

## **ОТЗЫВ**

официального оппонента

доктора технических наук, профессора Вишнякова Юрия Муссовича на диссертационную работу Мансур Али Махмуд «Модель, метод и алгоритмы Data Mining для интеллектуальной обработки и анализа текстов на естественном языке», представленную на соискание ученой степени кандидата технических наук по специальности 1.2.1. Искусственный интеллект и машинное обучение.

### **1. Актуальность темы диссертационной работы**

В условиях массового применения технологий искусственного интеллекта (ИИ) экспоненциальный рост объемов текстовых данных становится серьезной проблемой. Большая часть этих данных представлена неструктурированными текстами на естественном языке, генерируемыми в различных областях. Это требует разработки эффективных методов их обработки и анализа для выявления скрытых закономерностей, что, в свою очередь, позволит повысить качество алгоритмов машинного обучения. Ключевым аспектом здесь является создание низкоразмерных векторных представлений текстов с интерпретируемыми признаками. Решение этой задачи имеет важное научное и практическое значение, способствуя развитию технологий искусственного интеллекта и машинного обучения.

Таким образом, разработка модели, метода и алгоритмов Data Mining для интеллектуальной обработки и анализа текстовых данных в условиях их непрерывного роста позволяет повысить эффективность систем искусственного интеллекта. Это определяет **актуальность** темы диссертационного исследования как с научной, так и с прикладной точки зрения.

### **2. Оценка достоверности полученных результатов и новизны диссертационного исследования**

Представленные в диссертации научные положения, выводы и рекомендации являются обоснованными, достоверными и обладают

признаками научной новизны. Содержание диссертации изложено на высоком научно-техническом уровне и сопровождается убедительной аргументацией. В качестве основы для получения научных результатов автором использованы труды отечественных и зарубежных учёных, внесших значительный вклад в развитие исследуемой области.

**Достоверность** результатов диссертационной работы обеспечивается теоретическим обоснованием основных положений и экспериментальным подтверждением эффективности предлагаемых научных решений. Помимо этого, достоверность результатов диссертации подтверждается апробацией работы в высокорейтинговых научных изданиях и выступлениях с докладами на научно-технических конференциях.

**Научная новизна** рассматриваемой диссертации заключается в создании модели, метода и алгоритмов обработки и анализа текстов на естественном языке в условиях экспоненциального роста их объёмов, что позволяет повысить эффективность средств и инструментов систем искусственного интеллекта и машинного обучения.

В состав научной новизны исследования вошли следующие основные результаты:

1. Построена математическая модель векторизации текстов на основе концептов, отличающаяся применением новых правил построения эталонных концептов и новых функций определения их весов, позволяющая снизить размерность векторного пространства и улучшить дискриминационную способность результирующих векторов признаков;

2. Разработан модифицированный метод генерации векторных представлений документов на основе построенной модели векторизации, отличающийся применением интерпретируемых признаков при векторизации, позволяющий снизить частоту ошибок алгоритмов классификации и кластеризации документов;

3. Разработан алгоритм извлечения и фильтрации ключевых фраз на основе частоты их появления, отличающийся применением функции парсера

для разметки частей речи, что позволяет извлекать ключевые фразы с правильной грамматической структурой;

4. Разработан алгоритм построения концептов из семантически близких фраз, отличающийся решением задачи кластеризации фраз с учетом контекстуальной семантической близости, что позволяет повысить однородность кластеров, представляющих концепты.

**Теоретическая значимость** представленного исследования заключается в развитии научного аппарата обработки естественно-языковых текстов за счет создания новых моделей, методов и алгоритмов. Работа вносит вклад в развитие области искусственного интеллекта и машинного обучения, предлагая решения для эффективного анализа текстов, формирования интерпретируемых векторных представлений и улучшения качества классификации и кластеризации текстов. Полученные результаты имеют важное значение для развития методов работы с семантическими текстовыми представлениями в современных информационных системах.

**Практическая значимость** работы обусловлена созданием программного приложения, реализующего разработанные модель, метод и алгоритмы обработки и анализа текстов на естественном языке в системах искусственного интеллекта для минимизации частоты появления ошибок при решении задач классификации и кластеризации с учётом условий снижения размерности векторного пространства и сохранения его интерпретируемости. Подтверждением практической значимости работы является внедрение результатов диссертации в информационные процессы организации ООО «Ит-Эффект» (г. Москва) и в учебный процесс Южного федерального университета, о чем свидетельствуют представленные акты о внедрении.

### **3. Оценка содержания диссертации, степени ее завершенности, подтверждение публикаций автора**

Структура диссертации соответствует теме и задачам исследования.

Работа состоит из введения, четырех разделов, заключения, списка литературы из 146 наименований источников, 2-х приложений, изложена (без учета приложений) на 150 страницах машинописного текста, содержит 31 рисунок и 12 таблиц. Работа имеет логичную структуру, поясняющие рисунки и таблицы оформлены в соответствии с действующими требованиями ГОСТ.

Диссертация является завершенным научно-исследовательским трудом, в ней изложены предлагаемые автором модель, метод и алгоритмы обработки и анализа текстов на естественном языке в системах искусственного интеллекта. Достоинством работы является ее теоретический и прикладной характер. Судя по содержанию работы, поставленные автором задачи решены полностью, цель исследования достигнута. Диссертационная работа обладает научной новизной, теоретической значимостью и практической ценностью.

Результаты диссертационного исследования Мансур А.М., выводы и сформулированные положения изложены в 17 печатных работах, 3 из которых опубликованы в журналах, входящих в Перечень ВАК РФ, 2 в журналах, индексируемых в базах Scopus и/или Web of Science. Имеется 2 свидетельства о государственной регистрации программ для ЭВМ. Также полученные результаты представлены в 9 работах, опубликованных в трудах всероссийских и международных научно-технических конференций.

#### **4. Соответствие содержания автореферата диссертации**

Автореферат удовлетворяет требованиям положения ВАК, так как отражает основные положения, результаты и выводы диссертации, новизну и значимость результатов исследований, показывает структуру диссертации.

#### **5. Замечания по диссертационной работе**

При изучении материалов диссертации отмечены следующие недостатки:

1. Поскольку в диссертации утверждается, что концепты извлекаются из самого текста без применения внешних источников, таких как базы знаний и онтологии, то представляется целесообразным очертить класс документов, для которых предлагаемые методы классификации наиболее эффективны.
2. Целесообразным было бы привести оценку временной трудоемкости предложенных алгоритмов при обработке больших объемов данных.
3. Следовало бы исследовать возможность применения предлагаемых решений для обработки коротких текстов, а также документов, включающие многоязычные данные.

Тем не менее, отмеченные недостатки носят более рекомендательный характер и не влияют на общую положительную оценку диссертационного исследования Мансур А.М.

## **6. Заключение**


Оппонируемая диссертация Мансур А.М. представляет собой законченную научно-квалификационную работу, посвященную решению актуальной задачи искусственного интеллекта, имеющей важное хозяйственное значение. Диссертация характеризуется несомненной научной новизной, является самостоятельной и полностью завершенной, имеет теоретическую и практическую значимость, а также в полной мере соответствует пунктам 4 и 5 паспорта специальности 1.2.1. Искусственный интеллект и машинное обучение (технические науки).

Несмотря на отмеченные замечания, работа Мансур А.М. по своей актуальности, научной новизне, объему выполненных исследований и практической значимости полученных результатов, соответствует требованиям, предъявляемым ВАК РФ к диссертациям на соискание ученой степени кандидата технических наук в пп. 9-14 Положения о присуждении ученых степеней, утверждённого Постановлением Правительства РФ от 24 сентября 2013 г. № 842 «О порядке присуждения ученых степеней» (с

изменениями и дополнениями) в редакции от 16 октября 2024 года, и, вне всякого сомнения, заслуживает высокой оценки. Считаю, что Мансур Али Махмуд достоин присуждения ученой степени кандидата технических наук по специальности 1.2.1. Искусственный интеллект и машинное обучение (технические науки).

Официальный оппонент,  
доктор технических наук, профессор,  
(н.с. 05.13.12 Системы автоматизации проектирования, 05.13.14 Системы обработки информации и управления),  
профессор кафедры вычислительных технологий  
федерального государственного бюджетного  
образовательного учреждения высшего образования  
«Кубанский государственный университет» (КубГУ),  
350040, г. Краснодар, ул. Ставропольская, д. 149,  
Тел.: +7 (861) 219-95-19, e-mail: jury.vishnyakov@gmail.com,  
<https://www.kubsu.ru/ru/public-portfolio/23146>

«2» июня 2025 г.

 Юрий Муссович Вишняков

Подлинность подписи Вишнякова Ю.М.  
ЗАВЕРЯЮ  
Специалист по кадрам  
Климаньченко

