

Документ подписан простой электронной подписью
Информация о владельце:

ФИО: Исаев Игорь Магомедович

Должность: Проректор по безопасности и общим вопросам

Дата подписания: 30.01.2023 16:41:18

Уникальный программный ключ:

d7a26b9e8ca85e98ac3de2ab454b4659d961f749

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 машинное обучение

Закреплена за подразделением

Кафедра бизнес-информатики и систем управления производством

Направление подготовки

09.04.03 ПРИКЛАДНАЯ ИНФОРМАТИКА

Профиль

Прикладная информатика в цифровой экономике

Квалификация

Магистр

Форма обучения

очная

Общая трудоемкость

5 ЗЕТ

Часов по учебному плану

180

Формы контроля в семестрах:

в том числе:

экзамен 3

аудиторные занятия

34

курсовая работа 3

самостоятельная работа

110

часов на контроль

36

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	3 (2.1)		Итого	
	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	110	110	110	110
Часы на контроль	36	36	36	36
Итого	180	180	180	180

Программу составил(и):

ктн, Старший преподаватель, Котеленко Сергей Анатольевич; Ассистент, Кузнецов Денис Сергеевич

Рабочая программа

Технологии анализа данных и машинное обучение

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 09.04.03 ПРИКЛАДНАЯ ИНФОРМАТИКА (приказ от 05.03.2020 г. № 95 о.в.)

Составлена на основании учебного плана:

09.04.03 Прикладная информатика, 09.04.03-МПИ-22-2.plx Прикладная информатика в цифровой экономике, утвержденного Ученым советом ФГАОУ ВО НИТУ "МИСиС" в составе соответствующей ОПОП ВО 22.09.2022, протокол № 8-22

Утверждена в составе ОПОП ВО:

09.04.03 Прикладная информатика, Прикладная информатика в цифровой экономике, утвержденной Ученым советом ФГАОУ ВО НИТУ "МИСиС" 22.09.2022, протокол № 8-22

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.В
2.1	Требования к предварительной подготовке обучающегося:	
2.1.1	Технологии BIG DATA	
2.1.2	Управление информационной безопасностью	
2.1.3	Цифровые модели и технологии в управлении материалопотоком (SCM)	
2.1.4	Организация и технология научных исследований и педагогической деятельности	
2.1.5	Системно-архитектурные решения в корпоративном управлении	
2.2	Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:	
2.2.1	Подготовка к процедуре защиты и защита выпускной квалификационной работы	
2.2.2	Преддипломная практика	

3. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ФОРМИРУЕМЫМИ КОМПЕТЕНЦИЯМИ

ПК-4: Способен обеспечить процесс проектирования и дизайна ИС принятым в организации стандартам и технологиям, обеспечить эффективное распределение ресурсов и контроль за их использованием	
Знать:	
ПК-4-31 - принципы визуального программирования в области машинного обучения	
ОПК-3: Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	
Знать:	
ОПК-3-31 -современное состояние и тенденции в области развития анализа больших данных;	
ПК-4: Способен обеспечить процесс проектирования и дизайна ИС принятым в организации стандартам и технологиям, обеспечить эффективное распределение ресурсов и контроль за их использованием	
Владеть:	
ПК-4-В1 - методологией построения процессных моделей анализа данных в системах управления производством	
ОПК-3: Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	
Владеть:	
ОПК-3-В1 - навыками проведения анализа в бизнесе, в том числе посредством современных информационных технологий: методами анализа, прогноза и оценки развития систем управления предприятия	

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
	Раздел 1. Бизнес-аналитика в бизнес-процессах предприятия.Основной процесс анализа данных в Data Science							

1.1	Процесс анализа данных. DATA SCIENCE. Типы данных /Лек/	3	2	ОПК-3-31 ПК-4-31	Л1.1 Л1.2 Л1.4 Л1.5Л2.1 Л2.3		КМ1	
1.2	Технологии сбора и хранения данных - Data Warehouse, NOSQL, OLAP, Графовые БД /Лек/	3	2	ОПК-3-31 ПК-4-31	Л1.1 Л1.1 Л1.2Л2.1 Л2.3		КМ1	
1.3	Динамическая визуализация в DATA SCIENCE. QLIK-SENSE /Пр/	3	6	ОПК-3-В1 ПК-4-В1	Л1.1 Л1.2 Л1.4Л2.1 Л2.3			Р1
1.4	Основные типы категорий визуализации в BIG DATA /Лек/	3	2	ОПК-3-31	Л1.1 Л1.2 Л1.3 Л1.5Л2.1 Л2.3		КМ1	
1.5	Изучение платформы QLIK-SENSE /Ср/	3	10	ОПК-3-31 ОПК-3-В1 ПК-4-В1	Л1.1 Л1.2Л2.1 Л2.3		КМ1	
	Раздел 2. Технологии сбора и хранения данных – Data Warehouse и Data Lake.							
2.1	Выполнение домашней работы /Ср/	3	15	ОПК-3-31 ОПК-3-В1 ПК-4-В1	Л1.1 Л1.2Л1.1			Р5
2.2	Модели и методы Data Mining, Machine Learning /Лек/	3	4	ОПК-3-31 ПК-4-31	Л1.1 Л1.1 Л1.2Л2.1 Л2.3		КМ1	
2.3	Data Lake. Архитектура и построение современных "озер" данных /Лек/	3	2	ОПК-3-31 ПК-4-31	Л1.1 Л1.2 Л1.3Л1.1		КМ1	
	Раздел 3. Содержание и методы анализа и прогнозирования бизнес-процессов (деятельности предприятий) как объект автоматизации.							
3.1	Выполнение домашней работы /Ср/	3	15	ОПК-3-В1 ПК-4-В1	Л1.1 Л1.2Л2.4			Р5
3.2	Подготовка реферата /Ср/	3	20	ОПК-3-31 ОПК-3-В1 ПК-4-В1	Л1.1 Л1.2Л1.1			Р4
3.3	Методы и технологии BIG DATA в телекоме. Лучшие практики /Лек/	3	2	ОПК-3-31 ПК-4-31	Л1.1 Л1.2 Л1.3Л2.1 Л2.3		КМ1	
3.4	Задача классификации и прогноза. Orange. /Пр/	3	6	ОПК-3-В1 ПК-4-В1	Л1.1Л2.1 Л2.3			Р2
	Раздел 4. Методы и технологии BIG DATA, лучшие практики. Методы и техники обработки и хранения данных Hadoop.							
4.1	Методы и техники обработки и хранения данных Hadoop /Пр/	3	5	ОПК-3-В1 ПК-4-В1	Л1.1 Л1.2 Л1.3Л2.1 Л2.3			Р3
4.2	Подготовка к тестированию /Ср/	3	20	ОПК-3-31 ОПК-3-В1 ПК-4-В1	Л1.1 Л1.2 Л1.3Л2.1 Л2.3		КМ2	
4.3	Изучение вопросов коллоквиума по теме /Ср/	3	30	ОПК-3-31 ОПК-3-В1 ПК-4-В1	Л1.1 Л1.2 Л1.3Л2.1 Л2.3		КМ2	

4.4	Методы и техники обработки и хранения данных Hadoop /Лек/	3	3	ОПК-3-31 ПК-4-31	Л1.1 Л1.2 Л1.3Л2.1 Л2.3		КМ1	
-----	---	---	---	------------------	-------------------------------	--	-----	--

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	ОПК-3-31	Прохождение тестирования на платформе LMS Canvas

KM2	Вопросы к коллоквиуму	ОПК-3-31	<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	ОПК-3-B1	Динамическая визуализация в DATA SCIENCE. QLIK-SENSE
P2	Практическая работа №2	ОПК-3-B1	Реализовать задачу классификации и прогноза в системе Orange
P3	Практическая работа №3	ОПК-3-B1	Методы и техники обработки и хранения данных Hadoop

P4	Реферат	ОПК-3-31	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	ОПК-3-В1	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	Мыльников Л. А., Краузе Б., Кютц М., Баде К., Шмидт И. А.	Интеллектуальный анализ данных в управлении производственными системами (подходы и методы): монография	Электронная библиотека	Москва: Библио-Глобус, 2017

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л1.3	Бутаков Н. А., Петров М. В., Насонов Д.	Обработка больших данных с Apache Spark: учебно- методическое пособие	Электронная библиотека	Санкт-Петербург: Университет ИТМО, 2019
Л1.4	Алпайдин Э.	Машинное обучение: новый искусственный интеллект: пер. с англ.	Библиотека МИСиС	М.: Альпина Паблишер, 2017
Л1.5	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	Чубукова И. А.	Data Mining: учебное пособие	Электронная библиотека	Москва: Интернет- Университет Информационных Технологий (ИНТУИТ) Бином. Лаборатория знаний, 2008
Л2.3	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
Л2.4	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	https://lms.misis.ru/enroll/T77HWD
----	--	---

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	https://data.gov.ru – Портал открытых данных РФ
И.2	http://www.gks.ru – Росстат – Федеральная служба государственной статистики
И.3	www.economy.gov.ru – Базы данных Министерства экономического развития и торговли России
И.4	https://archive.ics.uci.edu/ml – Репозиторий наборов данных для машинного обучения Irvine Machine Learning Repository
И.5	https://www.kaggle.com/datasets – Платформа для обмена пользователей наборами данных для задач машинного обучения.
И.6	https://toolbox.google.com/datasetsearch – 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