

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

Учебный курс «Наука о данных и аналитика больших данных» предназначен для студентов направления подготовки 38.04.01 Экономика, магистерская программа «Государственный финансовый контроль».

Дисциплина «Наука о данных и аналитика больших 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, 2 семестре.

Изучение дисциплины «Наука о данных и аналитика больших данных» базируется на знаниях и навыках, полученных при освоении дисциплин «Микроэкономика», «Эконометрика», «Информационные технологии в профессиональной деятельности», «Экономика и управление: адаптационный курс» и позволяет подготовить студентов к освоению ряда таких дисциплин, как «Макроэкономика», «Методы исследования», «Финансовый мониторинг теневых экономических процессов».

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 технологий 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 курса:

1. Ознакомление с современными технологиями обработки и анализа данных;

2. Формирование навыков извлечения закономерностей из данных и обработки неструктурированных данных;

3. Выработка навыков выполнения работ по созданию (модификации) и сопровождению информационных систем, автоматизирующих задачи организационного управления и бизнес-процессы;

4. Формирование способности анализировать потребности заинтересованных лиц и подразделений организации и подходы к исследованию больших данных;

5. Формирование навыков разработки продуктов на основе встроенной аналитики больших данных.

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

- способность использовать основы экономических знаний в различных сферах деятельности;
- способность к самоорганизации и самообразованию;
- способность осуществлять сбор, анализ и обработку данных, необходимых для решения профессиональных задач;
- способность на основе типовых методик и действующей нормативно-правовой базы рассчитать экономические и социально-экономические показатели, характеризующие деятельность хозяйствующих субъектов;
- способность критически оценить предлагаемые варианты управленческих решений, разработать и обосновать предложения по их совершенствованию с учетом критериев социально-экономической эффективности, рисков и возможных социально-экономических последствий.

В результате изучения дисциплины у обучающихся формируются следующие профессиональные компетенции.

Код и формулировка компетенции	Этапы формирования компетенции	
ПК-8 -способность оценивать эффективность проектов с учетом фактора неопределенности	Знает	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. Масштабирование и многоуровневое хранение данных: Теорема CAP. Парадигма NoSQL. Классификация NoSQL баз данных
	Умеет	осуществлять высокопроизводительные вычисления масштабирование и многоуровневое хранение данных
	Владеет	навыками применения высокопроизводительных вычислений, масштабирования и многоуровневого хранения данных для оценки эффективности проектов с учетом фактора неопределенности в сфере

		государственного финансового контроля
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.
ПК-17 способность самостоятельно осуществлять разработку проектных решений с учетом фактора неопределенности, разработку	Знает	жизненный цикл аналитики данных, технологии и средства распределенной обработки и хранения данных, базовые методы аналитики больших объемов данных, техники визуализации данных для принятия рациональных экономических решений для осуществления функций государственного

соответствующих методических и нормативных документов, а также предложений и мероприятий по реализации разработанных проектов и программ		финансового контроля
	Умеет	критически оценивать применение технологии и средств распределенной обработки и хранения данных, базовые методы аналитики больших объемов данных, техники визуализации данных для принятия рациональных экономических решений для осуществления функций государственного финансового контроля
	Владеет	навыками критической оценки возможных экономических эффектов от технологии и средств распределенной обработки и хранения данных, базовые методы аналитики больших объемов данных, техники визуализации данных для принятия рациональных экономических решений осуществления функций государственного финансового контроля

Для формирования указанных компетенций в рамках дисциплины «Наука о данных и аналитика больших данных» применяются методы онлайн-обучения с использованием видеолекций, промежуточного тестирования и прокторинга.