



**Федеральное государственное автономное образовательное учреждение  
высшего образования  
"Национальный исследовательский университет  
"Высшая школа экономики"**

Факультет **Коммуникаций, медиа и дизайна**  
Департамент Интегрированных Коммуникаций

**Рабочая программа дисциплины  
Открытые данные**

для образовательной магистерской программы «Журналистика данных»  
направления подготовки **42.04.02 «Журналистика»**  
Уровень: **магистр**

Разработчики программы  
А.А. Незнанов, aneznanov@hse.ru  
А.А. Паринов, aparinov@hse.ru

Одобрена на заседании Департамента Интегрированных Коммуникаций  
«\_\_»\_\_\_\_\_ 2016 г.

Руководитель Департамента Интегрированных Коммуникаций  
С.А. Зверев \_\_\_\_\_

Утверждена Академическим советом образовательной программы  
«\_\_»\_\_\_\_\_ 2016 г., № протокола \_\_\_\_\_

Академический руководитель образовательной программы  
М.А. Пильгун \_\_\_\_\_

\_\_\_\_\_, 2016

*Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения подразделения-разработчика программы.*



## 1 Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает требования к образовательным результатам и результатам обучения студента и определяет содержание и виды учебных занятий и отчётности.

Программа предназначена для преподавателей, ведущих дисциплину «**Открытые данные**», учебных ассистентов и студентов направления подготовки/специальности **42.04.02 «Журналистика»**, обучающихся по образовательной программе «**Журналистика данных**».

Программа учебной дисциплины разработана в соответствии с образовательным стандартом НИУ ВШЭ по направлению подготовки **42.03.02 Журналистика** для квалификации «магистр».

## 2 Цели освоения дисциплины

Цель освоения дисциплины «**Открытые данные**» – овладение знаниями и навыками, необходимыми для:

- поиска, получения и преобразования данных из источников открытых данных;
- интеграции данных, оценки их качества и пригодности к дальнейшему использованию, как в конкретном исследовании, так и в качестве новой порции открытых данных;
- подготовки данных для эффективного анализа, как методами математической статистики, так и методами интеллектуального анализа.

Отметим, что взаимодействие с источниками открытых данных и последующая обработка данных требует не только некоторого минимума технических знаний. Специалисту следует опираться на знания и представления о функционировании предметных областей (экономики, социума, политической сферы и т.п.); хорошо представлять модели и методы информационного поиска, статистики, машинного обучения и майнинга данных, извлечения знаний и онтологического моделирования. Это позволит развить общие компетенции исследователя данных [*data scientist*].

Дисциплина фокусируется на феномене открытых данных [*Open Data*], современном состоянии инструментов работы с ними и специфике явления, получившего названия «журналистика данных» [*Data Journalism*].

## 3 Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

- знать специфичные свойства открытых данных и способы доступа к ним; понимать базовые модели, методы и инструменты статистического анализа, машинного обучения и майнинга данных применительно к открытым и связным данным;
- уметь адекватно применять современные модели, методы и инструменты (программные средства), позволяющие целостно решить задачи журналистики с использованием источников открытых данных;
- иметь навыки (приобрести опыт) получения, оценки, интеграции и анализа реальных данных из источников открытых данных с помощью изученных методов.

В результате освоения дисциплины студент осваивает следующие компетенции.



Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
Способен к самостоятельному освоению новых методов	СК-3	РБ	Студент успешно выполняет домашние задания	Лекции. Семинары	Домашние задания. Экзамен
Способен использовать методы, методики и приёмы для презентации результатов проектно-аналитических, научно-исследовательских, аналитических, экспертно-консультационных задач	ПК-18	РБ	Студент активно участвует в работе семинаров	Лекции. Семинары	Презентации
Способен работать с базами данных и размещёнными в открытом доступе материалами государственных структур, международных организаций, корпоративной отчётностью фирм и финансовых институтов.	ПК-21	РБ, СД	Студент успешно выполняет домашние задания	Лекции, семинары, коллоквиумные проекты	Выполняет домашние задания

#### 4 Место дисциплины в структуре образовательной программы

Для специализаций подготовки магистров настоящая дисциплина является неотъемлемой дисциплиной в рамках направления подготовки 42.04.02 "Журналистика".

Изучение данной дисциплины базируется на следующих дисциплинах:

- Введение в программирование

Для освоения учебной дисциплины, студенты должны владеть следующими знаниями и компетенциями:

- Основы программирования и математика в объёме программы средней школы.
- Желательно владение основными концепциями математической статистики.

#### 5 Тематический план учебной дисциплины

№	Название раздела	Всего часов	Аудиторные часы			Самостоятельная работа
			Лекции	Семинары	Практические занятия	
1	Введение, основные понятия в области об-		4	2		



	работки данных.					
2	Открытые и связанные данные. Стандарты, протоколы доступа и форматы.		2	4		
3	Базовые методы визуализации данных. Инструменты визуализации и построения отчётов.		2	2		
4	Язык <i>Python</i> в анализе открытых данных.		4	4		
5	Статистический анализ числовых данных		6	6		
6	Моделирование данных и высокоуровневые языки манипулирования данными.		4	4		
7	Интеллектуальный анализ данных. Числовые данные.		8	8		
8	Интеллектуальный анализ данных. Тексты и сети.		4	4		



## 6 Формы контроля знаний студентов

Тип контроля	Форма контроля	1 год		Параметры **
		1	2	
Текущий (неделя)	Домашнее задание	3		срок выполнения и требования указываются в тексте задания
	Домашнее задание	6		срок выполнения и требования указываются в тексте задания
	Домашнее задание	9		срок выполнения и требования указываются в тексте задания
	Проект по анализу реальных данных		5	Предполагается промежуточный контроль в конце первого модуля и итоговая защита в рамках экзамена во втором модуле.
Промежуточный	Коллоквиум			В устной форме
Итоговый	Экзамен			В устной форме, а также собеседование по результатам выполнения проекта

## 7 Критерии оценки знаний, навыков

Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

Выполнение домашнего задания оценивается в баллах по 10-ти балльной шкале. В тексте задания могут быть указаны баллы для каждого из подзаданий. Основные критерии: корректность и полнота представленного решения.

Проект по анализу реальных данных подразумевает коллективное или индивидуальное выполнение задания, связанного с применением программных средств и методов разработки данных и машинного обучения на реальных данных. Защита проекта происходит в два этапа, на первом модуле происходит выбор данных, обоснование методов для решения задачи и первичное исследование данных. К экзамену студент предоставляет (высылает в электронной форме) отчет по проекту, а рамках устного собеседования на экзамене происходит его защита.

Осуществляется следующая дистанционная поддержка при проведении контроля: задания и материалы лекций, рассылаются по электронной почте, учебная литература и материалы лекций доступны в *DropBox* (или *LMS*, *GitHub*, *SageMath*, *Microsoft OneNote Class* и другие сервисы).

## 8 Содержание дисциплины

### Раздел 1. Введение, основные понятия в области обработки данных

Информация, сведения, данные и документы. Задачи обработки данных. Информационные процессы и информационные системы. Системный подход к анализу данных. Базовые инструменты обработки данных, интерфейсы и протоколы, аутентификация и авторизация.

Маленькие, большие и огромные данные. Форматы данных, основные типы атомарных данных, сериализация и десериализация. Данные и знания.

Наборы данных и источники данных. Открытость как свойство набора данных и источника данных. Этапы анализа данных. Примеры прикладных задач анализа данных и их типы: описательная статистика, проверка гипотез, классификация, регрессия, ранжирование, кластеризация, поиск структуры в данных.



## Раздел 2. Стандарты, протоколы доступа и форматы открытых данных

Определение и основные свойства открытости [*Openness*]. Открытые лицензии. Стандарты в области открытых данных. Историческая справка и исторически сложившиеся условности. Основные типы открытых данных. Популярность формата *CSV*. Языки *JSON* и *XML*. Идеология получения и анализа данных. Роли участников. Связные [*Linked*] данные. Инструменты исследователей и разработчиков. *API* доступа к открытым данным.

## Раздел 3. Базовые методы визуализации данных. Инструменты визуализации и построения отчётов

Принципы и парадигмы визуализации. Классификация основных видов диаграмм. Отображение сходства и отличия. Отображение динамики изменения. Геоинформация и карты. Способы и инструменты извлечения, преобразования данных для последующей визуализации. *Microsoft Excel* как пример инструмента. Интерактивная визуализация. *Microsoft PowerBI* и *Tableau Software Tableau* как примеры инструментов. Возможности использования в работе журналиста.

## Раздел 4. Язык Python в анализе открытых данных

Основные особенности языка *Python*. Типы данных и структуры данных. Управляющие конструкции. Функции. Ввод/вывод данных. Необходимые библиотеки и пакеты, управление пакетами. Пакеты *Matplotlib* и *Pandas*. Высокоуровневые средства анализа данных. Пакет *Scikit-learn*.

## Раздел 5. Статистический анализ числовых данных

Вероятность и вероятностные пространства. Дискретные и непрерывные случайные величины. Таблицы и гистограммы. Функция распределения и функция плотности. Основные распределения. Математическая статистика. Генеральная совокупность и выборка. Представление данных. Описательная статистика, визуализация средних, медианы, моды, разброса, дисперсии и других характеристик. Понятие корреляции. Проверка статистических гипотез. Ошибки первого и второго рода, *p-value*, теоретическая и практическая значимость. Множественные сравнения. Линейная регрессия. Основы факторного анализа. Работа с реальными данными. Неполнота и противоречивость данных. Шумы и выбросы. Методы поиска выбросов. Пропуски в данных, методы их восстановления. Несбалансированные выборки: проблемы и методы борьбы. Задача отбора признаков, примеры подходов. Примеры реализации статистических методов в *Pandas*. Оптимальная визуализация результатов статистического анализа.

## Раздел 6. Моделирование данных и высокоуровневые языки манипулирования данными

Модели данных, сетевые, реляционные, объектовые и другие модели. Реляционная модель с точки зрения аналитика. Язык манипулирования данными *SQL*. Подъязыки *DDL* и *DML*. Структура запросов на выборку (*SELECT*). Современные реляционные системы управления базами данных. Хранилища больших данных. *NoSQL* и конвергенция технологий доступа к данным. Способы построения выборок из гетерогенных данных.

## Раздел 7. Интеллектуальный анализ данных. Числовые данные.

Введение в машинное обучение и интеллектуальный анализ данных. Постановки задач машинного обучения. Классификация: от линейной регрессии к более мощным методам. Проблема переобучения.



Линейная классификация. Аппроксимация дискретной функции потерь. Отступ. Примеры аппроксимаций, их особенности. Градиентный спуск, регуляризация. Классификация и оценки принадлежности классам. Пример задачи кредитного скоринга. Логистическая регрессия. Максимизация зазора как пример регуляризации и устранения неоднозначности решения.

Логические методы и их интерпретируемость. Простейший пример: список решений. Пример решающего списка для задачи фильтрации нежелательных сообщений. Деревья решений. Проблема построения оптимального дерева решений и жадные алгоритмы. Критерий ветвления. Выбор оптимального разбиения. Примеры критериев: энтропийный (прирост информации), Джини и их модификации. Критерии завершения построения. Регуляризация и стрижка деревьев.

Композиции алгоритмов. Простейший пример: уменьшение дисперсии при усреднении алгоритмов методом бутстреп. Блендинг алгоритмов. Понятие смещения и разброса (иллюстрация на примере линейных методов и решающих деревьев). Уменьшение разброса с помощью усреднения. Случайный лес. Оценка out-of-bag.

Оценивание качества алгоритмов. Регрессия: квадратичные и абсолютные потери, абсолютные логарифмические отклонения. Классификация: доля верных ответов, ее недостатки. Точность и полнота, их объединение: арифметическое среднее, минимум, гармоническое среднее (F-мера).

Оценки принадлежности классам: площади под кривыми. *AUC-ROC*, *AUC-PRC*, их свойства.

Оценивание качества алгоритмов. Скользящий контроль. Кросс-валидация. Практические особенности кросс-валидации. Стратификация. Потенциальные проблемы с разбиением зависимой или динамической выборки.

Задача кластеризация. Простые эвристические подходы. Алгоритм *K-Means*. Проблема устойчивости результатов и важность грамотной инициализации, алгоритм *K-Means++*. Выбор числа кластеров. Оценка качества кластеризации.

Задачи обнаружения импликаций и построения ассоциативных правил. Анализ частых множеств признаков. Пример задачи анализа потребительской корзины. Поддержка и достоверность. Частые, замкнутые и максимальные частые множества. Алгоритм Априори.

## Раздел 8. Интеллектуальный анализ данных. Тексты и сети.

Коллекции неструктурированных текстов как исходные данные для анализа. Предобработка: фильтрация, стеммирование, индексация. Ключевые слова. Меры сходства текстов. От синтаксиса к семантике. Пример задачи определения тональности высказываний. Доступные инструменты анализа текстов.

Анализ сетей. Основные графовые модели. Значимость цепей и деревьев. Протяжённость и разветвлённость. Покрытия. Различные меры центральности. Пример задачи поиска сообществ в социальных сетях. Доступные инструменты анализа сетей.

## 9 Образовательные технологии

Необходимое для выполнения работ программное обеспечение, как правило, находится в свободном доступе и его можно загрузить в сети Интернет или скопировать из репозитория, предоставляемого курсу.

В число программных пакетов входят инструменты сообщества *PyData* (<http://pydata.org/downloads>):

- пакеты *scipy* и *numpy*;
- сборка среды разработки *Anaconda*;
- пакеты *Pandas* и *Scikit-learn*;
- и др.

Дополнительно к каждой из тем доступны слайды лекций, изложение которых адаптировано с учётом используемых программных пакетов.



В рамках дисциплины используется проектная форма работы в малых группах с использованием реальных данных.

### 9.1 Методические рекомендации преподавателю

Даются по желанию автора. Методические рекомендации (материалы) преподавателю могут оформляться в виде приложения к программе дисциплины и должны указывать на средства и методы обучения, применение которых для освоения тех или иных тем наиболее эффективно.

### 9.2 Методические указания студентам

Преподавателю рекомендуется использовать демонстрацию работы изучаемых методов анализа данных с помощью предустановленных программных продуктов (*PyData*) во время лекционных занятий.

В рамках семинаров возможно решение задач, а домашние задания предлагается составлять практико-ориентированными.

## 10 Оценочные средства для текущего контроля и аттестации студента

### 10.1 Оценочные средства для оценки качества освоения дисциплины в ходе текущего контроля

Примерные темы домашних заданий:

Домашнее задание 1. Разработка программы извлечения данных социальной сети.

Домашнее задание 2. Инструменты извлечения и визуализации открытых данных.

Домашнее задание 3. Линейные методы классификации и регрессии.

### 10.2 Примеры заданий промежуточной аттестации

Примерный перечень тем экзаменационных билетов.

1. Информация, сведения, данные и документы. Определения, базовые инструменты, форматы и протоколы, хранение.
2. Аутентификация и авторизация при доступе к источникам данных.
3. Понятие открытых данных. Принципы открытости. Требования к источникам открытых данных.
4. Связные данные, стандарты и указания.
5. Официальные открытые данные государственных органов.
6. Инструменты доступа к Интернет-данным. *Web*-службы и *Web-API*.
7. Основы теории вероятностей. Случайные величины вероятностных пространства. Дискретные и непрерывные распределения. Примеры.
8. Методы статистического анализа 1. Статистика как инструмент анализа. Описательная статистика. Генеральная совокупность и выборки. Основные характеристики выборки. Гистограммы и другие методы визуализации основных свойств выборок.
9. Методы статистического анализа 2. Статистические гипотезы и критерии. Уровень доверия и доверительный интервал. Статистическая практическая значимость.
10. Методы статистического анализа 3. Суммаризация данных. Основы регрессионного и корреляционного анализа.
11. Методы статистического анализа 4. Методы поиска выбросов в данных. Методы восстановления пропусков в данных. Работа с несбалансированными выборками.
12. Основные понятия машинного обучения. Основные постановки задач. Примеры прикладных задач.
13. Линейные пространства. Векторы и матрицы. Линейная независимость. Обратная матрица.
14. Производная и градиент функции. Градиентный спуск. Выпуклые функции.
15. Оценивание параметров распределений, метод максимального правдоподобия. Бутстрэппинг.





16. Задача кластеризации. Алгоритм K-Means. Оценки качества кластеризации.
17. Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения.
18. Метрики качества алгоритмов классификации. Оценивание качества алгоритмов. Отложенная выборка, её недостатки. Оценка полного скользящего контроля. Кросс-валидация.
19. Деревья решений. Методы построения деревьев. Их регуляризация. Композиции алгоритмов. Случайный лес, его особенности.
20. Задача анализа потребительской корзины. Поддержка и достоверность. Частые, замкнутые и максимальные частые множества. Алгоритм Априори.
21. Автоматическая обработка текстов. Основные модели, методы и инструменты.
22. Анализ сетей. Основные модели, методы и инструменты.
23. Интерактивная визуализация данных.

Примерный перечень вопросов для подготовки к экзамену.

1. Чем данные отличаются от информации? Что можно назвать обработкой данных?
2. Что такое источник данных? Каковы основные параметры источника данных?
3. Что такое документ? Есть ли стандарты на реквизиты документов?
4. Как соотносятся понятия интерфейса и протокола?
5. Раскройте понятие открытых данных. Какие определения «открытости» вы знаете?
6. Как соотносятся понятия открытых данных [*Open Data*] и связанных данных [*Linked Data*]?
7. Какие стандарты и указания в области работы с открытыми данными вы знаете?
8. Какими критериями определяется качество открытых данных?
9. Поясните на примерах понятие актуальности данных.
10. Как раскрывается аббревиатура ETL?
11. Как должны быть представлены данные для применения статистических методов анализа?
12. Как правильно формулировать статистическую гипотезу?
13. Какие данные называют объекто-признаковыми? Какова стандартная форма их представления?
14. Сформулируйте несколько вариантов постановки задачи классификации объектов.
15. Сформулируйте стандартную постановку задачи ранжирования.
16. Какие методы кластеризации вы знаете? Как выбрать метод, адекватный стоящей задаче и имеющимся данным?
17. Какова стандартная цепочка действий предобработки текстов на естественном языке.
18. Что такое объекты и признаки в машинном обучении? Для чего нужен функционал качества? Что такое алгоритм/модель?
19. Чем задача классификации отличается от задачи регрессии? Приведите примеры задач классификации и регрессии.
20. Что такое вещественные (числовые), бинарные, категориальные признаки? Приведите примеры.
21. В чём заключается обобщающая способность алгоритма машинного обучения? К чему приводит её отсутствие?
22. Что такое переобучение? Как оно связано с градиентным спуском?
23. Что такое градиент? Какое его свойство активно используется в машинном обучении?
24. Опишите алгоритм градиентного спуска.



25. Как работает наивный байесовский классификатор на задачах с двумя признаками? Запишите вид алгоритма, объясните все входящие в формулу элементы.
26. Чем дискретные распределения отличаются от непрерывных? Приведите примеры дискретных и непрерывных распределений.
27. Что такое матожидание, медиана, дисперсия случайной величины? Что характеризует корреляция двух случайных величин?
28. Что такое интерквартильный (межквартильный) размах? Чем он отличается от дисперсии?
29. Что такое среднеквадратичная ошибка? Как она используется в линейной регрессии?
30. Как обучается линейная регрессия?
31. Почему наличие линейно зависимых признаков представляет проблему при обучении линейной регрессии?
32. Что такое регуляризация? Как она помогает бороться с переобучением?
33. Чем L1-регуляризация отличается от L2-регуляризации?
34. Как использовать регуляризацию в логистической регрессии?
35. Что такое масштабирование (шкалирование) признаков? Как его проводить? Зачем это нужно?
36. Как можно использовать категориальные признаки в линейных моделях?
37. В чём заключается использование квадратичных признаков в линейных моделях? Для чего это нужно?
38. Как выглядят модели линейной классификации в случае двух классов?
39. Что такое отступ? Для чего он нужен?
40. Как измерить ошибку линейного классификатора?
41. Как обучаются линейные классификаторы (общая схема)?
42. Как обучается логистическая регрессия? В чём её особенность?
43. Для чего может понадобиться оценивать вероятности классов?
44. Что такое доля правильных ответов? В чём заключаются её проблемы?
45. Что такое точность и полнота?
46. В чём заключается разница между метриками Accuracy и Precision?
47. Что такое ROC-кривая? Что такое AUC-ROC? Для чего он используется?
48. Что такое PR-кривая? Что такое AUC-PRC? Для чего он используется?
49. Как можно свести задачу многоклассовой классификации к серии задач бинарной классификации?
50. Что такое гиперпараметр? Чем гиперпараметры отличаются от обычных параметров алгоритмов? Приведите примеры параметров и гиперпараметров в линейных моделях.
51. Что такое отложенная выборка? Что такое кросс-валидация (скользящий контроль)? Как ими пользоваться для выбора гиперпараметров?

## 11 Порядок формирования оценок по дисциплине

Итоговая оценка формируется из оценок домашних заданий, коллективного проекта и экзамена.

$$O_{результ} = 0,1 \cdot O_{ДЗ1} + 0,1 \cdot O_{ДЗ2} + 0,1 \cdot O_{ДЗ3} + 0,5 \cdot O_{Проект} + 0,2 \cdot O_{экс.}$$

Коллективный проект оценивается на основании защиты отчёта. Сроки сдачи, проверки и защиты отчётов сообщаются не позднее, чем за две недели.

На экзамене студент может получить дополнительный вопрос (дополнительную практическую задачу, решить к передаче домашнее задание), ответ на который оценивается в 1 балл.

В диплом выставляет результирующая оценка по учебной дисциплине.



## 12 Учебно-методическое и информационное обеспечение дисциплины

### 12.1 Базовый учебник

Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2014

(<http://www.dataminingbook.info/pmwiki.php/Main/BookDownload>)

### 12.2 Основная литература

1. Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2014  
(<http://www.dataminingbook.info/pmwiki.php/Main/BookDownload>)
2. Boris Mirkin. Core Concepts in Data Analysis: Summarization, Correlation, Visualization. 2010 ([http://www.hse.ru/data/2010/10/14/1223126254/Mirkin\\_All.pdf](http://www.hse.ru/data/2010/10/14/1223126254/Mirkin_All.pdf))
3. Janert P.K. Data Analysis with Open Source Tools. – O'Reilly Media, 2010. – 540 p.
4. Lutz M. Learning Python, 5th ed. – O'Reilly Media, 2013. – 1648 p.
5. Raschka S. Python Machine Learning. – Packt, 2015. – 454 p.

### 12.3 Дополнительная литература

1. Monino J.-L., Sedkaoui S. Big Data, Open Data and Data Development. – Wiley, 2016. – 170 p.
2. Boyd, Vandenberghe. Convex Optimization (<http://stanford.edu/~boyd/cvxbook>)
3. Dekking, F.M., Kraaikamp, C., Lopuhaä, H.P., Meester, L.E., A Modern Introduction to Probability and Statistics (<http://www.ewi.tudelft.nl/index.php?id=50508> и <http://www.springer.com/gp/book/9781852338961>)
4. Data Portals – A Comprehensive List of Open Data Portals from Around the World (<http://dataportals.org>)
5. Портал открытых данных Российской Федерации (<http://data.gov.ru>)
6. The Open Data Foundation (ODaF) (<http://www.opendatafoundation.org>)
7. Online Statistics Education: An Interactive Multimedia Course of Study (<http://onlinestatbook.com/2/index.html>)
8. Теория вероятностей и математическая статистика (<http://statistica.ru/theory>)

### 12.4 Справочники, словари, энциклопедии

1. The Open Definition (<http://opendefinition.org>)
2. Портал MachineLearning.ru (<http://www.machinelearning.ru>)

### 12.5 Программные средства

Для успешного освоения дисциплины студент использует следующие инструменты (программные средства).

1. Программные комплексы *Microsoft Excel 2016* и *Microsoft Power BI*.
2. Язык программирования *Python*, его библиотеки *NumPy*, *SciPy*, *Pandas*, *Scikit-Learn*.
3. Сервис *Tableau Software Tableau 10* (<http://www.tableau.com>).

### 12.6 Дистанционная поддержка дисциплины

Дистанционная поддержка может осуществляться с помощью *LMS*, хранилища слайдов и данных в *Microsoft Onenote Class*, папке автора курса (возможно также использование иных облачных сервисов).



### **13 Материально-техническое обеспечение дисциплины**

Используется проектор (для лекций или семинаров), слайды мультимедийных презентаций и компьютеры с предустановленным программным обеспечением и доступом в Интернет.