

Машинное обучение и большие данные

Машинное обучение определяется как построение компьютерных алгоритмов, которые могут автоматически улучшаться с помощью опыта и использования данных.

В контексте больших данных машинное обучение используется для обработки больших объемов данных, поиску закономерностей и извлечению из них ключевой актуальной информации.

Основная цель специалиста по машинному обучению и большим данным – создание информационных технологий нового поколения, обеспечивающих экономически эффективное извлечение полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, и применение этих технологий в информационно-аналитической деятельности, в системах управления и принятия решений, а также для разработки на их основе новых продуктов и услуг.

Задачи компетенции заключаются в постановке и нахождении путей решения прикладных задач с использованием современных инструментов работы с большими данными.

Машинное обучение (machine learning) лежит в основе многих инновационных технологий различных секторов экономики. Лучшие практики использования машинного обучения:

1 В промышленности (минимизация простоев на производстве, системы управления производством, выявление угроз безопасности);

2 В финансах (обнаружение мошенничества для безопасных транзакций, алгоритмическая торговля – анализ настроений, финансовые консультации и управление портфелем, оценка кредитоспособности);

3 В здравоохранении (диагностика заболеваний; автоматические роботизированные операции; улучшение клиентского сервиса);

4 В ритейле и маркетинге (прогнозирование действий покупателей, персонализированные предложения и реклама, прогнозирование спроса и автоматизация закупок);

5 В логистике (экономия топлива и повышение производительности транспорта, предотвращение сбоев в поставках).

Актуальность компетенции «Машинное обучение и большие данные» связана в первую очередь с растущим количеством информации, необходимостью её обработки и анализа, а также бурным развитием инноваций.

Используя язык программирования Python, среду разработки Jupyter Notebook, инструменты хранения и извлечения данных SQL, PySpark, инструменты визуализации данных и библиотеки numpy, Scikit-Learn, Keras, Catboost, специалисты этой компетенции применяют алгоритмы машинного обучения для задач классификации и регрессии, оценивают качество обученных моделей, настраивают параметры и применяют нейронные сети для решения прикладных задач.

