

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт экономики и менеджмента

УТВЕРЖДАЮ:
Директор Института
экономики и менеджмента


_____ Е.В. Нехода

« 20 » _____ 04 _____ 2023 г.

Рабочая программа дисциплины

Python и R для анализа данных

по направлению подготовки

38.04.01 Экономика

Направленность (профиль) подготовки:
«Экономика»

Форма обучения
Очная

Квалификация
Магистр

Год приема
2023

СОГЛАСОВАНО:

Руководитель ОП

 Н.А. Скрыльникова

Председатель УМК

 М.В. Герман

Томск – 2023

1. Цель и планируемые результаты освоения дисциплины (модуля)

Целью освоения дисциплины является формирование следующих компетенций:

– ПК-1 Способен определять направления развития организации.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ПК-1.3. Выявляет, анализирует и оценивает несоответствия между параметрами текущего и будущего состояния организации.

2. Задачи освоения дисциплины

– Освоить основы языков программирования R и Python.

– Научиться выполнять операции по управлению данными с применением языков программирования R и Python.

– Научиться выполнять анализ данных средствами языков R и Python, строить отчёты с результатами, содержащие диаграммы и сложно форматированный текст.

– Разрабатывать архитектуру web-приложений, позволяющие выполнять анализ данных и построение отчётов.

3. Место дисциплины (модуля) в структуре образовательной программы

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, является обязательной для изучения. Относится к профессиональному модулю «Бизнес аналитика»;

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Семестр 1, экзамен.

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются компетенции, сформированные в ходе освоения образовательных программ бакалавриата.

6. Язык реализации

Русский

7. Объем дисциплины (модуля)

Общая трудоемкость дисциплины составляет 6 з.е., 216 часов, из которых:

– лекции: 18 ч.;

– практические занятия: 28 ч.;

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины (модуля), структурированное по темам

Тема 1. Основные понятия языка R.

Синтаксис языка R. Типы данных. Структуры данных.

Тема 2. Графические возможности языка R.

Графическая подсистема языка R. Команды построения графиков низкого и высокого уровня.

Тема 3. Управляющие конструкции языка R: операторы ветвления, циклы, функции.

Управляющие конструкции языка R: циклы for, while, оператор ветвления if, оператор выбора switch, операторы break, continue.

Тема 4. Управление данными.

Агрегация данных. Объединение данных. Функции `aggregate` и `merge`.
Тема 5. Инструменты языка R для работы с текстовыми данными
Функции для поиска и замены подстрок, регулярные выражения в R.
Тема 6. Продвинутое инструменты обработки данных языка R: пакеты `dplyr`, `tidyr`.
Модель данных пакетов `dplyr` и `tidyr`, синтаксис команд обработки данных.
Тема 7. Продвинутое графические возможности языка R: пакеты `ggplot` и `lattice`.
Модель графического изображения в пакетах `ggplot` и `lattice`. Команды построения графиков средствами `ggplot` и `lattice`.
Тема 8. Основные понятия языка Python.
Синтаксис языка Python. Основные типы и структуры данных.
Тема 9. Функции и лямбда- функции.
Лямбда функции и другая "синтаксическая глазурь" языка Python.
Тема 10. Библиотека `numpy`.
Типы и структуры данных библиотеки `numpy`. Методы доступа к многомерным массивам `numpy`. Операции импорта/экспорта данных и линейной алгебры.
Тема 11. Библиотека `pandas`.
Типы и структуры данных библиотеки `pandas`. Методы доступа к данным объектов `Series` и `DataFrame`. Импорт/экспорт. Операции манипуляции данными.
Тема 12. Графическая библиотека `matplotlib`.
Команды для настройки и построения графиков в системе `matplotlib`.
Тема 13. Системы построения отчетов `Markdown`.
Команды формирования текста отчета на языке `Markdown`.
Тема 14. Инструмент разработки web-приложений `Shiny`
Компоненты системы `Shiny` и их взаимодействие.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости и выполнения домашних заданий и фиксируется в форме контрольной точки не менее одного раза в семестр.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Экзамен в первом семестре состоит из двух частей.

Первая часть предполагает письменный ответ на вопрос экзаменационного билета.

Вторая часть предполагает защиту подготовленного проекта по выбранной теме.

Билет содержит один теоретический вопрос. Продолжительность теоретического экзамена 1,5 часа.

Примерный перечень теоретических вопросов, проверяющих ПК -1.3.

1. Перечислите основные типы и структуры данных языка R.
2. Что такое атрибуты в языке R, каково их назначение? Перечислите основные атрибуты, их назначение и функции для работы с ними.
3. Перечислите основные отличия структур данных `atomic vector`, `matrix`, `list` и `data frame` языка R, назовите функции/методы, позволяющие преобразовать данные из одной структуры в другую.
4. Перечислите основные способы извлечения элементов из структур данных R, укажите их отличия и ограничения.
5. Перечислите основные инструменты импорта/экспорта данных в языке R.
6. Перечислите основные управляющие языковые конструкции языка R, приведите примеры их применения.
7. Перечислите основные способы генерации случайных данных и функции работы с распределениями случайных величин в языке R?
8. Перечислите основные команды пакета `Graphics Base` языка R, укажите их назначение и основные параметры.

9. Перечислите основные функции языка R для работы с данными (поиск и устранение пропущенных данных, поиск противоречий и дубликатов, извлечения подмножеств, фильтрации данных, соединения и агрегации данных).

10. Перечислите основные функции языка R для работы с символьными данными, отвечающими за решение следующих задач (форматированный вывод, поиск и замена строк).

11. Перечислите основные типы и структуры данных языка Python.

12. Перечислите способы определения функций на языке Python. Что такое lambda-функции, зачем они требуются?

13. Перечислите основные свойства и методы класса ndarray библиотеки NumPy.

14. Перечислите методы библиотеки NumPy, реализующие файловый ввод/вывод, методы линейной алгебры, работы со случайными числами и вероятностными распределениями.

15. Перечислите основные структуры данных библиотеки pandas.

16. Перечислите основные возможности библиотеки pandas в решении задач, связанных с манипулированием данными (индексация, сортировка, фильтрация, редукция, обработка пропусков и противоречий).

17. Перечислите основные возможности библиотеки pandas в решении задач, связанных с обработкой строк, дат, операций импорта/экспорта.

18. Перечислите основные возможности библиотеки pandas в решении задач, связанных с агрегированием данных, объединением данных из множества таблиц.

19. Перечислите основные возможности и методы графической библиотеки matplotlib.

20. Перечислите основные команды языка подготовки отчетов markdown.

Проект, выносимый на защиту, направлен на проверку ПК-1.3, является практически полезным и предполагает:

1. Самостоятельный выбор темы проекта.

2. Сбор и подготовку исходных данных по теме проекта.

3. Формулировку цели анализа данных.

4. Разработку программного решения на языке R или Python.

5. Анализ результатов и их бизнес-интерпретация.

6. Подготовку презентации и доклад с описанием основных элементов решения и полученных результатов.

Результаты экзамена определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Оценка «отлично» выставляется при выполнении следующих условий:

1. Ответ на теоретический вопрос является полным, содержит описание методов/конструкция языка R/Python и примеры применения.

2. Проект содержит оценку текущего состояния проблемы (формулировку в бизнес-терминах), формальную постановку задачи в терминах анализа данных, описание доступных источников данных (доступность, качество), решение проблемы на языке R или Python, описание результатов решения, включая оценку бизнес-возможностей организации, необходимых для проведения стратегических изменений.

Оценка «хорошо» выставляется при выполнении любого из следующих условий:

1. Ответ на теоретический вопрос является полным, содержит описание методов/конструкция языка R/Python, но отсутствуют примеры применения.

2. Проект содержит оценку текущего состояния проблемы (формулировку в бизнес-терминах), формальную постановку задачи в терминах анализа данных, описание доступных источников данных (доступность, качество), решение проблемы на языке R

или Python, но отсутствует описание результатов решения, включая оценку бизнес-возможностей организации, необходимых для проведения стратегических изменений.

Оценка «удовлетворительно» выставляется при выполнении любого из следующих условий:

1. Ответ на теоретический вопрос не является полным, содержит описание не всех методов/конструкция языка R/Python и отсутствуют примеры применения.

2. Проект содержит оценку текущего состояния проблемы (формулировку в бизнес-терминах), формальную постановку задачи в терминах анализа данных, описание доступных источников данных (доступность, качество), решение проблемы на языке R или Python имеется, но содержит ошибки и отсутствует описание результатов решения, включая оценку бизнес-возможностей организации, необходимых для проведения стратегических изменений.

Оценка «неудовлетворительно» выставляется при выполнении любого из следующих условий:

1. Ответ на теоретический вопрос не является полным или отсутствует.

2. Проект не содержит оценку текущего состояния проблемы (формулировку в бизнес-терминах), формальная постановка задачи в терминах анализа данных отсутствует, нет описания доступных источников данных, решение проблемы на языке R или Python выполнено неверно, отсутствует описание результатов решения.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle»; <https://moodle.tsu.ru/course/view.php?id=31766>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

– Мартелли А. Python. Справочник : полное описание языка / Алекс Мартелли, Анна Рейвенскрофт, Стив Холден ; перевод с английского А. Г. Гузикевича. - 3-е изд.. - Санкт-Петербург [и др.] : Диалектика, 2019. - 892 с.

– Бизли Д. Python. Книга рецептов / Бизли Д., Джонс Б. К.. - Москва : ДМК Пресс, 2019. - 646 с.

– Мастицкий С. Э. Статистический анализ и визуализация данных с помощью R / Мастицкий С. Э., Шитиков В. К.. - Москва : ДМК Пресс, 2015. - 496 с.

б) дополнительная литература:

– Персиваль Г. Python. Разработка на основе тестирования. Повинуйся Билли-тестировщику, используя Django, Selenium и JavaScript / Персиваль Г.. - Москва : ДМК Пресс, 2018. - 622 с.

– Лучано Р. Python. К вершинам мастерства / Лучано Р.. - Москва : ДМК Пресс, 2016. - 768 с.

в) ресурсы сети Интернет:

– <http://www.machinelearning.ru/wiki/>

– <http://citforum.ru>

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

- Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office On-eNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);
- публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

б) информационные справочные системы:

- Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>
- Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>
- ЭБС Лань – <http://e.lanbook.com/>
- ЭБС Консультант студента – <http://www.studentlibrary.ru/>
- Образовательная платформа Юрайт – <https://urait.ru/>
- ЭБС ZNANIUM.com – <https://znanium.com/>
- ЭБС IPRbooks – <http://www.iprbookshop.ru/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения практических занятий, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

15. Информация о разработчиках

Богданов Александр Леонидович, к.т.н., доцент, ИЭМ ТГУ, доцент кафедры Информационных технологий и бизнес-аналитики.