Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Максимов Алексей Борисович
Должность: директор департамента по образовательной политике
Дата подписания: 31.08.2023 14:58:14
Уникальный программный ключ:
8db180d1a3f02ac9e60521a5172742755c168d2a8

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования

«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

Декан факультета

«Информационные технологии»

_____/Д.Г.Демидов/

«16» мая 2022

Рабочая программа дисциплины

«Методы работы с большими данными»

Направление подготовки:

09.03.01 Информатика и вычислительная техника

Образовательная программа (профиль):

«Веб-технологии»

Год начала обучения:

2022

Уровень образования:

Бакалавриат

Квалификация (степень) выпускника:

Бакалавр

Форма обучения:

очная

Москва, 2022

Рабочая программа дисциплины «Методы работы с большими данными» составлена в соответствии с федеральным образовательным

стандартом высшего образования – бакалавриата по направлению подготовки 09.03.01 Информатика и вычислительная техника.

Заведующий кафедрой «Инфокогнитивные технологии»:



к.т.н., доцент /Е.А.Пухова /

Согласовано:

Руководитель образовательной программы:



/М.В.Даньшина/

Программу составили:

_____ / _____/

_____ / _____ /

_____ / _____ /

_____ / _____ /

_____ / _____ /

1. Цели освоения дисциплины

К основным целям освоения дисциплины «Методы работы с большими данными» относится: формирование у студентов необходимой теоретической базы и практических навыков, которые позволят всесторонне и системно понимать современные проблемы прикладной математики и информатики, проблемы обработки и анализа информации, а также разрабатывать и анализировать концептуальные и теоретические модели при решении научных и прикладных задач в области информационных технологий.

К основным задачам дисциплины «Методы работы с большими данными» относятся:

- сформировать целостное представление о современных проблемах анализа и обработки больших данных;
- помочь овладеть опытом разработки и анализа концептуальных и теоретических моделей прикладных задач анализа больших данных с применением моделей Data Mining.

2. Место дисциплины в структуре ОПОП

Дисциплина относится к числу учебных дисциплин формируемые участниками образовательных отношений части Дисциплины по выбору студента «Элективные дисциплины» основной профессиональной образовательной программы.

Дисциплина взаимосвязана логически и содержательно-методически со следующими дисциплинами и практиками ОПОП:

- Алгоритмическое программирование;
- Веб-разработка на стороне клиента;
- Веб-разработка;
- Серверная веб-разработка;
- Индексирование текстов и информационный поиск;
- Основы разработки виртуальной и дополнительной реальности;
- Разработка мобильных и веб-приложений;

- Основы разработки КИС.

3. Перечень планируемых результатов обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы

В результате освоения дисциплины 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
УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.2. Уметь: анализировать и систематизировать разнородные данные.
УК-2	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1. Знать: необходимые для осуществления профессиональной деятельности правовые нормы.
ПК-3	Способен осуществлять управление проектами в области ИТ на основе полученных планов проектов в условиях, когда проект не выходит за пределы утвержденных параметров	ПК-3.1. Знать: возможности информационных систем.
ПК-4	Способен выполнять работы по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы	ПК-4.2. Уметь: анализировать научно-техническую литературу, извлекать из нее сведения, необходимые для решения поставленной задачи.

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.

4. Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 5 зачетных единицы, т.е. 180 академических часа (из них 106 часов – самостоятельная работа студентов).

Разделы дисциплины изучаются на четвертом курсе в седьмом семестре, форма промежуточной аттестации - экзамен.

Содержание дисциплины

1. Обработка данных с использованием библиотеки Pandas

Основные возможности библиотеки Pandas: загрузка и выборка данных, агрегирование данных, нормализация данных.

2. Визуализация данных и их анализ

Основные виды графиков для отображения данных: диаграммы рассеяния, размаха, гистограммы, эпюры, отображение трехмерных данных, отображение последовательностей. Элементы графика: оси, легенда, надписи. Методы снижения размерности пространства признаков: метод главных компонент, многомерное шкалирование, t-SNE, UMAP.

3. Классификация данных

Линейные методы классификации данных: линейная и логистическая регрессия, SMV. Методы, основанные на деревьях принятия решения, в том числе, метод случайного леса. Метод k ближайших соседей. Методы бустинга и стеккинга. Методы оценки результатов классификации: точность, полнота, f-мера, ROC-AUC, матрица ошибок.

4. Кластеризация данных

Метод k-средних, спектральная кластеризация. Методы, основанные на оценке плотности распределения точек в пространстве. Методы оценки точности кластеризации.

5. Обработка текстов на естественном языке

Лексический и синтаксический анализ текстов. Понятия пространства признаков для текста. Задачи обработки текстов: выделение именованных сущностей, фактов. Классификация и кластеризация текстов. Технологии Word2Vec, Glove.

6. Обработка изображений

Основные форматы хранения изображений и их отличия: BMP, PNG, JPG, GIF. Методы внесения изменений в изображения. Методы выделения областей изображения при помощи кластеризации точек. Библиотеки Python я работы с изображениями. Библиотека OpenCV и методы обработки изображений в sklearn.

7. «Плотные» нейронные сети

Бионические и искусственные нейронные сети, нейрон МакКаллока и Питтса, персептрон, сети Кохонена. Понятия порогового элемента и функции. Применение нейронных сетей для решения задач классификации и преобразования данных.

8. Глубинное обучение нейронных сетей

Понятие свертки в нейронных сетях, сверточные нейронные сети. Введение обратной связи в нейронной сети, рекуррентные сети. Виды сверточных и рекуррентных нейронных сетей. Построение архитектуры нейронной сети для решения прикладных задач.

5. Образовательные технологии

Методика преподавания дисциплины «Методы работы с большими данными» предусматривает использование следующих форм проведения занятий:

- посещение лекций;
- выполнение лабораторных работ;
- индивидуальные и групповые консультации студентов с преподавателем.

Самостоятельная внеаудиторная работа студентов состоит из анализа материалов лекций, самостоятельному освоению части материала, а также подготовки к промежуточной аттестации во время экзаменационной сессии.

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:

- лабораторные работы, экзамен.

6.1. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине «Методы работы с большими данными»

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

В результате освоения дисциплины формируются следующие компетенции:

Код компетенции	В результате освоения образовательной программы обучающийся должен обладать
УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач
УК-2	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений
ПК-3	Способен осуществлять управление проектами в области ИТ на основе полученных планов проектов в условиях, когда проект не выходит за пределы утвержденных параметров

ПК-4	Способен выполнять работы по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы

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

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

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

Показатель	**Критерии оценивания**			
	2	**3**	**4**	**5**
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач				
УК-1.2. Уметь: анализировать и систематизировать разнородные данные.	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие материалу дисциплины знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3).	Обучающийся демонстрирует неполное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся	Обучающийся демонстрирует частичное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	Обучающийся демонстрирует полное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Свободно оперирует приобретенными знаниями.

		испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.		
УК-2**.** Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений				
УК-2.1. Знать: необходимые для осуществления профессиональной деятельности правовые нормы.	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие материалу дисциплины знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3).	Обучающийся демонстрирует неполное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.	Обучающийся демонстрирует частичное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	Обучающийся демонстрирует полное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Свободно оперирует приобретенными знаниями.
ПК-3. Способен осуществлять управление проектами в области ИТ на основе полученных планов проектов в условиях, когда проект не выходит за пределы утвержденных параметров				
ПК-3.1. Знать:	Обучающийся демонстрирует полное	Обучающийся демонстрирует неполное	Обучающийся демонстрирует частичное	Обучающийся демонстрирует полное

возможности информационных систем.	отсутствие или недостаточное соответствие материалу дисциплины знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3).	соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.	соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Свободно оперирует приобретенными знаниями.
ПК-4. Способен выполнять работы по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы				
ПК-4.2. Уметь: анализировать научно-техническую литературу, извлекать из нее сведения, необходимые для решения поставленной задачи.	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие материалу дисциплины знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3).	Обучающийся демонстрирует неполное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся	Обучающийся демонстрирует частичное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	Обучающийся демонстрирует полное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Свободно оперирует приобретенными знаниями.

		испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.		

Шкала оценивания результатов промежуточной аттестации и её описание:

Форма промежуточной аттестации: экзамен.

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

Шкала оценивания	**Описание**
Отлично	Выполнены все **обязательные условия подготовки студента к промежуточной аттестации**, предусмотренные программой дисциплины. Студент демонстрирует соответствие знаний, умений, навыков показателям, приведенным в таблицах, оперирует приобретенными знаниями, умениями, навыками, применяет их в ситуациях повышенной сложности. При этом могут быть допущены незначительные ошибки, неточности, затруднения при переносе знаний и умений на новые, нестандартные задачи.
Хорошо	Выполнены все **обязательные условия подготовки студента к промежуточной аттестации**, предусмотренные программой дисциплины. Студент демонстрирует соответствие знаний, умений, навыков показателям, приведенным в таблицах, оперирует приобретенными знаниями, умениями, навыками. В то же время при ответе

	допускает несущественные погрешности, задачи решает с недочетами, не влияющими на общий ход решения.
Удовлетворительно	Выполнены все **обязательные условия подготовки студента к промежуточной аттестации**, предусмотренные программой дисциплины. Студент демонстрирует соответствие знаний, умений, навыков показателям, приведенным в таблицах, оперирует приобретенными знаниями, умениями, навыками. Но показывает неглубокие знания, при ответе не допускает грубых ошибок или противоречий, однако в формулировании ответа отсутствует должная связь между анализом, аргументацией и выводами, в решении задач могут содержаться грубые ошибки. Для получения правильного ответа требуются уточняющие вопросы.
Неудовлетворительно	Не выполнены обязательные **условия подготовки студента к промежуточной аттестации**, предусмотренные программой дисциплины, ИЛИ студент демонстрирует неполное соответствие знаний, умений, навыков приведенным в таблицах, допускаются значительные ошибки, проявляется отсутствие знаний, умений, навыков по ряду показателей, студент испытывает значительные затруднения при оперировании знаниями и умениями.

7. Учебно-методическое и информационное обеспечение дисциплины

7.1. Основная литература

1. Чернышев, С. А. Основы программирования на Python : учебное пособие для вузов / С. А. Чернышев. — Москва : Издательство Юрайт, 2023. — 286 с. — (Высшее образование). — ISBN 978-5-534-14350-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/519949

2. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2023. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/511020

7.2. Дополнительная литература

1. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2023. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/511121

2. Черпаков, И. В. Основы программирования : учебник и практикум для вузов / И. В. Черпаков. — Москва : Издательство Юрайт, 2023. — 219 с. — (Высшее образование). — ISBN 978-5-9916-9983-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/511703

8. Материально-техническое обеспечение дисциплины

8.1 Требования к оборудованию и помещению для занятий

Лабораторные работы и самостоятельная работа студентов должны проводиться в специализированной аудитории, оснащенной современной оргтехникой и персональными компьютерами с программным обеспечением в соответствии с тематикой изучаемого материала. Число рабочих мест в аудитории должно быть достаточным для обеспечения индивидуальной работы студентов. Рабочее место преподавателя должно быть оснащены современным компьютером с подключенным к нему проектором на настенный экран, или иным аналогичным по функциональному назначению оборудованием.

8.2 Требования к программному обеспечению

Для выполнения лабораторных работ и самостоятельной работы необходимо следующее программное обеспечение:

1. Microsoft Windows.
2. Веб-браузер, Chrome.

9. Методические рекомендации для самостоятельной работы студентов

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

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.

Приложение 1

Структура и содержание дисциплины «Методы работы с большими данными»
Направление подготовки: 09.03.01 «Информатика и вычислительная техника»

Профиль подготовки Веб-технологии (Бакалавр)

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.	РГР	Реферат	К/р	Э	З
1	Обработка данных с использованием библиотеки Pandas	7	1-2			9	12								
2	Визуализация данных и их анализ	7	3-4			9	12								
3	Классификация данных	7	5-6			9	12								
4	Кластеризация данных	7	7-8			9	14								
5	Обработка текстов на естественном языке	7	9-10			9	14								
6	Обработка изображений	7	11-12			9	14								
7	«Плотные» нейронные сети	7	13-14			10	14								
8	Глубинное обучение нейронных сетей	7	15			10	14								
	Форма аттестации		**16-17**											**Э**	
	Всего часов по дисциплине					**74**	**106**								

Приложение 2

МИНИСТЕРСТВО НАУКИ ИВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
(МОСКОВСКИЙ ПОЛИТЕХ)

Направление подготовки 09.03.01 Информатика и вычислительная техника
профиль подготовки «Веб-технологии»
Форма обучения: очная

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

ПО ДИСЦИПЛИНЕ

Методы работы с большими данными

Состав:

1. **Показатель уровня сформированности компетенций.**
2. **Вопросы для экзамена.**
3. **Типовые практические задания.**
4. **Типовый экзаменационный билет.**

Москва, 2022 год

1. ПОКАЗАТЕЛЬ УРОВНЯ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ

«Методы работы с большими данными»					
ФГОС ВО 09.03.01 «Информатика и вычислительная техника» профиль подготовки «**Веб-технологии»**					
В процессе освоения данной дисциплины студент формирует и демонстрирует следующие **общепрофессиональные компетенции**:					
Компетенции		**Перечень компонентов**	**Технология формирования компетенций**	**Форма оценочного средства****	**Степени уровней освоения компетенций**
Индекс	**Индекс**				
УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.2. Уметь: анализировать и систематизировать разнородные данные.	Лабораторные работы, самостоятельная работа	УО П Экзамен	**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.
УК-2	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1. Знать: необходимые для осуществления профессиональной деятельности правовые нормы.			
ПК-3	Способен осуществлять управление проектами в области ИТ на основе полученных планов проектов в условиях, когда проект не выходит за пределы утвержденных параметров	ПК-3.1. Знать: возможности информационных систем.			
ПК-4	Способен выполнять работы по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы	ПК-4.2. Уметь: анализировать научно-техническую литературу, извлекать из нее сведения, необходимые для решения поставленной задачи.			

**- Сокращения форм оценочных средств см. в приложении 2 к РП.

2. ВОПРОСЫ ДЛЯ ЭКЗАМЕНА

1. Что такое библиотека Pandas? Для чего она используется? Какие возможности предоставляет?
2. Что такое визуализация данных?
3. Перечислите основные виды графиков для отображения данных и кратко опишите каждый из них.
4. Какие элементы графика существуют? Опишите их.
5. Перечислите методы снижения размерности пространства признаков и кратко опишите каждый из них.
6. Что такое классификация данных?
7. Какие линейные методы классификации данных существуют? Опишите их.
8. Опишите методы, основанные на деревьях принятия решения, в том числе, метод случайного леса.
9. Перечислите и опишите методы оценки результатов классификации.
10. Что такое кластеризация данных?
11. Для чего используется метод k-средних? Опишите его.
12. Для чего используется спектральная кластеризация? Опишите ее.
13. Опишите методы, основанные на оценке плотности распределения точек в пространстве.
14. Опишите методы оценки точности кластеризации.
15. Что такое лексический анализ текстов?
16. Что такое синтаксический анализ текстов?
17. Опишите технологии Word2Vec и Globe.
18. Перечислите основные форматы хранения изображений и опишите их отличия.
19. Что такое «плотные» нейронные сети?
20. Что такое нейрон МакКаллока и Питтса?
21. Что такое персептрон?

22. Что такое сети Кохонена?

23. Что такое свертка в нейронных сетях?

24. Что такое рекуррентные сети?

25. Опишите виды сверточных и рекуррентых нейронных сетей.

26. Перечислите основные базы знаний Wolfram Data Platform.

27. Что представляет из себя формат CDF?

28. Что такое канонический идентификатор WDF?

29. Что такое хеширование?

30. Как осуществляется хранение и сравнение идентификационных данных пользователей компьютерных сетей?

4. ТИПОВЫЕ ПРАКТИЧЕСКИЕ ЗАДАНИЯ

1. Оцените общий объем данных, который могло бы вместить наиболее емкое теоретически возможное хранилище данных.

2. Предложите алгоритм выявления подозрительных транзакций по банковской карте.

3. Предложите алгоритм выявления нарушения автомобилем скоростного режима на дороге на основе данных камер видеофиксации.

4. Оцените количество слов в орфографическом словаре русского языка.

5. Предложите стандарт формата цифрового следа пользователя интернета и метрику в пространстве таких следов.

5. ТИПОВОЙ ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
(МОСКОВСКИЙ ПОЛИТЕХ)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №1
по дисциплине
«Методы работы с большими данными»
направление подготовки 09.03.01 Информатика и вычислительная техника

ВОПРОСЫ:

1. Какие линейные методы классификации данных существуют? Опишите их.
2. Что такое персептрон?
3. Что такое канонический идентификатор WDF?

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