

ИСПОЛЬЗОВАНИЕ ПЕРСОНАЛЬНЫХ ДАННЫХ ДЛЯ РАЗВИТИЯ ТЕХНОЛОГИЙ МАШИННОГО ОБУЧЕНИЯ

И. А. Старовойтов

*студент, Белорусский государственный технологический университет, г. Минск,
Беларусь, e-mail: workplaceilstar@gmail.com*

Научный руководитель: М. М. Пилинога

*старший преподаватель, Белорусский государственный технологический
университет, факультет информационных систем и технологий, г. Минск,
Беларусь, e-mail: pilinoga@belstu.by*

С развитием технологий машинного обучения стало возможным использовать персональные данные с уникальной эффективностью. Эта статья исследует различные аспекты использования персональных данных в современном мире, обсуждая как позитивные, так и потенциально негативные последствия. Рассматриваются вопросы конфиденциальности, этики, безопасности и регулирования в контексте сбора и анализа персональных данных для машинного обучения.

Ключевые слова: персональные данные; этика; обучение; безопасность.

USE OF PERSONAL DATA FOR THE DEVELOPMENT OF MACHINE LEARNING TECHNOLOGIES

I. A. Starovoitov

Student, Belarusian State University, Minsk, Belarus, e-mail: workplaceilstar@gmail.com

Supervisor: M. M. Pilinoga

*Senior Lecturer, Belarusian State Technological University, Faculty of Information
Systems and Technologies, Minsk, Belarus, e-mail: pilinoga@belstu.by*

With the advancement of machine learning technologies, the use of personal data has become more efficient than ever before. This article explores various aspects of utilizing personal data in the modern world, discussing both the positive and potential negative consequences. It addresses issues of privacy, ethics, data security, and regulation in the context of collecting and analyzing personal data for machine learning.

Keywords: personal data; ethics; training; security.

В свете стремительных технологических изменений последних десятилетий машинное обучение приобрело важное значение как инструмент для обработки и анализа данных. Эффективность и точность алгоритмов машинного обучения зависят от качества данных, используемых в процессе обучения [1]. Плохие или несбалансированные данные могут привести к низкой точности и ненадежным прогнозам. Важно понимать, что машинное обучение не способно компенсировать недостатки в данных. Напротив, оно подчеркивает и усиливает эти недостатки, делая их более заметными.

Собирать и подготавливать данные для машинного обучения – задача, требующая внимания к множеству аспектов. Необходимо проводить анализ данных, выявлять выбросы и аномалии, а также корректно заполнять пропущенные значения. Кроме того, выбор подходящих признаков или характеристик данных имеет решающее значение. Некорректный выбор признаков может привести к избыточности или недостаточной информативности модели. Алгоритмы машинного обучения автоматически извлекают признаки из данных, но предварительный анализ и экспертное знание могут значительно улучшить качество модели.

С развитием технологий машинного обучения разработчики и аналитики начинают все чаще использовать персональные данные для обучения и валидации моделей, что позволяет улучшить метрики готового продукта. Этот тренд отражает важность использования реальных и репрезентативных данных для создания более точных и надежных моделей машинного обучения. Данный подход позволяет улучшить точность и обобщающую способность модели, а также более точно персонализировать продвигаемые продукты и услуги, однако при этом раскрывает некоторые негативные последствия использования персональных данных:

1. Проблемы конфиденциальности и безопасности: сбор и хранение больших объемов персональных данных влечет за собой значительные риски в области конфиденциальности и безопасности. Даже при строгом соблюдении нормативных требований и стандартов, существует риск утечки данных или несанкционированного доступа к ним, что может привести к серьезным нарушениям частной жизни и финансовым потерям.

2. Этические аспекты: использование персональных данных вызывает сложные этические вопросы. Вопросы о том, какие данные можно использовать, как долго их хранить и какие цели они могут служить, поднимают важные этические дилеммы. Неправильное обращение с данными или их злоупотребление может привести к доверительному кризису и негативным последствиям для компаний.

3. Законодательство и регулирование: законодательство о персональных данных становится все более строгим и комплексным. Компа-

нии, использующие персональные данные, должны вкладывать дополнительные ресурсы и усилия в соблюдение требований законодательства, что может увеличивать операционные расходы и юридические риски.

4. Доверие и репутация: негативные события, связанные с утечкой или ненадежной обработкой персональных данных, могут серьезно подорвать доверие клиентов и потребителей. Утрата доверия может иметь долгосрочные негативные последствия для репутации компании.

Использование персональных данных в машинном обучении – это мощный инструмент, но и с ним связаны серьезные обязанности и вызовы. Правильный и этичный подход к обработке и анализу данных может помочь сформировать более совершенные и надежные решения, способствующие развитию технологий и общества в целом, а соблюдение лучших практик и норм обработки персональных данных помогает укрепить доверие пользователей и обеспечить долгосрочный успех как в развитии моделей машинного обучения, так и во взаимодействии с пользователями готового программного продукта. Особое внимание следует уделять обработке персональных данных, соблюдая нормы конфиденциальности и законодательства о защите данных.

Библиографические ссылки

1. *Карпов А. А., Смирнов И. В.* Применение алгоритмов машинного обучения в задачах обработки и анализа персональных данных: преимущества и риски // Вестник Национального исследовательского университета «МИЭТ». 2020. С. 50–58.

2. *Орлов И. П., Гринчик Н. Н.* Этические аспекты использования персональных данных в машинном обучении // Научно-технический вестник информационных технологий, механики и оптики. 2018. С. 62–67.

3. *Солдаткин В. И., Кузнецова О. А.* Проблемы конфиденциальности и безопасности при обработке персональных данных в машинном обучении // Вестник Московского университета. 2019. Серия 15: Вычислительная математика и кибернетика. С. 79–88.