

Магистерская программа “Аналитика Больших  
данных”

Март 2017

# 1 Структура программы

Магистерская программа в области науки о данных предназначена для бакалавров со знаниями в области компьютерных наук, математических методов и статистики.

Программа содержит лекции, объясняющие основные понятия, методы и средства, необходимые для анализа Больших данных. Лабораторные занятия позволят использовать эти знания на практике путем реализации сценариев использования (use cases), заимствованных из реальных практических решений в различных областях.

Программа предоставляет студентам знания и понимание основополагающих принципов и технологий Больших данных, обеспечивает необходимую подготовку студентов к работе в компаниях или научных учреждениях. Программа состоит из модулей, курсовой работы (первый год), научных семинаров и магистерской диссертации (второй год). Вводные лекции позволят заполнить пробелы в знаниях.

Программа основана на сочетании освоения практических технологий, таких как Hadoop, NoSQL и MapReduce, Спарк, технологий машинного обучения, использовании Python, R, C++, Ява, а также основополагающих знаний в математике и передовых вычислительных технологиях. Программа содержит все, что нужно знать для сбора, обработки и анализа больших, быстро меняющихся данных в науке или коммерции.

Программа учит извлекать скрытые знания и закономерности с помощью аналитических методов Больших данных, делать прогнозы и визуализировать результаты.

## 1.1 Цели программы

Цель программы – подготовка специалистов в области науки о данных (Data Science). Обеспечить возможность выбора оптимального решения в задачах с Большими данными, включая базы данных и облачные сервисы. Обеспечить понимание процесса анализа больших объемов данных, включая методы визуализации и автоматического обучения. Дать навыки программирования для создания простых решений с применением технологий Больших данных, таких как MapReduce и сценариев для NoSQL, научить писать параллельные алгоритмы для многоядерных процессоров.

## 1.2 Программа первого года обучения

Начальные модули программы посвящены основам фундаментальных знаний в данной области, а также в области программирования и математики, в том числе знаний в области структур данных и алгоритмов, статистики и машинного обучения. В течение первого года знания математики, программирования и анализа данных будет значительно углублены и расширены.

### Модули

- Математика для Больших данных.

- Комбинаторика и графы.
- Теория информации.
- Вероятность и правдоподобие.
- Статистический анализ данных.
- Машинное обучение.
- C++, Ява, Питон, R.
- Практический Unix.
- Структуры данных и алгоритмы.
- Параллельные и распределенные вычисления.
- Базы данных.
- Технические средства презентаций.
- Семинары и практикумы.
- Курсовой проект.

### 1.3 Программа второго года обучения

Модули второго года программы направлены на изучение ключевых приложений науки о данных, передовым методам математики и анализа данных. Студенты научатся проводить анализ данных любого масштаба, разрабатывать программное обеспечение, необходимое для анализа данных и представления результатов.

#### Модули

- Data Mining. Нейронные и байесовские сети. Деревья решений.
- Кластеризация и классификация.
- Машинное обучение и искусственный интеллект.
- Оптимизация. Локальный поиск. Генетические алгоритмы. Оптимизация роя.
- Статистический анализ данных.
- Методы извлечения знаний.
- Облачные вычисления.
- Визуализация и прогнозирование данных.
- Научные и бизнес-приложения Больших данных.

- Большие базы данных и NoSQL, включая MongoDB, Cassandra и Neo4j.
- Кластерные вычисления с Hadoop, Spark, Hive и MapReduce.
- Веб-аналитика и анализ социальных сетей. Сентимент анализ.
- Анализ изображений и видео.
- Семантический анализ.
- Машинный перевод.
- Семинары и практикумы.
- Магистерская диссертация.