



Кейс ЦБ РФ

Анализатор текстовых
пресс-релизов

10 сентября 2023 года

mic

ДЕКОМПОЗИЦИЯ КЕЙСА

При решении кейса мы определили 3 задачи которые нам надо решить



Задача 1

Реализация классификации "Категории" и "Уровень рейтинга" для пресс-релиза



Задача 2

Выделение ключевых конструкций для соответствующего кредитного рейтинга



Задача 3

Создание веб-интерфейса для нашего продукта

Классификация

Стек: tensorflow

Метрики:

F1w (7 классов): 0.74

F1w (17 классов): 0.59

F1w (Комбинированная): 0.64

Text

Tokenizer

Token and Positional
Embedding

Transformer Encoder

Global Average Polling

Target 1

Target 2

Получение ключевых слов

В модель передается текст

Вычисляются веса внимания на последнем слое энкодера трансформера, усредняются по оси многоголового внимания

Полученная матрица весов нормализуется так, чтобы мат. ожидание было равно нулю

Выбираются элементы и их индексы из этой матрицы, которые \geq чем максимальное значение этой матрицы минус одно стандартное отклонение

Из тензора закодированного текста выбираются ключевые слова ранее полученными индексами

Производится декодировка текста по словарю

Полученные слова и тензор их весов используются для дальнейшей обработки визуального представления этих слов в тексте

Веб-сервис



Введите текст статьи или загрузите файл

Ведущий рейтинговый аналитик представил членам рейтингового комитета факторы, влияющие на рейтинг, члены комитета выразили свои мнения и предложения. Председатель рейтингового комитета предоставил возможность каждому члену рейтингового комитета высказать свое мнение до начала процедуры голосования. АО «Эксперт РА»

Choose File No file chosen

Отправить

Рейтинг: А

Уровень: AA

**Триггер слова: Аналитик, процедуры
голосования, кредитные рейтинги ...**

Выход с пометками

Ведущий рейтинговый **аналитик** представил членам рейтингового комитета факторы, влияющие на рейтинг, члены комитета выразили свои мнения и предложения. Председатель рейтингового комитета предоставил возможность каждому члену рейтингового комитета высказать свое мнение до начала **процедуры голосования**. АО «Эксперт РА» в течение последних 12 месяцев не оказывало ООО «ВГК» дополнительных услуг. **Кредитные рейтинги,**

Скачать

ИТОГ

Что мы успели сделать

Во-первых

Построили и обучили модель классификатор

Во-вторых

Создали алгоритм выделения ключевых слов и конструкций

В-третьих

Создали веб-интерфейс

× mlc = ❤

Введите текст статьи или загрузите файл

Choose File No file chosen

Отправить

Рейтинг: А

Уровень: АА

Триггер слова: Аналитик, процедуры голосования, кредитные рейтинги ...

Macbook Air

ИДЕИ ДЛЯ РАЗВИТИЯ

Как говорить, хотели но не успели:

Во-первых

Расширить датасет

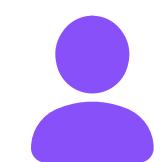
Во-вторых

Создать модель seq-to-seq для выделения ключевых слов

В-третьих

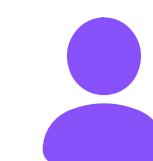
Использовать обратную связь пользователя для дообучения модели

Наша команда



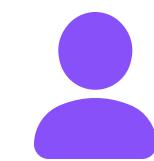
Викулин Дмитрий

tg: @shiz_vd40



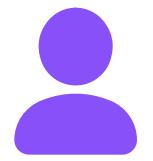
Кирилл Сидорский

tg: @ggwriuf



Маршев Иван

tg: @ivan_marshev



Маршев Илья

tg: @Mar_shevs



Глеб Овчинников

tg: @gllebatti

mic