

Федеральное государственное образовательное бюджетное учреждение
высшего образования

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

**Кафедра бизнес-информатики
Факультет информационных технологий и анализа больших данных**

Документ подписан усиленной неквалифицированной электронной подписью
Организация: Финансовый университет при Правительстве РФ
Утверждено: Проректор по учебной и методической работе Е.А. Каменева
Сертификат: StEfDrBRNemfTn9GHWunsA7DurPlzl1j
Дата: 18.03.2025 г.

П. А. Сахнюк

Технологии продвинутой аналитики

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки

38.03.05 - Бизнес-информатика

Образовательная программа:

Цифровая трансформация управления бизнесом

Профиль

Технологии цифровых бизнес-моделей

Рекомендовано

*Факультет информационных технологий и анализа больших данных
(протокол №56 от 17.06.2025)*

Одобрено

*Кафедра бизнес-информатики
(протокол №9 от 28.04.2025)*

Москва 2025

Содержание

Наименование разделов РПД		Стр.
1.	Наименование дисциплины	5
2.	Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	5
3.	Место дисциплины в структуре образовательной программы	13
4.	Объем дисциплины(модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	13
5.	Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий	13
5.1.	Содержание дисциплины	13
5.2.	Учебно-тематический план	15
5.3.	Содержание семинаров, практических занятий	16
6.	Перечень учебно-методического обеспечения для	17

	самостоятельной работы обучающихся по дисциплине	
6.1.	Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы	17
6.2.	Перечень вопросов, заданий, тем для подготовки к текущему контролю	18
7.	Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине	19
8.	Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	26
9.	Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	28
10.	Методические указания для обучающихся по освоению дисциплины	28
11.	Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных	28

	справочных систем	
12.	Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине	29

1. Наименование дисциплины

«Технологии продвинутой аналитики».

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

Код компетенции	Наименование компетенции	Индикатор достижения компетенции	Результаты обучения
ПКН-11	Способность управлять ИТ-финансами и ИТ-бюджетом	Демонстрирует знания об основных экономических методах, используемых в экономике ИТ	Знать: - как методы и концепции, применяемые в сфере информационных технологий, такие как расчет стоимости владения ИТ-решениями (ТСО), оценка возврата инвестиций (ROI), анализ затрат и выгод (CBA), а также определение жизненного цикла продукта (PLC) могут быть использованы для оценки эффективности внедрения цифровых технологий, таких как системы машинного обучения или инструменты анализа данных;

			- особенности экономики больших данных, включая затраты на сбор, хранение и обработку данных, а также стоимость масштабирования решений; - влияние технологий анализа данных (например, Polar, Vaex, Dask) на снижение операционных издержек и повышение конкурентоспособности бизнеса. Уметь: - анализировать затраты на разработку, внедрение и поддержку IT-решений, а также оценивать их экономическую выгоду; - использовать данные и аналитические инструменты (например, SQL, Pandas, Dask) для моделирования сценариев, прогнозирования эффекта от внедрения технологий и подготовки
--	--	--	---

			рекомендаций для руководства; - представлять результаты экономического анализа в виде четких отчетов или презентаций.
		Применяет основные экономические методы для оценки ИС	Знать: - основные экономические методы, используемые для оценки информационных систем (ИС); - как эти методы применяются для анализа затрат и выгод от внедрения ИС, включая системы, основанные на анализе данных и машинном обучении; - особенности оценки ИС, работающих с большими объемами данных, включая затраты на хранение, обработку и масштабирование данных, а также стоимость использования современных инструментов (например, Polar,

			Vaex, Dask). Уметь: - применять экономические методы для оценки информационных систем, используя данные и инструменты анализа (Python, SQL, Polar, Vaex, Dask). - использовать инструменты анализа данных для моделирования сценариев, прогнозирования эффекта от внедрения ИС и подготовки рекомендаций.
ПКП-1	Способность создавать новые бизнес-модели в условиях цифровизации бизнеса	Выявляет особенности, ограничения и возможности текущей бизнес-модели компании	Знать: - современные технологии и инструменты для анализа больших объемов структурированных данных, такие как библиотеки Python (Polar, Vaex, Dask, Modin, Ray) и SQL; - принципы работы с большими данными, включая их предобработку, агрегацию и параллельную обработку;

			- основные методы машинного обучения, применимые к анализу данных (регрессия, классификация, кластеризация и др.), их ограничения (например, зависимость от качества данных, вычислительные затраты). - как данные могут быть использованы для выявления особенностей бизнес-процессов, таких как оптимизация цепочек поставок, прогнозирования спроса или анализ клиентского поведения.; - какие типы данных (структурированные, полуструктурированные) подходят для анализа и как их можно интегрировать в существующие бизнес-модели. Уметь: - применять
--	--	--	--

			инструменты продвинутой аналитики (Python, SQL, Polar, Vaex, Dask, Modin, Ray) для анализа больших объемов данных и выявления особенностей бизнес-модели компании; - проводить предобработку данных (фильтрацию, нормализацию, обработку пропусков), выполнять агрегацию и визуализацию данных для выявления скрытых закономерностей; - строить модели машинного обучения для решения бизнес-задач (например, прогнозирования или сегментации клиентов) и оценивать их эффективность; - интерпретировать результаты анализа, формулировать выводы о возможностях и ограничениях
--	--	--	---

			текущей бизнес-модели и предлагать пути их улучшения.
		Разрабатывает и обосновывает варианты целевых бизнес-моделей компании, в основе которых лежит применение цифровых технологий	Знать: - современные цифровые технологии и инструменты анализа данных (Polar, Vaex, Dask, Modin, Ray), которые могут быть использованы для создания новых или оптимизации существующих бизнес-моделей; - как машинное обучение и анализ больших данных могут стать основой для разработки решений, таких как персонализация услуг, прогнозирование спроса, автоматизация процессов и создание продуктов на основе данных; - принципы работы с большими объемами структурированных данных, включая их предобработку, агрегацию и параллельную обработку, а

			также осознавать ограничения этих технологий (например, зависимость от качества данных, вычислительные затраты). Уметь: - применять инструменты продвинутой аналитики (Python, SQL, Polar, Vaex, Dask, Modin, Ray) для анализа данных и разработки целевых бизнес-моделей, основанных на цифровых технологиях. - выявлять скрытые закономерности в данных, строить модели машинного обучения для решения бизнес-задач (например, прогнозирования, классификации или кластеризации) и интерпретировать их результаты. - формулировать гипотезы о новых бизнес-моделях,
--	--	--	--

			обосновывать их экономическую эффективность и предлагать пути внедрения.
--	--	--	--

3. Место дисциплины в структуре образовательной программы

Дисциплина «Технологии продвинутой аналитики» относится к «Модулю "Информационно-аналитические технологии"»

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

Очная форма обучения

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 6 (в часах)
Общая трудоемкость дисциплины	3/108	108
Контактная работа - Аудиторные занятия	66	66
<i>Лекции</i>	16	16
<i>Семинары, практические занятия</i>	50	50
Самостоятельная работа	42	42
Вид текущего контроля	Контрольная работа;	Контрольная работа
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. Машинное обучение на Python и SQL

Применение pandas для анализа и обработки данных. Библиотеки визуализации данных Matplotlib, Seaborn, Altair, Plotly Express. Exploratory data analysis (EDA) - анализ основных свойств данных, нахождение в них общих закономерностей, распределений и аномалий, построение начальных моделей, зачастую с использованием инструментов визуализации. Автоматизация исследовательского анализа данных с использованием pandas profiling, библиотеки Mitosheet, Bamboolib, Dataprep.

Методика CRISM-DM, машинное обучение для решения задач Data Mining. Градиентный спуск в машинном обучении. Деревья решений. Бэггинг, бустинг, стекинг. Фреймворки машинного обучения: XGBoost, LightGBM, CatBoost, h2o.ai. Кластерный анализ, алгоритм k-means. Поиск ассоциативных правил. Анализ временных рядов. Автоматическое машинное обучение (AutoML): PyCaret, AutoGluon, LAMA.

Машинное обучение и технологии на базе искусственного интеллекта (ИИ) в платформах бизнес-аналитики и науки о данных.

Машинное обучение в корпоративных хранилищах данных средствами SQL.

Тема 2. Библиотеки python для анализа больших объемов структурированных данных: Polar, Vaex, Dask, Modin, Ray.

Многопоточный механизм запросов фреймворка Polars написан на Rust и разработан для эффективного параллелизма. Его векторизованная и столбчатая обработка обеспечивает алгоритмы, согласованные с кэшем, и высокую производительность на современных процессорах.

Dask: планировщик низкого уровня и частичная замена Pandas высокого уровня, ориентированный на выполнение кода на вычислительных кластерах.

Ray: низкоуровневый фреймворк для распараллеливания кода Python между процессорами или кластерами. Modin: замена Pandas на базе Dask или Ray.

Vaex: частичная замена Pandas, использующая ленивые вычисления, чтобы позволить работать с большими наборами данных на стандартных компьютерах. Его замена Pandas охватывает некоторые API Pandas, но больше ориентирована на исследование данных и визуализацию.

5.2. Учебно-тематический план

Очная форма обучения

№ п/п	Наименование тем(разделов) дисциплины	Трудоемкость в часах				
		Всего	Контактная работа - Аудиторная работа			Самостоятельная работа
			Общая, в т.ч.:	Лекции	Семинары, практические занятия	
1	Машинное обучение на Python и SQL	50	28	8	20	22
2	Библиотеки python для анализа больших объемов структурированных данных: Polar, Vaex, Dask,	58	38	8	30	20

	Modin, Ray.					
	Итого	108	66	16	50	42

5.3. Содержание семинаров, практических занятий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарах, практических занятиях	Формы проведения занятий
Машинное обучение на Python и SQL	1. Решение регрессионной задачи с помощью библиотек scikit-learn, XGBoost, LightGBM, CatBoost, H2O 2. Исследование алгоритмов построения ансамблей деревьев решений. 3. Сравнение промышленных фреймворков машинного обучения. 4. Использование кластерного анализа для решения задач Data Mining. 5. Сравнение фреймворков машинного обучения, используемых для анализа и прогнозирования временных рядов. 6. Основные фреймворки машинного обучения с функцией AutoML. 7. Применение технологий машинного обучения и искусственного интеллекта в платформах BI и Data Science. 8. Машинное обучение в Google BigQuery.	Выполнение и защита практических заданий

Библиотеки python для анализа больших объемов структурированных данных: Polar, Vaex, Dask, Modin, Ray.	1. Базовая аналитика и описательные статистики на больших объемах структурированных данных с использованием библиотек Polar, Vaex, Dask, Modin, Ray. 2. Визуальная аналитика больших объемах структурированных данных 3. Машинное обучение на больших объемах данных.	Выполнение и защита практических заданий
--	---	--

6. Перечень учебно-методического обеспечения 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
Машинное обучение на Python и SQL	Особенности фреймворков AutoML: FLAML; TPOT; Auto-sklearn.	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Подготовка к семинарам
Библиотеки python для анализа больших объемов структурированных данных: Polar, Vaex, Dask, Modin, Ray.	Машинное обучение с Dask	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и

		дополнительной литературы, интернет – источников. Подготовка к семинарам
--	--	--

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

Студенту в начале семестра предлагается выбрать кейс, на основе которого разрабатывается аналитический проект. Для этого можно использовать датасеты с Kaggle <https://www.kaggle.com/datasets?search=customer> , содержащие клиентские данные (всего 2466 различных Datasets).

Например:

1. Customer Personality Analysis.
2. Telco Customer Churn.
3. Shop Customer Data.
4. Customer Segmentation .
5. Mall Customer Segmentation Data .
6. Credit Card customers .
7. Starbucks Customer Survey.
8. Airlines Customer satisfaction.
9. Loan Prediction Based on Customer Behavior.
10. Brazilian E-Commerce Public Dataset by Olist.

Задание:

1. Провести исследовательский анализ данных: правильно прочесть данные, выполнить предварительную обработку данных (preprocessing), рассмотреть описательную статистику, визуализировать данные, выдвинуть предварительные гипотезы.
2. Обогащить датасет, проведя клиентскую аналитику: выполнить ABC - XYZ (RFM , когортный) -анализ, используя библиотеки Python .

3. Создать, обучить, оценить по метрикам качества, объяснить модели машинного обучения библиотеки Python , создать прогноз на новых данных.

Дополнительная информация:

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Кафедры бизнес-информатики.

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

ПKN-11 Способность управлять ИТ-финансами и ИТ-бюджетом

1) Демонстрирует знания об основных экономических методах, используемых в экономике ИТ

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: - как методы и концепции, применяемые в сфере информационных технологий, такие как расчет стоимости владения ИТ-решениями (TCO), оценка возврата инвестиций (ROI), анализ затрат и выгод (CBA), а также определение жизненного цикла продукта (PLC) могут быть использованы для оценки эффективности внедрения цифровых технологий, таких как системы машинного обучения или инструменты анализа данных; - особенности экономики больших данных, включая затраты на сбор,

хранение и обработку данных, а также стоимость масштабирования решений; - влияние технологий анализа данных (например, Polar, Vaex, Dask) на снижение операционных издержек и повышение конкурентоспособности бизнеса.

Уметь: - анализировать затраты на разработку, внедрение и поддержку IT-решений, а также оценивать их экономическую выгоду; - использовать данные и аналитические инструменты (например, SQL, Pandas, Dask) для моделирования сценариев, прогнозирования эффекта от внедрения технологий и подготовки рекомендаций для руководства; - представлять результаты экономического анализа в виде четких отчетов или презентаций.

Типовые контрольные задания

Задание 1:

Оцените экономическую эффективность внедрения системы прогнозирования спроса на базе машинного обучения для розничной сети. Используйте SQL и библиотеки Python (например, Pandas, Dask) для анализа исторических данных о продажах и расчета метрик, таких как TCO (стоимость владения системой), ROI (возврат инвестиций) и CBA (анализ затрат и выгод). Подготовьте отчет с выводами и рекомендациями для руководства компании.

Задание 2:

Проанализируйте экономическую модель проекта по автоматизации обработки больших объемов структурированных данных с использованием библиотеки Polars. Оцените выгоды от автоматизации, такие как снижение трудозатрат и повышение точности анализа. Подготовьте презентацию для заинтересованных сторон, демонстрирующую экономическую эффективность проекта и предлагая стратегию его масштабирования.

2) Применяет основные экономические методы для оценки ИС

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: - основные экономические методы, используемые для оценки информационных систем (ИС); - как эти методы применяются для анализа затрат и выгод от внедрения ИС, включая системы, основанные на анализе данных и машинном обучении; - особенности оценки ИС, работающих с большими объемами данных, включая затраты на хранение, обработку и масштабирование данных, а также стоимость использования современных инструментов (например, Polar, Vaex, Dask).

Уметь: - применять экономические методы для оценки информационных систем, используя данные и инструменты анализа (Python, SQL, Polar, Vaex, Dask). - использовать инструменты анализа данных для моделирования сценариев, прогнозирования эффекта от внедрения ИС и подготовки рекомендаций.

Типовые контрольные задания

Задание 1:

Оцените экономическую эффективность внедрения информационной системы для анализа клиентских данных с использованием библиотеки Polars. Проанализируйте, как использование системы повлияет на бизнес-процессы (например, улучшение качества маркетинговых кампаний). Подготовьте отчет с выводами и рекомендациями для руководства компании.

Задание 2:

Проанализируйте экономическую целесообразность внедрения системы прогнозирования отказов оборудования на производстве с использованием машинного обучения и библиотеки Vaex. Опишите ограничения текущей системы мониторинга и предложите стратегию внедрения новой ИС. Подготовьте презентацию для заинтересованных сторон, демонстрирующую экономическую эффективность проекта и предлагая план его реализации.

ПКП-1 Способность создавать новые бизнес-модели в условиях цифровизации бизнеса

1) Выявляет особенности, ограничения и возможности текущей бизнес-модели компании

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: - современные технологии и инструменты для анализа больших объемов структурированных данных, такие как библиотеки Python (Polar, Vaex, Dask, Modin, Ray) и SQL; - принципы работы с большими данными, включая их предобработку, агрегацию и параллельную обработку; - основные методы машинного обучения, применимые к анализу данных (регрессия, классификация, кластеризация и др.), их ограничения (например, зависимость от качества данных, вычислительные затраты). - как данные могут быть использованы для выявления особенностей бизнес-процессов, таких как оптимизация цепочек поставок, прогнозирование спроса или анализ клиентского поведения.; - какие типы данных (структурированные, полуструктурированные) подходят для анализа и как их можно интегрировать в существующие бизнес-модели.

Уметь: - применять инструменты продвинутой аналитики (Python, SQL, Polar, Vaex, Dask, Modin, Ray) для анализа больших объемов данных и выявления особенностей бизнес-модели компании; - проводить предобработку данных (фильтрацию, нормализацию, обработку пропусков), выполнять агрегацию и визуализацию данных для выявления скрытых закономерностей; - строить модели машинного обучения для решения бизнес-задач (например, прогнозирования или сегментации клиентов) и оценивать их эффективность; - интерпретировать результаты анализа, формулировать выводы о возможностях и ограничениях текущей бизнес-модели и предлагать пути их улучшения.

Типовые контрольные задания

Задание 1:

Проанализируйте данные о продажах крупной розничной сети с использованием библиотек Pandas, Dask и SQL. Выявите ключевые закономерности в данных (например, сезонные колебания спроса, популярные категории товаров) и оцените ограничения текущей бизнес-модели, такие как неэффективное управление запасами или недостаточная персонализация маркетинговых кампаний. Предложите стратегии улучшения бизнес-процессов на основе анализа данных, например, внедрение системы прогнозирования спроса или оптимизация ценообразования. Подготовьте отчет с визуализациями и рекомендациями.

Задание 2:

Используйте библиотеку Vaex для анализа больших объемов данных о клиентах онлайн-сервиса (например, история покупок, поведение на сайте). Постройте модель машинного обучения (например, кластеризации) для сегментации клиентов на группы с похожим поведением. Определите возможности для улучшения бизнес-модели, такие как персонализация предложений или удержание клиентов с высоким риском оттока. Опишите ограничения используемых данных (например, неполнота или шум) и предложите способы их устранения.

2) Разрабатывает и обосновывает варианты целевых бизнес-моделей компании, в основе которых лежит применение цифровых технологий

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: - современные цифровые технологии и инструменты анализа данных (Polar, Vaex, Dask, Modin, Ray), которые могут быть использованы для создания новых или оптимизации существующих бизнес-моделей; - как машинное обучение и анализ больших данных могут стать основой для разработки решений, таких как персонализация услуг, прогнозирование спроса, автоматизация процессов и создание продуктов на основе данных; - принципы работы с большими объемами структурированных данных, включая их предобработку, агрегацию и параллельную обработку, а также осознавать ограничения этих технологий (например, зависимость от качества данных, вычислительные затраты).

Уметь: - применять инструменты продвинутой аналитики (Python, SQL, Polar, Vaex, Dask, Modin, Ray) для анализа данных и разработки целевых бизнес-моделей, основанных на цифровых технологиях. - выявлять скрытые закономерности в данных, строить модели машинного обучения для решения бизнес-задач (например, прогнозирования, классификации или кластеризации) и интерпретировать их результаты. - формулировать гипотезы о новых бизнес-моделях, обосновывать их экономическую эффективность и предлагать пути внедрения.

Типовые контрольные задания

Задание

1:

Проанализируйте данные о клиентах онлайн-платформы (например, история покупок, поведение на сайте) с использованием

библиотек Pandas, Dask и SQL. Разработайте новую бизнес-модель платформы, основанную на персонализации предложений с помощью рекомендательной системы, построенной с использованием машинного обучения. Оцените экономическую эффективность модели, рассчитав метрики, такие как увеличение среднего чека или конверсии. Обоснуйте выбор технологий и предложите план внедрения модели.

Задание 2:

Используйте библиотеку Vaex для анализа данных о производственных процессах крупного предприятия (например, данные о времени выполнения операций, количестве брака). Разработайте бизнес-модель, основанную на внедрении системы прогнозирования отказов оборудования с использованием машинного обучения. Предложите стратегию монетизации модели (например, снижение издержек за счет предотвращения простоев). Опишите ограничения используемых данных и технические требования к внедрению модели. Подготовьте презентацию для руководства компании с четкими рекомендациями по реализации проекта.

Примеры практико-ориентированных заданий

Задание 1

Выполнить консолидацию данных кейса (если кейс состоит из нескольких таблиц),

провести очистку, трансформацию данных (ETL), выполнить preprocessing данных провести исследовательский анализ данных с использованием библиотек g python Numpy , Pandas , проанализировать данные временных рядов с использованием библиотеки D - Tale .

Задание 2

Решить задачу анализа временных рядов с использованием библиотеки Prophet

Задание 3

В Colab выполнить консолидацию данных кейса (если кейс состоит из нескольких таблиц),

провести очистку, трансформацию данных (ETL), выполнить preprocessing данных, провести исследовательский анализ данных с использованием библиотек python Numpy, Pandas, Scikit-learn

Задание 4

В Colab провести очистку, трансформацию данных, выполнить preprocessing данных, провести исследовательский анализ данных с использованием библиотек визуализации данных Matplotlib, Seaborn и Plotly Express

Задание 5

В Colab сравнивать алгоритмы регрессии, градиентного бустинга, случайного леса, используемые для создания модели машинного обучения, применяя фреймворк scikit-learn. Провести прогнозирование ответа лучшей модели на новых данных

Задание 6

В Colab используя библиотеку Vaex провести исследовательский анализ

больших объемов структурированных данных

Примерные вопросы для подготовки к Зачету

1. Охарактеризуйте отличия понятий искусственный интеллект, машинное обучение и глубокое обучение.
2. Охарактеризуйте процесс подготовки данных и разработки функций в машинном обучении.
3. Охарактеризуйте принципы работы машинного обучения: виды машинного обучения и минимизация функции потерь.
4. Охарактеризуйте метрики качества классификации: матрица ошибок (неточностей), Accuracy, Precision и Recall, F1-мера.
5. Охарактеризуйте основные возможности библиотеки Polar.

6. Охарактеризуйте ограничения библиотеки Polar.
7. Охарактеризуйте основные возможности библиотеки Vaex.
8. Охарактеризуйте ограничения библиотеки Vaex.
9. Охарактеризуйте основные возможности библиотеки Dask.
10. Охарактеризуйте ограничения библиотеки Dask.
11. Охарактеризуйте основные возможности библиотеки Modin.
12. Охарактеризуйте ограничения библиотеки Modin.
13. Охарактеризуйте основные возможности библиотеки Ray.
14. Охарактеризуйте ограничения библиотеки Ray.

Примерные практические задания к зачету:

1. Разработка сценариев ABC-XYZ, RFM, когортного - анализ в выбранной среде.
2. Визуализация результатов аналитики: в библиотеки Altair, Plotly, а также встроенными средствами платформ Deepnote (или Datalore, Mode).
3. Базовая аналитика и описательные статистики на больших объемах структурированных данных с использованием библиотек Polar, Vaex, Dask, Modin, Ray.
4. Визуальная аналитика больших объемах структурированных данных с использованием библиотек Polar, Vaex, Dask, Modin, Ray.

Методические материалы, определяющие процедуры оценивания знаний, умений

Приказ от 01.10.2024 №2187/о «Об утверждении Положения о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся по образовательным программам высшего образования в Финансовом университете».

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

Нормативно-правовые акты:

1. Федеральный Закон Российской Федерации «Об информации, информационных технологиях и о защите информации» № 149-ФЗ от 13.07.2015 г.
2. Паспорт национальной программы «Цифровая экономика Российской Федерации» (утв. президиумом Совета при Президенте РФ по стратегическому развитию и национальным проектам, протокол от 24.12.2018 N 16).
3. Государственная программа Российской Федерации "Информационное общество (2011-2020 годы)" (в ред. Постановления Правительства РФ от 18.05.2011 N 399).

Основная литература:

1. Большие данные и NoSQL базы данных [Электронный ресурс] : учебное пособие для вузов / Мамедли Р. Э., Казиахмедов Т. Б. ; Казиахмедов Т. Б. Санкт-Петербург : Лань , 2024 92 с. Книга из коллекции Лань - Информатика <https://e.lanbook.com/book/434051> ISBN 978-5-507-49873-4
2. Протоdjяконов , Андрей Владимирович Алгоритмы Data Science и их практическая реализация на Python : Учебное пособие / Кузбасский государственный технический университет имени Т. Ф. Горбачева Вологда : Инфра-Инженерия , 2022 392 с. ВО - Бакалавриат <https://znanium.com/catalog/document?id=417222> ISBN 978-5-9729-1006-9
3. Языки программирования. Python: решение сложных задач [Электронный ресурс] : учебное пособие для вузов / Борзунов С. В., Кургалин С. Д. Санкт-Петербург : Лань , 2023 192 с. Книга из коллекции Лань - Информатика <https://e.lanbook.com/book/319394> ISBN 978-5-507-45923-0

Дополнительная литература:

1. Параскевов , Александр Владимирович Большие данные : Учебное пособие / Кубанский государственный аграрный университет им.

И.Т. Трубилина Вологда : Инфра-Инженерия , 2024 148 с.
Профессиональное образование <https://znanium.ru/catalog/document?id=451729> ISBN 978-5-9729-2120-1

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

1. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/> (<http://library.fa.ru/files/elibfa.pdf>)
2. Образовательная платформа "ЮРАЙТ" <https://urait.ru/>
3. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
4. Информационно-образовательный портал Финуниверситета: <https://org.fa.ru>
5. 7.Электронные ресурсы БИК:
6. Школа больших данных [Электронный ресурс]: сайт. – Режим доступа: <https://www.bigdataschool.ru/>
7. Онлайн-учебник по машинному обучению. -- [Электронный ресурс]: сайт. – Режим доступа: <https://academy.yandex.ru/handbook/ml>
8. MachineLearning.ru. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. – [Электронный ресурс]: сайт. – Режим доступа: <http://machinelearning.ru>

10. Методические указания для обучающихся по освоению дисциплины

Студентам необходимо руководствоваться «Методическими рекомендациями по планированию и организации внеаудиторной самостоятельной работы по образовательным программам бакалавриата и магистратуры в Финансовом университете» (Приказ ректора № 1040_о от 11.05.2021) и данной рабочей программой дисциплины.

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

- Комплект лицензионного программного обеспечения:

1. Python 3.8
2. Jupyter Notebook (Windows 10)
3. Anaconda
4. PyTorch
5. библиотеки Keras
6. Yandex Cloud
7. Mail Cloud
8. Kaspersky

- Современные профессиональные базы данных и информационные справочные системы:

1. Язык программирования Python 3. <https://pythonworld.ru/>

- Сертифицированные программные и аппаратные средства защиты информации

1. не предусмотрены

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

1. **Компьютерный класс** для проведения учебных занятий, предусмотренных программой, в том числе групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенный оборудованием и техническими средствами обучения: мебель аудиторная (столы,

стулья, доска аудиторная), персональные компьютеры, набор демонстрационного оборудования (проектор, экран)