# Введение

Трудно представить себе жизнь современного человека без машинного обучения и искусственного интеллекта. Данные технологии создают новые возможности. Например, системы безопасности в метро используют машинное обучение для распознавания и поиска лиц, которые находятся в розыске. Компьютерное зрение используется в беспилотных автомобилях и летательных аппаратах для коммерческих и военных целей.

Машинный интеллект - это направление технологических исследований. Оно связано с созданием системы, которая способна обучаться самостоятельно.

Искусственный интеллект широко применяется в самых разных областях, как:

1. lT-сфера: разработка приложений;

2. Рекламные компании;

3. Маркетинговые исследования;

4. Медицинская диагностика;

5. Техническая диагностика;

6. Автопилотирование;

7. Биоинформатика и в другие сферы;

8. Обработка текста (в том числи кластеризация).

Все это и многое другое работает при помощи машинного интеллекта.

С одной стороны, такой технический прогресс приносит человеку огромную пользу, взять хотя бы медицинскую диагностику: многие исследования стали быстрыми и доступными, а с другой стороны такой технический прогресс бросает вызов каждому из нас: ведь совсем скоро не нужны будут машинисты в электропоездах, многий обслуживающий персонал заменят компьютеры, такcистов заменят беспилотные автомобили. Поэтому тему машинного обучения я считаю актуальной в настоящее время с научной и практической точки зрения. Нужно не только знать, что существует так называемое машинное обучение, но и понимать, как это работает.

Машинное обучение избавляет программиста от необходимости писать большой код, т.е объяснять компьютеру, как нужно решить какую-нибудь проблему. В процессе машинного обучения компьютер учат самостоятельно находить правильное решение. То есть у нас есть какие-либо известные данные и на основе их мы собираем статистику и уже на новых данных мы учимся что-то понимать (находим новые, ещё не описанные закономерности). И задачи классификации и кластеризации документов так же являются частью машинного обучения.

В связи с наблюдаемым на протяжении последних десятилетий стремительным ростом накапливаемых объемов электронных документов особое значение приобретает разработка программных средств поиска информации.

Традиционными подходами к решению проблемы поиска информации в коллекциях полнотекстовых документов, являются поиск по ключевым словам. Однако, традиционные механизмы классификационного поиска не успевают изменяться вслед за темпом развития науки и техники или требуют высоких затрат как на адаптацию классификаторов.

Таким образом, в настоящее время существует потребность в разработке методов классификации, которые способны на основе анализа текстов и внутренних связей между ними автоматически строить рубрикаторы коллекций полнотекстовых документов. Среди известных методов автоматического анализа текстовых данных потенциально способных решить представленную проблему следует выделить методы кластеризации, которые автоматически разбивают документы на группы (кластеры) на основе анализа тематической близости между ними.

Таким образом, **объектом исследований** в дипломной работе являются способы кластеризации и классификации документов на базе методов машинного обучения.

**Целью моей** дипломной работы является, создания программного средства, способного кластеризовать и классифицировать различные наборы документов.

Исходя из поставленной цели, были выделены следующие задачи:

1. Провести аналитический обзор предметной области.
2. Выбрать математическую модель и методов ее решения
3. Создать программный комплекс, реализующий выбранную математическую модель
4. Провести ряд экспериментов, показывающих эффективность разработанного программного комплекса.