

доступа: <https://habr.com/ru/articles/802435/> (дата обращения: 17.01.2025).

6. Методы искусственного интеллекта: учебное пособие / Н.В. Андреянов, Т.С. Евдокимова, А.Д. Павлов, А.С. Сытник, М.П. Шлеймович – Казань, 2023. – 342 с. – URL: https://elibs.kai.ru/_docs_file/623/HTML/index.html

7. Достоинства и недостатки нейронных сетей [Электронный ресурс]. Режим доступа: <https://bewave.ru/blog/dostoinstva-i-nedostatki-neyronnykh-setey/> (дата обращения: 19.01.2025).

УДК 004.05

В.В. Гедранович, доц. (БГЭУ, г. Минск)

ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ В ТЕСТИРОВАНИИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Современные технологии анализа данных играют важную роль в оптимизации процессов тестирования. Они позволяют собирать, обрабатывать и анализировать огромные объемы информации о поведении программного обеспечения (ПО), выявлять аномалии, которые могут указывать на возможные ошибки или недостатки. Внедрение алгоритмов машинного обучения и искусственного интеллекта в процессы тестирования способствует не только автоматизации рутинных задач, но и повышению точности и эффективности выявления дефектов. Анализ данных помогает тестировщикам принимать более обоснованные решения, улучшая качество конечного продукта и сокращая время вывода его на рынок.

Рассмотрим следующие технологии анализа данных, применяемые в тестировании ПО: Big Data, Машинное обучение, Data Mining, Аналитика в режиме реального времени.

Big Data, или большие данные, представляет собой большие объемы сложных и неструктурированных данных, которые постоянно генерируются и поступают из различных источников. Они характеризуются тремя основными аспектами, известными как три «*V*»: объем (*Volume*), скорость (*Velocity*) и разнообразие (*Variety*). Также иногда добавляют четвёртое «*V*» – достоверность (*Veracity*), что указывает на необходимость проверки качества и надежности данных, и пятое «*V*» – ценность (*Value*), чтобы подчеркнуть значимость извлечения полезной информации.

Использование технологий Big Data позволяет оптимизировать тестовые процессы:

– анализируя исторические данные о дефектах, можно определить наиболее уязвимые области кода, что помогает сосредоточить тестирование на них (*приоритизация тестирования*);

– наличие обширных массивов данных помогает создать более реалистичные и разнообразные тестовые сценарии, отражающие реальные условия работы систем (*улучшение качества тестовых данных*).

Big Data активно используется для моделирования нагрузочных тестов и оценки производительности. Сбор и анализ данных о пиковых нагрузках и поведении пользователей помогает моделировать реалистичные сценарии нагрузки, которые система испытывает в реальной эксплуатации. Примеры использования приведены в табл. 1.

Таблица 1 – Примеры использования Big Data в реальных проектах

Индустрия финансовых услуг	Электронная коммерция	Телекоммуникационные сети
Финансовые компании используют Big Data для тестирования систем управления рисками и обнаружения аномалий в реальном времени. Это помогает им своевременно выявлять мошеннические операции.	Платформы электронной коммерции применяют анализ больших данных для оптимизации взаимодействия с пользователями, что позволяет персонализировать предложения и улучшать пользовательский опыт в условиях роста трафика.	В телекоммуникационной отрасли Big Data помогает моделировать и тестировать сетевые нагрузки, улучшая качество обслуживания за счет своевременного распознавания и устранения узких мест в сети.

Машинное обучение (ML) – это область искусственного интеллекта, которая разрабатывает алгоритмы и статистические модели, позволяющие компьютерам выполнять задачи без явного программирования. В последние годы машинное обучение стало особенно популярным в тестировании ПО благодаря своей способности автоматизировать и улучшать процессы предсказания и обнаружения дефектов.

Рост применения машинного обучения в тестировании обусловлен несколькими факторами: увеличением объема данных, усложнением программных систем, требованием к скорости и качеству.

Машинное обучение применяется для прогнозирования и обнаружения дефектов на разных этапах разработки ПО. Одним из ключевых аспектов является использование моделей прогнозирования, основанных на исторических данных. Эти модели обучаются на прошлых данных о дефектах, ошибках и их исправлениях, чтобы выявлять вероятные области риска в новом и измененном коде. Они помогают предсказать, где в коде могут возникать ошибки и на какие участки следует

обратить внимание в первую очередь при тестировании. Прогнозирование позволяет сосредоточить ресурсы тестирования на критичных областях и снизить вероятность появления дефектов в продуктивной среде.

Для предсказания дефектов в тестировании ПО могут использоваться различные алгоритмы и инструменты машинного обучения: нейронные сети, линейные и нелинейные модели регрессии, ансамблевые методы (например, Random Forest, Gradient Boosting). Инструменты машинного обучения (библиотеки), такие как TensorFlow, PyTorch и Scikit-learn, часто используются для реализации этих алгоритмов в контексте предсказательного тестирования ПО [1, 2].

Data Mining – это процесс поиска и извлечения полезной информации из больших объемов данных. Он включает использование методов статистики, машинного обучения и баз данных для выявления скрытых шаблонов и значимых корреляций. В контексте тестирования ПО Data Mining помогает анализировать сложные наборы данных, что позволяет более эффективно обнаруживать дефекты и оптимизировать процессы тестирования.

Методы Data Mining играют важную роль в тестировании, так как позволяют:

- выявлять тенденции и закономерности в сложных и объемных данных;
- осуществлять прогнозирование и обнаружение аномалий;
- улучшать процессы путем оптимизации и автоматизации тестовых сценариев.

Примеры использования Data Mining представлены в табл. 2.

Таблица 2 – Кейсы успешного применения Data Mining в тестировании

Телекоммуникационные компании	Разработка ПО	Розничная торговля
Используют Data Mining для анализа логов сети и выявления аномалий, что позволяет оптимизировать процессы тестирования сетевой инфраструктуры и обеспечивать более высокое качество услуг.	Многие крупные software-компании внедрили Data Mining для анализа bag-trace-ров. Это позволяет выявлять часто повторяющиеся дефекты и автоматизировать некоторые аспекты тестирования на основании собранных данных.	Ритейлеры применяют Data Mining для анализа данных с помощью кластеризации и выявления аномалий, чтобы улучшить тестирование своих систем управления запасами и веб-сайтов.

Аналитика в режиме реального времени – это процесс обработки и анализа данных сразу после их получения, без задержек. В тестировании ПО аналитика в реальном времени играет важную роль, обеспечивая непрерывный мониторинг, возможность быстро реагировать на аномалии, неисправности и изменения в поведении систем, поддерживать высокий уровень качества, стабильности работы ПО и улучшать взаимодействие с пользователями (табл. 3).

Таблица 3 – Кейсы использования аналитики в режиме реального времени

Е-commerce компании	Финансовые учреждения	Платформы облачного стриминга
Обнаружение повышенной нагрузки на серверы в распродажи и акционные дни позволяет быстро масштабировать ресурсы и обеспечивать бесперебойный доступ для всех пользователей.	Мгновенный мониторинг транзакций и анализа поведения пользователей помогает во время выявлять подозрительную активность, тем самым предотвращая мошенничество.	Онлайн-сервисы, такие как Netflix, используют аналитику в реальном времени для обеспечения качественного потокового видео без задержек, анализируя параметры сети и устройства для оптимизации потоков.

Каждая из технологий анализа данных, используемая в тестировании ПО, обладает своими преимуществами и ограничениями, которые необходимо учитывать при их применении (табл. 4).

Таблица 4 – Преимущества и недостатки технологий

Характеристика	Big Data	Машинное обучение	Data Mining	Аналитика в реальном времени
Преимущества	Способность обрабатывать разнообразные данные	Обучение на опыте, постоянное улучшение	Глубокий анализ данных	Быстрая реакция на изменения
Недостатки	Требует значительных вычислительных ресурсов	Необходимость качественных данных	Возможность получения ложных выводов	Высокие требования к инфраструктуре

Выбор подходящей технологии анализа данных для тестирования ПО зависит от конкретных целей и ограничений проекта. Резюмируя, можно отметить, что технологии анализа данных существенно трансформируют подходы к тестированию ПО. Каждая из рассмотренных технологий – Big Data, машинное обучение, Data Mining и аналитика в режиме реального времени – предоставляет уникальные возможности для улучшения качества и эффективности тестирования.

Будущее технологий анализа данных в тестировании ПО обещает

быть динамичным и инновационным. По мере дальнейшего развития искусственного интеллекта и машинного обучения, интеграция этих решений в тестовые процессы будет только углубляться.

ЛИТЕРАТУРА

1. Pandey, R. K. Predicting Software Errors Using Neural Networks / R. K. Pandey, J. Chand, A. K. Srivastava, N. P. Singh, R. K. Gautam // 2023 12th International Conference on System Modeling & Advancement in Research Trends (SMART). – Moradabad, India, 2023. – P. 197-202. – doi: 10.1109/SMART59791.2023.10428165.
2. Ndlovu, L. Software Fault Detection Algorithms Using Artificial Intelligence: A Review and Classification / L. Ndlovu et al. // 2024 International Conference on Electrical, Computer and Energy Technologies (ICE-CET). – Sydney, Australia, 2024. – P. 1-6. – doi: 10.1109/ICE-CET61485.2024.10698379.

УДК 004

Ю.В. Евдокимова, доц. (МГУУ Правительства Москвы, г. Москва, Россия);
А.А. Евдокимова, адм. (МГТУ им. Н.Э. Баумана, г. Москва, Россия)

АВТОМАТИЗАЦИЯ ПРОЦЕССОВ ОБРАБОТКИ ДОКУМЕНТОВ В РОССИЙСКОЙ ПРАКТИКЕ

Предпринимательские структуры, стремящиеся достичь высокой результативности, сфокусированы на оптимизации бизнес-процессов, ориентируясь на рост их эффективности и сокращение затрат. Автоматизация процессов – продуманное стратегическое решение, дающая возможность компаниям адаптироваться к быстро меняющимся условиям рынка. ИТ-решения выполняют генеральную роль в данном процессе, реализуя конкретные задачи и обеспечивая достижение поставленных целей бизнеса.

Автоматизация обработки документов в рамках осуществления рабочих процессов становится всё более актуальной в современном обществе. Использование современных цифровых технологий показывает хорошую результативность во всех отраслях [1], позволяя значительно упростить процессы, оптимизировать временной ресурс и снизить вероятность ошибок.

Разработчикам и пользователям автоматизированных систем обработки документов необходимо учитывать определенные функциональные требований, обеспечивающие в дальнейшем эффективную работу системы. К ним отнесены: