

| Группа умений            | Умения   | Инструменты и знания   |
|--------------------------|--|--|
| <b>Работа с данными</b>  | Писать код на языке Python                     | Язык Python (синтаксис, типы данных, циклы, условия, функции, ООП)<br>Знание основных алгоритмов (сортировка и поиск, хэширование, графы, деревья и др.)<br>Jupyter notebook<br>Git<br>Bash  |
|                          | Работать с фреймворками для работы с данными   | NumPy, Pandas  |
|                          | Проводить предобработку данных                 | NumPy, Pandas  |
|                          | Проводить разведывательный анализ данных       | Алгоритмы и методы EDA, Статистические тесты, Проектирование экспериментов, Pandas, Seaborn, Matplotlib, Numpy   |
|                          | Визуализировать данные                         | Seaborn, Matplotlib, BI-аналитика  |
|                          | Извлекать фичи из данных                       | NumPy, Pandas  |
| <b>Машинное обучение</b> | Знать высшую математику для машинного обучения | Линейная алгебра<br>Математический анализ<br>Теория вероятностей<br>Дискретная математика<br>Алгоритмы на основе деревьев решений<br>Бустинг, стекинг  |
|                          | Работать с классическими алгоритмами МЛ        | Подходы к обучению: с учителем, без учителя, с подкреплением, semisupervised (смешанный)<br>Задачи классификации и кластеризации<br>Регрессионные задачи: линейная и логистическая регрессия<br>Снижение размерности<br>Градиентный бустинг<br>Оценка качества работы модели<br>Временные ряды<br>scikit-learn |

|                          |   |  |
|--------------------------|---|--|
|                          | Работать с продвинутыми алгоритмами машинного обучения и глубокого обучения   | xgboost (XGBoost)<br>catboost (CatBoost)<br>scikit-learn   |
|                          | Работать с нейронными сетями и глубоким обучением   | Сверточные и рекуррентные нейронные сети<br>Рекомендательные системы<br>Обучение с подкреплением<br>PyTorch, TensorFlow<br>Keras |
|                          | Проводить эксперименты для выбора оптимального алгоритма  | MLflow<br>scikit-learn<br>Keras  |
|                          | Проводить тестирование работы модели  | A/B-тестирование   |
|                          | Замерять качество работы модели в зависимости от типа   |  |
|                          | Оптимизировать гиперпараметры модели  | Инструменты оптимизации моделей  |
|                          | Выводить ML-модели в продуктивную среду   | Жизненный цикл ML-моделей<br>Docker<br>Сервисная архитектура   |
| <b>Инжиниринг данных</b> | Автоматизировать процессы разработки, тестирования, развертывания и мониторинга моделей машинного обучения                                | Jenkins, GitLab CI/CD  |
|                          | Поддержка версий моделей, данных и конфигураций для обеспечения воспроизводимости и контроля изменений                                    | MLflow, Git  |
|                          | Создание и управление инфраструктурой для развертывания моделей, включая вычислительные ресурсы, хранилища данных и инструменты для CI/CD | Docker<br>Kubernetes   |
|                          | Создавать hive-таблицы  | HiveQL   |
|                          | Строить витрины данных  | DWH (Data Warehouse)<br>Greenplum  |
|                          | Строить пайплайны выгрузки и доступа к данным для обучения алгоритмов ML  | Apache Airflow<br>NiFi   |

|  |   |   |
|--|---|---|
|  | Обеспечивать хранение данных в соответствии с ETL-процессами          |   |
|  | Владеть инструментами для обработки большого объема данных (big data) | Apache Airflow, Apache Spark, Hadoop, Kafka, Virtual box, Облачные хранилища, MapReduce |
| <b>Организация и проведение исследовательской работы</b> | Владеть принципами и методами проведения исследований                 |   |
|  | Анализировать и структурировать научную информацию                    |   |
|  | Работать с источниками  |   |
|  | Готовить научные доклады  |   |
|  | Писать ВКР  |   |
| <b>Софт скиллы</b>                                       | Работать и вести коммуникацию в команде                               |   |
|  | Презентовать идеи и решения   |   |
|  | Владеть приемами тайм-менеджмента                                     |   |
|  | Владеть навыками саморегуляции и самодисциплины                       |   |
|  | Аналитически мыслить  |   |
|  | Адаптироваться к новым условиям                                       |   |
|  | Непрерывно обучаться  |   |