

# АННОТИРОВАНИЕ

---

- Постановки задачи
- Какие бывают методы
- Простые методы без учителя
- Трудности

# ПОСТАНОВКИ ЗАДАЧИ

---

- › Сократить текст
- › Выделять в тексте ключевые предложения
- › Получить краткую аннотацию для коллекции текстов

Промышленная группа «Базовый Элемент» объединяет около 100 российских и международных предприятий, работающих в энергетической, горнодобывающей, металлургической, машиностроительной, авиационной, финансовой, сельскохозяйственной и других отраслях.

Обширный и диверсифицированный портфель активов «Базового Элемента» представляет собой единую бизнес-структуру, предприятия которой эффективно взаимодействуют между собой, реализуя партнерские программы, позволяющие усилить конкурентные преимущества отдельных компаний группы и холдинга в целом.

В состав «Базового Элемента» входят лидеры крупнейших промышленных отраслей. Среди них – ведущий мировой производитель алюминия РУСАЛ, крупнейший в России частный производитель электроэнергии «ЕвроСибЭнерго» (входят в группу En+), автомобильный холдинг номер один в России «Группа ГАЗ» (входит в холдинг «Русские машины»), а также компания «Главстрой», лидер строительного рынка Москвы и Санкт-Петербурга.

Масштабная и активная деятельность предприятий «Базового Элемента» вносит существенный вклад в развитие российской промышленности, экономики и социальной инфраструктуры. Бизнес-группа обеспечивает около 1% ВВП России и инвестирует значительные средства в развитие регионов страны. «Базовый Элемент» – один из крупнейших работодателей в России. В течение последних пяти лет компания создала более 15 тыс. новых рабочих мест и планирует создать еще несколько десятков тысяч рабочих мест к 2025 году.

Стратегия роста «Базового Элемента» направлена на укрепление лидерских позиций предприятий группы за счет повышения эффективности производства, расширения промышленной базы и реализации крупных инновационных проектов, а также на содействие социально-экономическому развитию территорий своего присутствия и обеспечение экологической безопасности производства.

- Обучение с учителем (supervised learning)
  - ▶ Нужна разметка: какие предложения оставить (решаем задачу классификации)
  - ▶ Признаки предложений: длина, количество слов с заглавной буквы/терминов, встречаемость слов из предложения во всем тексте и др.
  - ▶ Любой классификатор

- Обучение без учителя (unsupervised learning)
  - ▶ Подсчет «значимости» предложений на основе их содержания
  - ▶ Выделение групп предложений, относящихся к одной общей идее

- › Рассматривается корпус, состоящий из отдельных предложений документа и самого документа
- › Признаки — частоты слов
- › Cosine similarity между документом и предложениями — ранг предложений
- › Аннотация — предложения с рангом выше заданного порога (или предложения с самым высоким рангом)

# КЛАСТЕРИЗАЦИЯ ПРЕДЛОЖЕНИЙ

---

- Признаки предложений — частоты слов
- Кластеризуем предложения k-Means
- Оставляем ближайшие к центрам кластеров предложения



# ПОНИЖЕНИЕ РАЗМЕРНОСТИ

---

- › Во всех предыдущих примерах мы получали представление предложений в модели мешка слов
- › Можно пробовать преобразовывать пространство признаков

# БОЛЕЕ СЛОЖНЫЕ МЕТОДЫ

---

- › TextRank
- › Аннотирование с помощью нейросетей

- Постановки задачи
- Какие бывают методы
- Простые методы без учителя
- Трудности