928837", "date":"2017-06-03 19:44", "author":"Татьяна", "path":"Главная > Отрадное > Детская городская поликлиника № 110 > Отзывы", "positive":true} {"title":"Разочарована","text":"Записывалась на консультацию, первое впечатление хорошее, но последующие визиты разочаровали, то врач задержалась, то долгое ожидание. Созвониться с врачом было очень и очень проблематично. А после процедуры на проверку проходимости маточных труб, врач просто забыла назначить антибиотик и противогрибковые препараты, от чего началась жуткая молочница, хотя после она свою вину признала, но мне то от этого не легче.", "url":"https://www.spr.ru/forum_vyvod.php?id_tema=4001115", "date":"2017-07-07 12:00", "author":"Евгения", "path":"Главная > Измайлово > Фертимед > Отзывы", "positive":false}
```

# Corpus 3

`{"sideEffect": "After 4 1/2 years on Cymbalta(30mg)_ and still dealing with Depression from PTSD_ I had a 24 hr psychosis event and ended up in the emergency room. My doctor then thought it best to bump my dosage to 60 mg. That extra 30mg of Cymbalta _ almost was the last nail in my coffin. Within a few months I was always irritable and angry verbally abusive to my family. My anxiety was on full blast and I couldn't turn it off_ causing tremors and restless legs. I slept 3 to 4 days a week_ and I had extreme depression. I had migraines and lived in a fog and couldn't even comprehend basic responsibilities _ like paying my bills or working my cell phone. I showed paranoia symptoms and my blood pressure would raise or drop at any given moment_ I was often dizzy or light headed_ probably from the raising and falling of my blood pressure. My heart would beat rapidly much of the time_ and my sleep patterns way off. Because many of these symptoms resembled extreme PTSD symptoms and my VA doctor so many mil", "rating": 1, "tags": [50.0, "M"], "text": "I am experiencing residue side effect, after a few months without Cymbalta, mostly in tremors and restless legs. Also there is alway an amount of anxiety and nervousness that I never had before, and now rarely goes away.. After 4 1/2 years on Cymbalta(30mg), and still dealing with Depression from PTSD, I had a 24 hr psychosis event and`

# Python libraries

- ❏ pandas
- ❏ sklearn



# Texterra

Textterra automatically performs linguistic analysis of texts in Russian and English and extracts factual information, as well as analyzes the opinions presented in the texts.

The screenshot displays the Texterra web application interface. On the left is a sidebar menu with the following sections:

- Извлечение информации** (Information Extraction)
  - Именованные сущности (Named Entities) - highlighted in blue
  - Викификация (Wikification)
  - Ключевые сущности (Key Entities)
  - Луркификация (Lurkification)
- Лингвистика** (Linguistics)
  - Морфология (Morphology)
  - Синтаксис (Syntax)
  - Разрешение кореферентности (Coreference Resolution)
  - Исправление ошибок (Error Correction)
- Анализ мнений** (Opinion Analysis)
  - Определение тональности (Sentiment Determination)

Several 'BETA' labels are visible next to the 'Именованные сущности', 'Разрешение кореферентности', and 'Определение тональности' items.

The main content area shows a news article snippet:

Следующая новость \$ [social media icons] [edit icon]

Карло Анчелотти назначен главным тренером «Наполи».

Итальянский специалист Карло Анчелотти стал главным тренером итальянского футбольного клуба "Наполи". Об этом сообщает пресс-служба клуба.

Яндекс Новости # 23.05.2018 20:55

Below the article, the 'Найденные сущности' (Found Entities) section is displayed:

- ORGANIZATION\_OTHER**
  - "Наполи"
- PERSON**
  - Карло Анчелотти

# Data analysis

The total number of:

- ☐ reviews (positive/negative)
- ☐ words
- ☐ unique words
- ☐ PMI metrics

# Analyzing corpus2

- number of reviews:  
59937
- number of words:  
3555677
- size of dictionary:  
80877
- number of positive reviews:  
32220
- number of negative reviews:  
27717



# Analyzing corpus3

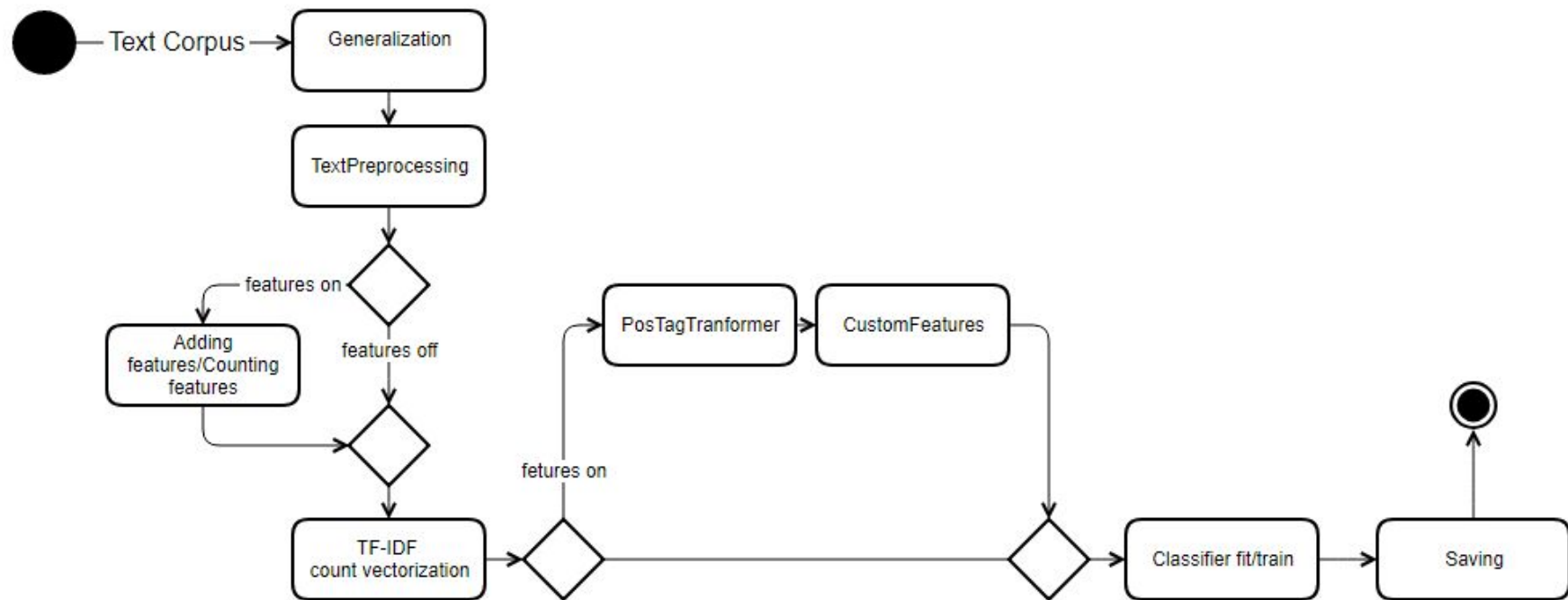
- number of reviews:  
113836
- number of words:  
6105654
- size of dictionary:  
72523
- number of positive reviews:  
46140
- number of negative reviews:  
49090
- number of neutral reviews:  
18606

# PMI

$$PMI(w, positive) = \log_2 \frac{freq(w, positive) * N}{freq(w) * freq(positive)}$$

- ❑  $freq(w, positive)$  is the number of times a term  $w$  occurs in positive reviews
- ❑  $freq(w)$  is the total frequency of term  $w$  in the corpus
- ❑  $freq(positive)$  is the total number of tokens in positive reviews
- ❑  $N$  is the total number of tokens in the corpus.

$PMI(w, negative)$  is calculated in a similar way.



# Features

- ❑ the number of tokens with  $\text{score}(w) \leq 0$ ;
- ❑ the total score =  $\sum_{w \in \text{review}} (\text{score}(w))$ ;
- ❑ the maximal score =  $\max_{w \in \text{review}} \text{score}(w)$ ;
- ❑ the score of the last token in the review.
- ❑ the sum of positive scores for review tokens in affirmative contexts;
- ❑ the sum of negative scores for review tokens in affirmative contexts;
- ❑ the sum of positive scores for review tokens in negated contexts;
- ❑ the sum of negative scores for review tokens in negated contexts

# Classifiers

- ❑ SVC
- ❑ LogisticRegression
- ❑ MultinomialNB
- ❑ LGBMClassifier

# Results: baseline corpus2

	precision	recall	f1-score	support	positive	negative
negative	0.92	0.96	0.94	3847	<ul style="list-style-type: none"><li>vectorizer__спасибо</li><li>vectorizer__благодарность</li><li>vectorizer__очень</li><li>vectorizer__поблагодарить</li><li>vectorizer__быстро</li><li>vectorizer__хочу</li><li>vectorizer__огромное</li><li>vectorizer__большое</li><li>vectorizer__довольна</li><li>vectorizer__приятно</li><li>vectorizer__выразить</li><li>vectorizer__профессионализм</li><li>vectorizer__благодарна</li><li>vectorizer__всем</li><li>vectorizer__благодарю</li><li>vectorizer__внимательный</li><li>vectorizer__хороший</li><li>vectorizer__отличный</li><li>vectorizer__побольше</li><li>vectorizer__нравится</li></ul>	<ul style="list-style-type: none"><li>vectorizer__сказала</li><li>vectorizer__деньги</li><li>vectorizer__почему</li><li>vectorizer__вообще</li><li>vectorizer__это</li><li>vectorizer__итоге</li><li>vectorizer__трубку</li><li>vectorizer__вопрос</li><li>vectorizer__хамское</li><li>vectorizer__сказали</li><li>vectorizer__ужас</li><li>vectorizer__такое</li><li>vectorizer__работать</li><li>vectorizer__дозвониться</li><li>vectorizer__невозможн</li><li>vectorizer__сегодня</li><li>vectorizer__ответ</li><li>vectorizer__хамство</li><li>vectorizer__грубо</li><li>vectorizer__часа</li></ul>
positive	0.95	0.91	0.93	3645		
avg / total	0.93	0.93	0.93	7492		

# Results: features corpus2

	precision	recall	f1-score	support	positive	negative
negative	0.91	0.94	0.92	3847	<ul style="list-style-type: none"><li>• vectorizer__спасибо</li><li>• vectorizer__очень</li><li>• vectorizer__благодарность</li></ul>	<ul style="list-style-type: none"><li>• vectorizer__сказала</li><li>• vectorizer__почему</li><li>• vectorizer__вообще</li></ul>
positive	0.93	0.90	0.92	3645	<ul style="list-style-type: none"><li>• vectorizer__хочу</li><li>• vectorizer__большое</li><li>• vectorizer__быстро</li></ul>	<ul style="list-style-type: none"><li>• vectorizer__невозможно</li><li>• vectorizer__дозвониться</li><li>• vectorizer__трубку</li></ul>
avg / total	0.92	0.92	0.92	7492	<ul style="list-style-type: none"><li>• vectorizer__огромное</li><li>• vectorizer__поблагодарить</li><li>• vectorizer__выразить</li><li>• vectorizer__всем</li><li>• vectorizer__довольна</li><li>• vectorizer__хороший</li><li>• vectorizer__лечение</li><li>• vectorizer__понравилось</li><li>• vectorizer__профессионализм</li><li>• vectorizer__здоровья</li><li>• vectorizer__доктор</li><li>• vectorizer__хорошее</li><li>• vectorizer__помощь</li><li>• vectorizer__внимательный</li></ul>	<ul style="list-style-type: none"><li>• vectorizer__сказали</li><li>• vectorizer__часа</li><li>• vectorizer__такое</li><li>• vectorizer__деньги</li><li>• vectorizer__это</li><li>• vectorizer__итоге</li><li>• vectorizer__вопрос</li><li>• vectorizer__нужно</li><li>• vectorizer__врача</li><li>• vectorizer__т</li><li>• vectorizer__направление</li><li>• vectorizer__регистратуре</li><li>• vectorizer__сегодня</li><li>• vectorizer__записаться</li></ul>

# Result: features 3c corpus2

	positive				negative			
	precision	recall	f1-score	support				
negative	0.92	0.95	0.93	3847	•	vectorizer__спасибо	•	vectorizer__?
					•	vectorizer__благодарность	•	vectorizer__сказала
					•	vectorizer__очень	•	vectorizer__деньги
					•	vectorizer__поблагодарить	•	vectorizer__вообще
positive	0.95	0.91	0.93	3645	•	vectorizer__хочу	•	vectorizer__(
					•	vectorizer__быстро	•	vectorizer__итоге
avg / total	0.93	0.93	0.93	7492	•	vectorizer__огромное	•	vectorizer__``
					•	vectorizer__большое	•	vectorizer__трубку
					•	vectorizer__довольна	•	vectorizer__"
					•	vectorizer__выразить	•	vectorizer__хамское
					•	vectorizer__приятно	•	vectorizer__ужас
					•	vectorizer__профессионализм	•	vectorizer__сказали
					•	vectorizer__благодарна	•	vectorizer__это
					•	vectorizer__всем	•	vectorizer__дозвониться
					•	vectorizer__благодарю	•	vectorizer__почему
					•	vectorizer__внимательный	•	vectorizer__невозможно
					•	vectorizer__хороший	•	vectorizer__работать
					•	vectorizer__отличный	•	vectorizer__такое
					•	vectorizer__побольше	•	vectorizer__грубо



# Results: baseline corpus3

	precision	recall	f1-score	support	positive	negative
negative	0.75	0.86	0.80	6531	<ul style="list-style-type: none"><li>• vectorizer__saved</li><li>• vectorizer__love</li><li>• vectorizer__great</li></ul>	<ul style="list-style-type: none"><li>• vectorizer__never</li><li>• vectorizer__poison</li><li>• vectorizer__worse</li></ul>
neutral	0.48	0.10	0.17	2236	<ul style="list-style-type: none"><li>• vectorizer__none</li><li>• vectorizer__best</li><li>• vectorizer__miracle</li></ul>	<ul style="list-style-type: none"><li>• vectorizer__market</li><li>• vectorizer__stopped</li><li>• vectorizer__worst</li></ul>
positive	0.71	0.82	0.76	5463	<ul style="list-style-type: none"><li>• vectorizer__works</li><li>• vectorizer__amazing</li><li>• vectorizer__slight</li></ul>	<ul style="list-style-type: none"><li>• vectorizer__horrible</li><li>• vectorizer__fda</li><li>• vectorizer__ruined</li></ul>
avg / total	0.69	0.73	0.69	14230	<ul style="list-style-type: none"><li>• vectorizer__wonderful</li><li>• vectorizer__wonders</li><li>• vectorizer__helped</li><li>• vectorizer__happy</li><li>• vectorizer__life</li><li>• vectorizer__saver</li><li>• vectorizer__years</li><li>• vectorizer__godsend</li><li>• vectorizer__excellent</li><li>• vectorizer__mild</li><li>• vectorizer__gone</li></ul>	<ul style="list-style-type: none"><li>• vectorizer__took</li><li>• vectorizer__dangerous</li><li>• vectorizer__drug</li><li>• vectorizer__boniva</li><li>• vectorizer__severe</li><li>• vectorizer__terrible</li><li>• vectorizer__vomiting</li><li>• vectorizer__system</li><li>• vectorizer__doctors</li><li>• vectorizer__damage</li><li>• vectorizer__extreme</li></ul>

# Results: baseline corpus3

neutral

- vectorizer\_\_however
- vectorizer\_\_somewhat
- vectorizer\_\_see
- vectorizer\_\_sure
- vectorizer\_\_recently
- vectorizer\_\_whether
- vectorizer\_\_hoping
- vectorizer\_\_norvasc
- vectorizer\_\_yet
- vectorizer\_\_help
- vectorizer\_\_hope
- vectorizer\_\_making
- vectorizer\_\_helping
- vectorizer\_\_know
- vectorizer\_\_weight
- vectorizer\_\_wondering
- vectorizer\_\_stand
- vectorizer\_\_helped
- vectorizer\_\_unsure
- vectorizer\_\_seems

# Results: features corpus3

	precision	recall	f1-score	support	positive	negative
negative	0.77	0.84	0.80	6531	<ul style="list-style-type: none"><li>vectorizer__saved</li><li>vectorizer__none</li><li>vectorizer__love</li></ul>	<ul style="list-style-type: none"><li>vectorizer__never</li><li>vectorizer__poison</li><li>vectorizer__market</li></ul>
neutral	0.45	0.15	0.22	2236	<ul style="list-style-type: none"><li>vectorizer__great</li><li>vectorizer__life</li></ul>	<ul style="list-style-type: none"><li>vectorizer__stopped</li><li>vectorizer__worse</li></ul>
positive	0.70	0.82	0.76	5463	<ul style="list-style-type: none"><li>vectorizer__miracle</li><li>vectorizer__slight</li><li>vectorizer__best</li></ul>	<ul style="list-style-type: none"><li>vectorizer__took</li><li>vectorizer__fda</li><li>vectorizer__ruined</li></ul>
avg / total	0.69	0.72	0.69	14230	<ul style="list-style-type: none"><li>vectorizer__amazing</li><li>vectorizer__happy</li><li>vectorizer__works</li><li>vectorizer__wonders</li><li>vectorizer__years</li><li>vectorizer__wonderful</li><li>vectorizer__saver</li><li>vectorizer__godsend</li><li>vectorizer__gone</li><li>vectorizer__occasional</li><li>vectorizer__mild</li><li>vectorizer__helps</li></ul>	<ul style="list-style-type: none"><li>vectorizer__dangerous</li><li>vectorizer__worst</li><li>vectorizer__drug</li><li>vectorizer__recommend</li><li>vectorizer__horrible</li><li>vectorizer__boniva</li><li>vectorizer__worth</li><li>vectorizer__please</li><li>vectorizer__system</li><li>vectorizer__doctors</li><li>vectorizer__became</li><li>vectorizer__work</li></ul>

# Results: features corpus3

neutral

- vectorizer\_\_however
- vectorizer\_\_somewhat
- vectorizer\_\_see
- vectorizer\_\_sure
- vectorizer\_\_recently
- vectorizer\_\_whether
- vectorizer\_\_hoping
- vectorizer\_\_yet
- vectorizer\_\_norvasc
- vectorizer\_\_hope
- vectorizer\_\_help
- vectorizer\_\_helping
- vectorizer\_\_weight
- vectorizer\_\_making
- vectorizer\_\_helped
- vectorizer\_\_rate
- vectorizer\_\_know
- vectorizer\_\_unsure
- vectorizer\_\_hard
- vectorizer\_\_afternoon

# Result: features 3c corpus3

	precision	recall	f1-score	support	positive		negative	
negative	0.75	0.87	0.80	6531	•	vectorizer__love	•	vectorizer__never
					•	vectorizer__saved	•	vectorizer__poison
neutral	0.49	0.10	0.16	2236	•	vectorizer__great	•	vectorizer__worse
					•	vectorizer__none	•	vectorizer__stopped
positive	0.71	0.82	0.76	5463	•	vectorizer__best	•	vectorizer__worst
					•	vectorizer__miracle	•	vectorizer__horrible
					•	vectorizer__works	•	vectorizer__took
avg / total	0.70	0.73	0.69	14230	•	vectorizer__slight	•	vectorizer__fda
					•	vectorizer__helped	•	vectorizer__ruined
					•	vectorizer__life	•	vectorizer__drug
					•	vectorizer__wonderful	•	vectorizer__severe
					•	vectorizer__amazing	•	vectorizer__boniva
					•	vectorizer__wonders	•	vectorizer__market
					•	vectorizer__happy	•	vectorizer__dangerous
					•	vectorizer__years	•	vectorizer__extreme
					•	vectorizer__saver	•	vectorizer__terrible
					•	vectorizer__mild	•	vectorizer__became
					•	vectorizer__occasional	•	vectorizer__vomiting
					•	vectorizer__first	•	vectorizer__please
					•	vectorizer__excellent	•	vectorizer__damage

# Result: features 3c corpus3

neutral

- vectorizer\_\_however
- vectorizer\_\_somewhat
- vectorizer\_\_see
- vectorizer\_\_sure
- vectorizer\_\_hoping
- vectorizer\_\_recently
- vectorizer\_\_whether
- vectorizer\_\_hope
- vectorizer\_\_yet
- vectorizer\_\_help
- vectorizer\_\_norvasc
- vectorizer\_\_weight
- vectorizer\_\_helping
- vectorizer\_\_making
- vectorizer\_\_know
- vectorizer\_\_seems
- vectorizer\_\_hard
- vectorizer\_\_helped
- vectorizer\_\_rate
- vectorizer\_\_withdrawal

# Our repository on GitHub

[https://github.com/ILMAZ13/NLP\\_TECH](https://github.com/ILMAZ13/NLP_TECH)