

«УТВЕРЖДАЮ»
Директор Федерального
государственного учреждения
«Федеральный исследовательский
центр «Информатика и управление»
Российской академии наук»,
член-корреспондент РАН

__ М.А. Посыпкин

«03» 04 2026 г.

МП

ОТЗЫВ

ведущей организации – Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» – на диссертационную работу Перминова Андрея Игоревича на тему «Доверенный байесовский классификатор для данных малой размерности на основе многослойного перцептрона», представленную на соискание учёной степени кандидата физико-математических наук по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»

Актуальность темы

В последние годы наблюдается активное внедрение методов машинного обучения в критически важные области, включая медицину, финансы, управление критической инфраструктурой. Однако широко известна проблема отсутствия формальных гарантий корректности принимаемых решений. Модели часто демонстрируют избыточную уверенность на объектах, нехарактерных для обучающей выборки, и не способны корректно оценивать границы своей компетенции.

В диссертационной работе Перминова А.И. предлагается новый подход к построению доверенного классификатора на основе многослойного перцептрона, который обеспечивает статистически обоснованную оценку апостериорных вероятностей и реализует механизм отказа от классификации вне носителя обучающего распределения. Предложенный метод унарной классификации позволяет принципиально устранить проблему дисбаланса классов, а разработанный подход к генерации синтетических данных открывает возможности для безопасного расширения обучающих выборок без риска раскрытия конфиденциальной информации. Таким образом, тема диссертационной работы Перминова А.И. является актуальной.

Характеристика содержания диссертационной работы

Диссертация имеет общий объём 155 страниц и состоит из введения, пяти глав, заключения и списка литературы из 102 источников.

Во введении обосновывается актуальность проблемы, формулируются цель и задачи диссертационной работы, указана научная новизна результатов исследования, раскрыта их

теоретическая и практическая значимость и приводятся основные положения, выносимые на защиту.

В первой главе приводится обзор работ по теме диссертации, рассматриваются существующие методы классификации в контексте требований доверенного искусственного интеллекта. Анализируются подходы к оценке неопределённости, обнаружению объектов вне распределения и обработке дисбаланса классов. На основе проведённого анализа делается вывод о необходимости разработки интегрированного статистически обоснованного подхода.

Вторая глава посвящена разработке модифицированного байесовского классификатора. Автором предлагается метод добавления искусственного фонового класса с равномерным распределением, что позволяет обеспечить отказ от классификации вне носителя обучающей выборки. Доказывается теорема о свойствах такого классификатора. Для аппроксимации предложенного подхода используется многослойный перцептрон, а для интерпретации его решений разработан метод построения объясняющего дерева eXVTree. Проведена экспериментальная оценка эффективности подхода и его устойчивости к состязательным атакам.

В третьей главе рассматривается разработанный метод унарной классификации, который полностью устраняет проблему чувствительности к дисбалансу классов. Для каждого класса строится независимый классификатор, отделяющий его от равномерного фона. Формулируется и доказывается теорема о состоятельности предложенного подхода. Предложены специализированные метрики для оценки качества унарных классификаторов, учитывающие их специфику. Экспериментальная оценка на наборах данных из репозитория UCI демонстрирует конкурентоспособность метода по сравнению с XGBoost и его устойчивость к дисбалансу.

Четвёртая глава содержит описание разработанного метода генерации синтетических табличных данных на основе унарной классификации. Предложенный подход позволяет создавать синтетические выборки, сохраняющие геометрические и статистические свойства исходного распределения. Проведён сравнительный анализ с методом CTGAN на наборах данных UCI. Результаты демонстрируют высокую полезность (utility) предложенного метода для данных малой и средней размерности.

В пятой главе приводится описание разработанной интеллектуальной системы машинного обучения – автономного веб-приложения, реализующего все предложенные методы. Описываются архитектура системы, реализация вычислительного ядра с оптимизацией разворачивания циклов, подсистема визуализации и примеры использования системы для реализации разработанных методов.

В заключении сформулированы основные результаты диссертации.

Новизна и достоверность полученных результатов

Основные научные результаты, полученные в диссертационной работе, являются новыми и заключаются в следующем:

1. Разработан метод построения доверенного классификатора на основе многослойного перцептрона, обеспечивающий статистически обоснованную оценку апостериорной вероятности и формальный механизм отказа от классификации вне носителя обучающего распределения.
2. Разработан метод унарной классификации, устойчивый к дисбалансу классов и позволяющий генерировать синтетические данные, сохраняющие геометрические свойства исходного распределения.
3. Разработан метод построения объясняющего дерева eXVTree для интерпретации решений многослойного перцептрона с кусочно-линейной функцией активации.

Достоверность результатов подтверждается строгими математическими доказательствами сформулированных теорем (о свойствах модифицированного байесовского классификатора, о временной сложности построения объясняющего дерева, о состоятельности унарной классификации), экспериментальной оценкой точности предложенных методов на

наборах данных из открытых репозиториях, а также сравнительным анализом с существующими подходами (XGBoost, CTGAN). Качественная оценка методов подтверждена обнаруженными свойствами устойчивости к состязательным атакам и корректного поведения вне носителя распределения.

Теоретическая значимость и практическая ценность

Теоретическая значимость результатов диссертации заключается в развитии статистических основ доверенного искусственного интеллекта для нейросетевых моделей. Сформулированные и доказанные теоремы обосновывают асимптотическую связь между нейросетевой и гистограммной оценками апостериорной вероятности, поведение классификатора вне носителя распределения и состоятельность предложенного унарного подхода. Полученные результаты расширяют теоретическую базу непараметрического оценивания и создают основу для построения математически строгой теории доверенного искусственного интеллекта.

Практическая значимость полученных результатов заключается в том, что предложенные методы позволяют создавать надёжные классификаторы для критических приложений с гарантированным поведением в условиях неопределённости. Разработанные методы могут быть применены специалистами в области анализа данных и машинного обучения, разработчиками программного обеспечения для критических инфраструктур, исследователями в области доверенного искусственного интеллекта, а также инженерами, решающими задачи классификации в условиях дисбаланса классов и высокой неопределённости данных в таких сферах, как медицинская диагностика, финансовый скоринг и системы поддержки принятия решений. Представленные методы реализованы в виде программной системы и используются в Исследовательском Центре Доверенного Искусственного Интеллекта ИСП РАН.

Замечания по работе

Несмотря на общую положительную оценку диссертационной работы, следует отметить несколько недостатков:

1. Анализ современного состояния области в первой главе носит избирательный характер. В частности, обзоры методов классификации (раздел 1.2) и подходов к обработке дисбаланса классов (раздел 1.3) не охватывают актуальные методы адаптации функций потерь, например, фокальную, и архитектурные модификации, включая механизмы внимания и ансамблевые подходы. Кроме того, при переходе к нейросетевым моделям недостаточное внимание уделено трудностям перехода от многослойного перцептрона к более сложным глубоким сетям, что могло бы дополнительно усилить мотивацию исследования.

2. В разделе 2.3 для формального определения маргинального распределения признаков более математически корректно было бы использовать не саму меру Лебега, а задаваемое ей вероятностное распределение на компакте.

3. В теореме 4 (глава 3) существенным образом используется функция активации на основе модуля. Выбор немонотонного и недифференцируемого в нуле симметричного функционала требует более детальной аргументации. В отличие от гладких немонотонных функций (например, GeLU или Swish), предложенная функция характеризуется разрывным скачком производной в точке излома, а также может препятствовать формированию разреженных активаций, что потенциально снижает обобщающую способность моделей.

4. В работе было бы уместно продемонстрировать преимущество разработанных подходов относительно классических статистических методов, например, дискриминантного анализа или непараметрического восстановления плотностей, являющихся в условиях малой размерности признакового пространства как вычислительно эффективными, так и интерпретируемыми.

Заключение по работе

Отмеченные недостатки не снижают положительной оценки диссертационной работы. Диссертация Перминова А.И. является законченной научно-квалификационной работой, в которой решена актуальная задача разработки статистически обоснованных методов построения доверенных классификаторов на основе многослойного перцептрона для данных малой размерности.

Диссертация соответствует паспорту научной специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей» (пп. 4, 7, 8, 9). Результаты диссертации в полной мере отражены в публикациях автора, в том числе 6 – в журналах из перечня ВАК, 4 – в журналах, индексируемых Web of Science и Scopus, и апробированы на крупных научных конференциях. Автореферат содержит в краткой форме все основные результаты, полученные в диссертации, и соответствует содержанию диссертации.

Диссертационная работа Перминова А.И. полностью соответствует всем требованиям ВАК РФ, предъявляемым к диссертациям на соискание учёной степени кандидата физико-математических наук, а Перминов Андрей Игоревич заслуживает присуждения учёной степени кандидата физико-математических наук по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Отзыв на диссертационную работу рассмотрен и утвержден на заседании секции Учёного совета ФИЦ ИУ РАН «02» апреля 2026 г., протокол № 3. Присутствовало на заседании 14 человек. Результаты голосования: принято единогласно.

Заместитель директора ФИЦ ИУ РАН
по научной работе, д.ф.-м.н.

Разумчик Р.В.

Научный руководитель Отделения ФИЦ ИУ РАН,
главный научный сотрудник, д.т.н., профессор

Зацаринный А.А.

Федеральное государственного учреждение «Федеральный исследовательский центр
«Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН)
Адрес: 119333, Москва, Вавилова, д.44, кор. 2
<http://www.frccsc.ru/>
Тел: + 7 (499) 135-62-60
E-mail: frccsc@frccsc.ru