

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

УТВЕРЖДЕНО

**Директор по цифровизации
образования**

Д.И. Гриц

	Рабочая программа дисциплины (модуля)
по дисциплине:	Большие данные в финансах
по направлению:	Бизнес-информатика
профиль подготовки:	Финансовые технологии и аналитика центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск" центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
курс:	1
квалификация:	магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

Аудиторных часов: 20 всего, в том числе:

лекции: 8 час.

семинары: 12 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 70 час.

Всего часов: 90, всего зач. ед.: 2

Программу составили:

Е.А. Савицкая, начальник отдела

О.А. Культепина, методист

Программа обсуждена на заседании центра дополнительного, дополнительного профессионального и онлайн-образования "Пуск" 13.06.2022

Аннотация

В рамках дисциплины «Большие данные в финансах» обучающиеся познакомятся с методами и подходами, используемых в программных системах Big Data, изучат технологии хранения и обработки больших данных, научатся применять их для решения прикладных задач в сфере финансов.

1. Цели и задачи

Цель дисциплины

Овладение практическими навыками использования технологий Big Data, для подготовки, хранения и обработки больших данных.

Задачи дисциплины

- ознакомление с основными методами хранения и обработки больших данных;
- изучение принципов работы с большими данными;
- формирование навыков применения полученных знаний и умений для решения прикладных задач в финансовой сфере.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ПК-16 Способен готовить аналитические материалы для оценки мероприятий и выработки стратегических решений в области ИКТ	ПК-16.1 Эффективно применяет в ходе профессиональной деятельности методы и инструментарий анализа данных
	ПК-16.2 Знает методы подготовки аналитических материалов для оценки мероприятий и выработки стратегических решений в области ИКТ

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- этапы жизненного цикла обработки больших данных;
- методы хранения и обработки больших объемов данных;
- принципы извлечения и обновления больших данных;
- методы и прикладные языки для разработки программных решений.

уметь:

- осуществлять сбор и обработку больших данных;
- выбирать подходящие технологии хранения больших данных;
- выбирать методы обеспечения безопасности и целостности данных;
- применять современные языки и платформы для работы с большими данными.

владеть:

- навыками сбора, обработки и интерпретации больших данных в сфере финансов;
- навыками применения современных языков программирования;
- методами обеспечения безопасности и целостности данных.

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	Методы многомерного статистического анализа и анализа нечисловой информации	2	2		16
2	Технологии хранения и обработки больших данных	2	4		18
3	Программирование обработки и загрузки больших данных	2	4		20
4	Аналитика в больших данных	2	2		16
Итого часов		8	12		70
Подготовка к экзамену		0 час.			
Общая трудоёмкость		90 час., 2 зач.ед.			

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

Семестр: 2 (Весенний)

1. Методы многомерного статистического анализа и анализа нечисловой информации

Факторный анализ. Дискриминантный анализ. Кластерный анализ. Многомерное шкалирование. Методы контроля качества.

2. Технологии хранения и обработки больших данных

Основные направления развития методов обработки и хранения данных. Проблема хранения неструктурированных данных. Проблема преобразования данных. Семантические анализаторы. Самообучающиеся автоматы.

3. Программирование обработки и загрузки больших данных

Языки программирования для больших данных. Фреймворки. Базы данных. Аналитические платформы.

4. Аналитика в больших данных

Аналитика Big Data — реалии и перспективы. Технологии и методы анализа, которые используются для работы с большими данными. Применение и возможности Big Data. Решения на основе больших данных. Big Data в области финансов.

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

Занятия по учебной дисциплине проводятся с использованием дистанционных образовательных технологий. Каждый обучающийся обеспечен доступом к образовательной платформе <https://netology.ru/>.

6. Перечень рекомендуемой литературы

Основная литература

1. Введение в многомерный статистический анализ [Текст] = An introduction to multivariate statistical analysis/Т. Андерсон, -М., Физматгиз, 1963

Дополнительная литература

1. Структуры и базы данных [Текст] : [учеб. пособие для вузов] / М. Нагао, Т. Катаяма, С. Уэмура ; пер. с яп. В. Ю. Акифьева ; под ред. В. И. Скворцова .— М. : Мир, 1986 .— 198 с.

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

1. Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 2-е изд., стер. — Санкт-Петербург: Лань, 2022. — 188 с. — ISBN 978-5-8114-9690-7. — Текст: электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/198599>
2. Толстобров, А. П. Управление данными: учебное пособие для вузов / А. П. Толстобров. — 3-е изд., перераб. и доп. — Москва: Издательство Юрайт, 2022. — 272 с. — (Высшее образование). — ISBN 978-5-534-14162-7. — Текст: электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/496748>
3. Дадян, Э. Г. Данные: хранение и обработка: учебник / Э.Г. Дадян. — Москва : ИНФРА-М, 2019. — 205 с. — (Высшее образование: Бакалавриат). — www.dx.doi.org/10.12737/textbook_5cf8c7f2b8cdb8.06963680. - ISBN 978-5-16-014903-5. - Текст: электронный. - URL: <https://znanium.com/catalog/product/989190>
4. Железнов, М. М. Методы и технологии обработки больших данных: учебно-методическое пособие / М. М. Железнов. — Москва: МИСИ – МГСУ, 2020. — 46 с. — ISBN 978-5-7264-2193-3. — Текст: электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/145102>
5. Leskovec J. Mining of Massive Datasets [Electronic Resource] / Jure Leskovec, Anand Rajaraman, Jeffrey David Ulman. — 2nd. ed. - Cambridge: Cambridge University Press, 2014. — 482 p. - Authorized access: <http://library.books24x7.com/toc.aspx?bookid=74213> (Online Digital Library "Books24x7").

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

1. Образовательная платформа <https://netology.ru/>
2. Webinar.ru
3. Zoom
4. Google Drive
5. Scientific bibliographic databases in the field of Big data, available on the Internet in free mode - Science Citation Index (Web of Science), Medline (PubMed), Scientific Electronic Library (NEB).
<https://predictivesolutions.ru/soft/ibm-spss-statistics/>

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

Студент, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения дисциплины, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

- посещения всех занятий, предусмотренных учебным планом по дисциплине;
- ведения конспекта занятий;
- напряжённой самостоятельной работы студента.

Самостоятельная работа включает в себя:

- чтение рекомендованной литературы;
- проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий текущей и промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к преподавателю.

Возможен промежуточный контроль знаний студентов в виде решения задач в соответствии с тематикой занятий.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Бизнес-информатика		
профиль подготовки:	Финансовые технологии и аналитика	▲	▲
	онлайн-образования "Пуск"	▲	▲
	онлайн-образования "Пуск"		
курс:	1		
квалификация:	магистр		

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

Разработчики:

Е.А. Савицкая, начальник отдела

О.А. Культепина, методист

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ПК-16 Способен готовить аналитические материалы для оценки мероприятий и выработки стратегических решений в области ИКТ	ПК-16.1 Эффективно применяет в ходе профессиональной деятельности методы и инструментарий анализа данных
	ПК-16.2 Знает методы подготовки аналитических материалов для оценки мероприятий и выработки стратегических решений в области ИКТ

2. Показатели оценивания компетенций

В результате изучения дисциплины «Большие данные в финансах» обучающийся должен:

знать:

- этапы жизненного цикла обработки больших данных;
- методы хранения и обработки больших объемов данных;
- принципы извлечения и обновления больших данных;
- методы и прикладные языки для разработки программных решений.

уметь:

- осуществлять сбор и обработку больших данных;
- выбирать подходящие технологии хранения больших данных;
- выбирать методы обеспечения безопасности и целостности данных;
- применять современные языки и платформы для работы с большими данными.

владеть:

- навыками сбора, обработки и интерпретации больших данных в сфере финансов;
- навыками применения современных языков программирования;
- методами обеспечения безопасности и целостности данных.

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

Во время текущего контроля студент должен уметь ответить на следующие вопросы:

1. Определение больших данных, ключевые характеристики
2. Факторный анализ.
3. Дискриминантный анализ.
4. Кластерный анализ.
5. Многомерное шкалирование.
6. Методы контроля качества.
7. Основные направления развития методов обработки и хранения данных.
8. Хранение неструктурированных данных.
9. Преобразование данных.
10. Семантические анализаторы.
11. Самообучающиеся автоматы.
12. Языки для Big Data: R.
13. Языки для Big Data: Python.
14. Языки для Big Data: Julia.
15. Языки для Big Data: Java.
16. Фреймворки для Big Data: Hadoop.
17. Фреймворки для Big Data: Spark.
18. Фреймворки для Big Data: Storm.
19. Базы данных для Big Data: Hive.
20. Базы данных для Big Data: Impala.
21. Базы данных для Big Data: Presto.
22. Базы данных для Big Data: Drill.

23. Oracle Big Data Preparation.
24. Аналитика Big Data — реалии и перспективы.
25. Data Mining.
26. Краудсорсинг.
27. Смешение и интеграция данных.
28. Машинное обучение.
29. Искусственные нейронные сети.
30. Распознавание образов.
31. Прогнозная аналитика.
32. Имитационное моделирование.
33. Пространственный анализ.
34. Статистический анализ.
35. Визуализация аналитических данных.
36. Big data: применение и возможности.
37. Решения на основе Big data.
38. Big data в финансах.

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

Пример итогового задания для дифференцированного зачета:

1. Дан набор заданной структуры.
2. Построить модель прогнозирования отклика с использованием процедуры `impstat` с алгоритмом `random forest` с заданным числом деревьев.
3. Применить полученную модель к тестовому набору данных той же структуры.
4. Визуализировать полученный график Lift.
5. Построить на том же наборе модель с использованием высокопроизводительной версии метода GLM.
6. Применить к тестовому набору.
7. Сравнить результаты GLM и Random Forest по AUC.

Критерии оценивания

Оценка отлично (10 баллов) - выставляется обучающемуся, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (9 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (8 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо (7 баллов) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо (6 баллов) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

Оценка хорошо (5 баллов) - выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно (4 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно (3 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

Оценка неудовлетворительно (2 балла) - выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

Оценка неудовлетворительно (1 балл) - выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

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

Дифференцированный зачет по дисциплине проводится в форме выполнения итогового задания.