

**Федеральное государственное образовательное
бюджетное учреждение высшего образования**

**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ
ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

**Департамент анализа данных и машинного обучения
Факультета информационных технологий и анализа больших данных**

СОГЛАСОВАНО

Газпромбанк

(наименование организации)

Начальник управления

(должность представителя работодателя)

_____ Дашковский В.А.

(подпись)

(ФИО)

18.04.2023 г.

УТВЕРЖДАЮ

Проректор по учебной и
методической работе

_____ Е.А. Каменева

25.04.2023 г.

С.В. Макрушин

Машинное обучение в семантическом и сетевом анализе

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки

01.03.02 - Прикладная математика и информатика,

ОП «Прикладная математика и информатика»

(Анализ данных и принятие решений в экономике и финансах),

ОП «Прикладное машинное обучение» (Прикладное машинное обучение)

Рекомендовано Ученым советом

*Факультета информационных технологий и анализа больших данных
(протокол №31 от 18.04.2023г.)*

Одобрено Советом учебно-научного

*Департамента анализа данных и машинного обучения
(протокол №2 от 29.03.2023г.)*

Москва 2023

Содержание

1. Наименование дисциплины.....	2
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	2
3. Место дисциплины в структуре образовательной программы.....	3
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	4
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий.....	4
5.1. Содержание дисциплины.....	4
5.2. Учебно-тематический план.....	7
5.3. Содержание семинаров, практических занятий.....	9
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.....	11
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы.....	11
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю.....	12
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	14
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	18
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.....	19
10. Методические указания для обучающихся по освоению дисциплины.	20
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем.....	21
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	22

1.Наименование дисциплины

«Машинное обучение в семантическом и сетевом анализе».

2.Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ОП «Прикладная математика и информатика»			
ОПК-4	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	1.Использует информационно-коммуникационные технологии при поиске необходимой информации, сбора, визуализации и обработки данных.	Знать: алгоритмы визуализации информации представленной в виде сетей. Уметь: использовать алгоритмы визуализации информации представленной в виде сетей.
		2.Осуществляет рациональный выбор программного продукта в зависимости от поставленной задачи.	Знать: возможности современных библиотек для сетевого анализа. Уметь: уметь решать типовые задачи сетевого анализа с помощью современных библиотек сетевого анализа.
		3.Владеет навыками обеспечения информационной безопасности автоматизированных систем.	Знать: подходы к формулировке типичных задач обеспечения ИБ с помощью сетевого анализа. Уметь: решать типичные задачи обеспечения ИБ с помощью сетевого анализа.
ПКП-5	Способность применять методы и инструменты анализа данных и машинного обучения при подготовке аналитического обоснования финансово-	1.Осуществляет обоснованный выбор методов и инструментов обработки данных для решения экономических и финансовых задач.	Знать: способы решения выделения сообществ в графе. Уметь: на практике выполнять выявления сообществ в графе.

	экономических решений	2. Владеет навыками решения прикладных задач с использованием методов и инструментов анализа данных и машинного обучения.	Знать: способы подготовки и загрузки данных для решения задач МО на графах. Уметь: на практике подготавливать и загружать данные для решения задач МО на графах.
ОП «Прикладное машинное обучение»			
ПКП-2	Способность выполнять анализ качества данных, выявлять и корректировать отклонения в данных и выполнять визуализацию данных	1. Демонстрирует знание методов анализа качества данных для различных типов значений.	Знать: подходы к созданию решений задач МО на графах. Уметь: уметь реализовывать решения задач МО на графах.
		2. Владеет методами нормализации данных, в том числе работы с пропусками и выбросами.	Знать: способы подготовки и загрузки данных для решения задач МО на графах. Уметь: на практике подготавливать и загружать данные для решения задач МО на графах.
		3. Владеет современным инструментарием и практическими навыками визуализации данных.	Знать: инструментальные средства для первичного анализа графовых (сетевых) данных. Уметь: использовать инструментальные средства для первичного анализа графовых (сетевых) данных.

3. Место дисциплины в структуре образовательных программ

Дисциплина «Машинное обучение в семантическом и сетевом анализе» является дисциплиной Профиля «Анализ данных и машинное обучение в экономике и финансах» по направлению подготовки 01.03.02 - Прикладная математика и информатика, ОП «Прикладная математика и информатика».

Дисциплина «Машинное обучение в семантическом и сетевом анализе» является дисциплиной Профиля «Прикладное машинное обучение» по направлению подготовки 01.03.02 - Прикладная математика и информатика, ОП «Прикладное машинное обучение».

4. Объем дисциплины(модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

ОП «Прикладная математика и информатика»

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Модуль 6 (в часах)
Общая трудоемкость дисциплины (в том числе курсовая работа)	4/144	144 (курсовая работа – 24 часа)
Контактная работа – Аудиторные занятия	68	68
<i>Лекции</i>	16	16
<i>Семинары, практические занятия</i>	52	52
Самостоятельная работа	76	76
Вид текущего контроля	-	-
Вид промежуточной аттестации	Зачет	Зачет

ОП «Прикладное машинное обучение»

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Модуль 6 (в часах)
Общая трудоемкость дисциплины (в том числе курсовая работа)	4/144	144 (курсовая работа -24 часа)
Контактная работа – Аудиторные занятия	50	50
<i>Лекции</i>	16	16
<i>Семинары, практические занятия</i>	34	34
Самостоятельная работа	94	94
Вид текущего контроля	-	-
Вид промежуточной аттестации	Зачет	Зачет

5.Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Базовые понятия теории сетей и введение в библиотеку NetworkX.

Введение в предметную область теории сложных сетей: объект исследования, история возникновения, прикладные области и основные задачи теории. Соотнесения понятий теории графов и теории сложных сетей, базовые понятия теории сложных сетей.

Технологический стек работы с сетевыми моделями, построенный на базе языка программирования Python и библиотеки NetworkX. Общие положения о языке программирования Python и знакомство с используемым программным инструментарием (версией языка программирования Python, дистрибутивом Anaconda). Интерактивная оболочка Jupyter notebook: принципы работы и применение для решения задач анализа сетей.

Знакомство с библиотекой NetworkX. Базовые операции в NetworkX: создание сети, оперирование с узлами и связями, базовая визуализация сетей. Обзор структуры библиотеки NetworkX.

Тема 2. Сетевые модели: модели роста сетей, модель тесного мира.

Модели формирования случайных сетей: модель Эрдёша-Реньи (Erdos-Renyi) и родственные модели. Распределения степеней узлов для случайных сетей. Фазовый переход при возникновении связанной компоненты сети. Диаметр и коэффициент кластеризации для случайных сетей.

Модели роста сетей: модель Барабаши-Альберта (Barabasi-Albert) и модель случайного роста сети. Распределения степеней узлов для рассматриваемых моделей, безмасштабные сети, содержательная интерпретация хвоста распределения степеней узлов. Универсальность моделей роста сетей для большого класса эмпирически наблюдаемых сложных сетей.

Феномен «тесного мира» и модель Ваттса-Строгатца (Watts-Strogats). Переход от регулярного графа к случайному: динамика изменения кластерного коэффициента и средней длины пути. Проблема построения критерия принадлежности сети к сетям «тесного мира». Алгоритм латтисизации и его адаптации.

Тема 3. Меры центральности узлов и выявление сообществ в сетях.

Понятие центральности узлов в сети. Модельные графы для сравнения различных метрик центральности. Центральность по степени (degree centrality), центральность по близости (closeness centrality), центральность по посредничеству (betweenness centrality), центральность по собственному вектору (eigenvector centrality).

Понятие сетевых сообществ (network communities). Плотность связей в сети.

Разделение сети на части (graph partitioning), разрезы в графе, минимальные разрезы в графе. Кластеризация и корреляционная матрица.

Степень посредничества связей, алгоритм Гирвана-Ньюмана (Girvin- Newman).
Спектральные методы.

Мера Модулярности. Алгоритм Лоувэйна.

Тема 4. Случайные блуждания на графах, PageRank, метод распространения меток.

Случайное блуждание в сети.

Алгоритм page rank: принцип алгоритма, ограничения на структуру графа для существования решения. Взгляд на задачу с позиции случайного блуждания на графе. Снятие ограничений на сеть для применения Pagerank.

Метод распространения меток

Задача классификации узлов. Обучение с частичным привлечением учителя. Расстановка меток на основе случайного блуждания. Процесс случайного блуждания с ловушками. Алгоритмы расстановки меток с использованием случайного блуждания.

Тема 5. Векторное представление (эмбединг) узлов графа.

Мотивация для построения эмбединга. Векторное представление слов. Дистрибутивная гипотеза. Матрица совместной встречаемости. Представление слов в виде векторов малой размерности. Сингулярное разложение (Singular Value Decomposition, SVD). Алгоритм word2vec. Модели CBOW и Skip-Gram.

Проблематика графов как объектов для эмбединга. Постановка задачи эмбединга. Определение меры схожести узлов. Неглубокие методы эмбединга.

Подход к эмбедингу узлов на основе случайного блуждания. Оптимизационная задача построения эмбедингов на основе случайного блуждания. Использование негативного семплирования. Общий алгоритм node2vec.

Тема 6. Графовые нейронные сети.

Введение в графовые нейронные сети. Общая схема построения графовых нейронных сетей без учителя. Задача semi-supervised learning.

Фреймворк Message Passing Neural Networks (MPNN). Фрэймворк GCN

(сверточные нейронные сети для решеток и графов). Спектральные модели GNN.

Общая схема конвейера для построения модели GNN (графовой нейронной сети). Типология моделей распространения GNN.

Тема 7. Графы знаний и эмбединги на них.

Графы знаний. Модели данных, используемые для представления информации и структурированных знаний. Представление сложных взаимосвязей между данными, связей сущностей, атрибутов и отношений. Семантические сети. Формализация с помощью графов знаний.

Эмбединги для графов знаний. Трансляционные эмбединги TransE, TransH.

Тема 8. Онтологии в графах знаний и онтологические эмбединги.

Понятие онтологии при создании графов знаний. Использование графов знаний и онтологий в обработке текстов на естественном языке и приложениях искусственного интеллекта.

Использование онтологий для повышения качества эмбедингов графов. Эмбединг Joie и современное состояние исследований в области разработок эмбедингов, использующих онтологии.

5.2. Учебно-тематический план

ОП «Прикладная математика и информатика»

№ п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	Контактная работа - Аудиторная работа			Самостоят ельная работа	
			Общая, в т.ч.:	Лекции	Семинары, практическ ие занятия		
1	Базовые понятия теории сетей и введение в библиотеку NetworkX.	18	8	2	6	10	Самостоятельн ые работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
2	Сетевые модели: модели роста сетей, модель тесного мира.	18	8	2	6	10	
3	Меры центральности	18	8	2	6	10	

	узлов и выявление сообществ в сетях.						
4	Случайные блуждания на графах, PageRank, метод распространения меток.	18	8	2	6	10	Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
5	Векторное представление (эмбединг) узлов графа.	18	8	2	6	10	
6	Графовые нейронные сети.	18	10	2	8	8	
7	Графы знаний и эмбединги на них.	18	10	2	8	8	
8	Онтологии в графах знаний и онтологические эмбединги.	18	8	2	6	10	
	В целом по дисциплине	144	68	16	52	76	Согласно учебному плану: -
	Итого в %		47	24	76	53	

ОП «Прикладное машинное обучение»

№ п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	Контактная работа - Аудиторная работа			Самостоят ельная работа	
			Общая, в т.ч.:	Лекции	Семинары, практическ ие занятия		
1	Базовые понятия теории сетей и введение в библиотеку NetworkX.	18	6	2	4	12	Самостоятельн ые работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
2	Сетевые модели: модели роста сетей, модель тесного мира.	18	6	2	4	12	
3	Меры центральности узлов и выявление сообществ в сетях.	18	6	2	4	12	

4	Случайные блуждания на графах, PageRank, метод распространения меток.	18	6	2	4	12	Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
5	Векторное представление (эмбединг) узлов графа.	18	6	2	4	12	
6	Графовые нейронные сети.	20	8	2	6	12	
7	Графы знаний и эмбединги на них.	18	6	2	4	12	
8	Онтологии в графах знаний и онтологические эмбединги.	16	6	2	4	10	
	В целом по дисциплине	144	50	16	34	94	Согласно учебному плану: -
	Итого в %		35	32	68	65	

5.3. Содержание семинаров, практических занятий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Базовые понятия теории сетей и введение в библиотеку NetworkX.	Знакомство с базовыми понятиями теории сложных сетей информационными технологиями анализа сложных сетей Входной контроль. Изучение технологического стека анализа сложных сетей, построенного на базе языка программирования Python и библиотеке NetworkX. <i>Рекомендуемые источники:</i> 8.1; 9.1; 9.2; 9.3; 9.4	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

Сетевые модели: модели роста сетей, модель тесного мира.	Модели роста сетей, содержательная интерпретация хвоста распределения степеней узлов для этих моделей. Ассортативность, коэффициент ассортативности. Решение задач построения сетей по моделям Барабаши-Альберта (Barabasi-Albert) и модель случайного роста сети. Построение распределения степеней узлов для этих моделей и поиск хабов в сети. <i>Рекомендуемые источники: 8.1- 8.5</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Меры центральности узлов и выявление сообществ в сетях.	Решение задач поиска значений метрик центральности для центральности по степени (degree centrality), центральности по близости (closeness centrality), центральности по посредничеству (betweenness centrality), центральности по собственному вектору (eigenvector centrality). Понятие сетевых сообществ. Плотность связей в сети. Разделение сети на части. Степень посредничества связей, алгоритм Гирвана-Ньюмана и спектральные методы. Решение задач на выявление сетевых сообществ в сетях. Расчет плотности сети. <i>Рекомендуемые источники: 8.2; 8.3; 8.5.</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Случайные блуждания на графах, PageRank, метод распространения меток.	Решение задач на использования метода PageRank, визуализация результатов. Расстановка меток на графе на основе случайного блуждания. Применение алгоритма расстановки меток с использованием случайного блуждания. <i>Рекомендуемые источники: 8.1- 8.5</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Векторное представление (эмбединг) узлов графа.	Построение эмбединга узлов сети с помощью метода node2vec. <i>Рекомендуемые источники: 8.1- 8.5</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

Графовые нейронные сети.	Решение задач кластеризации узлов сети с помощью нескольких алгоритмов графовых нейронных сетей. <i>Рекомендуемые источники: 8.1- 8.5</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Графы знаний и эмбединги на них.	Формирование графа знаний, загрузка графа знаний из внешних источников. Построение эмбединга для узлов графа знаний. <i>Рекомендуемые источники: 8.1- 8.5</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Онтологии в графах знаний и онтологические эмбединги.	Формирование графа знаний с онтологией. Использование реализации алгоритма JOIE для построения эмбединга. <i>Рекомендуемые источники: 8.1- 8.5</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

6. Учебно-методическое обеспечение для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Базовые понятия теории сетей и введение в библиотеку NetworkX.	Решения задач определения степеней узлов сети, построения функции распределения степеней узлов сети и визуализации этого распределения.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Сетевые модели: модели роста сетей, модель тесного мира.	Решение задач определения ассортативности в моделях роста сетей. Визуализация ассортативности на диаграммах.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Меры центральности узлов и выявление сообществ в сетях.	Решение задач на выявление сообществ алгоритмом Гирвана-Ньюмена.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.

Случайные блуждания на графах, PageRank, метод распространения меток.	Реализация метода PageRank. Проверка результатов его работы. Решение задач на применение процесса случайного блуждания с ловушками	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Векторное представление (эмбединг) узлов графа.	Построение эмбединга узлов сети с дополнительными признаками.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Графовые нейронные сети.	Решение задач кластеризации узлов сети с дополнительными признаками с помощью нескольких алгоритмов графовых нейронных сетей.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Графы знаний и эмбединги на них.	Построение графа знаний на основе данных из базы знаний Wikidata.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Онтологии в графах знаний и онтологические эмбединги.	Формирование онтологии для заданного графа знаний.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерная тематика курсовой работы

1. Создание согласованной базы знаний по узкой предметной области на основе крупномасштабной базы знаний.
2. Построение согласованной базы знаний по предметной области на основе нескольких источников данных.
3. Создание и анализ графового набора данных с одним типом ребер для заданной предметной области на основе крупномасштабной базы знаний.
4. Создание и анализ графового набора данных о заданной предметной области на основе крупномасштабной базы знаний.
5. Создание и анализ графового набора данных о таксономии товаров заданной предметной области с помощью веб-скрейпинга интернет-магазина.

6. Создание и анализ двудольного графового набора данных о художественных произведениях и их создателях с помощью веб-скрейпинга.
7. Эмпирическое сравнение различных техник сэмплирования для задач анализа крупного графа.
8. Сравнение различных методов выделения сообществ на искусственно сгенерированных графах
9. Анализ качества и устойчивости выделения сообществ с помощью дискретного алгоритма распространения меток на искусственно сгенерированных данных
10. Реализация различных методов выделения сообществ, основанных на дискретном алгоритме распространения меток и их сравнительный анализ.
11. Реализация и сравнение различных спектральных методов выделения сообществ и их сравнительный анализ.
12. Реализация методов визуализации структуры сети с помощью спектральных методов анализа сети.
13. Оптимизация гиперпараметров методов выделения сообществ для методов, основанных на случайных блужданиях.
14. Анализ динамики модулярности в многошаговых алгоритмах поиска сообществ в сетях.
15. Реализация алгоритма PageRank и анализ его применения для реальных сетей.
16. Построение рекомендательной системы с помощью векторного представления графа на основе методов случайного блуждания.
17. Построение рекомендательной системы с помощью векторного представления графа знаний.
18. Построение рекомендательной системы с помощью графовых нейронных сетей использующих атрибуты товаров и пользователей.
19. Сравнение решений задачи классификации узлов сети с использованием различных методов векторных представлений узлов.
20. Классификация узлов сети с помощью графовой нейронной сети, использующей атрибуты узлов.

21. Решение задачи предсказания частично заполненной числовой характеристики узлов сети с помощью графовой нейронной сети.
22. Предсказание центральности узлов по PageRank с помощью графовых нейронных сетей.
23. Реализация и применение алгоритма TransE для построения векторного представления графов знаний.
24. Реализация и применение алгоритма TransH для построения векторного представления графов знаний.
25. Сравнительный анализ качества решения задачи предсказания связи в графах знаний с помощью различных трансляционных алгоритмов.

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Департамента анализа данных и машинного обучения Факультета информационных технологий и анализа больших данных.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2. «Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний

Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
ОП «Прикладная математика и информатика»			

ОПК-4. Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	1.Использует информационно-коммуникационные технологии при поиске необходимой информации, сбора, визуализации и обработки данных.	<u>Знать:</u> алгоритмы визуализации информации представленной в виде сетей. <u>Уметь:</u> использовать алгоритмы визуализации информации представленной в виде сетей.	Визуализируйте сеть с помощью алгоритма спринг лейаут.
	2.Осуществляет рациональный выбор программного продукта в зависимости от поставленной задачи.	<u>Знать:</u> возможности современных библиотек для сетевого анализа. <u>Уметь:</u> уметь решать типовые задачи сетевого анализа с помощью современных библиотек сетевого анализа.	Загрузить набор данных о графе «Карате клуб» в библиотеку NetworkX.
	3. Владеет навыками обеспечения информационной безопасности автоматизированных систем.	<u>Знать:</u> подходы к формулировке типичных задач обеспечения ИБ с помощью сетевого анализа. <u>Уметь:</u> решать типичные задачи обеспечения ИБ с помощью сетевого анализа.	С помощью библиотеки PyTorch решите задачу классификации элементов описанных атрибутами разных типов.
ПКП-5 Способность применять методы и инструменты анализа данных и машинного обучения при подготовке аналитического обоснования	1.Осуществляет обоснованный выбор методов и инструментов обработки данных для решения экономических и финансовых задач.	<u>Знать:</u> способы решения выделения сообществ в графе. <u>Уметь:</u> на практике выполнять выявления сообществ в графе.	Выделить сообщества в графе «Карате клуб».

финансово-экономических решений	2. Владеет навыками решения прикладных задач с использованием методов и инструментов анализа данных и машинного обучения.	<u>Знать:</u> способы подготовки и загрузки данных для решения задач МО на графах. <u>Уметь:</u> на практике подготавливать и загружать данные для решения задач МО на графах.	Загрузить набор данных о графе «Карате клуб» в библиотеку NetworkX.
ОП «Прикладное машинное обучение»			
ПКП-2 Способность выполнять анализ качества данных, выявлять и корректировать отклонения в данных и выполнять визуализацию данных	1. Демонстрирует знание методов анализа качества данных для различных типов значений.	<u>Знать:</u> подходы к созданию решений задач МО на графах. <u>Уметь:</u> уметь реализовывать решения задач МО на графах.	С помощью библиотеки PyTorch решите задачу классификации элементов описанных атрибутами разных типов.
	2. Владеет методами нормализации данных, в том числе работы с пропусками и выбросами.	<u>Знать:</u> способы подготовки и загрузки данных для решения задач МО на графах. <u>Уметь:</u> на практике подготавливать и загружать данные для решения задач МО на графах.	Загрузить набор данных о графе «Карате клуб» в библиотеку NetworkX.
	3. Владеет современным инструментарием и практическими навыками визуализации данных.	<u>Знать:</u> инструментальные средства для первичного анализа графовых (сетевых) данных. <u>Уметь:</u> использовать инструментальные средства для первичного анализа графовых (сетевых) данных.	Загрузить набор данных о графе «Карате клуб» в библиотеку NetworkX. Построить график распределения степеней узлов графа.

Примерные вопросы для подготовки к зачету

1. Теория сложных сетей: связь с другими областями знания, причины возникновения. Истоки междисциплинарной применимости теории.
2. Способы представления сети (графа) в памяти. Типизация сетей. Модель данных property graph.
3. Основные метрики сетей: средняя степень узлов, коэффициенты кластеризации. Метрики сети, основанные на путях.
4. Степенной закон распределения степеней узлов. Свойство безмасштабности. Степенной закон распределения степеней узлов для реальных сетей.
5. Степенной закон распределения степеней узлов. Проблематика визуализации хвоста распределения и подходы к решению проблемы. Корректный способ эмпирического определения параметров распределения.
6. Степенной закон распределения степеней узлов. Проблема эмпирической идентификации математического ожидания и дисперсии распределения. Корректный способ эмпирического определения параметров распределения.
7. Модель Барабаши-Альберта. Распределения степеней узлов в модели.
8. Модель случайного роста сети. Распределения степеней узлов в модели.
9. Эксперимент Мильграма. Феномен тесного мира. Сильные и слабые связи в сетях.
10. Модель тесного мира Уотса-Страттса. Динамика кластеризации и средней длины пути.
11. Проблематика идентификации тесного мира. Базовые и продвинутое критерии идентификации. Алгоритм латтисизации.
12. Проблема локальной навигации в сетях. Алгоритм близорукого поиска. Проблема поиска короткого пути в случайном графе и модели Клейнберга.
13. Модель Клейнберга. Локальная навигация в модели Клейнберга.
14. Центральность по степени. Определения, основные свойства, примеры.
15. Центральность по близости. Определения, основные свойства, примеры.
16. Центральность по посредничеству. Определения, основные свойства, примеры.

- 17.Центральность по собственному вектору. Определения, основные свойства, примеры.
- 18.Алгоритм PageRank: базовый принцип и полный алгоритм. Снятие ограничений базового принципа.
- 19.Алгоритм кластеризации сообществ в сети Гирвана-Ньюмана. Особенности алгоритма.
- 20.Мера модулярности и алгоритм кластеризации (выявления сообществ) в сети Лоувэйна. Особенности алгоритма.
- 21.Алгоритм кластеризации (выявления сообществ) в сети label propagation. Особенности алгоритма.
- 22.Спектральные методы кластеризация (выявления сообществ) в сети. Особенности подхода.
- 23.Задача распространения меток. Постановка, базовые алгоритмы решения.
- 24.Ядро и периферия графа. Алгоритм выделения ядра графа.
- 25.Спектральный метод наименьшего разреза графа

8.Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Основная литература:

- [1] Криволапов, С. Я. Математика на Python : учебник / С. Я. Криволапов, М. Б. Хрипунова. — Москва : КноРус, 2022. — 455. — ЭБС BOOK.ru. - URL:<https://book.ru/book/943665> (дата обращения: 13.06.2023). — Текст : электронный.
- [2] Криволапов, С. Я. Использование языка Python в теории вероятностей : учебник / С. Я. Криволапов. — Москва : Прометей, 2022. — 492 с. — ЭБС Лань. — URL: <https://e.lanbook.com/book/220814> (дата обращения: 13.06.2023). — Текст : электронный.

Дополнительная литература:

[3] Криволапов, С. Я. Статистические вычисления на платформе Jupyter Notebook с использованием Python: учебник / С. Я. Криволапов. — Москва : КноРус, 2022. — 431 с. — ЭБС BOOK.RU. - URL:<https://book.ru/book/943660> (дата обращения: 13.06.2023). — Текст : электронный.

[4] Рутковская, Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилиньский, Л. Рутковский; пер. с польск. И. Д. Рудинского. - 2-е изд., стереотип. – Москва : Гор. линия-Телеком, 2013. - 384 с. - ЭБС ZNANIUM.com. - URL: <https://new.znanium.com/catalog/product/414545> (дата обращения: 13.06.2023). - Текст : электронный.

[5] Ростовцев, В. С. Искусственные нейронные сети : учебник / В. С. Ростовцев. — Санкт-Петербург : Лань, 2021. — 216 с. — ЭБС Лань. — URL: <https://e.lanbook.com/reader/book/160142> (дата обращения: 13.06.2023). — Текст : электронный.

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>
2. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
3. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>
4. Электронно-библиотечная система Znanium <http://www.znanium.com>
5. Электронно-библиотечная система издательства «ЮРАЙТ» <https://urait.ru/>
6. Электронно-библиотечная система издательства Проспект <http://ebs.prospekt.org/books>
7. Электронно-библиотечная система издательства Лань <https://e.lanbook.com/>
8. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
9. Электронная библиотека Издательского дома «Гребенников» <https://grebennikon.ru/>

10. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
11. Национальная электронная библиотека <http://нэб.рф/>
12. Финансовая справочная система «Финансовый директор» <http://www.1fd.ru/>
13. Электронная документация по работе в pytorch (вариант 1):
<https://neurohive.io/ru/tutorial/glubokoeobuchenie-s-pytorch/>
14. Электронная документация по работе в pytorch (вариант 2):
<https://pytorch.org/tutorials/>
15. Дистрибутив anaconda: <http://anaconda.org/>
16. Электронная документация по работе в библиотеке pandas:
<http://pandas.pydata.org/>
17. Курс лекций по машинному и глубокому обучению (электронный учебник):
<http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf>

10. Методические указания для обучающихся по освоению дисциплины.

Основные этапы работы студента по дисциплине ***Машинное обучение в семантическом и сетевом анализе***

1. Предварительная ориентировка в подлежащем изучению учебном материале по программе.
2. Ознакомление с рекомендованной учебной литературой.
3. Слушание и конспектирование лекций, а также выполнение других видов учебной работы.
4. Планирование самостоятельной работы.
5. Обобщение и систематизация информации, почерпнутой из лекций и прочитанной литературы.
6. Выполнение контрольной работы.

Рекомендации по работе с учебным материалом:

1. Осознавайте наличный уровень полученных вами знаний.
2. В ситуации непонимания нужно выявить тот первичный уровень и факторы непонимания, которые стали препятствием понимания последующего.

3. Задавайте сами себе вопросы и пытайтесь ответить на них.

Рекомендации по работе на лекции и с лекционным материалом:

1. Основная задача на лекции – осмысление излагаемого в ней материала. Для этого необходимо слушать лекцию с самого начала, не упуская общих, ориентирующих в материале рассуждений и установок лектора.
2. Ведение записей на лекции важно и полезно для лучшего осмысливания материала, для сохранения информации, с целью ее дальнейшего использования.
3. Для облегчения записи рекомендуется применять сокращения повторяющихся терминов или хорошо известных понятий.

Рекомендации по работе с литературой:

1. Если возникли затруднения при разыскивании материала, по какому-либо конкретному вопросу, следует обратиться к предметному указателю, напечатанному, как правило, в конце каждого литературного источника.
2. Предметный указатель – это алфавитный список основных научных понятий (терминов), содержание которых раскрыто в книге, рядом с термином стоят числа, обозначающие номера страниц, на которых изложен материал, относящийся к данному понятию.

11.Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

11.1. Комплект лицензионного программного обеспечения:

1. Пакет офисных программ;
2. Антивирус Kaspersky;

11.2. Современные профессиональные базы данных и информационные справочные системы

1. Информационно-правовая система «Гарант»,
2. Информационно-правовая система «Консультант Плюс»,

3. Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>

4. Система комплексного раскрытия информации «СКРИН» -
<http://www.skrin.ru/>

11.3. Сертифицированные программные и аппаратные средства защиты информации:

- не предусмотрены.

11.4. Язык программирования Python 3.x в среде Windows.

11.5. Платформа для научных исследований, основанная на языке программирования Python, Anaconda.

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Наличие аудитории, оснащенной компьютерной техникой и проектором, с возможностью подключения к сети «Интернет».