Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Максимов Алексей Борисович
Должность: директор департамента по образовательной политике
Дата подписания: 24.10.2023 11:54:08
Уникальный программный ключ:
8db180d1a3f02ac9e60521a5672742735c18b1d6

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

Декан факультета
информационных технологий
/Д. Г. Демидов/

«28» 04 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Хранилища данных»

Направление подготовки
09.03.01 Информатика и вычислительная техника
Образовательная программа (профиль подготовки)
«Программное обеспечение информационных систем»

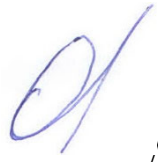
Квалификация (степень) выпускника
бакалавр

Форма обучения
заочная

Москва 2022

Программа дисциплины «Хранилища данных» составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению **09.03.01 Информатика и вычислительная техника** и профилю подготовки **«Программное обеспечение информационных систем».**

Программу составил



_____/А.В. Осипов/

Программа дисциплины утверждена на заседании кафедры «Прикладная информатика»
«____» августа 2022 г. протокол № _____

Заведующий кафедрой
доцент, к.э.н.



/С. В. Суворов/

Программа согласована с руководителем образовательной программы по направлению подготовки **09.03.01 Информатика и вычислительная техника** по профилю подготовки **«Программное обеспечение информационных систем».**



_____ /С. В. Суворов/
«____» августа 2022 г.

Программа утверждена на заседании учебно-методической комиссии факультета Информационных технологий

Председатель комиссии  _____ / Д. Г. Демидов/

«____»_____2022 г. Протокол:

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

К **основным целям** освоения дисциплины относится:

- расширенное формирование у студентов представления о принципах и методах машинного обучения;
- знакомство студентов с современными методами работы с большими данными;
- закрепление получаемых в семестре знаний и навыков на практике;
- формирование взаимосвязей, получаемых в семестре знаний и навыков с изученными ранее и изучаемых параллельно с данной дисциплиной;
- подготовка студентов к деятельности в соответствии с квалификационной характеристикой магистра.

К **основным задачам** дисциплины относятся:

- освоение методологии обработки больших данных;
- использование компьютерных технологий реализации методов машинного обучения;
- изучение и освоение теоретического материала, как в процессе контактной, так и в ходе самостоятельной работы;
- выполнение предоставленных практических заданий различных форм, как в процессе контактной, так и в ходе самостоятельной работы;
- самостоятельная работа над тематикой дисциплины для формирования компетенций основной образовательной программы (далее, ООП).

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП МАГИСТРАТУРЫ

Дисциплина «Хранилища данных» относится к вариативной части основной образовательной программы магистратуры.

Она взаимосвязана логически и содержательно-методически со следующими дисциплинами ООП:

- Современные проблемы прикладной математики и информатики;
- Математическое моделирование открытых данных;
- Непрерывные математические модели.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ), СООТНЕСЕННЫЕ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

В результате освоения дисциплины (модуля) у обучающихся формируются следующие компетенции и должны быть достигнуты следующие результаты обучения как этап формирования соответствующих компетенций:

Код компетенции	**Наименование компетенции**	**Индикаторы планируемых результатов обучения по дисциплине**
Категория (группа) «Общепрофессиональные компетенции»		
ОПК-3	способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять свое научное мировоззрение	*ОПК-3.1. Знать:* основные методы проектирования и эксплуатации хранилищ данных и технологий Big Data *ОПК-3.2. Уметь:* использовать Хранилища данных для решения научных задач *ОПК-3.3. Владеть:* методами проектирования и эксплуатации хранилищ данных и технологий Big Data для решения практических научных задач
Категория (группа) «Научно-исследовательская деятельность»		
ПК-2	способность разрабатывать и анализировать концептуальные и теоретические модели решаемых научных проблем и задач	*ПК-2.1. Знать:* концептуальные и теоретические подходы к построению хранилищ данных и использованию технологий Big Data *ПК-2.2. Уметь:* использовать методы проектирования и эксплуатации хранилищ данных и технологий Big Data для решения научных проблем и задач *ПК-2.3. Владеть:*

Код компетенции	Наименование компетенции	Индикаторы планируемых результатов обучения по дисциплине
		методами проектирования и эксплуатации хранилищ данных и технологий Big Data для проведения научных исследований
Категория (группа) «Организационно-управленческая деятельность»		
ПК-7	способность разрабатывать и оптимизировать бизнес-планы научно-прикладных проектов	*ПК-7.1. Знать:* концептуальные и теоретические подходы к разработке и оптимизации бизнес-планов научно-прикладных проектов *ПК-7.2. Уметь:* разрабатывать и оптимизировать бизнес-планы научно-прикладных проектов *ПК-7.3. Владеть:* методами разработки и оптимизации бизнес-планов

3.1. Связь компетенций с профессиональными стандартами

Определенные профессиональные компетенции и их связь с профессиональными стандартами, размещенными на сайте Министерства труда и социальной защиты Российской Федерации «Профессиональные стандарты» (http://profstandart.rosmintrud.ru)указаны ниже.

Профессиональная компетенция	Профессиональный стандарт / вид деятельности	Обобщенная трудовая функция (ОТФ)	Выделенные трудовые действия (ТД)
Тип задач профессиональной деятельности «Научно-исследовательский»			
ПК-2. Способен разрабатывать и анализировать концептуальные и	**06.022** «Системный аналитик»	С/6 Концептуальное, функциональное и логическое проектирование систем среднего и крупного масштаба и сложности	С/02.6 Анализ проблемной ситуации заинтересованных лиц

Профессиональная компетенция	**Профессиональный стандарт / вид деятельности**	**Обобщенная трудовая функция (ОТФ)**	**Выделенные трудовые действия (ТД)**
теоретические модели решаемых научных проблем и задач.		*ОТФ выделена частично (15%)*	C/05.6 Разработка концепции системы
Тип задач профессиональной деятельности «Организационно-управленческий»			
ПК-7. Способен разрабатывать и оптимизировать бизнес-планы научно-прикладных проектов.	**06.022** «Системный аналитик»	D/6 Управление аналитическими работами и подразделением *ОТФ выделена частично (80%)*	D/01.7 Разработка технико-коммерческого предложения и участие в его защите
			D/02.7 Разработка методик выполнения аналитических работ
			D/03.7 Планирование аналитических работ в информационно-технологическом (далее - ИТ) проекте
			D/04.7 Организация аналитических работ в ИТ-проекте
			D/05.7 Контроль аналитических работ в ИТ-проекте
			D/06.7 Составление отчетов об аналитических работах в ИТ-проекте
			D/09.7 Управление аналитическими ресурсами и компетенциями
			D/10.7 Управление инфраструктурой разработки и сопровождения требований к системе

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 5 зачетных единиц, т.е. 180 академических часов (из них 160 часов – самостоятельная работа студентов): лабораторные работы – 16 часов, лекции -4 часа, форма контроля – зачет.

Разделы дисциплины «Хранилища данных» изучаются на первом и втором курсах.

Содержание разделов дисциплины

Второй семестр

1. Область применения хранилищ данных и больших данных и их преимущества
2. Виды данных
2.1 Структурированные данные
2.2 Неструктурированные данные
2.3 Данные на естественном языке
2.4 Машинные данные
2.5 Графовые, или сетевые, данные
2.6 Аудио, видео и графика
2.7 Потоковые данные
3. Экосистема больших данных
3.1 Распределенные файловые системы
3.2 Инфраструктура распределенного программирования
3.3 Инфраструктура интеграции данных
3.4 Инфраструктуры машинного обучения
3.5 Базы данных NoSQL

Третий семестр

4. Технологии Data Mining
4.1 Задачи Data Mining

Обзор задач Data Mining. Стандартизация подхода к решению задач Data Mining. Процесс CRISP-DM. Виды данных. Кластеризация, классификация, регрессия. Понятие модели и алгоритма обучения.

4.2 Задача кластеризации и EM-алгоритм

Постановка задачи кластеризации. Функции расстояния. Критерии качества кластеризации. EM-алгоритм. K-means и модификации.

4.3 Различные алгоритмы кластеризации

Иерархическая кластеризация. Agglomerative и Divisive алгоритмы. Различные виды расстояний между кластерами. Stepwise-optimal алгоритм. Случай неэвклидовых пространств. Критерии выбора количества кластеров: rand, silhouette. DBSCAN.

4.4 Задача классификации

Постановка задач классификации и регрессии. Теория принятия решений. Виды моделей. Примеры функций потерь. Переобучение. Метрики качества классификации. MDL. Решающие деревья. Алгоритм CART.

4.5 Обработка текстов, Naive Bayes

Условная вероятность и теорема Байеса. Нормальное распределение. Naive Bayes: multinomial, binomial, gaussian. Сглаживание. Генеративная модель NB и байесовский вывод. Графические модели.

4.6 Линейные модели для классификации и регрессии

Обобщенные линейные модели. Постановка задачи оптимизации. Примеры критериев. Градиентный спуск. Регуляризация. Метод Maximum Likelihood. Логистическая регрессия.

4.7 Машина опорных векторов

Разделяющая поверхность с максимальным зазором. Формулировка задачи оптимизации для случаев линейно-разделимых и линейно-неразделимых классов. Сопряженная задача. Опорные векторы. KKT-условия. SVM для задач классификации и регрессии. Kernel trick. Теорема Мерсера. Примеры функций ядра.

4.8 Методы снижения размерности пространства

Проблема проклятия размерности. Отбор и выделение признаков. Методы выделения признаков (feature extraction). Метод главных компонент (PCA). Метод независимых компонент (ICA). Методы основанные на автоэнкодерах. Методы отбора признаков (feature selection). Методы основанные на взаимной корреляции признаков. Метод максимальной релевантность и минимальной избыточности (mRMR). Методы основанные на деревьях решений.

4.9 Алгоритмические композиции

Комбинации классификаторов. Модельные деревья решений. Смесь экспертов. Stacking. Стохастические методы построения ансамблей классификаторов. Bagging. RSM. Алгоритм RandomForest.

4.10 Глубокие нейронные сети

Трудности обучения многослойного персептрона. Предобучение используя РБМ. Глубокий автоэнкодер, глубокая многослойная нейросеть. Deep belief network и deep Boltzmann machine. Сверточные сети.

Календарный график дисциплины

n/n	Раздел	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов, и трудоемкость в часах					Виды самостоятельной работы студентов					Формы аттестации	
				Л	П/С	Лаб	СРС	КСР	К.Р.	К.П.	К-З	Реферат	К/р	Э	З
	девятый семестр														
1	Область применения хранилищ данных и больших данных и их преимущества	9	1	1		1	10				**4**				
2	Виды данных	9	2			1	10					10			
3	Экосистема больших данных	9	3-4			1	15				**40**				
	Форма аттестации	9	5	1											З
4	Задачи Data Mining	9	6			1	10								
5	Задача кластеризации и EM-алгоритм	9	8			1	15								
6	Различные алгоритмы кластеризации	9	9	1		1	10				20	10			
7	Задача классификации	9	10			1	10								
8	Обработка текстов, Naive Bayes	9	11			2	15								
9	Линейные модели для классификации и регрессии	9	12	1		2	15				10				
10	Машина опорных векторов	9	13			1	15					5			
11	Методы снижения размерности пространства	9	14			1	15				5	10			
12	Алгоритмические композиции	9	15			1	15					10			
13	Глубокие нейронные сети	9	16			2	15					20			
	Форма аттестации	9	17-18											Э	
	Всего часов по дисциплине в девятом семестре			4		16	160				35	55		2	

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Методика преподавания дисциплины «Хранилища данных» и реализация компетентного подхода в изложении и восприятии материала предусматривает использование следующих активных и интерактивных форм проведения групповых, индивидуальных, аудиторных занятий в сочетании с внеаудиторной работой с целью формирования и развития профессиональных навыков обучающихся:

– подготовка к выполнению практических работ в компьютерных классах вуза;

– подготовка, представление и обсуждение презентаций на семинарских занятиях;

– организация и проведение интерактивных форм текущего контроля знаний студентов в форме выполнения индивидуальных заданий.

Удельный вес занятий, проводимых в интерактивных формах, определен главной целью образовательной программы, особенностью контингента обучающихся и содержанием дисциплины «Хранилища данных» и в целом по дисциплине составляет 50% аудиторных занятий. Занятия лекционного типа составляют 50% от объема аудиторных занятий.

6. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

В процессе обучения используются следующие оценочные формы самостоятельной работы студентов, оценочные средства текущего контроля успеваемости и промежуточных аттестаций:

- индивидуальное задание на работу с хранилищами данных и большими данными;
- подготовка к выполнению практических работ и их защита.

6.1. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине (модулю)

6.1.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы.

В результате освоения дисциплины (модуля) формируются следующие компетенции:

Код компетенции	В результате освоения образовательной программы обучающийся должен обладать
ОПК-3	способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять свое научное мировоззрение
ПК-2	способность разрабатывать и анализировать концептуальные и теоретические модели решаемых научных проблем и задач
ПК-7	способность разрабатывать и оптимизировать бизнес-планы научно-прикладных проектов

В процессе освоения образовательной программы данные компетенции, в том числе их отдельные компоненты, формируются поэтапно в ходе освоения обучающимися дисциплин (модулей), практик в соответствии с учебным планом и календарным графиком учебного процесса.

6.1.2. Описание показателей и критериев оценивания компетенций, формируемых по итогам освоения дисциплины (модуля), описание шкал оценивания

Показателем оценивания компетенций на различных этапах их формирования является достижение обучающимися планируемых результатов обучения по дисциплине (модулю).

ОПК-3 - способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять свое научное мировоззрение				
Показатель	**Критерии оценивания**			
	2	**3**	**4**	**5**
знать: основные методы проектирования и эксплуатации хранилищ данных и технологий Big Data	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие следующих знаний: основные методы проектирования и эксплуатации хранилищ данных и технологий Big Data	Обучающийся демонстрирует неполное соответствие следующих знаний: основные методы проектирования и эксплуатации хранилищ данных и технологий Big Data. Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду методов, обучающийся испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.	Обучающийся демонстрирует частичное соответствие следующих знаний: основные методы проектирования и эксплуатации хранилищ данных и технологий Big Data, но допускаются незначительные ошибки, неточности, затруднения при рассмотрении новых ситуаций.	Обучающийся демонстрирует полное соответствие следующих знаний: основные методы проектирования и эксплуатации хранилищ данных и технологий Big Data, свободно оперирует приобретенными знаниями.
уметь: использовать методы проектирования и эксплуатации	Обучающийся не умеет или в недостаточной степени умеет использовать методы проектирования и	Обучающийся демонстрирует неполное соответствие следующих умений: использовать методы	Обучающийся демонстрирует частичное соответствие следующих умений:	Обучающийся демонстрирует полное соответствие следующих умений: использовать методы проектирования и эксплуатации хранилищ данных и технологий Big Data для решения

хранилищ данных и технологий Big Data для решения научных задач	эксплуатации хранилищ данных и технологий Big Data для решения научных задач	проектирования и эксплуатации хранилищ данных и технологий Big Data для решения научных задач. Допускаются значительные ошибки, проявляется недостаточность умений, обучающийся испытывает значительные затруднения при оперировании умениями при их переносе на новые ситуации.	использовать методы проектирования и эксплуатации хранилищ данных и технологий Big Data для решения научных задач. Умения освоены, но допускаются незначительные ошибки, неточности, затруднения при переносе умений на новые, нестандартные ситуации.	научных задач. Свободно оперирует приобретенными умениями, применяет их в ситуациях повышенной сложности.
владеть: методами проектирования и эксплуатации хранилищ данных и технологий Big Data для решения практических научных задач	Обучающийся не владеет или в недостаточной степени владеет методами проектирования и эксплуатации хранилищ данных и технологий Big Data для решения практических научных задач	Обучающийся владеет методами проектирования и эксплуатации хранилищ данных и технологий Big Data для решения практических научных задач в неполном объеме, допускаются значительные ошибки, проявляется недостаточность владения навыками по ряду задач. Обучающийся испытывает значительные затруднения при	Обучающийся частично владеет методами проектирования и эксплуатации хранилищ данных и технологий Big Data для решения практических научных задач, навыки освоены, но допускаются незначительные ошибки, неточности, затруднения при переносе умений на новые, нестандартные ситуации.	Обучающийся в полном объеме владеет методами проектирования и эксплуатации хранилищ данных и технологий Big Data для решения практических научных задач, свободно применяет полученные навыки в ситуациях повышенной сложности.

		применении навыков в новых ситуациях.		
ПК-2 - способность разрабатывать и анализировать концептуальные и теоретические модели решаемых научных проблем и задач				
знать: концептуальные и теоретические подходы к проектированию и эксплуатации хранилищ данных и технологий Big Data	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие следующих знаний: концептуальные и теоретические подходы к проектированию и эксплуатации хранилищ данных и технологий Big Data	Обучающийся демонстрирует неполное соответствие следующих знаний: концептуальные и теоретические подходы к проектированию и эксплуатации хранилищ данных и технологий Big Data. Допускаются значительные ошибки, проявляется недостаточность знаний, обучающийся испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.	Обучающийся демонстрирует частичное соответствие следующих знаний: концептуальные и теоретические подходы к проектированию и эксплуатации хранилищ данных и технологий Big Data, но допускаются незначительные ошибки, неточности, затруднения при рассмотрении новых ситуаций.	Обучающийся демонстрирует полное соответствие следующих знаний: концептуальные и теоретические подходы к проектированию и эксплуатации хранилищ данных и технологий Big Data, свободно оперирует приобретенными знаниями.

уметь: использовать методы проектирования и эксплуатации хранилищ данных и технологий Big Data для решения научных проблем и задач	Обучающийся не умеет или в недостаточной степени умеет использовать методы проектирования и эксплуатации хранилищ данных и технологий Big Data для решения научных проблем и задач	Обучающийся демонстрирует неполное соответствие следующих умений: использовать методы проектирования и эксплуатации хранилищ данных и технологий Big Data для решения научных проблем и задач. Допускаются значительные ошибки, проявляется недостаточность умений, обучающийся испытывает значительные затруднения при оперировании умениями при их переносе на новые ситуации.	Обучающийся демонстрирует частичное соответствие следующих умений: использовать методы проектирования и эксплуатации хранилищ данных и технологий Big Data для решения научных проблем и задач Умения освоены, но допускаются незначительные ошибки, неточности, затруднения при переносе умений на новые, нестандартные ситуации.	Обучающийся демонстрирует полное соответствие следующих умений: использовать методы проектирования и эксплуатации хранилищ данных и технологий Big Data для решения научных проблем и задач. Свободно оперирует приобретенными умениями, применяет их в ситуациях повышенной сложности.
владеть: методами проектирования и эксплуатации хранилищ данных и технологий Big Data для проведения научных исследований	Обучающийся не владеет или в недостаточной степени владеет методами проектирования и эксплуатации хранилищ данных и технологий Big Data для проведения научных исследований	Обучающийся владеет методами проектирования и эксплуатации хранилищ данных и технологий Big Data для проведения научных исследований в неполном объеме, допускаются значительные ошибки. Обучающийся испытывает	Обучающийся частично владеет методами проектирования и эксплуатации хранилищ данных и технологий Big Data для проведения научных исследований, навыки освоены, но допускаются незначительные	Обучающийся в полном объеме владеет методами проектирования и эксплуатации хранилищ данных и технологий Big Data для проведения научных исследований, свободно применяет полученные навыки в ситуациях повышенной сложности.

		значительные затруднения при применении навыков в новых ситуациях.	ошибки, неточности, затруднения при переносе умений на новые, нестандартные ситуации.	
ПК-7 - способность разрабатывать и оптимизировать бизнес-планы научно-прикладных проектов				
знать: концептуальные и теоретические подходы к разработке и оптимизации бизнес-планов научно-прикладных проектов	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие следующих знаний: концептуальные и теоретические подходы к разработке и оптимизации бизнес-планов научно-прикладных проектов	Обучающийся демонстрирует неполное соответствие следующих знаний: Обучающийся демонстрирует неполное соответствие следующих знаний: концептуальные и теоретические подходы к разработке и оптимизации бизнес-планов научно-прикладных проектов. Допускаются значительные ошибки, проявляется недостаточность знаний, обучающийся испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.	Обучающийся демонстрирует частичное соответствие следующих знаний: концептуальные и теоретические подходы к разработке и оптимизации бизнес-планов научно-прикладных проектов, но допускаются незначительные ошибки, неточности, затруднения при рассмотрении новых ситуаций.	Обучающийся демонстрирует полное соответствие следующих знаний: концептуальные и теоретические подходы к разработке и оптимизации бизнес-планов научно-прикладных проектов, свободно оперирует приобретенными знаниями

уметь: использовать методы разработки и оптимизации бизнес-планов научно-прикладных проектов	Обучающийся не умеет или в недостаточной степени умеет использовать методы разработки и оптимизации бизнес-планов научно-прикладных проектов	Обучающийся демонстрирует неполное соответствие следующих умений: использовать методы разработки и оптимизации бизнес-. Допускаются значительные ошибки, проявляется недостаточность умений, обучающийся испытывает значительные затруднения при оперировании умениями при их переносе на новые ситуации.	Обучающийся демонстрирует частичное соответствие следующих умений: использовать методы разработки и оптимизации бизнес-планов. Умения освоены, но допускаются незначительные ошибки, неточности, затруднения при переносе умений на новые, нестандартные ситуации.	Обучающийся демонстрирует полное соответствие следующих умений: использовать методы разработки и оптимизации бизнес-планов научно-прикладных проектов. Свободно оперирует приобретенными умениями, применяет их в ситуациях повышенной сложности.
владеть: методами разработки и оптимизации бизнес-планов научно-прикладных проектов	Обучающийся не владеет или в недостаточной степени владеет методами разработки и оптимизации бизнес-планов научно-прикладных проектов	Обучающийся владеет разработки и оптимизации бизнес-планов научно-прикладных проектов в неполном объеме, допускаются значительные ошибки. Обучающийся испытывает значительные затруднения при	Обучающийся частично владеет методами разработки и оптимизации бизнес-планов, навыки освоены, но допускаются незначительные ошибки, неточности, затруднения при переносе умений на	Обучающийся в полном объеме владеет методами разработки и оптимизации бизнес-планов научно-прикладных проектов, свободно применяет полученные навыки в ситуациях повышенной сложности.

		применении навыков в новых ситуациях.	новые, нестандартные ситуации.	

Шкалы оценивания результатов промежуточной аттестации и их описание:

Форма промежуточной аттестации: зачет.

Промежуточная аттестация обучающихся в форме зачёта проводится по результатам выполнения всех видов учебной работы, предусмотренных учебным планом по данной дисциплине (модулю), при этом учитываются результаты текущего контроля успеваемости в течение семестра. Оценка степени достижения обучающимися планируемых результатов обучения по дисциплине (модулю) проводится преподавателем, ведущим занятия по дисциплине (модулю) методом экспертной оценки. По итогам промежуточной аттестации по дисциплине (модулю) выставляется оценка «зачтено» или «не зачтено».

Шкала оценивания	**Описание**
Зачтено	Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, умений, навыков приведенным в таблицах показателей, оперирует приобретенными знаниями, умениями, навыками, применяет их в ситуациях повышенной сложности. При этом могут быть допущены незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
Не зачтено	Не выполнен один или более видов учебной работы, предусмотренных учебным планом. Студент демонстрирует неполное соответствие знаний, умений, навыков приведенным в таблицах показателей, допускаются значительные ошибки, проявляется отсутствие знаний, умений, навыков по ряду показателей, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.

Форма промежуточной аттестации: экзамен.

Промежуточная аттестация обучающихся в форме экзамена проводится по результатам выполнения всех видов учебной работы, предусмотренных учебным планом по данной дисциплине (модулю), при этом учитываются результаты текущего контроля успеваемости в течение семестра. Оценка степени достижения обучающимися планируемых результатов обучения по дисциплине (модулю) проводится преподавателем, ведущим занятия по дисциплине (модулю) методом экспертной оценки. По итогам промежуточной аттестации по дисциплине (модулю) выставляется оценка «отлично», «хорошо», «удовлетворительно» или «неудовлетворительно».

Шкала оценивания	***Описание***
Отлично	*Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, умений, навыков приведенным в таблицах показателей, оперирует приобретенными знаниями, умениями, навыками, применяет их в ситуациях повышенной сложности. При этом могут быть допущены незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.*
Хорошо	*Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует неполное, правильное соответствие знаний, умений, навыков приведенным в таблицах показателей, либо если при этом были допущены 2-3 несущественные ошибки.*
Удовлетворительно	*Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, в котором освещена основная, наиболее важная часть материала, но при этом допущена одна значительная ошибка или неточность.*
Неудовлетворительно	*Не выполнен один или более видов учебной работы, предусмотренных учебным планом. Студент демонстрирует неполное соответствие знаний, умений, навыков приведенным в таблицах показателей, допускаются значительные ошибки, проявляется отсутствие знаний, умений, навыков по ряду показателей, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.*

Экзаменационное задание выполняется студентом индивидуально, по итогам изучения дисциплины или ее части. При этом достижение порогового результата работы над экзаменационным заданием соответствует описанному в п. 3 данного документа этапу освоения соответствующих компетенций на базовом или продвинутом уровне.

Базовый уровень: способность выполнять полученное задание, применяя полученные знание и умения на практике, владеть соответствующими индикаторами компетенции при выполнении задания.

Продвинутый уровень: способность выполнять полученное задание и решать самостоятельно сформированные задачи, применяя полученные знание и умения на практике. Уверенно владеть соответствующими индикаторами компетенции при

выполнении задания, комбинировать их между собой и с индикаторами других компетенций для достижения проектных результатов.

Форма экзаменационного задания выбирается преподавателем и утверждается на заседании кафедры. Экзамен может проходить в следующих формах и с использованием следующих оценочных средств.

Форма	Представление оценочного средства в ФОС
Устная.	Банк контрольных вопросов, соответствующих отдельным темам дисциплины (см. п. 4 настоящего документа). Вопросы формируют экзаменационный билет (см. ниже), состоящий из теоретических вопросов и практических заданий (типовые практические задания представлены ниже). Билеты, включая вопросы и практические задания, формируются преподавателем и утверждаются на заседании кафедры. В них могут быть включены дополнительные контрольные вопросы и задания, не требующие у студентов наличия не формируемых данной дисциплиной компетенций или более высоких этапов сформированности формируемых. Для ответа на каждый вопрос и для решения любого практического задания студент должен находится на требуемом для данной дисциплине уровне сформированности всех соответствующих ей компетенций: каждый вопрос и задание проверяет уровень сформированности всех соответствующих данной дисциплине компетенций.
Письменная.	Оценочное средство полностью соответствует оценочным средствам устной формы задания.

Типовой экзаменационный билет

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

(МОСКОВСКИЙ ПОЛИТЕХ)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №1

по дисциплине

«ХРАНИЛИЩА ДАННЫХ»

направление подготовки 09.03.01 Информатика и вычислительная техника

ВОПРОСЫ:

1. Виды моделей. Примеры функций потерь. Переобучение. Метрики качества классификации.
2. Разделяющая поверхность с максимальным зазором. Формулировка задачи оптимизации для случаев линейно-разделимых и линейно-неразделимых классов.

3. Сверточные сети

Утверждено: _____ / _____ / «__» _____ 20__г.

Примеры оценочных средств:

Пример кейсового задания: Прогнозирование вредоносных URL-адресов

Загрузите данные по адресу http://sysnet.ucsd.edu/projects/url/#datasets и поместите их в рабочий каталог. Выберите данные в формате SVMLight.

Задача: создать стохастический градиентный классификатор сайтов. Целевая переменная может быть равна 1 или -1. "1": сайт безопасен для посещения, "-1": сайт небезопасен.

Результы выполнения оформить в виде:

Этап 1: Определение цели исследования

Этап 2: Сбор данных URL

Этап 3: Исследование данных

Этап 4: Построение модели

Пример кейсового задания: Построение рекомендательной системы внутри базы данных

Рекомендательная система должна выдавать результаты вида: «Клиенты, смотревшие этот фильм, также смотрели…»

Для обучения использовать метод “k-ближайших соседей“.

Исходные данные для построения рекомендательной системы создаются самими учащимися.

Данные должны быть вида:

Клиент	Фильм 1	Фильм 2	Фильм 3	...	Фильм 32
Jack Dani	1	0	0		1
Wilhelmson	1	1	0		1
...					
Jane Dane	0	0	1		0
Xi Liu	0	0	0		1
Eros Mazo	1	1	0		1
...					

Результаты выполнения оформить в виде:

Этап 1: Вопрос исследования

Этап 2: Подготовка данных
Этап 3: Построение модели
Этап 4: Отображение и автоматизация

Зачетные вопросы

1. Область применения хранилищ данных и больших данных и их преимущества
2. Виды данных. Структурированные данные
3. Виды данных. Неструктурированные данные
4. Виды данных. Данные на естественном языке
5. Виды данных. Машинные данные
6. Виды данных. Графовые, или сетевые, данные
7. Виды данных. Аудио, видео и графика
8. Виды данных. Потоковые данные
9. Понятие гиперкуба. OLAP.
10. ROLAP-системы
11. MOLAP-системы
12. HOLAP-системы
13. Распределенные файловые системы
14. Базы данных NoSQL

Экзаменационные вопросы

1. Обзор задач Data Mining. Кластеризация, классификация, регрессия.
2. Постановка задачи кластеризации. Функции расстояния. Критерии качества кластеризации.
3. EM-алгоритм.
4. K-means и модификации.
5. Иерархическая кластеризация. Agglomerative и Divisive алгоритмы.
6. Постановка задач классификации и регрессии. Теория принятия решений.
7. Виды моделей. Примеры функций потерь. Переобучение. Метрики качества классификации.
8. MDL. Решающие деревья.
9. Алгоритм CART.
10. Условная вероятность и теорема Байеса. Нормальное распределение.
11. Naive Bayes: multinomial, binomial, gaussian. Сглаживание. Генеративная модель NB и байесовский вывод.
12. Обобщенные линейные модели. Постановка задачи оптимизации. Примеры критериев. Градиентный спуск.
13. Разделяющая поверхность с максимальным зазором. Формулировка задачи оптимизации для случаев линейно-разделимых и линейно-неразделимых классов.
14. KKT-условия. SVM для задач классификации и регрессии.
15. Проблема проклятия размерности. Отбор и выделение признаков. Методы выделения признаков (feature extraction).

16. Метод главных компонент (PCA).
17. Метод независимых компонент (ICA).
18. Методы, основанные на автоэнкодерах.
19. Комбинации классификаторов. Модельные деревья решений.
20. Алгоритм RandomForest.
21. Глубокий автоэнкодер, глубокая многослойная нейросеть.
22. Deep belief network и deep Boltzmann machine.
23. Сверточные сети.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

а) основная литература:

Дэви Силен, Арно Мейсман Основы Data Science и Big Data. Python и наука о данных. СПб, Питер, 2017. – 336 стр.

Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям [Электронный ресурс], СПб, Питер, 2013. – 704 стр. Режим доступа: http://www.padaread.com/?book=52105

б) дополнительная литература

Заенцев И.В. Нейронные сети: основные модели [Электронный ресурс], Воронеж, 1999, - 76 стр. Режим доступа: http://computersbooks.net/books/web-programmirovanie/zaencev-iv/1999/files/neyronnieseti1999.pdf

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Компьютерный класс Ауд. АВ4810, оснащенный методическими материалами по дисциплине (практические задания).

9. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

Изучение дисциплины осуществляется в строгом соответствии с целевой установкой в тесной взаимосвязи учебным планом. Основой теоретической подготовки студентов являются *аудиторные занятия, лекции, лабораторные работы.*

В процессе самостоятельной работы студенты закрепляют и углубляют знания, полученные во время аудиторных занятий, дорабатывают конспекты и записи, готовятся к проведению и обрабатывают результаты лабораторных работ, готовятся к промежуточной аттестации, а также самостоятельно изучают отдельные темы учебной программы.

На занятиях студентов, в том числе предполагающих практическую деятельность, осуществляется закрепление полученных, в том числе и в процессе самостоятельной работы, знаний. Особое внимание обращается на развитие умений

и навыков установления связи положений теории с профессиональной деятельностью будущего специалиста.

Самостоятельная работа осуществляется индивидуально. Контроль самостоятельной работы организуется в двух формах:

- самоконтроль и самооценка студента;
- контроль со стороны преподавателей (текущий и промежуточный).

Критериями оценки результатов самостоятельной работы студента являются:

- уровень освоения студентом учебного материала;
- умения студента использовать теоретические знания при выполнении практических задач;
- сформированность компетенций;

оформление материала в соответствии с требованиями.

Раздел дисциплины (тема)	Вид самостоятельной работы	Форма текущего контроля
Область применения хранилищ данных и больших данных и их преимущества	Самостоятельное изучение Изучение теоретического материала и решение задач	Письменное тестирование
Виды данных	Самостоятельное изучение Изучение теоретического материала и решение задач	Письменное тестирование
Экосистема больших данных	Самостоятельное изучение Изучение теоретического материала и решение задач	Письменное тестирование
Задачи Data Mining	Самостоятельное изучение Изучение теоретического материала и решение задач	Письменное тестирование
Задача кластеризации и EM-алгоритм	Самостоятельное изучение Изучение теоретического материала и решение задач	Письменное тестирование
Различные алгоритмы кластеризации	Самостоятельное изучение Изучение теоретического материала и решение задач	Письменное тестирование
Задача классификации	Самостоятельное изучение Изучение теоретического материала и решение задач	Письменное тестирование
Обработка текстов, Naive Bayes	Самостоятельное изучение Изучение теоретического материала и решение задач	Письменное тестирование
Линейные модели для классификации и регрессии	Самостоятельное изучение Изучение теоретического материала и решение задач	Письменное тестирование

Раздел дисциплины (тема)	Вид самостоятельной работы	Форма текущего контроля
Машина опорных векторов	Самостоятельное изучение Изучение теоретического материала и решение задач	Письменное тестирование
Методы снижения размерности пространства	Самостоятельное изучение Изучение теоретического материала и решение задач	Письменное тестирование
Алгоритмические композиции	Самостоятельное изучение Изучение теоретического материала и решение задач	Письменное тестирование
Глубокие нейронные сети	Самостоятельное изучение Изучение теоретического материала и решение задач	Письменное тестирование

10. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ ПРЕПОДАВАТЕЛЯ

1. При подготовке к занятиям следует предварительно проработать материал занятия, предусмотрев его подачу точно в отведенное для этого время занятия. Следует подготовить необходимые материалы – теоретические сведения, задачи и др. При проведении занятия следует контролировать подачу материала и решение заданий с учетом учебного времени, отведенного для занятия.
2. При проверке работ и отчетов следует учитывать не только правильность выполнения заданий, но и оптимальность выбранных методов решения, правильность выполнения всех его шагов.
3. При организации и проведения зачетов в практико-ориентированной форме следует использовать утвержденные кафедрой методические рекомендации.

Тема занятий	Виды учебных занятий	Средства обучения	Методы обучения	Форма оценочного средства**
Область применения хранилищ данных и больших данных и их преимущества	Практическое занятие	Компьютерный класс	Практическое занятие	К-З
Виды данных	Практическое занятие	Компьютерный класс	Практическое занятие	К
Экосистема больших данных	Практическое занятие	Компьютерный класс	Практическое занятие	К-З
Задачи Data Mining	Практическое занятие	Компьютерный класс	Практическое занятие	К-З
Задача кластеризации и EM-алгоритм	Практическое занятие	Компьютерный класс	Практическое занятие	К-З
Различные алгоритмы кластеризации	Практическое занятие	Компьютерный класс	Практическое занятие	К-З
Задача классификации	Практическое занятие	Компьютерный класс	Практическое занятие	К-З

Тема занятий	Виды учебных занятий	Средства обучения	Методы обучения	Форма оценочного средства**
Обработка текстов, Naive Bayes	Практическое занятие	Компьютерный класс	Практическое занятие	К
Линейные модели для классификации и регрессии	Практическое занятие	Компьютерный класс	Практическое занятие	К-З
Машина опорных векторов	Практическое занятие	Компьютерный класс	Практическое занятие	К-З
Методы снижения размерности пространства	Практическое занятие	Компьютерный класс	Практическое занятие	К-З
Алгоритмические композиции	Практическое занятие	Компьютерный класс	Практическое занятие	К-З
Глубокие нейронные сети	Практическое занятие	Компьютерный класс	Практическое занятие	К