



# Анализ тональности текста

Курсовая работа студента 361 группы  
Калмыкова Алексея  
Научный руководитель: Губанов Ю.А.

# Введение

- Классификация текста по эмоциональной окраске
  - Позитивная
  - Негативная
- Автоматизированный анализ отзывов о продуктах

# Существующие методы

- Машинное обучение
- Поиск эмотивной лексики
- Комбинированный подход

# Существующие решения

- Множество веб-сервисов
  - Платное API
- Open Source – только для английского

# Постановка задачи

- Принимать на вход файлы формата .doc, .docx, .pdf, .odt, .txt
- Анализировать эмоциональную окраску текста
- Сообщать пользователю о результатах
- Точность порядка 70%

# Реализация

- Машинное обучение
  - База отзывов с сайта Яндекс.Маркет
  - Стемминг
  - Матрица документов
- Логистическая регрессия

$$\ln\left(\frac{p}{1-p}\right) = word_1 + word_2 + \dots$$

# Оценка ошибок

- Обучающая выборка – 2000 отзывов
- Тестовая выборка – 1000 отзывов
- Точность на тестовой выборке 58.2%

# Что привело к низкой точности

- ➔ Маленькая обучающая выборка
- ➔ Количество положительных отзывов намного больше, чем негативных



# Результаты

- Создана размеченная база отзывов
- Реализован алгоритм стемминга
- Реализован программный продукт для анализа тональности текста из файлов .doc, .docx, .pdf, .odt