"НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО" (УНИВЕРСИТЕТ ИТМО)

ЗАДАНИЕ НА КУРСОВОЙ ПРОЕКТ (РАБОТУ)

Факультет ПИиКТ Группа P41071 Направление (специальность) 09.04.04 Программная инженерия Руководитель Государев И.Б., к.п.н., доцент (Фамилия, И., О., должность,) Дисциплина Проектирование и анализ языков веб-решений Наименование темы: Сравнение библиотек Руthon для анализа тональности текста на русском языке (Наименование темы) Задание Разработать и реализовать эксперимент, позволяющий сравнить библиотеки Руthon для анализа тональности текста на русском языке Краткие методические указания (задачи работы) Изучить основные понятия в области обработки естественного языка. Проанализировать существующие подходы к решению задач анализа тональности
Направление (специальность) Руководитель Тосударев И.Б., к.п.н., доцент (Фамилия, И., О., должность,) Дисциплина Проектирование и анализ языков веб-решений Наименование темы: Сравнение библиотек Руthоп для анализа тональности текста на русском языке (Наименование темы) Задание Разработать и реализовать эксперимент, позволяющий сравнить библиотеки Руthоп для анализа тональности текста на русском языке Краткие методические указания (задачи работы) Изучить основные понятия в области обработки естественного языка.
Руководитель — Государев И.Б., к.п.н., доцент — (Фамилия, И., О., должность,) — Проектирование и анализ языков веб-решений — Наименование темы: — Сравнение библиотек Руthоп для анализа тональности текста — на русском языке — (Наименование темы) — Задание Разработать и реализовать эксперимент, позволяющий сравнить библиотеки Руthоп — для анализа тональности текста на русском языке — Краткие методические указания (задачи работы) — Изучить основные понятия в области обработки естественного языка.
Дисциплина Проектирование и анализ языков веб-решений Наименование темы: Сравнение библиотек Python для анализа тональности текста на русском языке Вадание Разработать и реализовать эксперимент, позволяющий сравнить библиотеки Python для анализа тональности текста на русском языке Краткие методические указания (задачи работы) Изучить основные понятия в области обработки естественного языка.
Наименование темы: Сравнение библиотек Python для анализа тональности текста на русском языке Парусском языке (Наименование темы) Задание Разработать и реализовать эксперимент, позволяющий сравнить библиотеки Python для анализа тональности текста на русском языке Краткие методические указания (задачи работы) Изучить основные понятия в области обработки естественного языка.
На русском языке (Наименование темы) Задание Разработать и реализовать эксперимент, позволяющий сравнить библиотеки Python для анализа тональности текста на русском языке Краткие методические указания (задачи работы) Изучить основные понятия в области обработки естественного языка.
На русском языке (Наименование темы) Задание Разработать и реализовать эксперимент, позволяющий сравнить библиотеки Python для анализа тональности текста на русском языке Краткие методические указания (задачи работы) Изучить основные понятия в области обработки естественного языка.
Задание Разработать и реализовать эксперимент, позволяющий сравнить библиотеки Python для анализа тональности текста на русском языке Краткие методические указания (задачи работы) Изучить основные понятия в области обработки естественного языка.
для анализа тональности текста на русском языке Краткие методические указания (задачи работы) Изучить основные понятия в области обработки естественного языка.
Изучить основные понятия в области обработки естественного языка.
•
•
проанализировать существующие подходы к решению задач анализа тональности
текста на естественном языке.
Проанализировать существующие инструменты языка Python для анализа
тональности текста.
Разработать план эксперимента для сравнения библиотек Python для сентимент-
анализа текста на русском языке.
Провести сравнительный анализ библиотек Python для сентимент-анализа текста на
русском языке.
Проанализировать полученные результаты и сделать выводы.
Содержание пояснительной записки
Оглавление. Введение. Теоретическая часть. Практическая часть. Заключение.
Список использованной литературы. Приложение.
список использованной литературы. приложение.
Руководитель И.Б. Государев
(подпись)
Serv.
Студент <u>П.А. Железникова</u> (подпись) (Фамилия И.О.)

"НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО'' (УНИВЕРСИТЕТ ИТМО)

ГРАФИК КУРСОВОГО ПРОЕКТА (РАБОТЫ)

Студент	Железникова П.А.		
-		(Фамилия, И., О.)	
Факультет	ПИиКТ	Группа	P41071
Направление (спеці	иальность) 09.0	4.04 Программная инже	нерия
Руководитель	Государев И.Б.,	к.п.н., доцент	
		(Фамилия, И., О., должность)	
Дисциплина	Проектирование	и анализ языков веб-рец	пений
Наименование тем	ы: <u>Сравнени</u> на русско		анализа тональности текста
	па русско		менование темы)
		1	,

№ Наименование этапа		Дата завершения		Оценка и	
		Планируемая	Фактическая	подпись руководителя	
1.	Проведение исследования по теме обработки естественного языка, библиотек Python для анализа тональности текста на русском языке.	апрель	апрель		
2.	Осуществление классификации текста с помощью выбранных библиотек Python. Составление матрицы ошибок, расчет метрик. Сравнение результатов. Написание отчета. Защита проекта.	май	май		

Руководитель		И.Б. Государев
•	(подпись)	
Студент	See.	П.А. Железникова
•	(подпись)	(Фамилия И.О.)

"НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО" (УНИВЕРСИТЕТ ИТМО)

АННОТАЦИЯ К КУРСОВОМУ ПРОЕКТУ (РАБОТЕ)

студент жел	тезникова 11.А.		
Факультет	ПИиКТ	(Фамилия, И., О.) Группа Р4	1071
направление (специаль		.04 Программная инженерия	1071
Руководитель	Государев И.Б., к.		
i juese duriens		(Фамилия, И., О., должность)	
π	П	-	
Дисциплина	Проектирование и	анализ языков веб-решений	
Наименование темы:	Сравнение	библиотек Python для анали	за тональности текста
	на русском	•	
		(Наименование	сайта)
ХАРА 1. Цель и задачи рабо		РСОВОГО ПРОЕКТА (РАБОны студентом Попредел	ОТЫ) лены руководителем
. •	* *	енты, предназначенные для р	азмещения в
электронном документ	ге		
Задачи работы:			
=		обработки естественного язын	
		ходы к решению задач анали:	за тональности
текста на естестве		D . 4	
	существующие инс	трументы языка Python для аг	нализа тональности
текста.		~ ~ P 1	
-	-	авнения библиотек Python дл	я сентимент-анализа
текста на русском		D 4	
	эльныи анализ библи	иотек Python для сентимент-а	нализа текста на
русском языке.			
о. проанализировать	полученные резуль:	гаты и сделать выводы.	
2. Характер работы			
Расчет В	Сонструирование	Моделирование	Другое
3. Содержание работ	Ы		
		еден анализ тональности русс	скоязычных текстов.
		ие выявить эффективность ка	
классификатора.	<u> </u>		индет с
<u></u>			
4. Выводы			
Требования к проекту	пеапизованы		
тресовины к проскту	решинованы		
-			
Руководитель		И.Б. Государев	
т уководитель	(подпись)	п.р. г осударсь	
Студент	Cobs.	П.А. Железников	ด
Студонт	(подпись)	(Фамилия И.О.)	<u>u</u>

"НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО" (УНИВЕРСИТЕТ ИТМО)

Факультет программной инженерии и компьютерной техники

Направление (специальность) — 09.04.04 Программная инженерия
Образовательная программа — Веб-технологии
Дисциплина — Проектирование и анализ языков веб-решений

Курсовой проект (работа)

TEMA: Сравнение библиотек Python для анализа тональности текста на русском языке

ВЫПОЛНИЛ			
Студент группь	и <u>Р41071</u>	5535-	Железникова П.А
	№ группы	подпись, дата	ФИО
ПРОВЕРИЛ	к.п.н., доцент		Государев И.Б.
	ученая степень, должность	подпись, дата	ФИО

САНКТ-ПЕТЕРБУРГ 2022 г.

СОДЕРЖАНИЕ

СОДЕРЖАНИЕ4
ВВЕДЕНИЕ6
1 ОСНОВНЫЕ СВЕДЕНИЯ ОБ АНАЛИЗЕ ТОНАЛЬНОСТИ ТЕКСТА НА
ЕСТЕСТВЕННОМ ЯЗЫКЕ
1.2 Варианты классификации текста9
1.3 Подходы к решению задач анализа тональности текста
1.4 Оценка качества анализа тональности
1.5 Обзор инструментов для анализа тональности текста на естественном
языке
1.6 Описание эксперимента
2 СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИНСТРУМЕНТОВ ДЛЯ АНАЛИЗА
ТОНАЛЬНОСТИ ТЕКСТА НА ЕСТВЕСТВЕННОМ ЯЗЫКЕ16
2.1 Экспериментальное сравнение библиотек Python для анализа тональности
текста на русском языке
2.2 Оценка полученных результатов эксперимента
ЗАКЛЮЧЕНИЕ26
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ28
ПРИЛОЖЕНИЕ 129

ВВЕДЕНИЕ

В последнее время в научном сообществе популярна проблема классификации текстов по эмоциональной окраске, которая также известна как анализ тональности. Анализ тональности является частью компьютерной лингвистики, которая изучает мнения и эмоции в текстовых данных и представляет собой набор методов, предназначенных для автоматического определения реакции эмоций или отношения (настроения), выраженных в тексте.

Долгое время практическое применение анализа тональности текстов на русском языке не достигало такого же успеха, как на английском. Этот инструмент не был развит в достаточной степени для русскоязычных данных, имелись проблемы с качеством классификации информации.

На данный момент ситуация значительно улучшилась: проводится множество исследований, посвященных проблемам анализа текстов на русском языке, а также разрабатываются новые программные продукты, устраняющие выявленные недостатки.

Большая коллекция инструментов и множество библиотек по автоматической обработке языка становятся доступными при использовании языка Python. Именно поэтому он является одним из наиболее популярных языков, применяемых для решения задач анализа тональности текста. Множество инструментов и библиотек языка предоставляют возможность обработки русскоязычных текстов.

Сравнение библиотек Python для анализа тональности текста на русском языке, позволит проанализировать имеющиеся решения, выявить их преимущества и недостатки, выбрать инструмент, осуществляющий наиболее точную классификацию. Таким образом, этим и обосновывается актуальность выбранной темы.

Цель работы состоит в исследовании современных методов анализа тональности текста, сравнительном анализе инструментов для сентиментанализа текста на русском языке.

Задачи исследования:

- 1. Изучить основные понятия в области обработки естественного языка.
- 2. Проанализировать существующие подходы к решению задач анализа тональности текста на естественном языке.
- 3. Проанализировать существующие инструменты языка Python для анализа тональности текста.
- 4. Разработать план эксперимента для сравнения библиотек Python для сентимент-анализа текста на русском языке.
- 5. Провести сравнительный анализ библиотек Python для сентиментанализа текста на русском языке.
 - 6. Проанализировать полученные результаты и сделать выводы.

1 ОСНОВНЫЕ СВЕДЕНИЯ ОБ АНАЛИЗЕ ТОНАЛЬНОСТИ ТЕКСТА НА ЕСТЕСТВЕННОМ ЯЗЫКЕ

1.2 Уровни анализа тональности

Анализ тональности — это метод, который идентифицирует текст из различных изображений и наборов данных. Его целью является выделение в тексте тональных компонентов. Единица, из которой выделяется одно мнение, называется уровнем анализа тональности. Существуют следующие уровни [1]: уровень документа, уровень предложения, уровень сущности и аспекта.

На первом уровне оценивается весь документ. Предполагается, что каждый документ должен выражать мнение об одном объекте, и, следовательно, уровень документа неприменим для тех документов, которые оценивают или сравнивают несколько объектов.

На уровне предложения задача состоит в классификации каждого предложения, определяется содержание в нем положительной, отрицательной или нейтральной оценки. Нейтральность обычно означает отсутствие мнения.

На уровне сущностей и аспектов определяются сущности и связанные с ними аспекты, классифицируются мнения, связанные с этими сущностями и аспектами. Примеры сущностей включают в себя продукты, услуги, темы, проблемы, людей, организации или события, которые обычно имеют несколько аспектов. Поскольку сущность представляет собой иерархию всех аспектов, она также является общим аспектом [2].

В рамках данной работы будет использоваться уровень документа, так как он является одним из наиболее подходящих для обработки небольших текстов.

1.2 Варианты классификации текста

Количество классов, на которые делится анализируемый текст, может варьироваться. Выделяют три варианта классификации анализируемого текста: бинарную, плоскую и иерархическую [3].

В первом случае текст делится на «положительный» и «отрицательный», данный способ классификации наиболее простой и распространенный. При этом точность отнесения объекта к тому или иному классу достигает высокой отметки, так как вариантов определения всего два.

Во втором случае к предыдущим двум классам добавляется «нейтральный» класс. С появлением трех и более классов сложность работы заметно возрастает.

В третьем случае система обучается на основе «регрессии»: с повышением показателя — положительнее тональность. Схема данного варианта классификации приведена на рисунке 1.



Рисунок 1 – Иерархическая классификация текста

Точность данной системы выше, чем у системы с плоской классификацией, но процессы обработки и обучения достаточно ресурсоёмкие.

1.3 Подходы к решению задач анализа тональности текста

На текущий момент существует множество различных методов для определения тональности текста: подход, основанный на правилах, подход, основанный на лексике (тональные словари), подход, основанный на шаблонах, частотные методы, подход, основанный на машинном обучении и др. Данные подходы можно объединить в три группы [4]:

- подход машинного обучения;
- лексический подход;
- гибридный подход.

обучении, Подход, основанный использует на машинном классификационный подход для классификации текста; он состоит из двух данных: обучающего И тестового набора. Первый используется для обучения классификатора, в то время как второй набор необходим для проверки качества работы классификатора. Особенности подхода, основанного на машинном обучении, к решению задач сентиментанализа, заключаются в следующем: использование частотного словаря, использование грамматической разметки частей речи для устранения неоднозначности смысла слов.

Подход, основанный на словаре, использует тональные словари и сопоставляет слова, имеющие метки, c данными ДЛЯ определения полярности. Существует три метода создания тонального построение словаря вручную, использование в качестве основы корпуса текстов, методы на основе словаря. Корпусные методы могут давать тональную оценку словам с высокой точностью. В методах на основе словаря сначала вручную определяется тональность небольшого набора слов, а затем этот набор расширяется путем поиска в словаре WordNet их синонимов и антонимов [5].

В гибридном подходе сочетание машинного обучения и подходов, основанных на лексике, может улучшить производительность в задачах классификации настроений.

Основным преимуществом подходов машинного обучения является возможность адаптации и создания обученной модели для конкретных целей. Недостатком является трудная интеграция общих знаний, которые нельзя получить из обучающего набора. Кроме того, изученные модели часто имеют плохую адаптивность между предметными областями или разными текстовыми жанрами, поскольку они часто полагаются на особенности предметной области из своих обучающих данных.

Подходы на основе лексики имеют более широкий охват терминов, однако у этих подходов есть два основных ограничения. Во-первых, количество слов в словаре ограничено, что может представлять проблему при извлечении тональности из динамичной среды. Во-вторых, тональные словари зачастую оценивают тональность слов, не учитывая контекста.

Основными преимуществами гибридных подходов являются сочетание лексики и обучения, обнаружение и измерение тональности на уровне понятий и устойчивость к изменению предметной области. Основным ограничением является то, что отзывам, тональность которых алгоритм не может определить, часто присваивается нейтральная оценка.

1.4 Оценка качества анализа тональности

Традиционной оценкой качества анализа тональности текста является сравнение результатов автоматической классификации с результатами, полученными путем ручной разметки эмоций исследуемого текста.

Существует три основных инструмента, доступных для проверки качества модели классификации [6]: матрицы ошибок (confusion matricies), графики роста (lift charts), и кривые ошибок (ROC, receiver operator characteristic).

Эффективность классификации лучше всего описывается матрицей ошибок. В классическом варианте она представляет собой двумерную матрицу, как показано в таблице 1.

Таблица 1 – Матрица ошибок

		Фактический класс		
		Положительный	Отрицательный	
Прогнозируемый	Положительный	TP	FP	
класс	Отрицательный	FN	TN	

В качестве данных таблицы выступают значения, отражающие количество верно и неверно классифицированных объектов:

ТР – истинно-положительные значения;

FP – ложно-положительные значения;

FN – ложно-отрицательные значения;

TN – истинно-отрицательные значения.

После составления матрицы ошибок рассчитываются четыре основных метрики: точность (P), полнота (R), F-мера (F1 Score) и аккуратность (Accuracy). На рисунке 2 представлены их формулы.

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$F1 \, Score = \frac{2 * P * R}{P + R}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Рисунок 2 – Метрики качества модели классификации

Точность (Р) показывает, сколько правильно предсказанных ответов модели на самом деле оказались положительными, что определяет её надежность. Полнота (R) демонстрирует, какое количество действительно положительных случаев правильно предсказано моделью. При высоких значениях данной метрики большинство положительных значений будут помечены как положительные, что приведет к снижению точности.

F-мера (F1 Score) — гармоническое среднее между точностью и полнотой, дает общее представление о данных показателях. Максимальна, когда точность равна полноте. Наконец, аккуратность (Accuracy) — показатель того, сколько правильных прогнозов сделала модель для всего тестового набора данных. Данная метрика является хорошим базовым показателем для измерения производительности модели [6].

1.5 Обзор инструментов для анализа тональности текста на естественном языке

На текущий момент наиболее известными инструментами языка Python для обработки естественного языка являются библиотеки: NLTK, TextBlob, CoreNLP, Gensim, spaCy, Polyglot, Scikit–learn, Pattern, Flair. Для анализа настроений наиболее популярными инструментами являются: Textblob, VADER, Dostoevsky.

В основе библиотеки TextBlob [8] лежит лексический подход. Оценка тональности для каждого текста представляет собой две характеристики: «полярность» и «субъективность». Первая выражается вещественным числом от -1 до 1, отрицательные значения присваиваются словам с негативной окраской, положительные — с позитивной, ноль означает нейтральную окраску. «Субъективность» означает наличие в тексте личных предположений человека, данная характеристика выражается вещественным числом от 0 до 1 (0 — очень объективно, а 1 — очень субъективно).

VADER [9] — инструмент анализа тональности, основанный на лексике и правилах, хорошо справляется с текстовыми данными из социальных сетей. В результате классификации библиотека дает следующие характеристики: compound, pos, neu, neg. Наиболее полезной метрикой является compound, так как она является единой одномерной мерой тональности для текста. Её пороговые значения:

- при позитивном настроении compound равна или превышает 0,05;
- при нейтральном настроении compound превышает -0,05 и меньше 0,05;
 - при негативном настроении compound меньше или равна -0,05.

Сумма оценок pos, neu, neg равна 1, они показывают долю текста, попадающего в категорию [9].

Библиотека Dostoevsky [10] обучена на наборе данных RuSentiment, которые представляют собой свыше 30 000 комментариев из «ВКонтакте». В результате классификации тексту присваивается категория:

- skip, если текст не удалось распознать;
- positive позитивная тональность;
- neutral нейтральная тональность;
- negative негативная тональность;
- speech речевой акт.

Таким образом, были рассмотрены наиболее популярные инструменты Выбрать языка Python ДЛЯ сентимент-анализа текста. библиотеку, применение которой наиболее даст точные оценки тональности русскоязычных отзывов и комментариев пользователей среди остальных, может помочь их экспериментальное сравнение.

1.6 Описание эксперимента

Для проверки качества работы анализаторов тональности библиотек Dostoevsky, VADER, TextBlob был проведен эксперимент. Можно выделить его основные этапы:

- 1. Сначала необходимо собрать опубликованные отзывы пользователей из Интернета.
- 2. Затем будет осуществляться разметка полярности собранных отзывов вручную. Негативным отзывам будет присваиваться значение -1, нейтральным значение 0, а положительным значение 1.
- 3. Далее с помощью библиотек Dostoevsky, VADER, TextBlob будет производиться анализ тональности текста.
- 4. Затем, полученные с использованием выбранных библиотек значения будут сравниваться с результатами, полученными вручную.
- 5. Далее будет производиться оценка точности полученных результатов: составление матриц ошибок, расчет метрик точности, полноты и F-меры.
- 6. Завершительным этапом будет выбор библиотеки, превосходящей по качеству классификации.

В рамках данного эксперимента была произведена выборка из 50 отзывов пользователей об университете ИТМО с сервиса Яндекс. Карты.

2 СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИНСТРУМЕНТОВ ДЛЯ АНАЛИЗА ТОНАЛЬНОСТИ ТЕКСТА НА ЕСТВЕСТВЕННОМ ЯЗЫКЕ

2.1 Экспериментальное сравнение библиотек Python для анализа тональности текста на русском языке

Для разработки была выбрана среда Google Colab, она позволяет писать и выполнять код Python в браузере, не требуя настройки.

Сбор данных с сервиса Яндекс. Карты производился с помощью библиотек Beautiful Soup 4 (bs 4) и requests, реализация данного этапа представлена на рисунке 3.

```
import requests
    from bs4 import BeautifulSoup
    import csv
   url = 'https://yandex.ru/maps/org/universitet_itmo/1051701714/reviews/'
   # открытие/создание файла и запись в него заголовков столбцов таблицы
   with open('ZheleznikovaPA_nlp.csv', mode='w') as nlp_file:
       nlp_writer = csv.writer(nlp_file, delimiter=',', quotechar='"', quoting=csv.QUOTE_MINIMAL)
       nlp_writer.writerow(['Номер', 'Отзыв'])
    # создание GET запроса
   response = requests.get(url)
             = BeautifulSoup(response.text, 'lxml')
   soup
   # поиск всех элементов с тегом span и названием класса business-review_view_body-text
   reviews = soup.find_all('span', class_='business-review-view_body-text')
   dstvsk = []
   blob = []
   for j in range(0, len(reviews)):
           # открытие/создание файла и добавление в него строк таблицы
           with open('ZheleznikovaPA_nlp.csv', mode='a') as nlp_file:
                                                      delimiter=',', quotechar='"', quoting=csv.QUOTE_MINIMAL)
              nlp_writer = csv.writer(nlp_file,
              # построчная запись элементов таблицы
              nlp_writer.writerow([j,reviews[j].text])
              dstvsk.append(reviews[j].text)
              blob.append(reviews[j].text)
   print("Файл записан.")
[→ Файл записан.
```

Рисунок 3 — Фрагмент кода, осуществляющего сбор данных с Яндекс. Карт

В результате работы данной программы был сформирован текстовый файл в формате csv, который содержит отзывы пользователей. Далее была произведена оценка полярности отзывов. Фрагмент полученных данных с

произведенной ручной разметкой тональности продемонстрирован на рисунке 4.

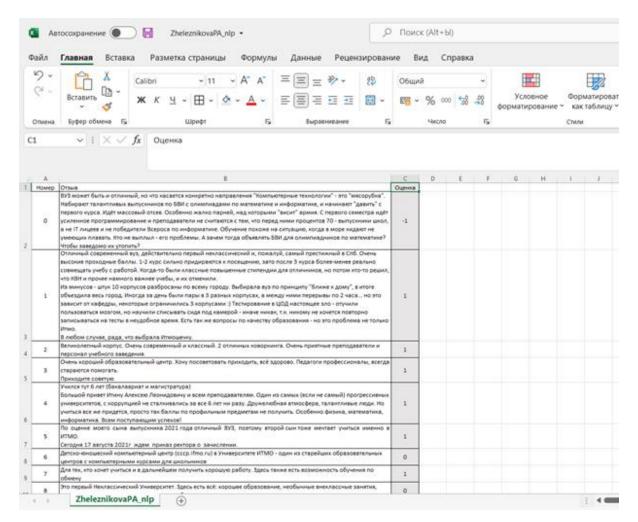


Рисунок 4 – Фрагмент данных с отзывами с размеченной полярностью

Негативным отзывам присвоено значение -1, нейтральным — значение 0, а положительным — значение 1.

Полученные данные были переданы в качестве входных параметров для классификаторов библиотек Dostoevsky, VADER, TextBlob с целью осуществления сентимент-анализа.

Фрагмент кода, отвечающий за классификацию тональности с помощью библиотеки Dostoevsky, приведен на рисунке 5.

```
from dostoevsky.tokenization import RegexTokenizer from dostoevsky.models import FastTextSocialNetworkModel

tokenizer = RegexTokenizer()
model = FastTextSocialNetworkModel(tokenizer=tokenizer)

results = model.predict(dstvsk, k = 2)
for dstvsk, sentiment in zip(dstvsk, results):
    with open('ZheleznikovaPA_dostoevsky.csv', mode='a') as nlp_file:
        nlp_writer = csv.writer(nlp_file, delimiter=',', quotechar='"', # построчная запись элементов таблицы nlp_writer.writerow([dstvsk,sentiment])
    print(dstvsk, '\n->', sentiment, '\n')
```

Рисунок 5 — Фрагмент кода, осуществляющего анализ тональности с использованием библиотеки Dostoevsky

Фрагмент текста, полученный в результате работы данной программы приведен на рисунке 6. Полный текст представлен в приложении 1.

```
Великолепный корпус. Очень современный и классный. 2 отличных коворкинга. Очень приятные преподаватели и персонал учебного заведения.

→ ('positive': 0.731686611450195, 'neutral': 0.07586817443370819)

Очень хороший образовательный центр. Хочу посоветовать приходить, всё здорово. Педагоги профессионалы, всегда старыктся помогать.
Приходите советую

→ ('neutral': 0.348655104637146, 'positive': 0.29422497749328613}

Учился тут 6 лет (бакалавриат и магистратура)

Большой привет Итину Алексею Леонидовичу и всем преподавателям. Один из самых (если не самый) прогрессивных университетов, с коррупцией не сталкивались за все 6 лет ни разу.

→ ('negative': 0.18714269995689392, 'neutral': 0.1732981950044632}

По оценке моего сына выпускника 2021 года отличный ВУЗ, поэтому второй сын тоже мечтает учиться именно в ИТМО.

Сегодня 17 августа 2021г ждем приказ ректора о зачислении.

→ ('neutral': 0.7879412174224854, 'positive': 0.41490885615348816)

Детско-вношеский компьютерный центр (ссср.ifmo.ru) в Университете ИТМО - один из старейших образовательных центров с компьютерными курсами для школьников

→ ('neutral': 0.8031837940216064, 'negative': 0.09889322447768}

Для тех, кто хочет учиться и в дальнейшем получить хорошую работу. Здесь также есть возможность обучения по обмену

→ ('neutral': 0.9099169969558716, 'positive': 0.061885979026556015)

Это первый Неклассический Университет. Здесь есть всё: хорошее образование, необычные внеклассные занятия, спорт, активная студенческая жизнь, невообразимые перспективы.

→ ('neutral': 0.6001983880996704, 'positive': 0.3276783227920532)

Хорошая программа, на моей кафедре заботятся о студентах и прислушиваются к мнению. Большинство предметов полезно для кругозора.

Начали давать возможность выбирать между предметами и это отличная тенденция.
```

Рисунок 6 – Результаты классификации отзывов, полученные с использованием библиотеки Dostoevsky

Результаты предсказаний были записаны в файл в формате csv, а также выведены в Google Colab.

Для использования классификаторов тональности библиотек VADER, TextBlob потребовалось перевести исходные данные на английский язык. Для была использована библиотека googletrans. Реализация перевода текста показана на рисунке 7.

```
from googletrans import Translator
 import googletrans
 translation = translator.translate(blob, src='ru', dest='en')
 for i in range(len(translation)):
    print(translation[i].text)
The university may be excellent, but with regard to the specific direction of "Computer Technology" - this is a "meat grinder". They recruit talented BVI graduate: An excellent modern university, indeed the first non-classical and, perhaps, the most prestigious in St. Petersburg. Very high passing scores. 1-2 courses are very of the minuses - about 10 buildings are scattered throughout the city. I chose the university according to the principle "closer to home", as a result I traveled in any case, I'm glad I chose Itmoshechka.
 Magnificent body. Very modern and classy. 2 great coworking spaces. Very nice teachers and staff.
Very good educational center. I want to advise you to come, everything is great. The teachers are professional and always willing to help.
 Come advise
Come advise
Studied here for 6 years (bachelor's and master's degree)
Big greetings to Aleksey Leonidovich Itin and all teachers. One of the most (if not the most) progressive universities, corruption has never been encountered in a According to my son, a graduate of 2021, an excellent university, so my second son also dreams of studying at ITMO.
Today, August 17, 2021, we are waiting for the order of the rector on enrollment.
Computer Center for Children and Youth (cccp.ifmo.ru) at ITMO University is one of the oldest educational centers with computer courses for schoolchildren For those who want to study and get a good job in the future. There is also an opportunity for exchange studies.
This is the first non-classical university. It has everything: a good education, unusual extra-curricular activities, sports, an active student life, unimaginable A good program, in my department they take care of students and listen to opinions. Most of the subjects are useful for the outlook.
 They began to give the opportunity to choose between subjects and this is a great trend.
The university is very bad, they don't teach anything, the canteen is expensive, the scholarship is small.
 Very cool university! The teachers clearly explain their subject.
 I do not regret that I transferred and finished it.
Everything is super!!!
Quality education + world-class events!!!!
 First in the country in ICT
 Cool youth atmosphere! The best programming club for children in St. Petersburg...
 My son attends preparatory courses at this university.
```

Рисунок 7 – Перевод текста с помощью библиотеки googletrans

Далее переведенный текст был передан на вход классификатору библиотеки TextBlob, с помощью которого был осуществлен сентиментанализ. Фрагмент кода, отвечающий за классификацию тональности с помощью библиотеки TextBlob, приведен на рисунке 8.

```
from textblob import TextBlob
from textblob.sentiments import NaiveBayesAnalyzer

for j in range(0, len(reviews)):
    # открытие/создание файла и добавление в него строк таблицы
    with open('ZheleznikovaPA_textblob.csv', mode='a') as nlp_file:
        text_one = TextBlob(translation[j].text).sentiment.subjectivity
        text_ = TextBlob(translation[j].text).sentiment.polarity
        nlp_writer = csv.writer(nlp_file, delimiter=',', quotechar='"',
        # построчная запись элементов таблицы
        nlp_writer.writerow([reviews[j].text, text_])
        print(reviews[j].text, "\nsubjectivity:", text_one,", ","polarity:",

print("Файл записан.")
```

Рисунок 8 — Фрагмент кода, осуществляющего анализ тональности с использованием библиотеки TextBlob

На рисунке 8 приведен фрагмент полученных данных в результате анализа тональности отзывов с помощью библиотеки TextBlob. Полный текст представлен в приложении 1.

```
subjectivity: 0.7114583333333333 , polarity: 0.2125
Отличный современный руз, действительно первый меклассический и, пожалуй, самый престижный в Спб. Очень высокие проходные баллы. 1-2 курс сильно придираются к посещению, из минусов — члух 10 колугор врабросане по всему городу. Выбирала вуз по принципу "ближе к дому", в итоге объездила весь город. Иногда за день были пары в 3 разных корпу В любом случае, рада, что выбрала Итмошечку.

зыбуестіvity: 0.8657906817406811, ројаліту: 0.20256547619047627
Великоленый корпус. Очень современный и классный. 2 отличных коворкинга. Очень приятные преподаватели и персонал учебного заведения. 
subjectivity: 0.885 ројасту 0.000 в 1.000 в 1.000
```

Рисунок 8 – Результаты классификации отзывов, полученные с использованием библиотеки TextBlob

Переведенный текст был передан также на вход классификатору библиотеки VADER, с помощью которого был осуществлен сентиментанализ. Фрагмент кода, отвечающий за классификацию тональности с помощью библиотеки VADER, приведен на рисунке 9.

```
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

analyzer = SentimentIntensityAnalyzer()

for j in range(0, len(reviews)):
    # открытие/создание файла и добавление в него строк таблицы
    with open('ZheleznikovaPA_vader.csv', mode='a') as nlp_file:
        nlp_writer = csv.writer(nlp_file, delimiter=',', quotechar='"',
        vs = analyzer.polarity_scores(translation[j].text)
        nlp_writer.writerow([reviews[j].text, vs])
        print(reviews[j].text,"->\n", vs)

print("Файл записан.")
```

Рисунок 9 – Результаты классификации отзывов, полученные с использованием библиотеки VADER

На рисунке 10 приведен фрагмент полученных данных в результате анализа тональности отзывов с помощью библиотеки VADER. Полный текст представлен в приложении 1.

```
    № 12 может бать и отличный, но что касемста конкратно направления "Компьютерные технологии" - это "масорубка", Набирают талантливых выпускников по 58И с олимпиадами по математ УМОВС » ("пед": 0.8.1"), торо: 10,968, "совроине" - 9,7755)
    Отличный совроменный вуз, действительно первый меклассический и, пожалуй, самый престижный в Слб. Очень высокие проходные балом. 1-2 курс сильно придираются к поседению, зато из менусов с штух 18 сморгов разбросьам по всему городу. Выбирала в уз по принципу "ближе к дому", в итоге объездила весь город. Иногда за день били пары в 3 разных корпусах, 8 лябом случае, рада, что выбразы Гитоцечку.
    УМОВС » ("пед": 0.8.9", пец": 0.3.8"), пурст: 0.119, "совроили": 0.3834)
    Великоленный корпус. Очень совроменный и классный. 2 отличных коворкинга. Очень приятные преподавателы и персонал учебного заведения.
    УМОВС » ("пед": 0.8.") пец": 0.3.8", рост: 0.519, "совуствовный корпус. Очень короший образовательный центр. Хочу посоветовать приходить, всё здорово. Педагоги профессионалы, всегда стараются помогать. Приходить советую
    УМОВС » ("пед": 0.8, "пец": 0.8, "рост: 0.155, "совуствовный в стара с педагоги профессионалы, всегда стараются помогать.
    Уморст » ("пед": 0.8, "пец": 0.8, "корт: 0.155, "совуствовный в уз по принципую профессионалы, всегда стараются помогать.
    Уморст » ("пед": 0.8, "пец": 0.8, "корт: 0.155, "совуствовный в уз по принципую профессиональ, всегда стараются помогать.
    Уморст » ("пед": 0.8, "пец": 0.8, "корт: 0.155, "совуствовный в уз по принципую профессиональ, всегда стараются помогать профессиональный с пр
```

Рисунок 9 – Результаты классификации отзывов, полученные с использованием библиотеки VADER

Таким образом, с помощью библиотек Dostoevsky, VADER, TextBlob был произведен анализ тональности отзывов об университете ИТМО с сервиса Яндекс.Карты. Работа с данными библиотеками Python в среде Goolge Colab не представляет сложностей, а код занимает небольшое количество строк.

2.2 Оценка полученных результатов эксперимента

Оценка точности классификации производилась с использованием библиотеки Python scikit-learn. Для построения матрицы ошибок необходимо вызвать функцию confusion_matrix, в качестве аргументов достаточно ввести правильные значения, которые получены вручную, а также показатели, возвращаемые классификатором. Функция classification_report рассчитывает все необходимы метрики для каждого класса, в качестве входных параметров принимает такие же аргументы, как и confusion_matrix.

Пример вызова функций и результаты их работы приведены на рисунке 10.

Рисунок 10 — Расчет точности результатов классификации, полученных с использованием библиотеки TextBlob

Оценкам тональности, полученным в результате работы классификаторов, в зависимости от показателей, были присвоены значения целые значения от -1 до 1, для возможности оценки точности результатов классификации.

Результаты расчетов точности классификаторов библиотек TextBlob, Dostoevsky, VADER, были перенесены в таблицы. В таблице 2 приведена матрица ошибок для библиотеки Dostoevsky.

Таблица 2 – Матрица ошибок для библиотеки Dostoevsky

		Фактический класс		
		-1	0	1
	-1	2	1	2
Прогнозируемый класс	0	1	1	10
	1	0	1	32

В таблице 3 приведена матрица ошибок для библиотеки TextBlob.

Таблица 3 – Матрица ошибок для библиотеки TextBlob

		Фактический класс		
		-1	0	1
	-1	5	0	0
Прогнозируемый класс	0	0	9	3
	1	0	6	27

В таблице 4 приведена матрица ошибок для библиотеки VADER.

Таблица 4 – Матрица ошибок для библиотеки VADER

		Фактический класс		
		-1	0	1
Прогнозируемый класс	-1	3	1	1
	0	1	3	8
	1	0	4	29

Метрики точности (P), полноты (R), F-меры (F1 Score) и аккуратности (Accuracy) для каждой из библиотек были перенесены в таблицу 5 для сравнения результатов между собой.

Таблица 5 — Сравнительная характеристика работы библиотек Dostoevsky, TextBlob, VADER

Dostoevsky				
Метрика	-1	0	1	
P	0,67	0,33	0,73	
R	0,40	0,08	0,97	
F1 Score	0,5	0,13	0,83	
Accuracy	0,70			

Продолжение таблицы 5

TextBlob	TextBlob					
Метрика	-1	0 1				
P	1,00	0,60	0,90			
R	1,00	0,75	0,82			
F1 Score	1,00	0,67	0,86			
Accuracy	0,82					
VADER	VADER					
Метрика	-1	0	1			
P	0,75	0,38	0,76			
R	0,60	0,25	0,88			
F1 Score	0,67	0,30	0,82			
Accuracy	0,70					

Результаты, приведенные в таблице 5, показывают, что для негативных отзывов библиотека TextBlob дает более точные результаты. Точность, Полнота и F-мера равны единице, что говорит о том, что все негативные высказывания были определены верно.

Для нейтральных отзывов также наилучшие результаты дает TextBlob ($P=0,60,\ R=0,75,\ F1\ Score=0,67$). На втором месте стоит библиотека VADER, имея также достаточно высокие результаты ($P=0,38,\ R=0,25,\ F1\ Score=0,30$).

Положительные отзывы определили с высокой точностью все три классификатора. Самую высокую точность имеет библиотека TextBlob, у которой P=0.90, показатели точности библиотек VADER и Dostoevsky незначительно расходятся. Классификатор библиотеки Dostoevsky имеет полноту R=0.97, что выше, чем у двух других, что говорит о захвате лишних высказываний.

Показатели Accuracy у библиотек TextBlob, VADER и Dostoevsky составляют 0,82, 0,70 и 0,70 соответственно. Данная метрика показывает, сколько правильных прогнозов сделала модель для всего набора данных, представляющих собой выборку из 50 отзывов с сервиса Яндекс.Карты. Наибольшее значение имеет библиотека TextBlob.

Таким образом, качество классификации отзывов на русском языке преобладает у библиотеки TextBlob.

ЗАКЛЮЧЕНИЕ

В ходе выполнения курсовой работы был проведен сравнительный анализ библиотек Python для анализа тональности текста на русском языке. Рассмотрены основные уровни сентимент-анализа, а именно: уровень документа, уровень предложения, уровень сущности и аспекта. В данной работе использовался уровень документа. Исследованы варианты классификации текста: бинарная, плоская и иерархическая. Выявлено, что наиболее простой и распространенной является бинарная классификация, и при этом она достигает высокой отметки точности.

Проанализированы существующие подходы к решению задач анализа тональности текста на естественном языке: подход, основанный на машинном обучении, лексический подход и гибридный подход. В рамках выполнения данной работы был использован лексический подход. Также была рассмотрена оценка качества классификации, выделены четыре основных метрики: точность (Р), полнота (R), F-мера (F1 Score) и аккуратность (Ассигасу).

Рассмотрены существующие инструменты языка Python для анализа тональности текста, библиотеки: Textblob, VADER, Dostoevsky. Составлен план по их экспериментальному сравнению. С использованием каждой из вышеперечисленных библиотек был произведен анализ тональности 50 отзывов пользователей об университете ИТМО с сервиса Яндекс.Карты. Для оценки точности работы классификаторов была составлена матрица ошибок и измерены основные метрики.

Результаты исследования показали, качество классификации отзывов на русском языке преобладает у библиотеки TextBlob. Для положительных отзывов получены оценки: $P=0.90,\ R=0.82,\ F1$ Score $=0.86,\ для$ нейтральных: $P=0.60,\ R=0.75,\ F1$ Score $=0.67,\ и$ для негативных $P=1,\ 0,\ R=1.0,\ F1$ Score =1.0.

Проведенный сравнительный анализ инструментов для сентиментанализа текста на русском языке, полученный опыт знакомства с методами анализа тональности текста позволит применить знания и навыки в разработке программного продукта, который минимизирует появление проблем при классификации текстовых данных.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1. Bhatt R., Gupta P. Sentiment Analysis // Indian Journal of Science and Technology. Vol. 12. 2019. P. 1-6.
- 2. Dohaiha H., Prasad P., Maag A., Alsadoon A. Deep learning for aspect-based sentiment analysis: a comparative review // Expert Systems with Applications. 2018. P. 1-44.
- 3. Гринин И.Л. Разработка, тестирование и сравнение моделей сентиментального анализа коротких текстов // Инновации и инвестиции. 2020. №6. URL: https://cyberleninka.ru/article/n/razrabotka-testirovanie-i-sravnenie-modeley-sentimentalnogo-analiza-korotkih-tekstov (дата обращения: 06.06.2022).
- 4. D'Andrea A., Fernando F., Patriza G., Tiziana G., Approaches, Tools and Applications for Sentiment Analysis Implementation // International Journal of Computer Applications. Vol. 125. 2015. N 3. P. 26-33.
- 5. Сметанина О.Н., Сазонова Е.Ю., Сулейманов А.К., Селиванов С.Г., Андрушко Д.Ю., Габдиев Ф.Ф. Программное решение по построению классификаторов для анализа текстов на естественном языке // Современные наукоёмкие технологии. 2020. –N 12-1. С. 119-127.
- 6. Kotu V., Deshapande B. Model Evaluation // Predictive Analytics and Data Mining. 2015. P. 257-274.
- 7. Kenteris M., Gavalas D., Mpitziopoulos A. A mobile tourism recommender system. In Computers and Communications // Proceedings of the 15th IEEE Symposium on Computers and Communications. 2010. P. 840-845.
- 8. TextBlob: Simplified Text Processing. URL: https://textblob.readthedocs.io/ (дата обращения: 06.06.2022).
- 9. VADER. URL: https://pypi.org/project/vaderSentiment/ (дата обращения: 06.06.2022).
- 10. Dostoevsky. URL: https://pypi.org/project/dostoevsky/ (дата обращения: 06.06.2022).

ПРИЛОЖЕНИЕ 1

Результаты оценки тональности отзывов библиотеками Dostoevsky,

TextBlob, VADER

ВУЗ может быть и отличный, но что касается конкретно направления "Компьютерные технологии" - это "мясорубка". Набирают талантливых выпускников по БВИ с олимпиадами по математике и информатике, и начинают "давить" с первого курса. Идёт массовый отсев. Особенно жалко парней, над которыми "висит" армия. С первого семестра идёт усиленное программирование и преподаватели не считаются с тем, что перед ними процентов 70 - выпускники школ, а не ІТ лицеев и не победители Всероса по информатике. Обучение похоже на ситуацию, когда в море кидают не умеющих плавать. Кто не выплыл - его проблемы. А зачем тогда объявлять БВИ для олимпиадников по математике? Чтобы заведомо их утопить?

```
Dostoevsky -> {'neutral': 0.44553956389427185, 'negative': 0.23371636867523193}

TextBlob -> subjectivity: 0.711458333333333, polarity: 0.2125

VADER -> {'neg': 0.114, 'neu': 0.817, 'pos': 0.068, 'compound': -0.7756}
```

Отличный современный вуз, действительно первый неклассический и, пожалуй, самый престижный в Спб. Очень высокие проходные баллы. 1-2 курс сильно придираются к посещению, зато после 3 курса более-менее реально совмещать учебу с работой. Когда-то были классные повышенные стипендии для отличников, но потом кто-то решил, что КВН и прочее намного важнее учебы, и их отменили.

Из минусов — штук 10 корпусов разбросаны по всему городу. Выбирала вуз по принципу "ближе к дому", в итоге объездила весь город. Иногда за день были пары в 3 разных корпусах, а между ними перерывы по 2 часа... но это зависит от кафедры, некоторые ограничились 3 корпусами :) Тестирование в ЦОД настоящее зло — отучили пользоваться мозгом, но научили списывать сидя под камерой — иначе никак, т.к. никому не хочется повторно записываться на тесты в неудобное время. Есть так же вопросы по качеству образования — но это проблема не только Итмо.

В любом случае, рада, что выбрала Итмошечку.

```
Dostoevsky -> {'neutral': 0.4301573634147644, 'negative': 0.24509501457214355}

TextBlob -> subjectivity: 0.578399585921325, polarity: 0.20267701863354046

VADER -> {'neg': 0.093, 'neu': 0.788, 'pos': 0.119, 'compound': 0.8034}
```

Великолепный корпус. Очень современный и классный. 2 отличных коворкинга. Очень приятные преподаватели и персонал учебного заведения.

```
Dostoevsky -> {'positive': 0.7310686111450195, 'neutral': 0.07586817443370819}

TextBlob -> subjectivity: 0.808. polarity: 0.588000000000001

VADER -> {'neg': 0.0, 'neu': 0.436, 'pos': 0.564, 'compound': 0.9355}
```

Очень хороший образовательный центр. Хочу посоветовать приходить, всё здорово. Педагоги профессионалы, всегда стараются помогать. Приходите советую

```
Dostoevsky -> {'neutral': 0.348655104637146, 'positive': 0.29422497749328613}

TextBlob -> subjectivity: 0.4550000000000007, polarity: 0.3683333333333335

VADER -> {'neg': 0.0, 'neu': 0.65, 'pos': 0.35, 'compound': 0.8832}

Учился тут 6 лет (бакалавриат и магистратура)

Большой привет Итину Алексею Леонидовичу и всем преподавателям. Один из самых (если не самый) прогрессивных университетов, с коррупцией не сталкивались за все 6 лет ни разу. Дружелюбная атмосфера, талантливые люди. Но учиться все же придется, просто так баллы по профильным предметам не получить. Особенно физика, математика, информатика. Всем поступающим успехов!
```

По оценке моего сына выпускника 2021 года отличный ВУЗ, поэтому второй сын тоже мечтает учиться именно в ИТМО.

Сегодня 17 августа 2021г ждем приказ ректора о зачислении.

```
Dostoevsky -> {'neutral': 0.7879412174224854, 'positive': 0.41490885615348816}

TextBlob -> subjectivity: 0.5. polarity: 0.5

VADER -> {'neg': 0.0, 'neu': 0.842, 'pos': 0.158, 'compound': 0.7506}
```

Детско-юношеский компьютерный центр (cccp.ifmo.ru) в Университете ИТМО - один из старейших образовательных центров с компьютерными курсами для школьников

```
Dostoevsky -> {'neutral': 0.8031837940216064, 'negative': 0.0980893224477768}

TextBlob -> subjectivity: 0.175. polarity: 0.075

VADER -> {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

Для тех, кто хочет учиться и в дальнейшем получить хорошую работу. Здесь также есть возможность обучения по обмену

```
Dostoevsky -> {'neutral': 0.9099169969558716, 'positive': 0.061885979026556015}

TextBlob -> subjectivity: 0.3625000000000004. polarity: 0.35

VADER -> {'neg': 0.0, 'neu': 0.731, 'pos': 0.269, 'compound': 0.7184}
```

Это первый Неклассический Университет. Здесь есть всё: хорошее образование, необычные внеклассные занятия, спорт, активная студенческая жизнь, невообразимые перспективы.

```
Dostoevsky -> {'neutral': 0.6001983880996704, 'positive':
0.3276783227920532}
```

```
TextBlob -> subjectivity: 0.633333333333333. polarity:
0.25416666666666665
        -> {'neg': 0.0, 'neu': 0.709, 'pos': 0.291, 'compound':
VADER
0.7783}
Хорошая программа, на моей кафедре заботятся о студентах и прислушиваются
к мнению. Большинство предметов полезно для кругозора.
Начали давать возможность выбирать между предметами и это отличная
тенденция.
Dostoevsky -> {'neutral': 0.8354935646057129, 'negative':
0.13661839067935944}
TextBlob -> subjectivity: 0.4625. polarity: 0.575
VADER
           -> {'neq': 0.0, 'neu': 0.688, 'pos': 0.312, 'compound':
0.9423}
ВУЗ очень плохой, ничему не учат, столовая дорогая, стипендия маленькая.
Dostoevsky -> {'negative': 0.585111141204834, 'positive':
0.3415925204753876}
          -> subjectivity: 0.6555555555556. polarity: -
TextBlob
0.55333333333333333
          -> {'neg': 0.192, 'neu': 0.808, 'pos': 0.0, 'compound': -
VADER
0.5849}
Очень классный ВУЗ! Преподаватели доходчиво объясняют свой предмет.
Не жалею, что перевёлся и закончил его.
Dostoevsky -> {'negative': 0.3140605390071869, 'positive':
0.23371636867523193}
TextBlob -> subjectivity: 0.52055555555555, polarity:
0.1673611111111111
         -> {'neg': 0.0, 'neu': 0.669, 'pos': 0.331, 'compound':
VADER
0.7856}
Всё - супер!!!
Качественное образование + мероприятия мирового уровня!!!!
Dostoevsky -> {'positive': 0.8962613940238953, 'neutral':
0.061885979026556015}
VADER
          -> {'neg': 0.0, 'neu': 0.669, 'pos': 0.331, 'compound':
0.7856}
Первый в стране по ИКТ
Dostoevsky -> {'neutral': 0.9905974268913269, 'skip':
0.014967083930969238}
TextBlob -> subjectivity: 0.33333333333333. polarity: 0.25
           -> {'neg': 0.0, 'neu': 0.58, 'pos': 0.42, 'compound': 0.7243}
VADER
Классная молодежная атмосфера! Лучший в Питере кружок программирования для
детей...
Dostoevsky -> {'positive': 0.8774868249893188, 'neutral':
0.1824355274438858}
TextBlob -> subjectivity: 0.475. polarity: 0.71875
          -> {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
VADER
```

```
В этот университете, на подготовительные курсы, ходит мой сын. Его всегда тянет туда, потому что ему интересен предмети, очевидно, и то, как его преподают.
```

Ходит он уже 3й год и пока все хорошо, ему нравиться, и он намерен продолжать.

Я счетаю, что это хороший университет.

```
Dostoevsky -> {'negative': 0.26285186409950256, 'positive': 0.1824355274438858}

TextBlob -> subjectivity: 0.49047619047619045. polarity: 0.18571428571428572

VADER -> {'neg': 0.0, 'neu': 0.596, 'pos': 0.404, 'compound': 0.7777}
```

Тут был лучший горячий бутерброд в России. Он согревал, поднимал и направлял на путь истинный немало заблудших душ.

```
Dostoevsky -> {'negative': 0.2814156115055084, 'positive': 0.1824355274438858}

TextBlob -> subjectivity: 0.57500000000001. polarity: 0.525

VADER -> {'neg': 0.0, 'neu': 0.828, 'pos': 0.172, 'compound': 0.8481}
```

Самый лучший вуз!!! Бесконечно люблю и горжусь♥

```
Dostoevsky -> {'positive': 1.0000100135803223, 'skip': 0.024433093145489693}

TextBlob -> subjectivity: 0.36666666666664. polarity: 0.566666666666667

VADER -> {'neg': 0.084, 'neu': 0.584, 'pos': 0.332, 'compound': 0.7783}
```

Развитая внеучебная деятельность, стильные интерьеры

```
Dostoevsky -> {'negative': 0.5775054097175598, 'neutral': 0.4765896201133728}

TextBlob -> subjectivity: 0.65. polarity: 0.3

VADER -> {'neg': 0.0, 'neu': 0.352, 'pos': 0.648, 'compound': 0.9557}
```

Второй год посещаем занятия для младших школьников по основам компьютерной грамотности, изучают работу приложений. Сыну очень нравится.

```
Dostoevsky -> {'neutral': 0.7826725244522095, 'positive': 0.1732981950044632}

TextBlob -> subjectivity: 0.0866666666667. polarity: 0.086666666666666666667

VADER -> {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

Отличный вуз. Это административный корпус. Тут ректор – друг всего сообщества вуза, и следовательно царит позитивная атмосфера. Не спроста одна из ценностей вуза – Любовь. Тут действительно Любовь друг к другу, Любовь к вузу и к своему делу. Другое название вуза – итмо.фэмили. И это понятно почему!

```
Dostoevsky -> {'positive': 0.546748161315918, 'neutral':
0.19194278120994568}
```

```
TextBlob -> subjectivity: 0.6207792207792208. polarity:
0.38961038961038963
        -> {'neg': 0.0, 'neu': 0.896, 'pos': 0.104, 'compound':
VADER
0.4215}
Образование на уровне, внеучебная деятельность очень хорошо развита
Dostoevsky -> {'positive': 0.46102678775787354, 'negative':
0.3702353835105896}
-> {'neg': 0.0, 'neu': 0.695, 'pos': 0.305, 'compound':
VADER
0.9806}
Пишу отзыв только с точки зрения моего факультета. Вуз не плохой, но во
всех есть большие НО. Хорошо распиарен, нНО по факту простой неплохой
технический вуз. Неплохая программа, НО велась достаточно паршиво.
Теоретически много практики, НО ничем не отличалась от других пар с не
очень полезными лекциями. Тяжело совмещать с работой, если действительно
хочется учиться из-за расписания. В общем я получил здесь достаточно
неплохое самообразование.
Dostoevsky -> {'neutral': 0.4765896201133728, 'negative':
0.3702353835105896}
TextBlob -> subjectivity: 0.43169642857142854. polarity:
0.08237179487179486
          -> {'neg': 0.0, 'neu': 0.79, 'pos': 0.21, 'compound': 0.3384}
Один из лучших универов России (только фотоника чего стоит). Хорошая
материальная база, отличные стипендии сильным студентам
Dostoevsky -> {'positive': 0.3923468291759491, 'neutral':
0.11597072333097458}
TextBlob -> subjectivity: 0.6761904761904762. polarity:
0.3761904761904762
        -> {'neg': 0.12, 'neu': 0.681, 'pos': 0.199, 'compound':
VADER
0.8232}
Последний раз был внутри 16 лет назад. :))) Внешне преобразился.
Dostoevsky -> {'neutral': 0.7773098945617676, 'skip': 0.13297423720359802}
TextBlob -> subjectivity: 0.533333333333333. polarity: 0.25
VADER
          -> {'neg': 0.0, 'neu': 0.484, 'pos': 0.516, 'compound':
0.9432}
учился там в 2005 году. институт регулярно выигрывает олимпиады по
программированию, имеет высокие рейтинги среди других институтов мира.
финансируется из разных источников.
Dostoevsky -> {'neutral': 0.9706977605819702, 'negative':
0.07370626926422119}
TextBlob
          -> subjectivity: 0.40375. polarity: 0.0837499999999999
           -> {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

```
Не знаю как сейчас, а во времена моего студенчества Литмище было самым демократичным институтом. В буфете пиво продавали не только преподавателям, но и студентам ))

Dostoevsky -> {'neutral': 0.585111141204834, 'negative': 0.16452647745609283}

TextBlob -> subjectivity: 0.75. polarity: 0.25

VADER -> {'neg': 0.0, 'neu': 0.856, 'pos': 0.144, 'compound': 0.5719}
```

Развитая учебная и внеучебная деятельности. Качественны ое образование. Вежливое отношение.

```
Dostoevsky -> {'neutral': 0.3775506913661957, 'negative':
0.320831298828125}
TextBlob -> subjectivity: 0.275. polarity: 0.175
VADER -> {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

Много возможностей для самореализации и занятия тем чем хочешь заниматься на самом деле помимо учебного процесса.

Если не забивать на учебу и заниматься в кружках, на выпуске станешь успешным и реально востребованным.

```
Dostoevsky -> {'neutral': 0.7826725244522095, 'negative': 0.07808816432952881}

TextBlob -> subjectivity: 0.420000000000004. polarity: 0.38

VADER -> {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

Кто тот странный человек, что будет выбирать вуз по отзывам в Яндекс? Нормальный корпус. Столовая есть. Сделали кпп новое, теперь без пропуска фиг попадаешь. Охранники замечательные, кстати. Приятно пообщаться с ними, пока ждёшь разрешения пройти. В здании 6 этажей, но только избранные находят 5. Шестой — вообще тайная комната. Во внутреннем дворе можно покурить и ещё раз пообщаться с охранником. В столовой закажите котлету по-киевски с пюре. Вообще огонь.

```
Dostoevsky -> {'neutral': 0.4765896201133728, 'negative': 0.14805719256401062}

TextBlob -> subjectivity: 0.64045454545454. polarity: 0.028636363636366

VADER -> {'neg': 0.038, 'neu': 0.791, 'pos': 0.171, 'compound': 0.7363}
```

Относительно высокое качество образования, есть внеучебная деятельность, возможность поехать учиться по обмену.

```
Dostoevsky -> {'neutral': 0.8080772161483765, 'negative': 0.1778208613395691}

TextBlob -> subjectivity: 0.53999999999999 polarity: 0.16

VADER -> {'neg': 0.043, 'neu': 0.915, 'pos': 0.041, 'compound': -0.0516}
```

Классная детско-юношеская компьютерная школа.

```
отличный коворкинг, и даже женские туалеты в наличие!!
Dostoevsky -> {'positive': 0.6723417043685913, 'neutral':
0.22271016240119934}
           -> subjectivity: 0.575. polarity: 0.7125
TextBlob
           -> {'neg': 0.0, 'neu': 0.685, 'pos': 0.315, 'compound':
VADER
0.3182}
Приехали подорвать документы, причём планировали на платную основу
сразу.... Но знали что на80% попадаем на бюджет. На месте не было того
человека который принимает документы и нас попросили подождать где-то
около 30 минут,.... После пройденного времени нас ещё несколько раз
попросили подождать...!!!???((( но мы решили и примерно подумали, какая
тут дисциплинарность и уехали не подошва документов!!!!!!!! ((( А ДЛЯ ТЕХ
КТО ПЛАНИРУЕТ ПОСТУПАТЬ В ЭТОТ УЗ, ОЧЕНЬ СИЛЬНО И ПОДРОБНО ИЗУЧИТЬ ИХ
УСТА, И ОТНОШЕНИЯ К ПАСТУПАЮШИМ СТУДЕНТАМ (А ТАК ЖЕ ВНИМАТЕЛЬНО ПОЧИТАТЬ
ГДЕ-ТО ОТЗЫВЫ).!!!!!! НО Я БЫ ВАМ ЕГО НЕ РЕКОМЕНДОВАЛ!!!!!!!!!!!
Dostoevsky -> {'neutral': 0.538993239402771, 'negative':
0.3923468291759491}
           -> subjectivity: 0.5265608465608466. polarity:
0.061005291005291
           -> {'neg': 0.038, 'neu': 0.865, 'pos': 0.096, 'compound':
VADER
0.8663}
Сосредоточие мысли, позитивной энергии и огня в глазах студентов. Планы
руководства и их устремленность внушает уверенность в успехе.
Dostoevsky -> {'neutral': 0.585111141204834, 'negative':
0.11280541121959686}
TextBlob -> subjectivity: 0.27272727272727. polarity:
0.2636363636363636
          -> {'neg': 0.065, 'neu': 0.489, 'pos': 0.446, 'compound':
VADER
0.9325}
Ужасный вуз!!!
Преподаватели либо не ходють на лпкции, либо пьют, либо кусают студентов.
Меня Кузнецов укусил на впд. Это ужас!!!!
Уж лучше гумфр
Dostoevsky -> {'negative': 0.3007555902004242, 'neutral':
0.22816647589206696}
VADER
           -> {'neg': 0.238, 'neu': 0.673, 'pos': 0.089, 'compound': -
0.7067}
Один из лучших вузов страны.
Хоть и (неоправданно) сильно распиаренный, но хорошие направления есть
Dostoevsky -> {'positive': 0.4225146472454071, 'neutral':
0.0980893224477768}
TextBlob
          -> subjectivity: 0.4666666666666673. polarity: 0.5
           -> {'neg': 0.0, 'neu': 0.568, 'pos': 0.432, 'compound': 0.872}
VADER
Вуз оправдал все ожидания. Очень рад, что попал сюда
Dostoevsky -> {'negative': 0.4532718360424042, 'neutral':
0.23935934901237488}
```

TextBlob -> subjectivity: 1.0. polarity: 0.65

```
-> {'neg': 0.0, 'neu': 0.752, 'pos': 0.248, 'compound':
VADER
0.5095}
Я там учился было весело и интересно. Но это было давно и неправда.
Dostoevsky -> {'positive': 0.8670457601547241, 'neutral':
0.14805719256401062}
TextBlob -> subjectivity: 0.4375, polarity: 0.14375
subjectivity: 0.7114285714285715. polarity: 0.562142857142857
          -> {'neg': 0.13, 'neu': 0.696, 'pos': 0.174, 'compound':
VADER
0.0005}
Очень хорошее образование студентов. Вуз не изгажен реформами. Умное
руководство.
Dostoevsky -> {'neutral': 0.3415925204753876, 'positive':
0.22816647589206696}
TextBlob
         -> subjectivity: 0.7114285714285715. polarity:
0.562142857142857
VADER
          -> {'neg': 0.0, 'neu': 0.584, 'pos': 0.416, 'compound':
0.8504}
Вроде чему-то даже учат, но я даже не знаю, ведь на пары ходить не
обязательно, а я пользуюсь этой возможностью на максимум
Dostoevsky -> {'positive': 0.275139719247818, 'neutral':
0.275139719247818}
TextBlob -> subjectivity: 1.0. polarity: 0.0
           -> {'neq': 0.0, 'neu': 0.856, 'pos': 0.144, 'compound': 0.657}
Хороший университет, много продвинутой молодежи, будущие ученые страны .
Dostoevsky -> {'neutral': 0.8354935646057129, 'negative':
0.0695517510175705}
          -> subjectivity: 0.4416666666666667. polarity:
TextBlob
0.3666666666666667
VADER
         -> {'neq': 0.0, 'neu': 0.692, 'pos': 0.308, 'compound':
0.5994}
Как и во всех других вузах качество обучения разнится от факультета к
факультету, но в среднем очень даже ничего.
Dostoevsky -> {'neutral': 0.2568419873714447, 'negative':
0.24509501457214355}
         -> subjectivity: 0.5183333333333334. polarity:
TextBlob
0.21166666666666664
           -> {'neg': 0.0, 'neu': 0.816, 'pos': 0.184, 'compound':
VADER
0.6474}
Если вы поступили в магистратуру с другого университета. об вас вытрут
ноги. потому что действует фраза « вы не НАШИ бакалавриаты»
Dostoevsky -> {'neutral': 0.6584275364875793, 'skip': 0.287777841091156}
TextBlob -> subjectivity: 0.0. polarity: 0.0
VADER -> {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

Хороший вуз, одно из немногих мест в России, где бюджет не распиливают, а видно что и куда тратится.

```
Dostoevsky -> {'neutral': 0.5078218579292297, 'negative':
0.287777841091156}
           -> subjectivity: 0.266666666666666. polarity:
0.13333333333333333
           -> {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
VADER
Ни про один ВУЗ, в котором училась, не могу сказать ничего плохого.
Dostoevsky -> {'neutral': 0.585111141204834, 'negative':
0.4225146472454071}
TextBlob -> subjectivity: 0.66666666666666. polarity: -
0.699999999999998
VADER
          -> {'neq': 0.0, 'neu': 0.877, 'pos': 0.123, 'compound':
0.3307}
Лучшее место во всем Питере 💜 💜
Dostoevsky -> {'positive': 0.7663036584854126, 'neutral':
0.0191340371966362}
TextBlob -> subjectivity: 0.3. polarity: 1.0
VADER
           -> {'neg': 0.0, 'neu': 0.778, 'pos': 0.222, 'compound': 0.431}
Один из лучших университетов страны где готовят ИТ специалистов, прямой
конкурент Политеха!!!
Dostoevsky -> {'neutral': 0.6370407938957214, 'negative':
0.1259327530860901}
TextBlob -> subjectivity: 0.35. polarity: 0.59765625
VADER
           -> {'neg': 0.0, 'neu': 0.789, 'pos': 0.211, 'compound':
0.7249}
Очень открытый и небюрократизированный университета фоне остальных в
городе
Dostoevsky -> {'neutral': 0.5000100135803223, 'negative':
0.33459946513175964}
TextBlob -> subjectivity: 0.65. polarity: 0.0
VADER
           -> {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
Тут медведев учился, наверно все хорошо,
Dostoevsky -> {'positive': 0.46102678775787354, 'skip':
0.11921291798353195}
TextBlob -> subjectivity: 0.5. polarity: 0.4166666666666667
VADER
           -> {'neg': 0.0, 'neu': 0.769, 'pos': 0.231, 'compound':
0.2023}
```

"НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО'' (УНИВЕРСИТЕТ ИТМО)

ОТЗЫВ РУКОВОДИТЕЛЯ о выполнении курсового проекта (работы)

Студент	Железникова П.А.					
-	(Фамилия, И., О.)					
Факультет	ПИиКТ		Группа	P41071		
Направление (спеці	иальность) <u>09.</u>	09.04.04 Программная инженерия				
Руководитель	Государев И.Б.,	к.п.н., доцент	-			
		(Фамилия, И.,	О., должность)			
Дисциплина	Проектирование	е и анализ языі	ков веб-реш	п ений		
TT			D-41			
Наименование тем	ы: <u>Сравнені</u>	ие оиолиотек	Pytnon для	анализа тональности текста		
	на русско	ом языке				
			(Наи	менование темы)		

ОЦЕНКА КУРСОВОГО ПРОЕКТА (РАБОТЫ)

No	Помолого ти		Оценка				
п/п	Показатели	5	4	3	0		
1.	Проект создан обучающимся самостоятельно						
2.	Созданные элементы сайта раскрывают тематику и название фирмы						
3.	Проект технологически грамотный						
4.	Оформление отвечает требованиям к отчету						
5.	Во время защиты обучающийся показал умение кратко, доступно представить результаты работы, умение анализировать, аргументировать свою точку зрения, делать обобщение и выводы, адекватно ответить на поставленные вопросы.						
	ИТОГОВАЯ ОЦЕНКА						

Отмеченные достоинства:
В отчете студента отражены полученные в ходе выполнения проекта навыки
соответствующие компетенциям по данной тематике. Студент показал себя личностьк
пунктуальной, ответственной, готовой к изучению нового материала.
Отмеченные недостатки:
Заключение:
Студент подтвердил навыки, полученные за время обучения по указанной специальности.
_
Руководитель И.Б. Государев
(подпись)