

Аннотация к рабочей программе дисциплины «Наука о данных и аналитика больших объёмов данных»

Учебный курс «Наука о данных и аналитика больших объёмов данных» предназначен для студентов направления подготовки 38.04.01 Экономика, магистерская программа «Транспорт и пространственная логистика».

Дисциплина «Наука о данных и аналитика больших объёмов данных» включена в состав вариативной части блока «Дисциплины (модули)».

Общая трудоёмкость освоения дисциплины составляет 3 зачётных единицы, 108 часов. Учебным планом предусмотрены лекционные занятия (8 часов), самостоятельная работа студентов (100 часов). Дисциплина реализуется на 1 курсе в 1 семестре.

Изучение дисциплины «Наука о данных и аналитика больших объёмов данных» базируется на знаниях, умениях и навыках, полученных при освоении дисциплин «Микроэкономика (продвинутый уровень)», «Эконометрика 1», «Экономика и управление: адаптационный курс» и позволяет подготовить студентов к освоению ряда таких дисциплин, как «Макроэкономика», «Методы исследования», «Институциональная экономика».

Содержание дисциплины охватывает следующий круг вопросов:

1. Введение в большие данные: Определение больших данных и причины их появления. Примеры возможностей для бизнеса. Различие между Business Intelligence и Big Data

2. Жизненный цикл аналитики данных: Понятие жизненного цикла аналитики данных. Роли, необходимые для успешного создания проекта по аналитике данных

3. Высокопроизводительные вычисления: Распределенные вычисления на нескольких серверах, вычислительная парадигма MapReduce. Проект Apache Hadoop и его экосистема. Apache Spark и его компоненты. Вычисления в реальном времени, Apache Storm, Flink

4. Масштабирование и многоуровневое хранение данных: Теорема CAP. Парадигма NoSQL. Классификация NoSQL баз данных

5. Визуализация данных и результатов анализа: Техники визуализации данных, введение в язык R. Визуализация данных в R

6. Сложные методы аналитики: Классификация задач анализа: Text, Data, Web, Social Mining. Применение машинного обучения в аналитике. K-means и C-means кластеризация, классификация. Логистическая регрессия, ассоциации, алгоритм Априори.

7. Анализ текста: Поисковые механизмы: Lucene, Solr, ElasticSearch. Алгоритмы Word2Vec и Glove.

Цель - расширить и углубить знания студентов по следующим вопросам: жизненный цикл аналитики данных, технологии и средства распределенной обработки и хранения данных, базовые методы аналитики больших объемов данных, техники визуализации данных; использование типовых технологии и средств аналитики данных, таких как MapReduce, Hadoop, NoSQL, язык R; использование и внедрение современных технологий и инструментальных средств анализа, и работы с большими объемами данных.

Задачи:

- ознакомление с современными технологиями обработки и анализа данных;
- формирование навыков извлечения закономерностей из данных и обработки неструктурированных данных;
- выработка навыков выполнения работ по созданию (модификации) и сопровождению информационных систем, автоматизирующих задачи организационного управления и бизнес-процессы;
- формирование способности анализировать потребности заинтересованных лиц и подразделений организации и подходы к исследованию больших данных;

- формирование навыков разработки продуктов на основе встроенной аналитики больших данных.

Для успешного изучения дисциплины «Наука о данных и аналитика больших объёмов данных» у обучающихся по программе должны быть сформированы следующие предварительные компетенции:

- умение быстро осваивать новые предметные области, выявлять противоречия, проблемы и вырабатывать альтернативные варианты их решения;

- способность генерировать идеи в научной и профессиональной деятельности;

- способность принимать организационно-управленческие решения.

В результате изучения данной дисциплины у обучающихся формируются следующие профессиональные компетенции (элементы компетенций):

Код и формулировка компетенции	Этапы формирования компетенции	
ПК-10 способность готовить аналитические материалы для оценки мероприятий в области экономической политики и принятия стратегических решений на микро- и макроуровне	Знает	основы визуализации данных и результатов анализа: техники визуализации данных, введение в язык R, визуализация данных в R
	Умеет	применять техники визуализации данных, в том числе визуализация данных в R
	Владеет	навыками подготовки аналитических материалов с применением техники визуализации данных, в том числе визуализация данных в R для оценки мероприятий в области экономической политики и принятия стратегических решений на микро- и макроуровне
ПК-11 способность анализировать и использовать различные источники информации для проведения экономических расчетов	Знает	современные технологии обработки и анализа данных, используемые для исследования социально-экономических процессов, сравнительного анализа и эффективности государственного финансового контроля
	Умеет	использовать современные технологии обработки и анализа данных, используемые для исследования социально-экономических процессов, сравнительного анализа и эффективности государственного финансового контроля
	Владеет	навыками применения современных технологий обработки и анализа данных, используемых для

		исследования социально-экономических процессов, сравнительного анализа и эффективности государственного финансового контроля
ПК-13 способность использовать современные методы и инструменты исследования социально-экономических процессов, сравнительного анализа национальных моделей экономики	Знает	типовые технологии и средства аналитики данных, такие как MapReduce, Hadoop, NoSQL, Поисковые механизмы: Lucene, Solr, ElasticSearch. Алгоритмы Work2Vec и Glove.
	Умеет	использовать типовые технологии и средства аналитики данных, такие как MapReduce, Hadoop, NoSQL, Поисковые механизмы: Lucene, Solr, ElasticSearch. Алгоритмы Work2Vec и Glove.
	Владеет	навыками применения и обоснования использования типовых технологии и средств аналитики данных, таких как MapReduce, Hadoop, NoSQL, Поисковые механизмы: Lucene, Solr, ElasticSearch. Алгоритмы Work2Vec и Glove.

Для формирования вышеуказанных компетенций в рамках дисциплины «Наука о данных и аналитика больших объёмов данных» применяются следующие методы активного / интерактивного обучения: методы онлайн-обучения с использованием видеолекций, промежуточного тестирования и прокторинга.