

Разработка системы рекомендаций по выбору алгоритмов машинного обучения

**Руководитель: Соболевская Евгения
Юрьевна к.т.н., доцент кафедры ИТС**

Проектная команда:
БИС-23-1 Камалова В. А
БИС-23-1 Конева А. Е

MACHINE

ЗАКАЗЧИК

Информация о заказчике:

Кафедра информационных
систем и технологий

Соболевская Евгения Юрьевна
к.т.н., доцент кафедры ИТС

<<<<



MACHINE LEARNING

СТЕЙКХОЛДЕРЫ



ПРОЕКТНАЯ КОМАНДА

An illustration of two simple stick figures in business suits shaking hands, symbolizing a partnership or agreement.

ЗАКАЗЧИК

The logo for the Faculty of Information Technology (ИТС) at Volgograd State University (ВВГУ). It features the text 'ВВГУ' and 'КАФЕДРА ИТС' in a stylized, bold font.

«««« Проблема, на
решение которой
направлен проект

Сложность в выборе наиболее
эффективного алгоритма машинного
обучения: существующие ресурсы не
содержат четких рекомендаций по
выбору алгоритмов машинного обучения



ГИПОТЕЗА

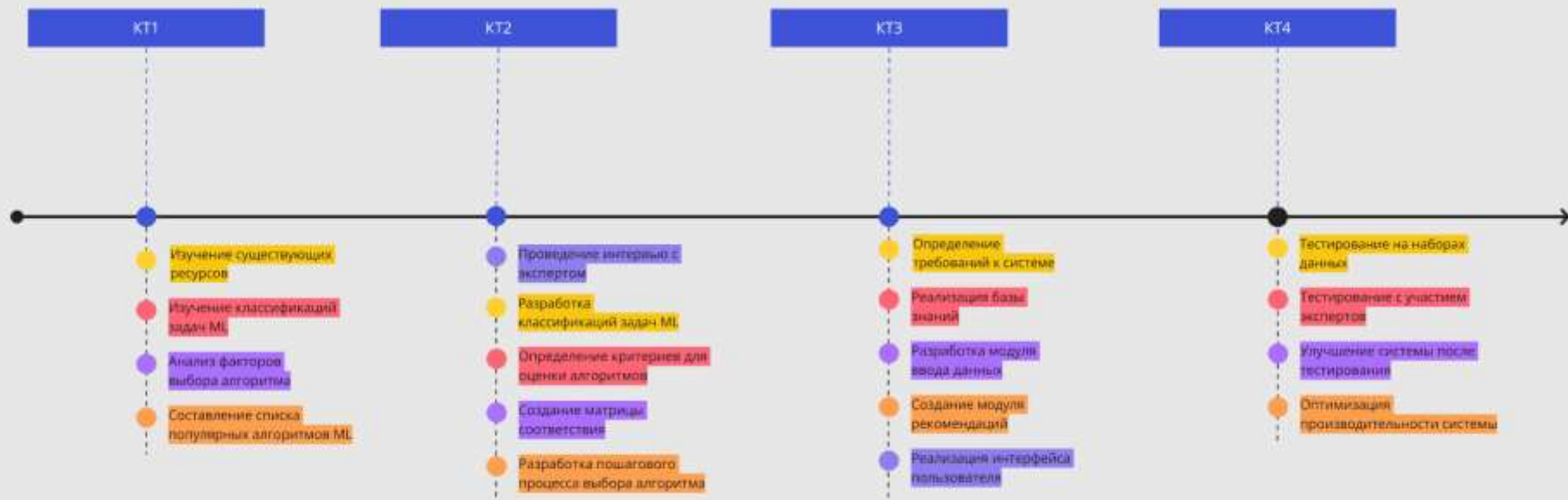
Разработка системы рекомендаций на основе характеристик задач и данных повысит эффективность выбора алгоритмов машинного обучения и улучшит качество прогнозов по сравнению с традиционным подходом.

Цель проекта

Создать систематизированный подход и практическое руководство, которое поможет эффективно выбирать алгоритмы машинного обучения (ML) для решения конкретных задач

MACHINE LEARNING

Дорожная карта



(TW) / (TW) /

ЗАДАЧИ ПРОЕКТА

на контрольную точку 1

Провести обзор литературы и существующих ресурсов по выбору алгоритмов машинного обучения

Изучить существующие классификации задач машинного обучения

Проанализировать факторы, которые влияют на выбор алгоритма

Составить список наиболее популярных алгоритмов ML с их характеристиками



>>>> Проектный инструментарий

01.

DISCORD



02.

MIRO



03.

WORD



04.

POWERPOINT



05.

TELEGRAM



MACHINE LEARNING (ML)

МАТРИЦА ОТВЕТСТВЕННОСТИ РЕАЛИЗАЦИИ ПРОЕКТА

	Провести обзор литературы и существующих ресурсов	Изучить существующие классификации задач машинного обучения	Проанализировать факторы, которые влияют на выбор алгоритма	Составление списка популярных алгоритмов ML
Камалова В.А				
Конева А.Е				

КАЛЕНДАРНЫЙ ГРАФИК ВСТРЕЧ ПО ПРОЕКТУ



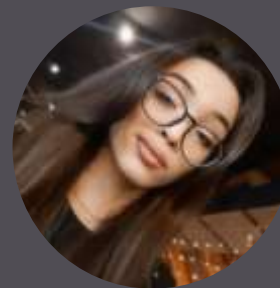
miro

<<<<

КАЛЕНДАРНЫЙ ГРАФИК ВСТРЕЧ ПО ПРОЕКТУ

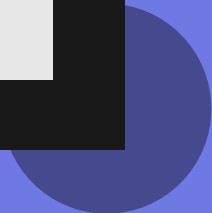
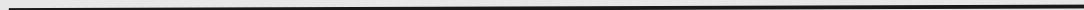
Каждые вторник, четверг, суббота и
воскресенье с 18.03.2025 начали
заниматься проектом дистанционно
через DISCORD

>>>



/ [A1]

РЕЗУЛЬТАТЫ
НАШЕГО
ИССЛЕДОВАНИЯ



СПИСОК ЛИТЕРАТУРЫ И СУЩЕСТВУЮЩИХ РЕСУРСОВ

01. KAGGLE.COM



— платформа для соревнований по data science, анализа данных и совместной работы над ML-проектами.

HABR.COM .02

— русскоязычная IT-платформа для публикации статей, обсуждений и новостей в сфере технологий



03. COMMUNITY.OPENAI.COM



— сообщество разработчиков, работающих с OpenAI.

.04.

ML.MIT.EDU



Massachusetts
Institute of
Technology

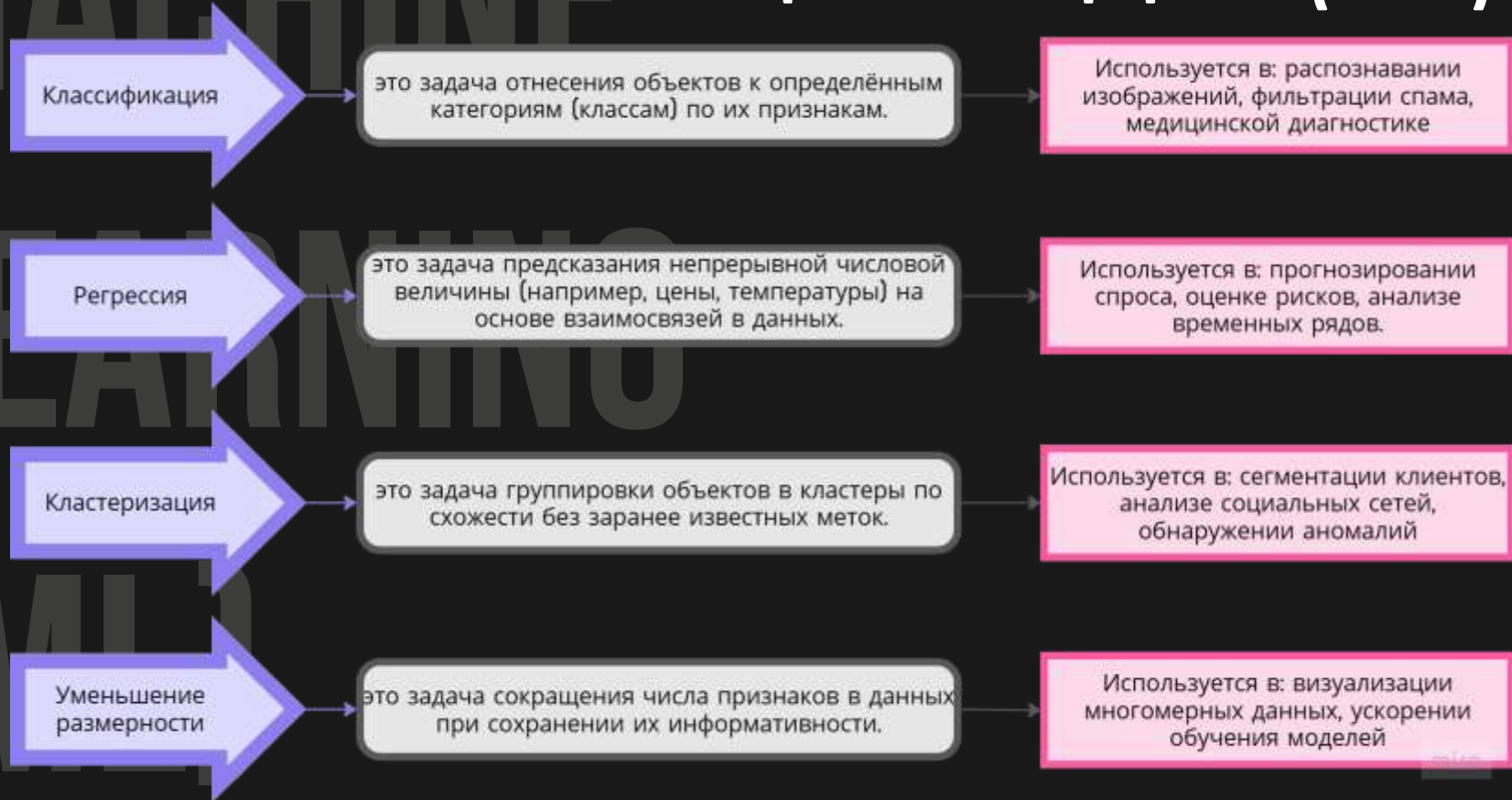
— статьи и проекты от MIT по машинному обучению

GITHUB .05

— веб-сервис для хостинга IT-проектов и их совместной разработки.



КЛАССИФИКАЦИИ ЗАДАЧ (ML)



ФАКТОРЫ, КОТОРЫЕ ВЛИЯЮТ НА ВЫБОР АЛГОРИТМА

Тип задач	Объем и структура данных	Точность	Скорость и интерпретируемость	Ограничения по вычислительным ресурсам
Нужно понять, какую задачу нужно решить и с какими данными работать	Влияет на производительность, надёжность и возможности обобщения моделей машинного обучения	Алгоритмы делятся на точные и примерные. Для больших данных предпочтительны примерные (снижают риск переобучения), для малых — точные	Время обучения зависит от алгоритма и связано с точностью: чем сложнее модель, тем дольше обучение. Ограничение по времени критично для больших выборок. Интерпретируемость важна для прозрачности решений (например, в медицине или финансах).	Ограничения ресурсов увеличивают время обучения/вывода. Решение: оптимизация данных и выбор алгоритмов с низкой сложностью (например, линейные модели вместо глубоких сетей).

MACHINE

TOP-10

ПОПУЛЯРНЫХ АЛГОРИТМОВ

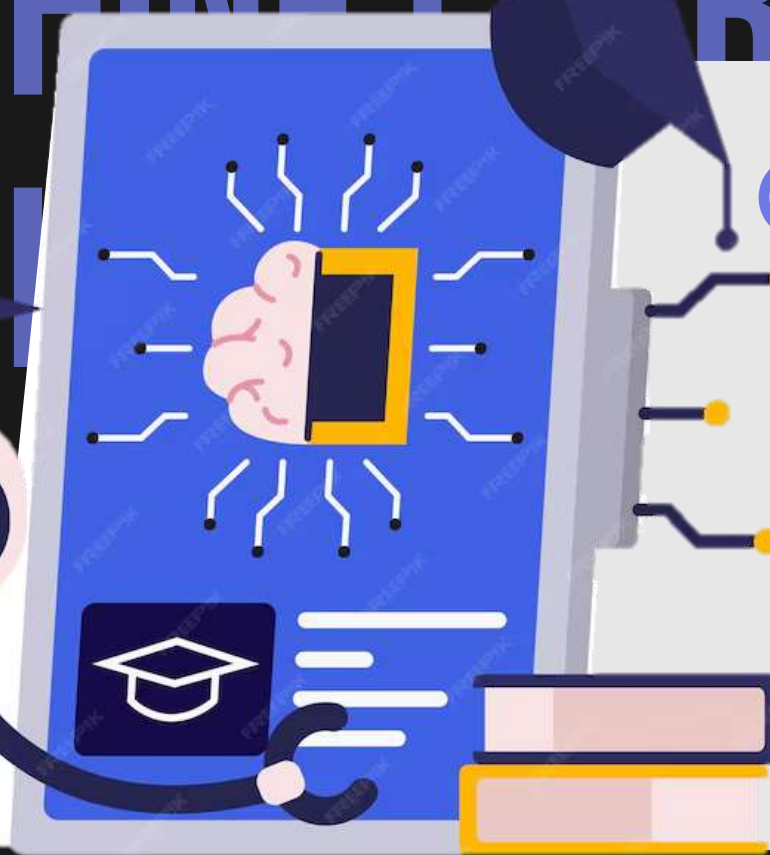
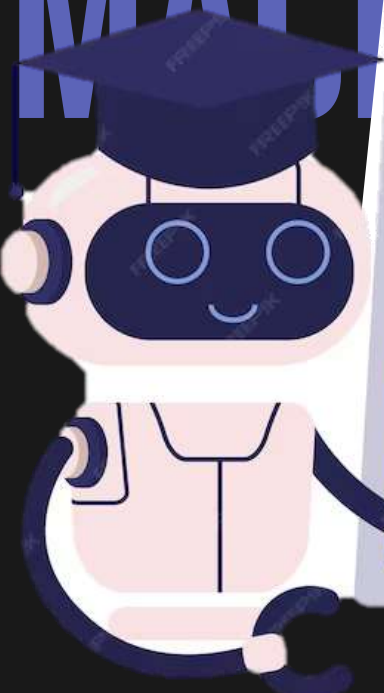
ML



MACHINE

Линейная регрессия	Прогнозирует числовые значения через линейную зависимость переменных (например, цена дома от площади).
Логистическая регрессия	Предсказывает бинарные исходы (да/нет) с помощью логистической функции (спам/не спам, диагноз болезни).
Деревья решений	Строит прозрачные правила для классификации или регрессии (кредитный скоринг, отток клиентов).
RANDOM FOREST	Ансамбль деревьев решений: голосование для повышения точности и снижения переобучения (прогноз акций, рекомендации).
SVM (SUPPORT VECTOR MACHINE)	Разделяет данные гиперплоскостью для классификации (распознавание изображений, спам-фильтры).
KNN (к-ближайших соседей)	Классифицирует объекты по большинству среди К ближайших точек (рекомендации, медицинская диагностика).
PCA	Сокращает размерность данных, сохраняя главные признаки (визуализация, ускорение моделей).
K-MEANS	Группирует данные в К кластеров по близости (сегментация клиентов, сжатие изображений).
Градиентный бустинг	Последовательно улучшает предсказания слабых моделей (часто побеждает в KAGGLE).
Глубокое обучение	Нейросети для сложных задач: изображения, текст, речь (распознавание лиц, переводы, голосовые ассистенты).

MACHINE LEARNING (ML)



ОЖИДАЕМЫЕ РЕЗУЛЬТАТЫ

разработанная система рекомендаций в виде прототипа или готового приложения



MACHINE

LEAR

(ML)

Благодарим
за внимание!

ГОТОВЫ ОБСУДИТЬ НАШУ РАБОТУ



(ML)