

# Автоматическое определение поддерживающего тона в педагогической обратной связи

- Елена Учайкина
- Пилотное NLP-исследование

- Проблема: комментарии преподавателей различаются по тону.
- Цель: проверить, можно ли автоматически выявлять поддерживающий тон по тексту.

# Данные

- ~189 русскоязычных реплик
- Ручная разметка по 3 критериям
- В финале используется только crit3 — поддерживающий тон
- (остальные исключены из-за дисбаланса классов)

# Метод

- TF-IDF (униграммы + биграммы, 5000 признаков)
- Модели:
  - - Logistic Regression
  - - LinearSVC
  - - ComplementNB
  - - Random Forest
- Оценка: 5-fold stratified CV, F1

# Результаты

- Logistic Regression —  $0.84 \pm 0.02$
  - LinearSVC —  $0.83 \pm 0.02$
  - Random Forest —  $0.77 \pm 0.01$
  - ComplementNB —  $0.75 \pm 0.05$
- 
- Линейные модели лучше работают на TF-IDF



# Почему crit1 / crit2 исключены

- Крайний дисбаланс классов
- Меньше 10 положительных примеров
- Это ограничение данных, а не моделей.  
Проблема в корпусе

# Лингвистический анализ

- Поддержка:
  - - молодец, великолепно
  - - очень, значительно
  - - ты, горжусь
- Отсутствие поддержки:
  - - институциональная лексика
  - - условия и сравнения
  - - отрицательные формулировки

# Интерпретация

- Два регистра педагогического дискурса:
-  Аффективно-оценочный
-  Административно-инструктивный
- Поддержка хорошо кодируется лексически



# Выводы

- Поддерживающий тон надёжно определяется
- Даже небольшой корпус даёт устойчивый baseline
- Линейные модели интерпретируемы и эффективны

# Дальнейшая работа

- Расширение корпуса
- Контекстные модели
- Применение в edtech