**Разработка модели для анализа эмоциональной окраски отзывов в сети.**

**Сфера**: онлайн-данные

**SMART-модель:**

**Специфичность (Specific)**

Цель: Создать модель машинного обучения для анализа эмоциональной окраски (сентимент-анализа) отзывов в сети, которая сможет автоматически определять положительные, негативные и нейтральные отзывы с высокой точностью.

Задачи:

* Исследовать и выбрать подходящие наборы данных для обучения и тестирования модели.
* Определить ключевые параметры и метрики для оценки эффективности модели (например, точность, полнота, F1-мера).
* Выбрать подходящие технологии и инструменты для разработки модели (например, TensorFlow, PyTorch).

**Измеримость (Measurable)**

Цель: Достигнуть точности классификации отзывов не менее 85% на тестовом наборе данных.

Задачи:

* Разработать систему логирования и мониторинга процесса обучения для отслеживания прогресса.
* Регулярно проводить валидацию модели на валидационном наборе данных для предотвращения переобучения.

**Достижимость (Achievable)**

Цель: Убедиться, что имеются необходимые ресурсы и знания для создания и обучения модели.

Задачи:

* Провести обзор доступных образовательных материалов и курсов по машинному обучению и сентимент-анализу.
* Оценить доступность вычислительных мощностей для обучения модели.
* При необходимости привлечь дополнительных специалистов в команду проекта.

**Релевантность (Relevant)**

Цель: Убедиться, что разработка модели соответствует общим бизнес-целям и задачам.

Задачи:

* Определить, как анализ эмоциональной окраски отзывов может помочь в улучшении продуктов или услуг компании.
* Проанализировать, как результаты работы модели могут быть интегрированы в существующие бизнес-процессы.

**Ограниченность по времени (Time-bound)**

Цель: Запустить рабочую версию модели для анализа эмоциональной окраски отзывов в течение 6 месяцев.

Задачи:

* Разработать подробный план проекта с промежуточными этапами и дедлайнами.
* Регулярно проводить встречи команды проекта для отслеживания прогресса и корректировки плана при необходимости.

**Анализ существующих решений:**

* Изучение моделей машинного обучения применяемых для сентимент-анализа текстовых данных. Выбор наиболее оптимальной и эффективной модели. Изучение средств разработки модели.
* Поиск современных моделей сентимент-анализа и анализ качества данных моделей.
* Оценка результатов использования существующих моделей, мониторинг внедрения новых моделей сентимент-анализа.

**Шаги по реализации:**

* Поиск датасета отзывов клиентов на услуги заведений/компаний.
* Предобработка данных (очистка от пустых значений, дубликатов, формирование конкретных классов для обучения).
* Обработка корпуса текстов средствами NLP.
* Классификация текстов с помощью моделей машинного обучения.
* Оценка качества модели на метриках точности, полноты и F-меры (precision, recall, F-score).
* Внедрение и анализ результатов различных моделей, поиск наиболее точной и оптимальной модели.