

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ,
МЕХАНИКИ И ОПТИКИ»

Проект «ИТ-классы в московских школах»

РАБОЧАЯ ПРОГРАММА НАПРАВЛЕНИЯ

Большие данные

(10, 11 класс, 128 часа)

Руководитель авторского коллектива:

Калюжная Анна Владимировна, к. т. н., доцент института дизайна и урбанистики, старший научный сотрудник национального центра когнитивных разработок

**Санкт-Петербург
2019 г.**

РАБОЧАЯ ПРОГРАММА НАПРАВЛЕНИЯ

Разработана:

Руководитель авторского коллектива:

Калюжная Анна Владимировна, к. т. н., доцент института дизайна и урбанистики, старший научный сотрудник национального центра когнитивных разработок,
подпись _____

Список авторов:

Боченина Клавдия Олеговна, к. т. н., доцент института дизайна и урбанистики, старший научный сотрудник национального центра когнитивных разработок,
подпись _____

Бутаков Николай Алексеевич, к. т. н., старший научный сотрудник национального центра когнитивных разработок,
подпись _____

Петров Максим Владимирович, инженер национального центра когнитивных разработок, аспирант института дизайна и урбанистики, подпись _____

Гладилин Пётр Евгеньевич, к.ф.-м.н, научный сотрудник национального центра когнитивных разработок, подпись _____

Деева Ирина Юрьевна, аспирант института дизайна и урбанистики, инженер национального центра когнитивных разработок, подпись _____

1. ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

ВВЕДЕНИЕ

Курс «Большие данные» является частью образовательной программы для ИТ-классов средней школы.

Элементы обучения основам обработки и анализа данных вводятся с первого полугодия 10 класса с постепенным усложнением содержания соответственно возрасту обучающегося и заканчиваются во втором полугодии 11-го класса.

Курс носит междисциплинарный характер и может быть фактически разнесен между часами, отведенными на элективные дисциплины и внеурочную деятельность.

Предлагаемая программа соответствует положениям федерального государственного образовательного стандарта среднего общего образования.

Программа курса отражает способы формирования универсальных учебных действий, составляющих основу для профессионального самоопределения, саморазвития и непрерывного образования, выработки коммуникативных качеств, целостности общекультурного, личностного и познавательного развития учащихся.

Рабочая программа составлена на основе:

- Закона об образовании Российской Федерации
- Федерального государственного образовательного стандарта среднего общего образования.
- Профессионального стандарта «06.001 Разработка программного обеспечения»
- Профессионального стандарта «08.022 Статистическая деятельность»

Программа соответствует требованиям к структуре программ, заявленным в ФГОС, и включает следующие разделы:

- Пояснительная записка, в которой уточняются общие цели образования с учетом специфики курса.
- Общая характеристика курса, содержащая ценностные ориентиры образования по профилю «Большие данные».
- Место данного курса в учебном плане.
- Результаты освоения курса (личностные, метапредметные и предметные), соответствующие глобальным целям образования по профилю «Большие данные» и принципу развивающего обучения, лежащему в основе предлагаемой программы.
- Содержание курса по направлению «Большие данные» в 10 и 11 классах.
- Тематическое планирование, которое дает представление об основных видах учебной деятельности в процессе освоения курса в 10-11 классах основной школы.
- Рекомендации по учебно-методическому и материально-техническому обеспечению образовательного процесса.
- Планируемые результаты освоения программы.

ЦЕЛИ И ЗАДАЧИ РЕАЛИЗАЦИИ ОСНОВНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ ОСНОВНОГО ОБЩЕГО ОБРАЗОВАНИЯ ПО КУРСУ

Глобальными целями курса являются формирование у обучающегося аналитического мышления и, соответственно, знаний и умений, необходимых для успешного развития в отраслях, связанных со сложной аналитикой данных. Для достижения поставленных целей образование по данному направлению должно обеспечить решение следующих задач:

- **овладение** реальными и практическими знаниями методов статистического анализа данных;
- **формирование** навыков построения математических моделей (от нейронных сетей до кластеризации, от факторного до корреляционного анализов);
- **формирование** навыков работы с большими массивами данных;
- **осознание** практической важности нахождения уникальной закономерности в данных.

ПРИНЦИПЫ И ПОДХОДЫ К ФОРМИРОВАНИЮ ПРОГРАММЫ

Стандарт второго поколения (ФГОС) в сравнении со стандартом первого поколения предполагает деятельностный подход к обучению, где главная цель: развитие личности учащегося. Система образования отказывается от традиционного представления результатов обучения в виде знаний, умений и навыков. Формулировки стандарта указывают реальные виды деятельности, которыми следует овладеть к концу обучения, т. е. обучающиеся должны уметь учиться, самостоятельно добывать знания, анализировать, отбирать нужную информацию, уметь контактировать в различных по возрастному составу группах. Data scientist – это профессия, которая требует реальных и практических знаний методов статистического анализа данных, навыков построения математических моделей (от нейронных сетей до кластеризации, от факторного до корреляционного анализов), работы с большими массивами данных и уникальной способности находить закономерности. Будущему специалисту важно обладать не только фундаментальными, инженерными знаниями, но и разбираться в предметной области, в которой будет идти работа.

СОСТАВ УЧАСТНИКОВ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

Программа основного общего образования рассчитана на реализацию в 10 - 11 классах общеобразовательных учреждений и учреждений с углубленным изучением отдельных предметов, и нацелена на возрастную категорию учащихся 15 – 18 лет.

ОБЩАЯ ХАРАКТЕРИСТИКА УЧЕБНОГО КУРСА

Представленная программа направления «Большие данные (10-11 класс)» предназначена для практического освоения учащимися работы с технологиями информационного поиска и обработки больших данных, работы с инструментами анализа данных, основ математической статистики и теории вероятностей, основ математического моделирования. Программа рассчитана на 2 года (10-11 класс), при этом обучение можно разделить на 6 модулей:

□ Модуль «Введение в вероятностное моделирование» содержит основы исчисления вероятностей, вероятностного анализа данных и начальные сведения о вероятностных моделях, используемых для решения задач машинного обучения. В рамках модуля излагаются примеры применения изучаемых моделей, методов и алгоритмов, а также типовые алгоритмы решения задач реального мира с использованием вероятностных методов и моделей.

□ Модуль «Анализ и визуализация данных на Python» предполагает изучение основных методов, подходов и инструментов для анализа и визуализации данных с использованием возможностей Python и его основных библиотек.

□ Модуль «Параллельная обработка и управление большими данными» предполагает изучение теории баз данных, а также современных инструментов и технологий для решения задач, связанных с параллельной обработкой и анализом больших данных.

□ Модуль «Машинное обучение» предполагает изучение основных методов и моделей машинного, а также их реализацию на Python. В рамках модуля даются алгоритмы решения типовых проблем машинного обучения с примерами вариантов их применения в реальных задачах.

Содержание курса направлено на формирование универсальных учебных действий, обеспечивающих развитие познавательных и коммуникативных качеств личности. Обучающиеся включаются в проектную и исследовательскую деятельность, основу которой составляют такие учебные действия, как умение видеть закономерности в данных, строить на этих закономерностях модели, а также проводить валидацию и последующую доработку модели. Обучающиеся включаются в коммуникативную учебную деятельность, где преобладают такие ее виды, как умение полно и точно выражать свои мысли, аргументировать свою точку зрения, работать в группе, представлять и сообщать информацию в устной и письменной форме, вступать в диалог. Решение практических задач аналитики данных позволит ученикам освоить умение применять разработанные решения на практике, тем самым размышляя не только в контексте теоретической модели, но и охватывая предметную область.

Программа курса «Большие данные» для средней школы предусматривает реализацию следующих принципов:

Придать развитию знаний динамичный характер: уметь работать с технологиями информационного поиска, а также уметь анализировать большие объёмы данных.

Сконцентрировать учебный материал, укрупнив комплектные единицы знаний, что создает дидактические условия для развития системного мышления у учащихся: освободить учебный материал от деталей, имеющих специальное значение, но излишних для общего образования, группируя при этом частные понятия, необходимые для общего образования, вокруг ведущих понятий.

Формировать у обучающихся аналитическое мышление, сочетая его с активной познавательной и исследовательской деятельностью обучающихся.

Учитывать возрастные, индивидуальные особенности и возможности обучающихся, предлагая им задания по выбору, а также возможность самостоятельно собирать данные и разрабатывать модели на них.

ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ УЧЕБНОГО КУРСА

Деятельность образовательного учреждения в обучении по направлению «Большие данные» должна быть направлена на достижение обучающимися следующих **личностных результатов**:

готовность и способность к самостоятельной, творческой и ответственной деятельности;

навыки сотрудничества со сверстниками, детьми младшего возраста, взрослыми в образовательной, общественно полезной, учебно-исследовательской, проектной и других видах деятельности;

готовность и способность к образованию, в том числе самообразованию, на протяжении всей жизни;

сознательное отношение к непрерывному образованию как условию успешной профессиональной и общественной деятельности;

осознанный выбор будущей профессии и возможностей реализации собственных жизненных планов; отношение к профессиональной деятельности как возможности участия в решении личных, общественных, государственных, общенациональных проблем;

Метапредметными результатами освоения программы по направлению «Большие данные» являются:

1) умение самостоятельно определять цели и составлять планы в различных сферах деятельности, осознавая приоритетные и второстепенные задачи; самостоятельно осуществлять, контролировать и корректировать учебную, внеурочную и внешкольную деятельность с учётом предварительного планирования; использовать различные ресурсы для достижения целей;

выбирать успешные стратегии в трудных ситуациях;

2) умение продуктивно общаться и взаимодействовать с коллегами по совместной деятельности, учитывать позиции другого (совместное целеполагание и планирование общих способов работы на основе прогнозирования, контроль и коррекция хода и результатов совместной деятельности), эффективно разрешать конфликты;

3) владение навыками исследовательской и проектной деятельности (определение целей и задач, планирование проведения исследования, формулирование гипотез и плана их проверки; осуществление наблюдений и экспериментов, использование количественных и качественных методов обработки и анализа полученных данных; построение доказательств в отношении выдвинутых гипотез и формулирование выводов; представление результатов исследования в заданном формате, составление текста отчёта и презентации с использованием информационных и коммуникационных технологий);

Предметными результатами освоения программы по направлению «Большие данные» являются:

1) Владение базовыми элементами теории вероятностей, методов математической статистики и методов машинного обучения;

2) Умение находить закономерности в данных, разрабатывать математические модели и модели машинного обучения на эти данных;

3) Умение выполнять численный анализ данных и визуализировать полученные результаты на языке Python;

4) Владение практическим опытом решения задач с применением методов математической статистики и машинного обучения.

ЗАКЛЮЧЕНИЕ

Таким образом, данная учебная программа удовлетворяет требованиям Федерального государственного образовательного стандарта среднего общего образования по изучению дополнительных учебных предметов, курсов по выбору обучающихся.

2. СОДЕРЖАНИЕ КУРСА

Рабочей программой предусмотрен следующий тематический план, который представлен в таблице 1.

Таблица 1 Тематический план

№ п/п	Модуль	Наименование раздела	Количество часов
1 полугодие 10 класса			
1.	Введение в вероятностное моделирование	Вводное занятие. Что такое математическая модель?	2
2.		Интуитивные понятия теории вероятностей.	2
3.		Исчисление вероятностей и элементы комбинаторики. Текущий контроль.	2
4.		Условная и полная вероятность.	2
5.		Понятие случайной величины.	2
6.		Обработка результатов наблюдений. Понятие статистической оценки. Текущий контроль.	2
7.		Числовые оценки выборочных характеристик.	2
8.		Вероятностные модели случайной величины.	2
9.		Оценка параметров распределения случайной величины. Текущий контроль.	2
10.		Интервальные оценки и проверка статистических гипотез.	2
11.		Базовые понятия из линейной алгебры.	2
12.		Элементы многомерного статистического анализа и моделирования. Базовые элементы корреляционного анализа и регрессионного анализа. Текущий контроль.	2
13.		Понятие классификации и кластеризации. Как связаны эти две задачи? Чем классификация отличается от регрессии?	2
14.		Понятие градиента. Текущий контроль.	2
15.		Реализация итогового проекта.	2
16.		Презентация результатов итогового проекта.	2
2 полугодие 10 класса			
17.	Анализ и визуализация данных на Python	Анализ данных. Примеры и задачи	2
18.		Одномерный анализ данных. График функции. Гистограммы. Распределения.	2
19.		Вектора и матрицы. Текущий контроль.	2
20.		Введение в Python. Базовые операции.	2
21.		Библиотека numpy. Примеры.	2
22.		Библиотека pandas. Примеры. Текущий контроль.	2
23.		Библиотека matplotlib. Примеры.	2
24.		Понятие корреляции. Примеры на pandas и numpy	2
25.		Обучение с учителем. Примеры. Текущий контроль.	2
26.		Обучение без учителя. Примеры.	2

27.		Кластеризация данных на Python	2
28.		Линейная регрессия на Python. Текущий контроль.	2
29.		Логистическая регрессия на Python.	2
30.		Работа с изображениями в Python. Текущий контроль.	2
31.		Реализация итогового проекта.	2
32.		Презентация результатов итогового проекта.	2
1 полугодие 11 класса			
33.	Параллельная обработка и управление большими данными	Понятие обработки данных. Виды обработки данных. Виды баз данных.	2
34.		Типы данных, таблицы и отношения между ними. Реляционная модель данных.	2
35.		Введение в SQL. Примеры в PostgreSQL. Текущий контроль.	2
36		Понятие индекса. Виды индексов.	2
37		Проектирование баз данных. Цели проектирования. Нормализация данных. Проектирование базы данных в PostgreSQL.	2
38		Текстовые и бинарные форматы хранения данных json, csv, parquet. Текущий контроль.	2
39		Обработка данных в памяти. Продвинутый pandas. Знакомство с dataframe'ами. Примеры.	2
40		Колоночные базы данных (NoSQL для больших данных): HBase, ClickHouse	2
41		Основные понятия распределенной обработки данных. Текущий контроль.	2
42		Знакомство с Apache Spark (PySpark).	2
43		Парадигма MapReduce. Сравнение с Hadoop.	2
44		Параллельная и распределенная обработка больших данных средствами PySpark.	2
45		Разработка итогового проекта. Постановка задачи организации хранения и обработки данных. Текущий контроль.	2
46		Проектирование хранилища и процесса обработки данных.	2
47		Реализация итогового проекта.	2
48		Презентация результатов итогового проекта.	2
2 полугодие 11 класса			
49.	Введение в машинное обучение	Введение в машинное обучение	2
50.		Типология и метрики качества алгоритмов машинного обучения	2
51.		Метрические алгоритмы классификации. Текущий контроль.	2
52.		Логические алгоритмы классификации Введение в ансамблевые методы.	2
53.		Модели смесей распределений. Текущий контроль.	2

54.	Методы кластеризации и детектирования аномалий	2
55.	Методы снижения размерности многомерных данных	2
57.	Обучение с подкреплением. Текущий контроль.	2
58.	Введение в нейронные сети	2
59.	Многослойный перцептрон.	2
60.	Свёрточные нейронные сети. Текущий контроль.	2
61.	Рекуррентные нейронные сети	2
62.	Глубокое обучение без учителя. Текущий контроль.	2
63.	Постановка задачи для итогового проекта. Разработка итогового проекта	2
64.	Презентация результатов итогового проекта	2

Примечание. Разделы, относящиеся к одному модулю, могут быть реализованы в различных полугодиях. В том числе, возможно параллельное изучение материала нескольких модулей, если это обосновано логикой освоения материала.

Краткое содержание разделов:

1-е полугодие 10 класса.

- 1. Вводное занятие. Что такое математическая модель?** Основные принципы и подходы к моделированию. Классификация математических моделей. Бытовое понятие о вероятности.
- 2. Интуитивные понятия теории вероятностей.** Элементарные события. Случайные события. Алгебра событий. Примеры. Упражнения.
- 3. Исчисление вероятностей и элементы комбинаторики.** Сложение вероятностей совместных и несовместных событий. Перестановки, выборки и сочетания. Примеры. Упражнения. Текущий контроль.
- 4. Условная и полная вероятность.** Понятие условной вероятности. Формула Байеса. Теорема о полной вероятности. Упражнения.
- 5. Понятие случайной величины.** Дискретная случайная величина. Схемы повторения испытаний. Формула Пуассона. Законы распределения дискретной случайной величины. Непрерывная случайная величина. Функция распределения. Плотность распределения. Равномерный закон распределения. Нормальный закон распределения. Упражнения.
- 6. Обработка результатов наблюдений. Понятие статистической оценки.** Что такое статистические оценки и чем занимается математическая статистика? Эмпирическая функция распределения. Принципы построения гистограмм. Ядерная оценка плотности распределения. Упражнения. Текущий контроль.
- 7. Числовые оценки выборочных характеристик.** Выборочное среднее. Выборочная дисперсия. Выборочное среднее квадратическое отклонение. Упражнения.
- 8. Вероятностные модели случайной величины.** Выбор функции распределения как вероятностной модели случайной величины. Вероятностная модель как смесь распределений. Смесь распределений Гаусса. Примеры. Упражнения.

9. **Оценка параметров распределения случайной величины.** Точечные оценки параметров распределения. Метод максимального правдоподобия. Упражнения. Текущий контроль.
10. **Интервальные оценки и проверка статистических гипотез.** Понятие доверительного интервала как модели для ошибки оцененных параметров. Понятие статистической гипотезы и критериев для ее проверки. Непараметрические и параметрические критерии. Упражнения.
11. **Базовые понятия из линейной алгебры.** Понятие вектора и матрицы. Операции над матрицами. Матричные произведения. Специальные виды матриц. Обратная матрица. Понятие СЛАУ. Методы решения СЛАУ (обзорно). Метод Гаусса. Упражнения.
12. **Элементы многомерного статистического анализа и моделирования. Базовые элементы корреляционного анализа и регрессионного анализа.** Выборочные коэффициенты корреляции. Корреляционная матрица. Уравнение прямой и задача регрессии. Множественная регрессия. Оценка качества регрессионной модели. Понятие вероятностного интервала. Примеры. Упражнения. Текущий контроль.
13. **Понятие классификации и кластеризации. Как связаны эти две задачи? Чем классификация отличается от регрессии?** Расстояние между объектами. Типы кластеров. Методы кластеризации (обзорно). Метод ближайшего соседа. Метод k-means. Кластеризация с помощью вероятностных моделей: разделение смеси Гауссовых распределений (дискриминантный анализ). Методы классификации. Логистическая регрессия. Упражнения.
14. **Понятие градиента.** Использование градиента в задачах оптимизации и машинного обучения. Метод градиентного спуска. Стохастический градиентный спуск. Примеры. Упражнения. Текущий контроль.
15. **Реализация итогового проекта.** Постановка задачи по построению одномерной вероятностной модели. Рекомендации по выполнению.
16. **Презентация результатов итогового проекта.**

2-е полугодие 10 класса.

17. **Анализ данных. Примеры и задачи**
Какие бывают данные. Понятия числовых, категориальных данных. Способы представления информации. Основные задачи анализа данных: классификация, регрессия, кластеризация (повторение).
18. **Одномерный анализ данных. График функции. Гистограммы. Распределения.**
Понятие функции и аргумента. Зависимость и независимость. Построение графика функции по табличным значениям. Понятие гистограммы как способа представления табличных данных, примеры (повторение). Понятие распределения (повторение) и способы визуализации различных распределений.
19. **Вектора и матрицы.**
Понятие вектора и понятие матрицы и их физический смысл. Размерность матриц. Связь матриц и таблиц данных. Линейная зависимость и линейная независимость векторов. Понятия коллинеарности и компланарности. Основные операции над матрицами и векторами: сложение, скалярное произведение, умножение матриц и их физический смысл (повторение и углубление). Транспонирование матриц. Обратные матрицы.

Системы линейных алгебраических уравнений и способы их решения (повторение и углубление). Текущий контроль.

20. Введение в Python. Базовые операции.

Базовые типы данных в Python: численные, строковые, логические переменные. Циклы. Функции. Структуры данных в Python: списки, множества и словари - примеры создания и основные операции с ними. Понятие list comprehension. Пример реализации функции одной переменной. Импорт модулей и функций.

21. Библиотека numpy. Примеры.

Основные конструкции библиотеки numpy как библиотеки для высокопроизводительных вычислений. Векторизация вычислений. Создание массивов, одномерные и многомерные массивы. Вычисление основных статистических показателей матрицы с помощью numpy: минимум, максимум, среднее, argmax и др. Примеры. Текущий контроль.

22. Библиотека pandas. Примеры.

Основные конструкции библиотеки pandas. Чтение файлов и запись в файл. Понятие pandas.DataFrame и pandas.Series. Выгрузка данных по условию. Создание таблиц. Агрегация и слияние имеющихся данных. Выполнение сложных запросов к датасету.

23. Библиотека matplotlib. Примеры.

Библиотека matplotlib и визуализация данных. Построение графика функции и создание своего стиля для графика. Линейные и логарифмические шкалы, выбор масштаба представления данных. Гистограммы в matplotlib. Примеры построения нескольких независимых графиков в одном окне: метод subplots(). Сохранение графиков в виде изображения. Текущий контроль.

24. Понятие корреляции. Примеры на pandas и numpy

Понятие корреляции (повторение). Ложные корреляции. Виды зависимостей данных друг от друга. Понятие кросс-корреляции, автокорреляции и свёртки. Понятие ранговых списков. Корреляция Пирсона и корреляция Спирмена. Вычисление попарных корреляций и корреляционных таблиц средствами numpy и pandas. Heatmap и графическое представление таблиц данных.

25. Обучение с учителем. Примеры.

Задача обучения с учителем. Обучение по прецедентам. Объекты и целевые переменные. Понятие функции ошибок. Тренировочная и тестовая выборка. Задачи классификации и регрессии – сходства и различия. Данные для обучения в виде таблиц значений. Текущий контроль.

26. Обучение без учителя. Примеры.

Задачи обучения без учителя и data mining. Обзор: кластеризация, корреляционный анализ, понижение размерности. Алгоритмы кластеризации (повторение).

27. Кластеризация данных на Python

Библиотека scikit-learn и её использование для кластеризации данных в Python. Изменение параметров методов кластеризации и проверка качества кластеризации. Метрики кластеризации и их реализация в Python. Реализация алгоритмов KNN, SVM и Kmeans в библиотеке scikit-learn. Примеры и визуализация. Текущий контроль.

28. Линейная регрессия на Python.

Понятие линейной регрессии (повторение). Понятие весовых коэффициентов и настройка параметров модели. Отбор признаков и работа с данными. Скалирование и центрирование данных. недообучение и переобучение. Понятие регуляризации. Регуляризация модели линейной регрессии – подходы Lasso и Ridge и их отличия. Случай нелинейной зависимости, полиномиальная регрессия. Примеры и упражнения на Python.

29. Логистическая регрессия на Python.

Логистическая регрессия как модель бинарной классификации. Целевая переменная и виды функции ошибок для задач классификации. Понятие функции активации и её виды: линейная, сигмоида, гиперболический тангенс, ReLU. Примеры и упражнения на Python. Текущий контроль.

30. Работа с изображениями в Python.

Изображение как матрица. Понятие RGB изображения и примеры других цветовых пространств. Понятие яркости и контраста. Основы обработки изображений: фильтрация, бинаризация, выделение границ, размытие. Загрузка изображений в Python и использование библиотеки matplotlib для работы с изображениями. Примеры и упражнения по обработке изображений в Python.

31. Подготовка итогового проекта.

Реализация итогового проекта по построению модели бинарной классификации в Python с помощью библиотек numpy, pandas, scikit-learn и matplotlib.

32. Презентация результатов итогового проекта.

1-е полугодие 11 класса.

33. Понятие обработки данных. Виды обработки данных. Виды баз данных. Обработка цифровой, символьной, текстовой и табличной информации. Реляционные и NoSQL базы данных, их отличия, области применения, примеры использования.

34. Типы данных, таблицы и отношения между ними. Реляционная модель данных. Строковые, целочисленные, дробные, дата, время. Понятие таблицы, ключа. Нормальные формы. Ключи, первичные и внешние ключи.

35. Введение в SQL. Примеры в PostgreSQL. Создание таблиц, вставка, выборка, удаление, изменение данных. Создание ключей на колонки. Выборка данных из таблиц, фильтрация, сортировки, группировки, слияние, подзапросы. Текущий контроль.

36. Понятие индекса. Виды индексов. Индексы в базах данных: назначение, влияние на производительность, принципы создания индексов. Индексы: По порядку сортировки, источнику данных, воздействию на источник данных, структуре, количественному составу, характеристике содержимого, механизму обновления, покрытию индексируемого содержимого,

37. Проектирование баз данных. Цели проектирования. Нормализация данных. Проектирование базы данных в PostgreSQL. Основные задачи и этапы проектирования баз данных. Концептуальное, Логическое, Физическое проектирование.

38. Текстовые и бинарные форматы хранения данных json, csv, parquet.

39. Обработка данных в памяти. Продвинутый pandas. Знакомство с dataframe'ами. Примеры. Чтение данных, обработка и запись в различные форматы.

40. **Колоночные базы данных (NoSQL для больших данных): HBase, ClickHouse.** Понятие колонки, способ представления, отличия от строкового представления. Достоинства и недостатки КБД.
41. **Основные понятия распределенной обработки данных. Архитектура «клиент-сервер», «мастер-воркер».** Достоинства и недостатки распределенной обработки данных. Способы распределения данных: централизованный, децентрализованный, смешанный
42. **Знакомство с Apache Spark (PySpark).** Компоненты экосистемы Apache Spark, Особенности Apache Spark, RDD и особенности использования, трансформации и действия.
43. **Парадигма MapReduce. Сравнение с Hadoop.** Понятие shuffle, виды реализации shuffle в spark.
44. **Параллельная и распределенная обработка больших данных средствами PySpark.** Знакомство со Spark-shell, Написание программ в Apache Spark, чтение и запись данных. Понятие DataFrame, Использование DataFrame вместо RDD, простые запросы, фильтрация и агрегация. Продвинутые операции: join, broadcast, udf, udaf.
45. **Разработка итогового проекта. Постановка задачи организации хранения и обработки данных.**
46. **Проектирование хранилища и процесса обработки данных.**
47. **Программная реализация проекта.**
48. **Презентация результатов итогового проекта.**

2-е полугодие 11 класса.

49. **Введение в машинное обучение.** Классификация моделей искусственного интеллекта по Расселу и Норвигу. Имитация когнитивных функций человека современными моделями машинного обучения. Определения машинного обучения. Опыт, задача, качество решения. Способы задания входных данных для алгоритма машинного обучения. Обобщающая способность модели. Дилемма смещения-разброса, понятие недообучения и переобучения.
50. **Типология и метрики качества алгоритмов машинного обучения.** Обучение с учителем (карта методов). Разметка данных. Функции потерь. Тренировочная, тестовая (контрольная), валидационная (проверочная) выборка. Кросс-валидация. Метрики качества бинарной классификации. ROC-AUC и Precision-Recall кривые. Метрики качества для несбалансированных выборок. Обучение без учителя. Метрики качества для оценки результатов кластеризации. Модулярность. Коэффициент силуэта. Semi-supervised обучение. Обучение с подкреплением.
51. **Метрические методы классификации.** Метрики расстояния (Манхэттенская, евклидово расстояние, косинусное расстояние). Метод ближайших соседей. Подбор числа соседей. Метод опорных векторов для случая линейно разделимой выборки.
52. **Логические классификаторы.** Решающие правила. Конструирование решающих правил. Решающие деревья. Метрики информативности. Подрезка решающих деревьев.
53. **Введение в ансамблевые методы.** Бэггинг. Случайный лес. Бустинг. Алгоритм AdaBoost.
54. **Модели смесей распределений.** Обучение смеси гауссианов. EM-алгоритм. Алгоритмы тематического моделирования. Вероятностный латентно-семантический анализ.
55. **Методы кластеризации и детектирования аномалий.** Алгоритм K средних. Применение EM-алгоритма для алгоритма K средних. Иерархическая кластеризация. Интерпретация дендрограмм. Кластеризация на графах. Методы детектирования аномалий.
56. **Методы снижения размерности многомерных данных.** Многомерное шкалирование. Метод главных компонент. Методы обучения представлений для текстовых данных. Методы обучения представлений для графовых данных.

- 57. Обучение с подкреплением.** Агент и среда. Система подкрепления. Способы обучения с подкреплением. Задача о многоруком бандите. Q-обучение.
- 58. Введение в нейронные сети.** Типология нейронных сетей. Однослойные модели нейронных сетей. Правило Хебба. Карты Кохонена.
- 59. Многослойный перцептрон.** Алгоритм обратного распространения ошибки. Способы борьбы с переобучением для нейронных сетей.
- 60. Сверточные нейронные сети.** Принцип построения иерархических признаков. Архитектура сверточной нейронной сети. Слои свертки и субдискретизации. Реализация операций пулинга. Современные архитектуры сверточных нейросетей: ImageNet, VGG16.
- 61. Рекуррентные нейронные сети.** Архитектура сети RNN. Архитектура сети LSTM. Архитектура сети GRU. Примеры использования рекуррентных сетей в области машинного перевода и прогнозирования временных рядов.
- 62. Глубокое обучение без учителя.** Ограниченная машина Больцмана. Автоэнкодеры. Сети глубокого доверия. Генеративно-сопоставительная сеть.
- 63. Постановка задачи для итогового проекта. Разработка итогового проекта.**
- 64. Презентация результатов итогового проекта**

3. ПОУРОЧНО-ТЕМАТИЧЕСКОЕ ПЛАНИРОВАНИЕ

№ п/п	Номер урока	Тема урока	Тип урока, форма проведения	Планируемые результаты (личностные, предметные, метапредметные)	Средства обучения	Примечания, домашние задания
1	1-4	Вводное занятие. Что такое математическая модель?	Тип: урок открытия нового знания. Форма: комбинированный урок. Предложить придумать свой пример модели объекта реального мира.	7.1.1. Базовые определения	Визуальные (слайды), ПК, электронные образовательные ресурсы	
2	5-8	Интуитивные понятия теории вероятностей.	Тип: урок систематизации, урок отработки умений. Форма: комбинированный урок. Разбор примеров случайных событий, предложить придумать свои примеры случайных и элементарных событий.	7.1.1. Базовые определения	Визуальные (слайды), ПК, электронные образовательные ресурсы	
3	9-12	Исчисление вероятностей и элементы комбинаторики.	Тип: урок открытия нового знания, урок отработки умений, урок развивающего контроля. Форма: комбинированный урок. Решение задач на классическое определение вероятности. Повторение формул комбинаторики – перестановки, выборки и сочетания.	7.1.2. Основы исчисления вероятностей случайных событий	Визуальные (слайды), ПК, электронные образовательные ресурсы	
4	13-16	Условная и полная вероятность.	Тип: урок открытия нового знания, урок систематизации. Форма: комбинированный урок. Разбор примера на условную вероятность. Сравнить два случая – независимые события и зависящие друг от друга. Решение задач на теорему Байеса, нахождение полной вероятности события.	7.1.1. Базовые определения 7.1.2. Основы исчисления вероятностей случайных событий	Визуальные (слайды), ПК, электронные образовательные ресурсы	

5	17-20	Понятие случайной величины.	Тип: урок открытия нового знания, урок систематизации. Форма: комбинированный урок. Предложить найти свои примеры дискретной и случайной величины. Понятие нормального закона распределения. Примеры законов распределения в реальной жизни. Понятия функции и плотности распределения.	7.1.3. Понятие случайной величины. Функция распределения	Визуальные (слайды), ПК, электронные образовательные ресурсы
6	21-24	Обработка результатов наблюдений. Понятие статистической оценки.	Тип: урок открытия нового знания, урок систематизации, урок развивающего контроля. Форма: комбинированный урок. Предложить построить гистограмму для своей случайной величины, которая может быть собрана самими школьниками (температура воздуха за месяц, вес одноклассников и т. д.). Примеры автоматического построения гистограмм в Excel.	7.2.1. Понятие статистической оценки. Оценки параметров распределений. Интервальные оценки	Визуальные (слайды), ПК, электронные образовательные ресурсы
7	25-28	Числовые оценки выборочных характеристик.	Тип: урок систематизации, урок отработки умений. Форма: комбинированный урок. Предложить найти выборочное среднее, дисперсию и среднее квадратическое отклонение для своей случайной величины.	7.2.1. Понятие статистической оценки. Оценки параметров распределений. Интервальные оценки	Визуальные (слайды), ПК, электронные образовательные ресурсы
8	29-32	Вероятностные модели случайной величины.	Тип: урок открытия нового знания, урок систематизации. Форма: комбинированный урок.	7.1.3. Понятие случайной величины. Функция распределения 7.2.1. Понятие статистической оценки. Оценки параметров распределений. Интервальные оценки	Визуальные (слайды), ПК, электронные образовательные ресурсы
9	33-36	Оценка параметров распределения случайной величины.	Тип: урок открытия нового знания, урок отработки умений, урок развивающего контроля. Форма: комбинированный урок. Предложить провести точечную оценку параметров распределения своей случайной величины.	7.1.3. Понятие случайной величины. Функция распределения 7.2.1. Понятие статистической оценки. Оценки параметров распределений. Интервальные оценки	Визуальные (слайды), ПК, электронные образовательные ресурсы

10	37-40	Интервальные оценки и проверка статистических гипотез.	Тип: урок открытия нового знания, урок систематизации. Форма: семинар. Предложить построить доверительный интервал для параметров распределения своей случайно величины.	7.2.1. Понятие статистической оценки. Оценки параметров распределений. Интервальные оценки	Визуальные (слайды), ПК, электронные образовательные ресурсы
11	41-44	Базовые понятия из линейной алгебры.	Тип: урок открытия нового знания, урок систематизации. Форма: комбинированный урок, конференция. Решение задач на операции с матрицами. Решение систем линейных алгебраических уравнений.	7.3.1. Основные понятия. Типология методов машинного обучения. Численные методы нахождения лучшего решения.	Визуальные (слайды), ПК, электронные образовательные ресурсы
12 45-48		Элементы многомерного статистического анализа и моделирования. Базовые элементы корреляционного анализа и регрессионного анализа.		7.2.2. Анализ многомерной случайной величины. Методы снижения размерности. 7.2.3. Основные понятия корреляционного анализа. Парный коэффициент корреляции. Корреляционная матрица. 7.2.4. Основные понятия линейного регрессионного анализа. Парная регрессия. Множественная регрессия.	Визуальные (слайды), ПК, электронные образовательные ресурсы
13 49-52		Понятие классификации и кластеризации. Как связаны эти две задачи? Чем классификация отличается от регрессии?		7.2.4. Основные понятия линейного регрессионного анализа. Парная регрессия. Множественная регрессия. 7.4.1. Понятие кластера. Типы расстояний. 7.5.2. Методы решения задачи классификации. Методы решения задачи регрессии.	Визуальные (слайды), ПК, электронные образовательные ресурсы
14	53-56	Понятие градиента.	Тип: урок открытия нового знания, урок систематизации, урок отработки умений. Форма: комбинированный урок. Визуализация метода градиентного спуска.	7.3.1. Основные понятия. Типология методов машинного обучения. Численные методы нахождения лучшего решения.	Визуальные (слайды), ПК, электронные образовательные ресурсы

					бные ресурсы	
15	57-60	Реализация итогового проекта.	Тип: урок отработки умений, урок развивающего контроля. Форма: комбинированный урок. Постановка задачи итогового проекта. Сбор данных для проекта. Предложить ученикам собрать измерения для проекта. Затем построить на данных измерениях модель, предсказывающую дальнейшие значения величин.	7.2.1. Понятие статистической оценки. Оценки параметров распределений. Интервальные оценки 7.3.1. Основные понятия. Типология методов машинного обучения. Численные методы нахождения лучшего решения.	Визуальные (слайды), ПК, электронные образовательные ресурсы	
16	61-64	Презентация результатов итогового проекта.	Тип: урок развивающего контроля. Форма: конференция. Подготовка презентации проекта. Выступление с презентацией.	7.2.1. Понятие статистической оценки. Оценки параметров распределений. Интервальные оценки 7.3.1. Основные понятия. Типология методов машинного обучения. Численные методы нахождения лучшего решения.	Визуальные (слайды), ПК, электронные образовательные ресурсы	
Анализ и визуализация данных в Python						
17	65-68	Анализ данных. Примеры и задачи	Тип: урок открытия нового знания. Форма: комбинированный урок. Предложить придумать примеры числовых и категориальных данных. Предложить привести примеры задач классификации, регрессии и кластеризации (можно примеры своего проекта из предыдущего модуля)	4.11.1 Искусственный интеллект 4.11.2 Машинное обучение 7.3.1. Основные понятия. Типология методов машинного обучения. Численные методы нахождения лучшего решения.		
18	69-72	Одномерный анализ данных. График функции. Гистограммы. Распределения.	Тип: урок открытия нового знания, урок отработки умений. Форма: комбинированный урок. Рассмотреть понятие функции и аргумента. Построить график функции по табличным значениям. Построить гистограмму по табличным значениям вручную и с помощью Excel (повторение)	3.2.6. Табличные величины. Массивы 7.1.3. Понятие случайной величины. Функция распределения		
19	73-76	Вектора и матрицы.	Тип: урок открытия нового знания, урок отработки умений, урок систематизации. Форма: комбинированный урок. Рассмотреть понятие вектора и матрицы (повторение). Решение задач на операции с матрицами и векторами.	3.2.6. Табличные величины. Массивы		

20	77-80	Введение в Python. Базовые операции.	Тип: урок открытия нового знания, урок отработки умений. Форма: комбинированный урок. Базовые понятия языка Python. Запуск и создание скрипта в имеющейся среде выполнения (Jupyter, PyCharm, Visual Studio, Eclipse). Написание скрипта, реализующего функцию одной переменной.	3.2.1. Интерфейс выбранного языка программирования. 3.2.2. Типы и структуры данных. Операции с ними. 3.2.3. Синтаксис и семантика выбранного языка программирования 3.2.4. Переменные и их описание 3.2.6. Табличные величины. Массивы 3.2.7. Средства работы с данными во внешней памяти. Файлы.		
21	81-84	Библиотека numpy. Примеры.	Тип: урок открытия нового знания, урок отработки умений. Форма: комбинированный урок. Основные конструкции библиотеки numpy. Написание скрипта на вычисление статистических показателей матрицы данных, использованных в предыдущем модуле (можно взять данные из индивидуального проекта).	3.2.6. Табличные величины. Массивы 3.5.8. - Построение и исследование моделей из различных предметных областей с использованием выбранных языков программирования и электронных таблиц.	numpy	
22	85-88	Библиотека pandas. Примеры.	Тип: урок открытия нового знания, урок отработки умений. Форма: комбинированный урок. Основные конструкции библиотеки pandas. Написание скрипта на чтение данных из файла, созданию таблиц и их слиянию между собой. Предложить сделать сравнение между библиотеками numpy и pandas.	3.5.8. - Построение и исследование моделей из различных предметных областей с использованием выбранных языков программирования и электронных таблиц.	pandas	
23	89-92	Библиотека matplotlib. Примеры.	Тип: урок открытия нового знания, урок отработки умений. Форма: комбинированный урок. Предложить визуализировать данные, которые использовались в индивидуальном проекте. Построить несколько независимых графиков в одном окне. Предложить сравнить возможности matplotlib с возможностями Excel.	3.5.8. - Построение и исследование моделей из различных предметных областей с использованием выбранных языков программирования и электронных таблиц.	numpy, matplotlib	
24	93-96	Понятие корреляции. Примеры на pandas и numpy	Тип: урок открытия нового знания, урок отработки умений, урок систематизации. Форма: комбинированный урок. Понятие корреляции (повторение). Посмотреть на корреляционные функции в Python. Написание скрипта на визуализацию корреляционной матрицы средствами matplotlib.	3.5.8. - Построение и исследование моделей из различных предметных областей с использованием выбранных языков программирования и электронных таблиц 7.2.3. Основные понятия корреляционного анализа. Парный коэффициент корреляции. Корреляционная матрица.	numpy, pandas	

25	97-100	Обучение учителем. Примеры.	с	<p>Тип: урок открытия нового знания, урок отработки умений. Форма: комбинированный урок.</p> <p>Предложить привести свои примеры задачи обучения с учителем для классификации и регрессии. Предложить поразмышлять, зачем делить данные на тренировочные и тестовые.</p>	<p>3.5.8. - Построение и исследование моделей из различных предметных областей с использованием выбранных языков программирования и электронных таблиц.</p> <p>7.3.2. Понятие обучения и тестирования модели. Обобщающая способность модели.</p> <p>7.5.1. Основные понятия. Методы оценивания качества. Кросс-валидация.</p> <p>7.5.2. Методы решения задачи классификации. Методы решения задачи регрессии.</p>	numpy, scikit-learn	
26	101-104	Обучение учителя. Примеры.	без	<p>Тип: урок открытия нового знания, урок отработки умений. Форма: комбинированный урок.</p> <p>Предложить привести свои примеры задачи обучения без учителя. Ответить на вопрос, почему задача обучения без учителя сложнее. Алгоритмы кластеризации (повторение).</p>	<p>3.5.8. - Построение и исследование моделей из различных предметных областей с использованием выбранных языков программирования и электронных таблиц.</p>	numpy, scikit-learn	
27	105-108	Кластеризация данных на Python		<p>Тип: урок открытия нового знания, урок отработки умений, урок систематизации. Форма: комбинированный урок.</p> <p>Написание скрипта на использование библиотеки scikit-learn для кластеризации данных. Визуальное представление кластеризации данных в Python. Предложить собрать данные (измерения) и кластеризовать их одним из алгоритмов.</p>	<p>3.5.8. - Построение и исследование моделей из различных предметных областей с использованием выбранных языков программирования и электронных таблиц.</p> <p>7.4.1. Понятие кластера. Типы расстояний.</p> <p>7.4.2. Метрические методы кластеризации. Центроидные методы.</p>	numpy, scikit-learn, matplotlib	
28	109-112	Линейная регрессия на Python.	на	<p>Тип: урок открытия нового знания, урок отработки умений, урок систематизации. Форма: комбинированный урок.</p> <p>Понятие линейной регрессии (повторение). Примеры полиномиальной регрессии. Рассмотреть примеры регуляризации модели – лассо и Ridge. Предложить датасет, разбить его на участки и посмотреть, какие коэффициенты оставляет лассо на каждом участке, сделать выводы.</p>	<p>3.5.8. - Построение и исследование моделей из различных предметных областей с использованием выбранных языков программирования и электронных таблиц.</p> <p>7.2.4. Основные понятия линейного регрессионного анализа. Парная регрессия. Множественная регрессия.</p>	numpy, scikit-learn, matplotlib	
29	113-116	Логистическая регрессия на Python.	на	<p>Тип: урок открытия нового знания, урок отработки умений. Форма: комбинированный урок.</p> <p>Понятие логистической регрессии. Рассмотреть примеры логистической регрессии с разной функцией активации.</p>	<p>3.5.8. - Построение и исследование моделей из различных предметных областей с использованием выбранных</p>	numpy, scikit-learn, matplotlib	

			Предложить применить разные функции активации для своих данных и визуализировать результаты.	языков программирования и электронных таблиц. 7.5.2. Методы решения задачи классификации. Методы решения задачи регрессии.		
30	117-120	Работа с изображениями в Python.	Тип: урок открытия нового знания, урок отработки умений. Форма: комбинированный урок. Предложить собрать свой датасет изображений, загрузить их средствами Python и применить основные функции обработки изображений.	3.3.5. Вспомогательные алгоритмы и программы. Процедуры (подпрограммы) и библиотеки подпрограмм.	numpy, matplotlib	
31	121-124	Реализация итогового проекта.	Тип: урок отработки умений, урок развивающего контроля. Форма: комбинированный урок. Постановка задачи. Взять данные из предыдущего модуля и проделать всю обработку данных, а также статистический анализ средствами языка Python.		numpy, pandas, scikit-learn, matplotlib	
32	125-128	Презентация результатов итогового проекта.	Тип: урок развивающего контроля. Форма: семинар. Презентация результатов.		numpy, pandas, scikit-learn, matplotlib	
Параллельная обработка и управление большими данными						
33	129-140	Понятие обработки и хранения данных.	Тип: урок открытия нового знания - основные понятия, принципы обработки и хранения данных. Урок отработки умений. Форма: комбинированный урок.	8.1.1. Базовые понятия.		
34	141-145	Реляционная модель данных	Тип: урок открытия нового знания - основные понятия реляционной модели данных, типы данных, таблицы и отношения между ними. Урок отработки умений. Форма: комбинированный урок.	8.1.2. Типы данных, таблицы и отношения между ними. Реляционная модель данных.		
35	146-150	Введение в SQL	Тип: урок открытия нового знания - основные понятия языка SQL, работа с Postgresql Форма: комбинированный урок.	8.1.3. SQL запросы	Postgresql	

36	151-153	Индексы	Тип: урок открытия нового знания - индексы в базах данных. Форма: комбинированный урок.	8.1.4. Понятие индекса. Виды индексов.	Postgresql	
37	154-156	Проектирование баз данных.		8.1.5. Основные задачи и этапы проектирования баз данных. Концептуальное, Логическое, Физическое проектирование.	Postgresql	
38	157-160	Форматы хранения данных	Тип: урок открытия нового знания - основные форматы данных, урок отработки умений. Форма: комбинированный урок.	8.2.1. Основные понятия. Текстовые и бинарные форматы хранения данных		
39	161-164	Обработка данных в памяти Pandas	Тип: урок открытия нового знания - основные возможности библиотеки Pandas, урок отработки умений. Форма: комбинированный урок.	8.2.2. Обработка данных в памяти с помощью pandas	Pandas	
40	165-168	NOSQL Базы данных	Тип: урок открытия нового знания - NOSQL базы данных, урок отработки умений. Форма: комбинированный урок.	8.3.1. Основные понятия. Типы NOSQL баз данных. Колоночные NOSQL БД.		
41	169-172	Распределенная обработка данных	Тип: урок открытия нового знания - Распределенная обработка данных, урок отработки умений. Форма: комбинированный урок.	8.4.1. Основные понятия. Архитектура «клиент-сервер», «мастер-воркер»		
42	173-176	Знакомство с Apache Spark	Тип: урок открытия нового знания - библиотека распределенной обработки данных Apache Spark, урок отработки умений. Форма: комбинированный урок.	8.4.2. Apache Spark (PySpark). Парадигма MapReduce.	Apache Spark	
43	177-183	Работа с PySpark	Тип: урок открытия нового знания - библиотека распределенной обработки данных Apache PySpark, урок отработки умений. Форма: комбинированный урок.	8.4.2. Apache Spark (PySpark). Парадигма MapReduce.	PySpark	

44	184-185	Разработка итогового проекта.	Тип: урок отработки умений, урок развивающего контроля. Форма: комбинированный урок.			
45	186-189	Проектирование хранилища и процесса обработки данных.	Тип: урок отработки умений, урок развивающего контроля. Форма: комбинированный урок.		PySpark	
46	190-196	Реализация итогового проекта.	Тип: урок отработки умений, урок развивающего контроля. Форма: комбинированный урок.		PySpark	
47	197-198	Презентация результатов итогового проекта.	Тип: урок развивающего контроля. Форма: семинар.			
Введение в машинное обучение						
	199-202	Введение в машинное обучение	Тип: урок открытия нового знания, урок систематизации. Форма: комбинированный урок. Предложить привести примеры машинного обучения в различных сервисах и сферах жизни.	7.3.1. Основные понятия. Типология методов машинного обучения. Численные методы нахождения лучшего решения.	Визуальные (слайды), ПК, электронные образовательные ресурсы	
49	203-206	Типология и метрики качества алгоритмов машинного обучения	Тип: урок открытия нового знания, урок систематизации, урок отработки умений. Форма: комбинированный урок. Рассмотреть основные метрики качества для задач классификации. Реализовать алгоритм классификации на своих данных (повторение) и оценить качество его работы с помощью изученных метрик. Сделать выводы.	7.3.2. Понятие обучения и тестирования модели. Обобщающая способность модели. 7.3.3. Оценка качества модели. Основные метрики. Дилемма смещения и разброса. 7.5.1. Основные понятия. Методы оценивания качества. Кросс-валидация.	Визуальные (слайды), ПК, электронные образовательные ресурсы	
50	207-210	Метрические алгоритмы классификации	Тип: урок открытия нового знания. Форма: комбинированный урок. Метрики расстояния. Реализовать алгоритм К ближайших соседей для разного числа соседей и сравнить их с помощью метрик качества.	7.5.2. Методы решения задачи классификации. Методы решения задачи регрессии.	Визуальные (слайды), ПК, электронные образовательные ресурсы	

51	211-214	Логические алгоритмы классификации	Тип: урок открытия нового знания. Форма: комбинированный урок. Решающие правила и решающие деревья. Разработать своё решающее дерево на своих данных.	7.5.2. Методы решения задачи классификации. Методы решения задачи регрессии.	Визуальные (слайды), ПК, электронные образовательные ресурсы
52	215-218	Введение в ансамблевые методы	Тип: урок открытия нового знания, урок систематизации. Форма: комбинированный урок. Бэггинг, случайный лес, бустинг – обзорно рассмотреть. Сравнить ансамблевые методы с рассмотренными прежде алгоритмами.	7.5.2. Методы решения задачи классификации. Методы решения задачи регрессии.	Визуальные (слайды), ПК, электронные образовательные ресурсы
53	219-222	Модели смесей распределений	Тип: урок открытия нового знания, урок отработки умений. Форма: комбинированный урок. Рассмотреть модель смеси распределений на примере задачи тематического моделирования – обзорно.	7.5.2. Методы решения задачи классификации. Методы решения задачи регрессии.	Визуальные (слайды), ПК, электронные образовательные ресурсы
54	223-226	Методы кластеризации и детектирования аномалий	Тип: урок открытия нового знания, урок систематизации. Форма: комбинированный урок. Методы кластеризации. Загрузить открытый датасет с температурами воздуха (например, с платформы Kaggle) и кластеризовать их изученными алгоритмами.	7.4.1. Понятие кластера. Типы расстояний. 7.4.2. Метрические методы кластеризации. Центроидные методы. 7.4.3. Вероятностные методы кластеризации. Дискриминантный анализ.	Визуальные (слайды), ПК, электронные образовательные ресурсы
55	227-230	Методы снижения размерности многомерных данных	Тип: урок открытия нового знания, урок отработки умений. Форма: комбинированный урок. Сравнить качество работы модели машинного обучения на полных исходных данных и после снижения размерности, сделать выводы.	7.2.2. Анализ многомерной случайной величины. Методы снижения размерности.	Визуальные (слайды), ПК, электронные образовательные ресурсы

56	213-234	Обучение с подкреплением	<p>Тип: урок открытия нового знания., Форма: комбинированный урок.</p> <p>Системы обучения с подкреплением – обзорно. Предложить привести свои примеры обучения с подкреплением.</p>	7.5.2. Методы решения задачи классификации. Методы решения задачи регрессии.	Визуальные (слайды), ПК, электронные образовательные ресурсы
57	235-238	Введение в нейронные сети	<p>Тип: урок открытия нового знания. Форма: комбинированный урок.</p> <p>Рассмотреть типологию нейронных сетей, сравнить с биологическим прототипом.</p>	7.5.2. Методы решения задачи классификации. Методы решения задачи регрессии. 7.5.3. Искусственные нейронные сети. Перцептрон. Сверточные нейронные сети. Рекуррентные нейронные сети.	Визуальные (слайды), ПК, электронные образовательные ресурсы
58	239-242	Многослойный перцептрон	<p>Тип: урок открытия нового знания, урок систематизации, урок отработки умений. Форма: комбинированный урок.</p> <p>Алгоритм обратного распространения ошибки. Реализация алгоритма на языке Python.</p>	7.5.2. Методы решения задачи классификации. Методы решения задачи регрессии. 7.5.3. Искусственные нейронные сети. Перцептрон. Сверточные нейронные сети. Рекуррентные нейронные сети.	Визуальные (слайды), ПК, электронные образовательные ресурсы
59	240-243	Сверточные нейронные сети	<p>Тип: урок открытия нового знания, урок систематизации, урок отработки умений. Форма: комбинированный урок.</p> <p>Архитектура свёрточных нейронных сетей – обзорно. Современные архитектуры и средства реализации. Предложить привести примеры задач для свёрточных нейронных сетей.</p>	7.5.2. Методы решения задачи классификации. Методы решения задачи регрессии. 7.5.3. Искусственные нейронные сети. Перцептрон. Сверточные нейронные сети. Рекуррентные нейронные сети.	Визуальные (слайды), ПК, электронные образовательные ресурсы
60	244-247	Рекуррентные нейронные сети	<p>Тип: урок открытия нового знания, урок систематизации, урок отработки умений. Форма: комбинированный урок.</p> <p>Архитектура рекуррентных нейронных сетей. Предложить привести примеры задач для рекуррентных нейронных сетей.</p>	7.5.2. Методы решения задачи классификации. Методы решения задачи регрессии. 7.5.3. Искусственные нейронные сети. Перцептрон. Сверточные нейронные сети. Рекуррентные нейронные сети.	Визуальные (слайды), ПК, электронные образовательные ресурсы

61	248-251	Глубокое обучение без учителя	<p>Тип: урок открытия нового знания, урок систематизации, урок отработки умений. Форма: комбинированный урок.</p> <p>Автоэнкодеры, сети глубокого доверия, GAN – обзорно.</p>	7.5.3. Искусственные нейронные сети. Перцептрон. Сверточные нейронные сети. Рекуррентные нейронные сети.	Визуальные (слайды), ПК, электронные образовательные ресурсы
62	252-255	Постановка задачи для итогового проекта. Разработка итогового проекта.	<p>Тип: урок отработки умений, урок развивающего контроля. Форма: комбинированный урок.</p> <p>Постановка задачи. На собственном датасете обучить модель машинного обучения, визуализировать данные и оценить качество работы обученной модели.</p>	<p>7.3.1. Основные понятия. Типология методов машинного обучения. Численные методы нахождения лучшего решения.</p> <p>7.5.2. Методы решения задачи классификации. Методы решения задачи регрессии.</p>	Визуальные (слайды), ПК, электронные образовательные ресурсы
63	256-259	Презентация результатов итогового проекта	<p>Тип: урок развивающего контроля. Форма: семинар.</p> <p>Презентация результатов.</p>	7.3.1. Основные понятия. Типология методов машинного обучения. Численные методы нахождения лучшего решения.	Визуальные (слайды), ПК, электронные образовательные ресурсы

4. УЧЕБНО-МЕТОДИЧЕСКОЕ И МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА.

Оборудование

1. Рабочая станция ученика (Intel i5, 8Гб ОЗУ, SSD 500Гб, видеокарта Radeon RX VEGA 56 или аналогичная с 8Гб видеопамяти, монитор с разрешением 1920x1080, клавиатура, мышь)
2. Рабочая станция учителя (Intel i7, 16Гб ОЗУ, SSD 500Гб, видеокарта Radeon RX VEGA 56 или аналогичная с 8Гб видеопамяти, монитор с разрешением 1920x1080, клавиатура, мышь)
3. Источник бесперебойного питания для сервера (ИБП)
4. Телекоммуникационный шкаф
5. Коммутаторы уровня доступа с поддержкой VLAN и IP-телефонии
6. Лазерный станок с ЧПУ
7. Копировальная станция
8. Интерактивная панель (например IRBIS “Панель МЭШ”)

Программное обеспечение (в том числе системное ПО)

1. ОС Windows 10
2. MS Office 2016
3. VisualStudio 2017
4. Statsoft Statistica
5. SPSS
6. Oracle SQL Developer
7. MySQL
8. Adobe Creative Suite
9. PyCharm CE
10. Python 3.7

Список учебной и методической литературы и другие источники:

1. Баврин И.И. Теория вероятностей и математическая статистика / И.И.Баврин. - М.: Высш. шк., 2005.— 160 с:
2. Вентцель Е. С. Задачи и упражнения по теории вероятностей: Учеб. пособие для студ. вузов / Е. С. Вентцель, Л. А. Овчаров. — 5-е изд., испр. — М.: Издательский центр «Академия», 2003. — 448 с.
3. Виленкин Н.Я. Комбинаторика / Н.Я. Виленкон, А.Н. Виленкин, П.А. Виленкин. - М.: ФИМА, МЦНМО, 2006. - 400 с.
4. Вуколов Э.А. Сборник задач по математике для вузов. В 4-х ч. ч. 4 / Э.А. Вуколов, А.В. Ефимов, В.Н. Земсков, А.С. Поспелов. - М., Физматлит, 2004- 432 с.
5. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике /В. Е. Гмурман. - М., Высш.шк., 2004.- 404 с.
6. Гмурман, Владимир Ефимович. Теория вероятностей и математическая статистика:учебное пособие для вузов /В. Е. Гмурман.-Изд. 12-е, перераб.- М.:Высшая школа,2009.-478с.

7. Гнеденко Б.В. Курс теории вероятностей: Учебник / Б.В. Гнеденко. - Изд. 8-е, испр. и доп. — М.: Едиториал УРСС, 2005. — 448 с.
8. Кибзун А.И. Теория вероятностей и математическая статистика. базовый курс с примерами и задачами / А.И. Кибзун.- М.: Физматлит, 2002. - 224 с.
9. Кибзун, Андрей Иванович. Теория вероятностей и математическая статистика: базовый курс с примерами и задачами : учебное пособие для втузов/А. И. Кибзун, Е. Р. Горяинова, А. В. Наумов ; под ред. А. И. Кибзуна.-Изд. 5-е, перераб. и доп.- М.:Физматлит,2007.-231 с.:
- 10.Кремер Н.Ш. Теория вероятностей и математическая статистика: Учебник для вузов / Н.Ш. Кремер.— М.: ЮНИТИ-ДАНА, 2004. - 573 с.
- 11.Максимов Ю.Д. Вероятностные разделы математики / Ю.Д. Максимов. - Изд.: Иван Федоров, 2001. - 592 с.
- 12.Математическая статистика: Учеб. для вузов / В. Б. Горяинов, И. В. Павлов, Г. М. Цветкова, О. И. Тескин.; Под ред. В.С. Зарубина, А.П. Крищенко. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2001. 424 с.
- 13.Письменный Д.Т. Конспект лекций по теории вероятностей, математической статистике и случайным процессам / Д.Т. Письменный. - 3-е изд. - М.: Айрис-пресс, 2008. -288 с.
- 14.Пугачев В.С. Теория вероятностей и математическая статистика: Учеб. Пособие / В.С. Пугачев.— 2-е изд., исправл. и дополи.— М.: Физматлит,2002.- 496 с.
- 15.Сидняев, Н. И. Теория планирования эксперимента и анализ статистических данных : учеб. пособие для магистров / Н. И. Сидняев. – М. : Юрайт, 2012. – 399 с.
- 16.Теория вероятностей : Учеб. для вузов. - 3-е изд., испр. / А.В. Печинкин, О.И. Тескин, Г.М. Цветкова и др.; Под ред. В.С. Зарубина, А.П. Крищенко. - М.: Изд-во МГТУ им. Н.Э.Баумана, 2004. - 456 с.
- 17.Рейзлин, В. И. Математическое моделирование. Учебное пособие / В.И. Рейзлин. - М.: Юрайт, 2016. - 128 с.
- 18.Юдович, В. И. Математические модели естественных наук / В.И. Юдович. - М.: Лань, 2011. - 336 с
- 19.Миркин, Б. Г. Введение в анализ данных : учебник и практикум для бакалавриата и магистратуры / Б. Г. Миркин. — М. : Издательство Юрайт, 2014. — 174 с. — Серия : Авторский учебник
- 20.Рузаков, А.А. Управление данными: учеб. пособие / А.А. Рузаков. – Челябинск: Изд-во Челяб. гос. пед. ун-та, 2015. – 132 с.
- 21.Флах П. Ф70 Машинное обучение. Наука и искусств.о построения алгоритмов, которые извлекают знания из данных/ пер. с англ. А. А. Слинкина. - М.: ДМК Пресс, 2015. - 400 с.: ил.
- 22.Джеймс Г. Уиттон Д., Хасти Т., Тибширани Р. Введение в статистическое обучение на языке R // ДМК Пресс, 2016. – 450 стр.

5. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ИЗУЧЕНИЯ УЧЕБНОГО КУРСА

Выпускник научится:

- 1) Основами математической статистики и теории вероятностей;
- 2) Находить закономерности в данных, разрабатывать математические модели и модели машинного обучения на эти данных;
- 3) Визуализировать полученные результаты моделирования;
- 4) Основами глубокого обучения.