

Документ подписан простой электронной подписью  
Информация о владельце:

ФИО: Исаев Игорь Магомедович

Должность: Проректор по учебной и научной работе

Дата подписания: 27.09.2023 15:23:36

Уникальный идентификатор документа:

d7a26b9e8ca85e98ec3de2eb454b4659d061f249

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Федеральное государственное автономное образовательное учреждение  
высшего образования**

**«Национальный исследовательский технологический университет «МИСИС»**

## Рабочая программа дисциплины (модуля)

# Технологии анализа данных и принятие решений

Закреплена за подразделением

Кафедра бизнес-информатики и систем управления производством

Направление подготовки

38.04.05 БИЗНЕС-ИНФОРМАТИКА

Профиль

Процессная аналитика в цифровой экономике

Квалификация

**Магистр бизнес-информатики**

Форма обучения

**очная**

Общая трудоемкость

**3 ЗЕТ**

Часов по учебному плану

108

Формы контроля в семестрах:

в том числе:

зачет с оценкой 1

аудиторные занятия

34

самостоятельная работа

74

### Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	1 (1.1)		Итого	
	УП	РП	УП	РП
Неделя	18			
Вид занятий	УП	РП	УП	РП
Лекции	17	17	17	17
Практические	17	17	17	17
Итого ауд.	34	34	34	34
Контактная работа	34	34	34	34
Сам. работа	74	74	74	74
Итого	108	108	108	108

Программу составил(и):

*ктн, Старший преподаватель, Котеленко Сергей Анатольевич*

Рабочая программа

**Технологии анализа данных и принятие решений**

Разработана в соответствии с ОС ВО:

Самостоятельно устанавливаемый образовательный стандарт высшего образования - магистратура Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский технологический университет «МИСИС» по направлению подготовки 38.04.05 БИЗНЕС-ИНФОРМАТИКА (приказ от 28.06.2023 г. № 292 о.в.)

Составлена на основании учебного плана:

38.04.05 БИЗНЕС-ИНФОРМАТИКА, 38.04.05-МБИ-23-2-ПП.plx Процессная аналитика в цифровой экономике, утвержденного Ученым советом НИТУ МИСИС в составе соответствующей ОПОП ВО 22.06.2023, протокол № 5-23

Утверждена в составе ОПОП ВО:

38.04.05 БИЗНЕС-ИНФОРМАТИКА, Процессная аналитика в цифровой экономике, утвержденной Ученым советом НИТУ МИСИС 22.06.2023, протокол № 5-23

Рабочая программа одобрена на заседании

**Кафедра бизнес-информатики и систем управления производством**

Протокол от 23.06.2020 г., №22

Руководитель подразделения д.т.н., доцент, Пятецкий Валерий Ефимович

**1. ЦЕЛИ ОСВОЕНИЯ**

1.1	Формирование знаний в области принятия управленческих решений на базе результатов, полученных посредством обработки аналитической информации современными автоматизированными информационными системами.
1.2	Изучение сущности и содержания следующих категорий: информация, анализ, информационное пространство, информационное хранилище, оперативный и интеллектуальный анализ данных.
1.3	Формирование у обучающихся целостного представления и основных теоретических знаний о назначении, принципах построения и использования информационно-аналитических систем; знакомство с идеями, теоретическими основами, программной реализацией и получение практических навыков в применении информационно-аналитических технологий для поддержки принятия бизнес-решений в процессе управления компанией.

**2. МЕСТО В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ**

Блок ОП:		Б1.О
2.1	<b>Требования к предварительной подготовке обучающегося:</b>	
2.2	<b>Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:</b>	
2.2.1	Научно-исследовательская работа	
2.2.2	Подготовка к процедуре защиты и защита выпускной квалификационной работы	

**3. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ФОРМИРУЕМЫМИ КОМПЕТЕНЦИЯМИ**

<b>ОПК-2: Способен принимать решения, осуществлять моделирование, анализ и эксперименты в целях проведения детального исследования для решения сложных задач в профессиональной области</b>	
<b>Знать:</b>	
ОПК-2-31 современное состояние и тенденции в области развития анализа больших данных;	
<b>УК-1: Способен использовать проектный метод в профессиональной области: разрабатывать систему взаимосвязанных целей; создавать программы их достижения, представляющие собой комплекс научно-исследовательских, опытно-конструкторских, производственных, организационных, финансовых, коммерческих и других мероприятий; организовывать мониторинг выполнения программ; определять достижение целевых показателей</b>	
<b>Знать:</b>	
УК-1-31 принципы визуального программирования в области машинного обучения	
<b>ОПК-2: Способен принимать решения, осуществлять моделирование, анализ и эксперименты в целях проведения детального исследования для решения сложных задач в профессиональной области</b>	
<b>Уметь:</b>	
ОПК-2-У1 применять визуализировать данных, строить дашборды	
<b>УК-1: Способен использовать проектный метод в профессиональной области: разрабатывать систему взаимосвязанных целей; создавать программы их достижения, представляющие собой комплекс научно-исследовательских, опытно-конструкторских, производственных, организационных, финансовых, коммерческих и других мероприятий; организовывать мониторинг выполнения программ; определять достижение целевых показателей</b>	
<b>Уметь:</b>	
УК-1-У1 строить модели анализа данных в системах управления предприятием	
<b>ОПК-2: Способен принимать решения, осуществлять моделирование, анализ и эксперименты в целях проведения детального исследования для решения сложных задач в профессиональной области</b>	
<b>Владеть:</b>	
ОПК-2-В1 навыками проведения анализа в бизнесе, в том числе посредством современных информационных технологий: методами анализа, прогноза и оценки развития систем управления предприятия	
<b>УК-1: Способен использовать проектный метод в профессиональной области: разрабатывать систему взаимосвязанных целей; создавать программы их достижения, представляющие собой комплекс научно-исследовательских, опытно-конструкторских, производственных, организационных, финансовых, коммерческих и других мероприятий; организовывать мониторинг выполнения программ; определять достижение целевых показателей</b>	
<b>Владеть:</b>	
УК-1-В1 методологией построения процессных моделей анализа данных в системах управления производством	

4. СТРУКТУРА И СОДЕРЖАНИЕ								
Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Формируемые индикаторы компетенций	Литература и эл. ресурсы	Примечание	КМ	Выполняемые работы
	<b>Раздел 1. Бизнес-аналитика в бизнес-процессах предприятия. Основной процесс анализа данных в Data Science</b>							
1.1	Процесс анализа данных. DATA SCIENCE. Типы данных. /Лек/	1	2	УК-1-31 ОПК-2-31	Л1.1 Л1.3 Л1.5 Л1.6Л2.1 Л2.2		КМ1	
1.2	Технологии сбора и хранения данных - Data Warehouse, NOSQL, OLAP, Графовые БД /Лек/	1	2	УК-1-31 ОПК-2-31	Л1.1 Л1.2 Л1.3Л2.1 Л2.2		КМ1	
1.3	Динамическая визуализация в DATA SCIENCE. QLIK-SENSE /Пр/	1	6	ОПК-2-У1 ОПК-2-В1	Л1.1 Л1.3 Л1.5Л2.1 Л2.2			Р1
1.4	Основные типы категорий визуализации в BIG DATA /Лек/	1	2	УК-1-31 ОПК-2-31	Л1.1 Л1.3 Л1.4 Л1.6Л2.1 Л2.2		КМ1	
1.5	Изучение платформы QLIK-SENSE /Ср/	1	10	ОПК-2-У1	Л1.1 Л1.3Л2.1 Л2.2		КМ1	
	<b>Раздел 2. Технологии сбора и хранения данных – Data Warehouse и Data Lake.</b>							
2.1	Выполнение домашней работы /Ср/	1	15	УК-1-У1 УК-1-В1	Л1.1 Л1.2 Л1.3			Р5
2.2	Модели и методы Data Mining, Machine Learning /Лек/	1	4	УК-1-31 ОПК-2-31	Л1.1 Л1.2 Л1.3Л2.1 Л2.2		КМ1	
2.3	Data Lake. Архитектура и построение современных "озер" данных /Лек/	1	2	УК-1-31 ОПК-2-31	Л1.1 Л1.2 Л1.3 Л1.4		КМ1	
	<b>Раздел 3. Содержание и методы анализа и прогнозирования бизнес-процессов (деятельности предприятий) как объект автоматизации.</b>							
3.1	Выполнение домашней работы /Ср/	1	15	УК-1-У1 УК-1-В1	Л1.1 Л1.3Л2.3			Р5
3.2	Подготовка реферата /Ср/	1	10	УК-1-31 ОПК-2-31	Л1.1 Л1.2 Л1.3			Р4
3.3	Методы и технологии BIG DATA в телекоме. Лучшие практики /Лек/	1	2	УК-1-31 ОПК-2-31	Л1.1 Л1.3 Л1.4Л2.1 Л2.2		КМ1	
3.4	Задача классификации и прогноза. Orange. /Пр/	1	6	ОПК-2-У1 ОПК-2-В1	Л1.2Л2.1 Л2.2			Р2
	<b>Раздел 4. Методы и технологии BIG DATA, лучшие практики. Методы и техники обработки и хранения данных Hadoop.</b>							

4.1	Методы и техники обработки и хранения данных Hadoop /Пр/	1	5	УК-1-У1 УК-1-В1	Л1.1 Л1.3 Л1.4Л2.1 Л2.2			Р3
4.2	Подготовка к тестированию /Ср/	1	10	УК-1-31 ОПК-2-31	Л1.1 Л1.3 Л1.4Л2.1 Л2.2		КМ2	
4.3	Изучение вопросов коллоквиума по теме /Ср/	1	10	УК-1-31 ОПК-2-31	Л1.1 Л1.3 Л1.4Л2.1 Л2.2		КМ2	
4.4	Методы и техники обработки и хранения данных Hadoop /Лек/	1	3	УК-1-31 ОПК-2-31	Л1.1 Л1.3 Л1.4Л2.1 Л2.2		КМ1	
4.5	Проведение зачета с оценкой /ЗачётСОц/	1	4	УК-1-31 ОПК-2-31	Л1.1 Л1.2 Л1.3 Л1.4 Л1.5 Л1.6Л2.1 Л2.2 Л2.3		КМ2	

### 5. ФОНД ОЦЕНОЧНЫХ МАТЕРИАЛОВ

#### 5.1. Контрольные мероприятия (контрольная работа, тест, коллоквиум, экзамен и т.п), вопросы для самостоятельной подготовки

Код КМ	Контрольное мероприятие	Проверяемые индикаторы компетенций	Вопросы для подготовки
КМ1	Тестирование		Прохождение тестирования на платформе LMS Canvas

KM2	Вопросы к коллоквиуму	<p>Применение технологий больших данных в телекоме. Основные кейсы</p> <p>Основные тенденции развития технологий по Gartner Group (отчет 2018)</p> <p>Появление технологий больших данных на рынке. Предпосылки и основные игроки</p> <p>Apache Hadoop: архитектура, основные компоненты, назначение</p> <p>Преимущества и недостатки hadoop</p> <p>Преимущества и недостатки MapReduce</p> <p>YARN как инструмент управления ресурсами кластера.</p> <p>Архитектура</p> <p>Системы десериализации данных. Apache AVRO</p> <p>Основные преимущества использования AVRO</p> <p>NoSQL key-value columnar DB. Основные характеристики и назначение</p> <p>Hbase: назначение и архитектура</p> <p>Cassandra: назначение и архитектура</p> <p>Accumulo: назначение и архитектура</p> <p>CAP-теорема</p> <p>Workflow management для больших данных. Основные инструменты</p> <p>Apache Oozie: назначение, примеры, интеграционные сервисы</p> <p>Apache Spark: назначение, архитектура</p> <p>Apache NIFI: назначение, архитектура</p> <p>Передача данных между системами SQL и NoSQL. Apache Sqoop</p> <p>Apache Hive: назначение, архитектура</p> <p>Системы обмена сообщениями. Apache Kafka</p> <p>Управление конфигурациями сервисов. Zookeeper</p> <p>Форматы хранения данных в Hadoop. Parquet, ORC, ORCfile, avro.</p> <p>Lambda-архитектура. Основные характеристики</p> <p>Карра-архитектура. Основные характеристики</p> <p>Data lake: назначение и основные системы</p> <p>Слой данных в Data lake</p> <p>Хранилища данных. Основные отличия от data lake</p> <p>Виды взаимодействий Data lake и хранилищ данных</p> <p>Основные NoSQL базы данных</p> <p>Структурированные данные</p> <p>Неструктурированные данные</p> <p>Машинные данные</p> <p>Естественные языки</p> <p>Данные, генерируемые компьютером</p> <p>Графовые данные ( семантические сети, онтологии, RDF)</p> <p>Аудио, видео и изображения</p> <p>Потоковые данные (streaming)</p> <p>Социальные сети</p> <p>Формирование целей исследования</p> <p>Извлечение данных</p> <p>Подготовка данных</p> <p>Анализ данных (EDA)</p> <p>Моделирование данных</p> <p>Представление и автоматизация</p> <p>Основные отличия оперативных и аналитических систем</p> <p>Визуализация данных</p> <p>Временные диаграммы</p> <p>Иерархические диаграммы</p> <p>Сетевые диаграммы</p> <p>Многомерные диаграммы</p> <p>Гео-пространственные диаграммы</p> <p>Столбиковая диаграмма (Bat Chart)</p> <p>Линейные диаграммы</p> <p>Точечная диаграмма (Scatter Plot)</p> <p>Sparkline</p> <p>Круговая диаграмма</p> <p>Индикатор (Gauge)</p> <p>Диаграмма водопад (Waterfall Chart)</p> <p>Диаграмма воронка (Funnel Chart)</p> <p>Heat Map</p> <p>Гистограмма</p>
-----	-----------------------	--

			<p>Box Plot  Карты  Таблицы  Индикаторы  Area chart  Радарная диаграмма  Диаграмма Дерево  Задачи и области машинного обучения  Модель машинного обучения  Различие между алгоритмом машинного обучения и моделью машинного обучения  Виды машинного обучения  Алгоритмы обучения, используемые в обучении с учителем  Контролируемое обучение  Алгоритмы обучения, используемые в самообучении  Обучение с подкреплением  Цель модели линейной регрессии  Допущения метода линейной регрессии  Обучение линейной модели  Градиентный спуск  Оценка качества модели  Пример задач, решаемых линейной регрессией  Архитектура аналитической платформы  Основные различия между ХД и транзакционными БД  Хранилища данных. Основные свойства ХД. Витрины данных  Данные и метаданные. Типы и функции Метаданных.  Структуры хранилищ данных: схемы «Звезда», «Снежинка»  Многомерная модель данных: измерения, факты, Олар-кубы  Агрегирование и консолидация данных. Очистка данных  Требования к OLAP-системам. Операции с данными в OLAP  Виды OLAP. Преимущества и недостатки различных видов OLAP  NumPy  Matplotlib  Pandas  Ski-kit learn  Примеры задач для логистической регрессии  Типы логистической регрессии  Сигмоидная функция  Теорема Байеса  Наивный байесовский классификатор  Простой Байесовский метод  Структура дерева принятия решений  Пример использования метода ДПР для двух классов  Визуальное представление дерева решений  Построение дерева решений  Пример расчета энтропии и прироста информации  Карра-архитектура. Основные характеристики  Data lake: назначение и основные системы  Слой данных в Data lake  Хранилища данных. Основные отличия от data lake  Виды взаимодействий Data lake и хранилищ данных  Основные NoSQL базы данных</p>
--	--	--	--

**5.2. Перечень работ, выполняемых по дисциплине (Курсовая работа, Курсовой проект, РГР, Реферат, ЛР, ПР и т.п.)**

Код работы	Название работы	Проверяемые индикаторы компетенций	Содержание работы
P1	Практическая работа №1		Динамическая визуализация в DATA SCIENCE. QLIK-SENSE
P2	Практическая работа №2		Реализовать задачу классификации и прогноза в системе Orange
P3	Практическая работа №3		Методы и техники обработки и хранения данных Hadoop

P4	Реферат		Сформировать реферат на одну из следующих тем: 1 HDFS VS Amazon S3 VS Azure Blob 2 Scylla db 3 Greenplum 4 Viktoria Metrics 5 Grafana VS Kibana 6 Tensorflow 7 NiFi VS Flume VS Storm 8 ActiveMQ 9 Spark VS Flink VS Samza VS Storm 10 Accumulo VS HBASE VS Cassandra 11 Avro VS Thrift VS Protobuf 12 ORC VS Parquet 13 Ignite VS Hazelcast VS Redis 14 Oozie VS Airflow VS Azkaban 15 Vertica VS Cassandra 16 Aerospike 17 Keepalived 18 Apache Ranger VS Knox 19 Spark RDD VS Spark DataFrame VS SparkDataset 20 ZooKeeper VS Consul 21 SparkSQL & SparkStreaming 22 Kafka VS RabbitMQ 23 Neo4J 24 Tableau VS QLIK VS Power BI 25 Kafka VS RabbitMQ 26 Elasticsearch VS TimescaleDB VS Prometheus 27 Tarantool VS Clickhouse 28 Distributed file systems 29 Conteinirization platforms
P5	Домашняя работа		Проанализировать данные и выполнить самостоятельное задание согласно варианту: Задание № 1: Наиболее важные факторы, связанные с увольнением и эффективностью сотрудников Задание № 2: Международный рейтинг университетов Задание № 3: Исследование факторов, влияющих на продолжительность жизни Задание № 4: Анализ факторов, влияющих на прекращение абонентом контракта с оператором телефонной связи Задание № 5: Анализ доходов ведущих спортсменов мира

### 5.3. Оценочные материалы, используемые для экзамена (описание билетов, тестов и т.п.)

Экзаменационный билет состоит из двух теоретических вопросов и одной задачи. Билеты хранятся на кафедре.

### 5.4. Методика оценки освоения дисциплины (модуля, практики. НИР)

Для получения оценки "отлично" обучающемуся необходимо полностью ответить на теоретические вопросы и без ошибок выполнить задачу

Для получения оценки "хорошо" обучающемуся необходимо частично ответить на теоретические вопросы и без ошибок выполнить задачу или полностью ответить на теоретические вопросы и с недочетами выполнить задачу

Для получения оценки "удовлетворительно" обучающемуся необходимо частично ответить на теоретические вопросы и с недочетами выполнить задачу

## 6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

### 6.1. Рекомендуемая литература

#### 6.1.1. Основная литература

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л1.1	Мхитарян С. В.	Бизнес-аналитика в менеджменте: практикум	Электронная библиотека	Москва: Евразийский открытый институт, 2011
Л1.2	Чубукова И. А.	Data Mining: учебное пособие	Электронная библиотека	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ)  Бином. Лаборатория знаний, 2008

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л1.3	Мыльников Л. А., Краузе Б., Кютц М., Бадэ К., Шмидт И. А.	Интеллектуальный анализ данных в управлении производственными системами (подходы и методы): монография	Электронная библиотека	Москва: Библио-Глобус, 2017
Л1.4	Бутаков Н. А., Петров М. В., Насонов Д.	Обработка больших данных с Apache Spark: учебно-методическое пособие	Электронная библиотека	Санкт-Петербург: Университет ИТМО, 2019
Л1.5	Алпайдин Э.	Машинное обучение: новый искусственный интеллект: пер. с англ.	Библиотека МИСиС	М.: Альпина Паблишер, 2017
Л1.6	Адлер Ю. П., Черных Е. А.	Статистическое управление процессами. "Большие данные" (N 2909): учеб. пособие	Электронная библиотека	М.: [МИСиС], 2016

### 6.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л2.1	Гудов А. М., Завозкин С. Ю., Рейн Т. С.	Базы данных и системы управления базами данных. Программирование на языке PL/SQL: учебное пособие	Электронная библиотека	Кемерово: Кемеровский государственный университет, 2010
Л2.2	Чурбанова О. В., Чурбанов А. Л.	Базы данных и знаний. Проектирование баз данных в Microsoft Access: учебно-методическое пособие	Электронная библиотека	Архангельск: Северный (Арктический) федеральный университет (САФУ), 2015
Л2.3	Шульц О.	Практикум по SAP. Руководство для новичков и конечных пользователей: пер. с англ.	Библиотека МИСиС	СПб.: Эксперт РП, 2015

### 6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

Э1	38.03.05 Информационные аналитические системы управления предприятием	<a href="https://lms.misis.ru/enroll/T77HWD">https://lms.misis.ru/enroll/T77HWD</a>
----	---	---

### 6.3 Перечень программного обеспечения

П.1	LMS Canvas
П.2	MS Teams
П.3	Microsoft Office
П.4	SAP (удаленный доступ предоставляет заказчик - IBS)
П.5	Hadoop
П.6	Python
П.7	Loginom
П.8	Orange 3.23.1
П.9	Qlik Sense Desktop 2019

### 6.4. Перечень информационных справочных систем и профессиональных баз данных

И.1	<a href="https://data.gov.ru">https://data.gov.ru</a> – Портал открытых данных РФ
И.2	<a href="http://www.gks.ru">http://www.gks.ru</a> – Росстат – Федеральная служба государственной статистики
И.3	<a href="http://www.economy.gov.ru">www.economy.gov.ru</a> – Базы данных Министерства экономического развития и торговли России
И.4	<a href="https://archive.ics.uci.edu/ml">https://archive.ics.uci.edu/ml</a> – Репозиторий наборов данных для машинного обучения Irvine Machine Learning Repository
И.5	<a href="https://www.kaggle.com/datasets">https://www.kaggle.com/datasets</a> – Платформа для обмена пользователей наборами данных для задач машинного обучения.
И.6	<a href="https://toolbox.google.com/datasetsearch">https://toolbox.google.com/datasetsearch</a> – Сервис поиска наборов данных от Google.
И.7	Инструментальные средства для проектирования: Loginom (Community Edition, Standard)

## 7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ

Ауд.	Назначение	Оснащение
------	------------	-----------

Б-434	Компьютерный класс	персональные компьютеры - 80 шт., пакет лицензионных программ MS Office, проектор, комплект учебной мебели
Б-1135	Компьютерный класс	персональные компьютеры - 30 шт., пакет лицензионных программ MS Office, проектор, комплект учебной мебели

### 8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ

Лекционные материалы и методические рекомендации к работам размещаются в начале семестра в LMS Canvas.  
Рекомендуемая в курсе литература доступна в Электронной библиотеке НИТУ "МИСиС".  
Электронные версии методических указаний находятся на кафедре