

Качество данных (Data quality)

Разделы: [Бизнес-задачи](#)

Качество данных — обобщенное понятие, отражающее степень их пригодности к решению определенной задачи. В соответствии со стандартом ISO 9000:2015 основными критериями качества являются полнота, достоверность, точность, согласованность, доступность и своевременность.

Оценка качества данных и действия по его повышению являются необходимым этапом любого аналитического проекта, поскольку аналитические алгоритмы или не смогут работать с некачественными данными либо будут давать некорректные результаты.

Приведение исходных «сырых» данных в соответствие с требуемыми критериями качества является важнейшей задачей [Data Mining](#) и образует целое направление, называемое [предобработкой](#).

В качестве основных проблем, вызывающих снижение качества данных, обычно выделяют следующие:

- пропущенные значения;
- [дубликаты](#);
- [противоречия](#);
- [аномальные значения](#) и выбросы;
- [шум](#);
- отсутствие полноты данных;
- нарушения [целостности данных](#);
- некорректные форматы и представления данных;
- [фиктивные значения](#);
- ошибки ввода данных;
- нарушения структуры.

Некоторые из этих проблем являются критическими в том смысле, что они блокируют работу аналитических моделей и алгоритмов (например, пропущенные значения и нарушения структуры). Другие (например, дубликаты, противоречия, шумы) не нарушают работу алгоритмов, но порождают некорректные результаты анализа.

Независимо от того, какие факторы снижения качества присутствуют в данных, с ними необходимо бороться. Это делается в два этапа:

1. [Профайлинг](#) — исследование данных с целью выявления проблем и выработки стратегии их решения.

2. Очистка — применение различных методов для разрешения обнаруженных проблем: восстановление пропущенных значений, редактирование аномалий, обработка дубликатов и противоречий и т.д.

В Loginom визуализатор качество данных позволяет производить комплексную оценку степени пригодности данных для решения определенной задачи. Решение по очистке и дедупликации данных Loginom Data Quality обеспечивает автоматизацию процесса получения стандартизированной и качественной информации о клиенте. Подробнее в деморолике «Очистка клиентских данных».