

УДК 60.608

**ПРИМЕНЕНИЕ МЕТОДОВ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА В
АНАЛИЗЕ ПОЛЬЗОВАТЕЛЬСКИХ НАСТРОЕНИЙ**

Решетникова Д. Р. (Университет ИТМО), Проскурин М. Н. (Университет ИТМО)

**Научный руководитель – старший преподаватель Волков А. Р.
(Университет ИТМО)**

Введение. Современный мир переживает эпоху цифровизации и информатизации, в которой огромное количество данных генерируется ежедневно в различных сферах человеческой деятельности. Одним из ключевых аспектов этого процесса является растущее значение текстовой информации, поступающей от пользователей через социальные сети, форумы, отзывы на товары и услуги. В этом контексте анализ пользовательских настроений, основанный на методах обработки естественного языка (Natural Language Processing, NLP), становится важным инструментом для компаний и организаций в понимании и предвидении потребительского поведения.

Основная часть. Основополагающие решения и принципы.

Данный блок будет рассматривать решения, где каждый из приведенных ниже аспектов представляет собой важный элемент исследования, который поможет глубоко понять и анализировать пользовательские настроения с использованием методов обработки естественного языка.

Создание облака слов для позитивных и негативных текстов:

Для каждой категории отзывов (позитивные и негативные) будет сформировано отдельное облако слов. Это предполагает предварительную классификацию текстов на основе их эмоциональной окраски. Облака слов будут визуализировать частоту употребления различных слов и фраз, помогая выявить ключевые элементы, которые определяют положительное или отрицательное восприятие сотрудниками своего рабочего места.

Связи между различными переменными, такими как возраст, стаж работы, должность и уровень удовлетворенности работой. Это поможет понять, какие факторы наиболее сильно влияют на настроение сотрудников.

Гипотеза о квалификации и удовлетворенности сотрудников:

Будет проведен анализ для проверки гипотезы о том, что более квалифицированные сотрудники испытывают большее удовлетворение от своей работы по сравнению с менее квалифицированными. Для этого данные будут сегментированы по уровню образования, квалификации и должности, с последующим сравнением уровней удовлетворенности в каждой группе.

Зависимость настроения сотрудников от отрасли:

Исследование будет направлено на выявление различий в уровне удовлетворенности сотрудников в зависимости от отрасли. Это может включать сравнение секторов, таких как IT, финансы, государственные учреждения, и др., для определения того, где сотрудники чувствуют себя наиболее комфортно.

Геолокация и анализ вакансий:

Будет проведена визуализация распределения вакансий и соответствующих отзывов с использованием картографических инструментов. Это позволит выявить географические тенденции в удовлетворенности работой и предпочтениях сотрудников.

Требовательность сотрудников в крупных городах:

Анализ будет сосредоточен на выяснении, отличается ли уровень требовательности

сотрудников в крупных городах по сравнению с другими регионами. Это включает в себя оценку ожиданий сотрудников от работодателей и их общее удовлетворение работой.

Модальный стаж сотрудников по сегментам бизнеса:

Исследование направлено на вычисление и анализ самого частого (модального) стажа работы сотрудников в различных секторах. Это поможет определить, есть ли отраслевые особенности в карьерной траектории и удовлетворенности работой.

Анализ авторов отзывов (текущие или бывшие сотрудники):

Будет проведен анализ для выяснения, оставляют ли отзывы в основном текущие или бывшие сотрудники, и как это влияет на общий тон и содержание отзывов.

Рейтинг должностей будет составлен на основе уровня удовлетворенности сотрудников. Это позволит выявить профессии с наиболее высокой и низкой удовлетворенностью среди работников.

Отраслевой анализ удовлетворенности сотрудников:

Анализ будет направлен на определение отраслей, где сотрудники наиболее и наименее удовлетворены своей работой, с последующим выявлением возможных причин такой динамики.

Сезонность написания отзывов и её влияние на оценки:

Будет проведен анализ сезонности написания отзывов и исследовано, как время года влияет на настроение и оценку, оставленную в отзыве.

Предсказание оценки компании из отзыва:

Разработка модели машинного обучения для предсказания оценки компании на основе текста отзыва. Это включает в себя обучение модели на существующих данных и последующую проверку её эффективности на синтетически созданных отзывах.

В ходе данного исследования были успешно применены передовые методы обработки естественного языка для анализа и понимания настроений сотрудников в различных отраслях и регионах. Создание облаков слов для позитивных и негативных отзывов позволило визуально оценить ключевые факторы, влияющие на удовлетворенность работой. Анализ корреляционной матрицы выявил значимые взаимосвязи между различными факторами, такими как квалификация сотрудников, их стаж работы и уровень удовлетворенности. Исследование показало, что уровень квалификации влияет на удовлетворенность работой, подтверждая гипотезу о том, что более квалифицированные сотрудники чувствуют себя более удовлетворенными на своих рабочих местах. Также была выявлена зависимость удовлетворенности от отрасли, что подчеркивает значимость отраслевых особенностей в восприятии рабочей среды сотрудниками. Геолокационный анализ отзывов выявил региональные различия в удовлетворенности сотрудников, особенно выделяя разницу между крупными городами и другими регионами. Было установлено, что сотрудники в крупных городах могут быть более требовательны к работодателям. Анализ модального стажа работы в различных секторах бизнеса позволил определить специфику карьерного роста и удовлетворенности в разных отраслях.

Выводы. В заключение, разработанная модель предсказания оценок компании на основе текста отзыва демонстрирует важность и эффективность применения методов машинного обучения и NLP для анализа пользовательских настроений. Это исследование не только подтверждает значимость обработки естественного языка в современном анализе данных, но и открывает новые перспективы для его применения в управлении человеческими ресурсами и улучшении рабочего климата.

Список использованных источников.

1. Nikolic D. The Human Side of AI // School of Finance and Management (DBIS). Frankfurt: Goethe-University, 2017.
2. Faggella D. What is Machine Learning? // Emerj. URL: <https://emerj.com/aiglossary-terms/what-is-machine-learning/> (accessed: 03.01.2024)
3. Giménez M., Palanca J., Botti V. Semantic-based padding in convolutional neural networks for improving the performance in natural language processing. A case of study in sentiment analysis // Neurocomputing. 2020. Vol. 78. P. 315-323.