**Слайд 1**

Здравствуйте уважаемая комиссия. Вашему вниманию представляется мой дипломный проект, тема которого – «Разработка системы кластеризации документов на базе методов машинного обучения».

**Слайд 2**

Целью дипломной работы является создание программного средства, способного кластеризовать и классифицировать различные наборы документов.

Для достижения этой цели были выполнены следующие задачи:

* проведен аналитический обзор предметной области;
* выбрана математическая модель и метод ее решения;
* создан программный комплекс, реализующий выбранную математическую модель;
* проведен ряд экспериментов, показывающих эффективность разработанного программного комплекса.

**Слайд 3**

Кластеризация — задача группировки множества объектов на подмножества (кластеры) таким образом, чтобы объекты из одного кластера были более похожи друг на друга, чем на объекты из других кластеров по какому-либо критерию.

Классификация — это система распределения предметов, явлений или понятий какой-нибудь области на классы, разделы и разряды.

**Слайд 4**

Кластеризация и классификация находят широкое применение во многих сферах:

* медицине (определение заболеваний)
* компьютерные науки (подбор рекомендаций для пользователя на основе предпочтений других пользователей в данном кластере)
* маркетинг (выделения типичных групп покупателей, разделения рынка для создания персонализированных предложений)
* интернет (повышение релевантности ответов поисковых запросов)

**Слайд 5**

Бывают различные виды алгоритмов кластеризации: