

**Федеральное государственное автономное образовательное  
учреждение высшего образования  
«Московский физико-технический институт  
(национальный исследовательский университет)»**

**УТВЕРЖДЕНО**

**Директор физтех-школы бизнеса  
высоких технологий**

**В.Ю. Григорьев**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Основы дата-ориентированных коммуникаций
<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Управление инновациями в бизнесе
	Физтех-школа бизнеса высоких технологий
	Физтех-школа бизнеса высоких технологий
<b>курс:</b>	2
<b>квалификация:</b>	бакалавр

Семестр, формы промежуточной аттестации: 3 (осенний) - Дифференцированный зачет

Аудиторных часов: 20 всего, в том числе:

лекции: 10 час.

семинары: 10 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 25 час.

Всего часов: 45, всего зач. ед.: 1

Количество контрольных работ, заданий: 2

Программу составил: А.В. Щербенок, канд. филол. наук

Программа обсуждена на заседании Физтех-школы бизнеса высоких технологий 15.05.2023

## Аннотация

Дисциплина знакомит студентов с основными направлениями использования больших данных для аналитики в разных областях профессиональной деятельности. В ходе обучения студенты изучают принципы работы с источниками данных и основной инструментарий для работы с большими данными, основные платформы для медиапланирования и визуализации данных.

### 1. Цели и задачи

#### Цель дисциплины

Дисциплина поможет студенту обрести системные представления об эволюции подходов к сбору и анализу данных для решения практических задач в будущей профессиональной деятельности.

#### Задачи дисциплины

1. Изучить основы работы с большими данными и инструменты их анализа,
2. Ознакомиться с практическим применением дата-ориентированного анализа,
3. Научиться классифицировать и выбирать релевантные методы прикладных исследований с использованием анализа данных.

### 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
	ОПК-2.2 Знает и умеет применять численные математические методы и прикладное программное обеспечение для решения научных задач в профессиональной области
	ОПК-2.3 Знает основные требования информационной безопасности
ПК-2 Способен анализировать полученные в ходе научно-исследовательской работы данные и делать научные выводы (заключения)	ПК-2.1 Владеет методами статистической обработки и анализа научных данных
	ПК-2.2 Умеет находить ключевые параметры, определяющие изучаемое явление, и производить численные оценки по порядку величины

### 3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны знать:

1. Основные направления использования больших данных для аналитики.
2. Источники данных и основной инструментарий для работы с большими данными.
3. Основные платформы для визуализации данных.

уметь:

1. Классифицировать и выбирать релевантные методы прикладных исследований с использованием анализа данных.
2. Выстраивать коммуникационную кампанию с учетом дата-ориентированного подхода.

владеть:

1. Основными навыками дата-ориентированного подхода в профессиональной сфере.
2. Техниками подбора релевантного инструментария при работе с большими данными.

### 4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

#### 4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Основные тренды в мире данных и коммуникаций	2	2		5
2	Традиционные методы сбора данных	2	2		7
3	Сбор и анализ данных	2	4		8
4	Data-driven коммуникации	4	2		5
Итого часов		10	10		25
Подготовка к экзамену		0 час.			
Общая трудоёмкость		45 час., 1 зач.ед.			

#### 4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 3 (Осенний)

##### 1. Основные тренды в мире данных и коммуникаций

Дата-ориентированный подход к построению коммуникационных кампаний. Правда и мифы о данных. Что могут и что не могут нам дать данные и их анализ. От big data к smart data. Интеграция big data + количественные измерения и качественные исследования.

##### 2. Традиционные методы сбора данных

Опросы, фокус-группы и др. Большие соц. данные, онлайн и офлайн-источники, интеллектуальный анализ данных (data mining). Откуда берутся данные.

##### 3. Сбор и анализ данных

Требования к данным. Традиционные методы сбора данных. New data. Big data. Small data.

##### 4. Data-driven коммуникации

Что такое data-driven коммуникации, решения, действия. Какие задачи можно решать с помощью данных.

#### 5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Для реализации дисциплины требуется следующее материально-техническое обеспечение:  
Аудитория для вмещения не менее 25-ти человек.  
Наличие доступа в электронно-коммуникационную сеть интернет, компьютер.

#### 6. Перечень рекомендуемой литературы

Основная литература

Основная литература предоставляется МШУ Сколково.

- Mayer-Schönberger, V. Big data: a revolution that will transform how we live, work, and think. - Boston: Houghton Mifflin Harcourt, 2013. URL: <<https://sklib.skolkovo.ru/cgi-bin/koha/opac-detail.pl?biblionumber=3149>>. (дата обращения: 22.07.2023).
- Hair J.H. Multivariate data analysis: a global perspective. - Upper Saddle River, N.J.; London: Pearson, 2010. URL: <<https://sklib.skolkovo.ru/cgi-bin/koha/opac-detail.pl?biblionumber=1287>>. (дата обращения: 22.07.2023)."

#### Дополнительная литература

Дополнительная литература предоставляется МШУ Сколково.

- "- Koch, I. Analysis of multivariate and high-dimensional data. - Cambridge series in statistical and probabilistic mathematics. URL: <<https://sklib.skolkovo.ru/cgi-bin/koha/opac-detail.pl?biblionumber=3680>>. (Дата обращения: 22.07.2023).
- Kutz, J. N. Data-driven modeling & scientific computation: methods for complex systems & big data. URL: <<https://sklib.skolkovo.ru/cgi-bin/koha/opac-detail.pl?biblionumber=4376>>. (Дата обращения: 22.07.2023)."

### **7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)**

- "- Интернет-ресурс ПостНаука (образовательная платформа): <https://postnauka.org/>
- Научно-популярное развлекательное издание о науке N+1: <https://nplus1.ru/>"

### **8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)**

На занятиях используются мультимедийные технологии, включая демонстрацию презентаций, просмотр видеофрагментов и прослушивание аудиозаписей.

### **9. Методические указания для обучающихся по освоению дисциплины (модуля)**

Студент должен овладеть основными понятиями, ключевыми концепциями и методологиями, составляющими основу дисциплины, и научиться применять их на практике, выполняя групповые и индивидуальные задания в аудитории, при выполнении домашней работы, на выездных мероприятиях дисциплины.

Для успешного овладения компетенциями, которые развивает дисциплина, студент должен внимательно изучать материалы курса и регулярно, посещать лекционные и семинарские занятия, участвовать в дискуссиях, выполнять групповые и индивидуальные задания, обсуждая результаты в классе и следуя рекомендациям преподавателя. Самостоятельная работа предусматривает работу с литературой.

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Управление инновациями в бизнесе Физтех-школа бизнеса высоких технологий Физтех-школа бизнеса высоких технологий
<b>курс:</b>	<u>2</u>
<b>квалификация:</b>	бакалавр

Семестр, формы промежуточной аттестации: 3 (осенний) - Дифференцированный зачет

**Разработчик:** А.В. Щербенок, канд. филол. наук

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
	ОПК-2.2 Знает и умеет применять численные математические методы и прикладное программное обеспечение для решения научных задач в профессиональной области
	ОПК-2.3 Знает основные требования информационной безопасности
ПК-2 Способен анализировать полученные в ходе научно-исследовательской работы данные и делать научные выводы (заключения)	ПК-2.1 Владеет методами статистической обработки и анализа научных данных
	ПК-2.2 Умеет находить ключевые параметры, определяющие изучаемое явление, и производить численные оценки по порядку величины

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Основы дата-ориентированных коммуникаций» обучающийся должен:

### знать:

- 1.Основные направления использования больших данных для аналитики.
- 2.Источники данных и основной инструментарий для работы с большими данными.
- 3.Основные платформы для визуализации данных.

### уметь:

- 1.Классифицировать и выбирать релевантные методы прикладных исследований с использованием анализа данных.
- 2.Выстраивать коммуникационную кампанию с учетом дата-ориентированного подхода.

### владеть:

- 1.Основными навыками дата-ориентированного подхода в профессиональной сфере.
- 2.Техниками подбора релевантного инструментария при работе с большими данными.

## 3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Текущий контроль применяется в формах:

- оценки преподавателем ответов на вопросы в процессе краткого (до 5 мин) выборочного устного опроса перед началом каждого занятия по материалам предыдущей лекции,
- оценки умения решать рассматриваемые на лекциях типовые примеры и задачи с применением разобранных цифровых инструментов.

## 4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

Дифференцированный зачет проводится в форме теста и решения задач.

примеры тестовых вопросов:

1. Для распределенного глубокого машинного обучения (Deep Learning) больше подходит фреймворк
  - PyTorch
  - TensorFlow
  - Flask
  - Scikit-learn

2. Автоматизировать запуск пакетных задач в рамках конвейера обработки больших данных по расписанию можно с помощью

- Apache Hadoop
- Apache AirFlow
- Apache Kafka
- Apache Hive

3. Для полнотекстового интеллектуального поиска и аналитики по полуструктурированным данным в формате JSON отлично подходит СУБД

- Elasticsearch
- Hive
- Cassandra
- HBase

4. Apache NiFi используется для

- визуализации результатов аналитики
- оптимизации SQL-запросов к DWH
- маршрутизации потоков Big Data и построения ETL-конвейеров
- эффективного хранения больших данных

5. Анализировать данные, хранящиеся в Apache Hadoop, с помощью стандартного инструментария SQL-запросов

- Можно
- Нельзя

6. Для реализации микросервисной архитектуры и интеграции разрозненных систем подходит

- Apache AirFlow
- Apache Kafka
- Apache Hadoop
- Apache Spark

7. Выберите технологию потоковой обработки событий в режиме реального времени

- Spark Streaming
- Apache Hadoop
- Apache Kafka
- MapReduce

8. Формат Parquet считается

- колоночным (столбцовым)
- строковым
- неструктурированным
- полуструктурированным

9. Для машинного обучения подходят данные

- Предварительно подготовленные, очищенные от ошибок, пропусков и выбросов, а также нормализованные и представленные в виде числовых векторов
- Любых форматов в цифровом виде
- Числовые типа int
- Бинарные

10. Повысить производительность Apache Kafka можно с помощью:

- Замены HDD-дисков на SSD
- Увеличения плотности разделов на каждом брокере
- Повышения коэффициента репликации

- Увеличения размера сообщений

примеры задач для решения на зачёте:

1. Как найти евклидово расстояние между двумя Series (точками) a и b, не используя встроенную формулу?

Ввод:

```
a = pd.Series([2, 4, 6, 8])
```

```
b = pd.Series([1, 3, 5, 7])
```

2. Как нормализовать все столбцы в DataFrame?

Нормализуйте все столбцы df путём вычитания среднего значения столбца и деления на стандартное отклонение.

Сделайте так, чтобы все значения в df находились в диапазоне от 0 до 1.

Ввод:

```
df = pd.DataFrame(np.random.randint(1, 100, 16).reshape(4, 4))
```

3. Как найти второе максимальное значение в массиве, который сгруппирован по другому массиву? Найдите значение второго самого длинного petallength вида setosa в таблице iris.

Ввод: url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data' iris = np.genfromtxt(url, delimiter=',', dtype='object')

4. Как сделать left join двух объектов datatable.Frame? Объедините заданные объекты по ключу A.

Ввод: import datatable as dt

```
df1 = dt.Frame(A=[1, 2, 3, 4], B=['a', 'b', 'c', 'd'])
```

```
df2 = dt.Frame(A=[1, 2, 3, 4, 5], C=['a2', 'b2', 'c2', 'd2', 'e2'])
```

5. Как найти повторяющиеся значения в массиве NumPy? Найдите повторяющиеся значения (начиная со второго вхождения) в заданном массиве и отметьте их как True. Первое вхождение отмечайте как False.

Ввод:

```
np.random.seed(10)
```

```
a = np.random.randint(0, 7, 10)
```

```
print(a)
```

```
# [1 5 4 0 1 3 4 1 5 0]
```

#### Критерии оценивания

оценка «отлично (10)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины при ответе экзаменационного билета и ответе на вопросы по программе дисциплины, а также по результатам контрольных работ;



оценка «отлично (9)» выставляется студенту, показавшему систематизированные, глубокие знания учебной программы дисциплины при ответе экзаменационного билета и ответе на вопросы по программе дисциплины, а также по результатам контрольных работ;

оценка «отлично (8)» выставляется студенту, показавшему систематизированные, знания учебной программы дисциплины при ответе экзаменационного билета и ответе на вопросы по программе дисциплины, а также по результатам контрольных работ;

оценка «хорошо (7)» выставляется студенту по результатам контрольных работ, если он твердо знает материал экзаменационного билета, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;

оценка «хорошо (6)» выставляется студенту по результатам контрольных работ, если он знает материал экзаменационного билета, по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе много неточностей;

оценка «хорошо (5)» выставляется студенту по результатам контрольных работ, если он знает материал экзаменационного билета, излагает его, умеет применять полученные знания на практике, не допускает в ответе грубых ошибок;

оценка «удовлетворительно (4)» выставляется студенту по результатам контрольных работ, а также, если во время ответа экзаменационного билета он показал фрагментарный, характер знаний, недостаточно правильные формулировки базовых понятий, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения;

оценка «удовлетворительно (3)» выставляется студенту по результатам контрольных работ, а также, если во время ответа экзаменационного билета он показал разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушение логической последовательности в изложении программного материала, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;

оценка «неудовлетворительно (2-1)» выставляется студенту по результатам контрольных работ, а также, если во время ответа экзаменационного билета, он показал что не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

## **5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности**

При проведении дифференцированного зачета обучающемуся предоставляется 30 минут на подготовку.

Во время проведения дифференцированного зачета обучающиеся могут пользоваться программой дисциплины, конспектами лекций и любой другой литературой.