I. Введение

* Обзор задачи поиска штрихкода по описанию товара
* Краткое введение в NLP и машинное обучение
* Цели и задачи работы

II. Обзор литературы по NLP и машинному обучению

* Основные методы машинного обучения в NLP
* Техники предобработки текстовых данных
* Алгоритмы классификации текстовых данных
* Работа с неструктурированными данными

III. Обзор литературы по задаче поиска штрихкода по описанию товара

* Подходы к решению задачи
* Анализ существующих методов и моделей
* Классификация товаров на основе текстовых данных

IV. Методология работы

* Выбор архитектуры модели машинного обучения
* Подготовка данных для обучения
* Описание параметров модели
* Обучение и оценка модели

V. Результаты работы

* Описание полученной модели машинного обучения
* Сравнение существующих методов и моделей
* Анализ полученных результатов

VI. Обсуждение и заключение

* Обсуждение достигнутых результатов
* Возможности дальнейшего улучшения модели
* Заключение работы

VII. Список литературы

РЕФЕРАТ

Изм.

Лист

№ докум.

Подпись

Дата

Лист

*\_*

Разраб.

Провер.

Реценз.

Н. Контр.

Утверд.

Лит.

Листов

Пояснительная записка \_\_ стр., \_\_ рис., \_\_ табл.

КЛЮЧЕВЫЕ СЛОВА.

Цель работы – разработка программы автоматического сопоставления названий товаров для продуктивного использования продающими компаниями. Объектом исследования являются исторические данные о продажах компании. Рассматриваются основы моделей обработки естественного языка, приводятся основные источники шума в экспериментальном наборе данных. Рассмотрена гибридная схема с использованием нескольких моделей машинного обучения на основе архитектуры трансформеров. Приведены алгоритмы работы моделей машинного обучения. Оценена точность приведенных алгоритмов.

СОДЕРЖАНИЕ

[РЕФЕРАТ 2](#_Toc129690728)

[СОДЕРЖАНИЕ 3](#_Toc129690729)

[ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ 4](#_Toc129690730)

[ВВЕДЕНИЕ 5](#_Toc129690731)

[ГЛАВА 1. ПРИНЦИП РАБОТЫ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ НА ОСНОВЕ АРХИТЕКТУРЫ ТРАНСФОРМЕРОВ 6](#_Toc129690732)

[7.11. ВЫВОДЫ 7](#_Toc129690733)

[ЗАКЛЮЧЕНИЕ 8](#_Toc129690734)

[СПИСОК ЛИТЕРАТУРЫ 9](#_Toc129690735)

[ПРИЛОЖЕНИЕ 1 10](#_Toc129690736)

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

*ML* - Machine Learning (машинное обучение)

*AI* - Artificial Intelligence (искусственный интеллект)

*ANN* - Artificial Neural Network (искусственная нейронная сеть)

*RF* - Random Forest (случайный лес)

*KNN* - K-Nearest Neighbors (метод ближайших соседей)

*NB* - Naive Bayes (наивный байесовский классификатор)

*DL* - Deep Learning (глубокое обучение)

*CNN* - Convolutional Neural Network (сверточная нейронная сеть)

*RNN* - Recurrent Neural Network (рекуррентная нейронная сеть)

*NLP* - Natural Language Processing (обработка естественного языка)

*NLTK* - Natural Language Toolkit (инструментарий для работы с естественным языком)

*POS* - Part of Speech (часть речи)

*NER* - Named Entity Recognition (распознавание именованных сущностей)

*LSA* - Latent Semantic Analysis (латентный семантический анализ)

*LDA* - Latent Dirichlet Allocation (латентное размещение Дирихле)

*SVM* - Support Vector Machine (метод опорных векторов)

*ANN* - Artificial Neural Network (искусственная нейронная сеть)

*MLP* - Multi-Layer Perceptron (многослойный персептрон)

*LSTM* - Long Short-Term Memory (долгая краткосрочная память)

ВВЕДЕНИЕ

Для современного бизнеса в России ежегодно проводятся реформы, направленные на увеличение эффективности компаний и роста их количества на территории государства. В условиях стремительно развивающегося бизнеса важно своевременно найти слабое место на пути к развитию и вовремя его устранить. И, если на уровне государства, можно быстро понять, что тормозит бизнес-процессы в целом, то на уровне отдельно взятой корпорации, при рассмотрении бизнес-процессов, все может быть неоднозначно. Множество факторов могут быть годами скрыты от глаз и из эффект на работы компании не явен. В частности, большое влияние оказывает осведомленность сотрудников компании о трудовых обязанностях друг друга, поскольку понять, что процесс выстроен неэффективно можно лишь зная чем занимается каждый участник процесса.

Поскольку бизнес строится людьми и в бизнес-процессах участвуют люди, на определенном уровне абстракции, можно считать, что развитие бизнеса ограничено именно человеческими способностями, например способностью долго сохранять концентрацию или эффективно вести коммуникацию в рамках рабочих процессов.

Все это подталкивает к тому, что необходимы способы увеличения человеческой продуктивности в задачах, где человека не может полностью заменить компьютер. Машинное обучение в задачах подобного рода наиболее органично встраивается в бизнес-процесс.

До сих пор, трудной для автоматизации областью является обработка тестовой информации. В данной работе предлагается эффективный метод снижения ресурсов на обработку информации о товарообороте компании. Главная особенность этой области для компании в том, что товары могут называться множеством различных способов, на это влияют не только маркетинговые потребности, например, более красочное звучание названия товара, но и особенности практической реализации, например, название может не помещаться на ценник товара, и, в результате сокращений слов, можно получить несколько новых названий одного и того же товара.

# ПРИНЦИП РАБОТЫ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ НА ОСНОВЕ АРХИТЕКТУРЫ ТРАНСФОРМЕРОВ

## ВЫВОДЫ

ЗАКЛЮЧЕНИЕ

СПИСОК ЛИТЕРАТУРЫ

ПРИЛОЖЕНИЕ 1