

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: zKpwFBnkveMIU3Q96bqFc5TKdD2n2uEV
Дата: 06.05.2025 г.

П. А. Сахнюк

Технологии продвинутой аналитики

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки
38.04.01 - Экономика

Направленность программы магистратуры
Анализ и стратегический менеджмент в бизнесе

Рекомендовано

*Факультет информационных технологий и анализа больших данных
(протокол №55 от 20.05.2025)*

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	8
4.	Объем дисциплины(модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	8
5.	Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий	8
5.1.	Содержание дисциплины	8
5.2.	Учебно-тематический план	9
5.3.	Содержание семинаров, практических занятий	10
6.	Перечень учебно-методического обеспечения для	11

	самостоятельной работы обучающихся по дисциплине	
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	13
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	20

	справочных систем	
12.	Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине	21

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
ПК-2	Способность применять современные инструменты анализа, в том числе, с использованием интеллектуальных информационно-аналитических систем	Проводит самостоятельные исследования бизнес-процессов в соответствии с разработанной производственной программой с использованием современных информационно-аналитических систем	Знать: основные методы и инструменты для проведения исследований бизнес-процессов с использованием современных информационно-аналитических систем; ключевые этапы исследования: определение целей, сбор данных, их предобработка, анализ и интерпретация результатов; как использовать машинное обучение для выявления закономерностей и оптимизации процессов. Уметь: самостоятельно проводить

			<p>исследования бизнес-процессов, используя современные аналитические инструменты;</p> <p>собирать данные о бизнес-процессах из различных источников (ERP-системы, CRM, базы данных) и выполнять их предобработку (фильтрация, агрегация, очистка) с использованием библиотек Python (например, Dask для параллельной обработки или Vaex для работы с большими объемами данных);</p> <p>анализировать данные для выявления неэффективных участков процессов, например, через построение отчетов, визуализацию трендов и выполнение статистического анализа.</p>
		Демонстрирует владение методами сбора,	Знать: современные методы и

		<p>анализа и обработки данных для принятия управленческих решений при моделировании бизнес-процессов</p>	<p>инструменты для сбора, анализа и обработки данных, которые применяются для моделирования бизнес-процессов;</p> <p>как данные могут быть использованы для моделирования бизнес-процессов, выявления узких мест и оптимизации операций.</p> <p>Уметь: собирать, анализировать и обрабатывать данные для моделирования бизнес-процессов и поддержки принятия управленческих решений;</p> <p>использовать SQL для извлечения данных из корпоративных баз данных и Python-библиотеки (например, Dask или Vaex) для их обработки и анализа; визуализировать данные и результаты</p>
--	--	--	---

			анализа для наглядного представления информации заинтересованным сторонам.
--	--	--	--

3. Место дисциплины в структуре образовательной программы

Дисциплина «Технологии продвинутой аналитики» относится к «Модулю направленности программы магистратуры»

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)	Модуль 6 (в часах)
Общая трудоемкость дисциплины	3/108	108
Контактная работа - Аудиторные занятия	40	40
<i>Лекции</i>	8	8
<i>Семинары, практические занятия</i>	32	32
Самостоятельная работа	68	68
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

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	54	20	4	16	34
2	Библиотеки python для анализа больших объемов структурированных данных: Polar, Vaex, Dask, Modin, Ray	54	20	4	16	34
	Итого	108	40	8	32	68

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	Выполнение и защита практических заданий

	<p>Data Mining.</p> <p>5. Сравнение фреймворков машинного обучения, используемых для анализа и прогнозирования временных рядов.</p> <p>6. Основные фреймворки машинного обучения с функцией AutoML.</p> <p>7. Применение технологий машинного обучения и искусственного интеллекта в платформах BI и Data Science.</p> <p>8. Машинное обучение в Google BigQuery.</p>	
<p>Библиотеки python для анализа больших объемов структурированных данных: Polar, Vaex, Dask, Modin, Ray</p>	<p>1. Базовая аналитика и описательные статистики на больших объемах структурированных данных с использованием библиотек Polar, Vaex, Dask, Modin, Ray.</p> <p>2. Визуальная аналитика больших объемах структурированных данных</p> <p>3. Машинное обучение на больших объемах данных.</p>	<p>Выполнение и защита практических заданий</p>

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 разрабатываться 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. Обогащать датасет, проводя 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

ПК-2 Способность применять современные инструменты анализа, в том числе, с использованием интеллектуальных информационно-аналитических систем

1) Проводит самостоятельные исследования бизнес-процессов в соответствии с разработанной производственной программой с использованием современных информационно-аналитических систем

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

Знать: основные методы и инструменты для проведения исследований бизнес-процессов с использованием современных информационно-аналитических систем; ключевые этапы исследования: определение целей, сбор данных, их предобработка, анализ и интерпретация результатов; как использовать машинное обучение для выявления закономерностей и оптимизации процессов.

Уметь: самостоятельно проводить исследования бизнес-процессов, используя современные аналитические инструменты; собирать данные о бизнес-процессах из различных источников (ERP-системы, CRM, базы данных) и выполнять их предобработку (фильтрация, агрегация, очистка) с использованием библиотек Python (например, Dask для параллельной обработки или Vaex для работы с большими объемами данных); анализировать данные для выявления неэффективных участков процессов, например, через построение отчетов, визуализацию трендов и выполнение статистического анализа.

Типовые контрольные задания

Задание 1:

Проведите исследование бизнес-процесса управления запасами в розничной сети. Используйте SQL для извлечения данных о продажах, остатках товаров и поставках, а затем выполните их анализ с помощью библиотеки Dask. Выявите проблемные зоны, такие как избыточные запасы или дефицит товаров, и предложите рекомендации для оптимизации процесса управления запасами.

Задание 2:

Исследуйте бизнес-процесс обслуживания клиентов в онлайн-сервисе. Используйте Vaex для анализа данных о поведении

клиентов (например, частота обращений, время ответа, уровень удовлетворенности). Постройте модели кластеризации для сегментации клиентов по их активности и предложите стратегии для улучшения качества обслуживания.

2) Демонстрирует владение методами сбора, анализа и обработки данных для принятия управленческих решений при моделировании бизнес-процессов

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

Знать: современные методы и инструменты для сбора, анализа и обработки данных, которые применяются для моделирования бизнес-процессов; как данные могут быть использованы для моделирования бизнес-процессов, выявления узких мест и оптимизации операций.

Уметь: собирать, анализировать и обрабатывать данные для моделирования бизнес-процессов и поддержки принятия управленческих решений; использовать SQL для извлечения данных из корпоративных баз данных и Python-библиотеки (например, Dask или Vaex) для их обработки и анализа; визуализировать данные и результаты анализа для наглядного представления информации заинтересованным сторонам.

Типовые контрольные задания

Задание 1:

Проведите анализ бизнес-процесса управления цепочками поставок в производственной компании. Используйте SQL для сбора данных о поставках, времени выполнения заказов и уровнях запасов, а затем выполните их предобработку и анализ с помощью библиотеки Dask. Постройте модель прогнозирования времени доставки с использованием машинного обучения и визуализируйте результаты. На основе анализа данных предложите рекомендации для оптимизации процесса поставок.

Задание 2:

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, проанализировать данные временных рядов с использованием библиотеки D-Tale.

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 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 - анализа 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. Языки программирования. Python: решение сложных задач [Электронный ресурс] : учебное пособие для спо / Борзунов С. В.,Кургалин С. Д. Санкт-Петербург : Лань , 2023 192 с. Книга из коллекции Лань - Информатика <https://e.lanbook.com/book/319391> ISBN 978-5-507-45922-3
2. 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
3. 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

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. <https://www.planetaexcel.ru/>
4. Электронно-библиотечная система издательства "Лань" <https://e.lanbook.com/>
5. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>

6. 7.Электронные ресурсы БИК:

7. Data Science Essentials / Microsoft. – <https://www.edx.org/course/data-science-essentials-microsoft-dat203-1x-3>

8. Специализация «Машинное обучение и анализ данных» / МФТИ и Яндекс. – <https://www.coursera.org/specializations/machine-learning-data-analysis>

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)