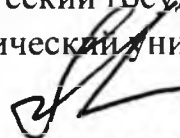


Учреждение образования «Белорусский государственный экономический университет»

УТВЕРЖДАЮ

Ректор учреждения образования
«Белорусский государственный
экономический университет»



В.Ю.Шутилин

« 29 » 06 2021 г.

Регистрационный № УД 4828-21 /уч.

АНАЛИЗ ДАННЫХ В RYTHON

Учебная программа учреждения высшего образования
по учебной дисциплине

для специальностей:

1-25 80 01 «Экономика» профилизация «Экономическая информатика»

1-25 80 10 «Статистика и анализ»

Учебная программа составлена на основе ОСВО 1-25 80 01-2019 специальность 1-25 80 01 «Экономика», ОСВО 1-25 80 10-2019 специальность 1-25 80 10 «Статистика и анализ» от 26.06.2019 № 81 и типовых учебных планов от 21.03.2019 рег. № Е 25-2-001/пр.-тип. специальность «Экономика» II ступени заочной формы получения высшего образования, № Е 25-2-007/пр.-тип. специальность «Статистика и анализ» II ступени дневной формы получения высшего образования.

СОСТАВИТЕЛИ:

Акулич В.А., доцент кафедры экономической информатики учреждения образования «Белорусский государственный экономический университет», кандидат экономических наук, доцент.

Оськин Д.А., ассистент кафедры экономической информатики учреждения образования «Белорусский государственный экономический университет».

РЕЦЕНЗЕНТЫ:

Шешолко В.К., доцент кафедры доцент кафедры управления информационными ресурсами Академии управления при Президенте Республики Беларусь, кандидат физико-математических наук, доцент.

Говядинова Н.Н., доцент кафедры информационных технологий учреждения образования «Белорусский государственный экономический университет», доцент.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой экономической информатики Учреждения образования «Белорусский государственный экономический университет» (протокол № 12 от 19.04.2021);

Научно-методическим советом Учреждения образования «Белорусский государственный экономический университет» (протокол № 5 от 22.06.2021).

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Целью преподавания учебной дисциплины «Анализ данных в Python» является развитие навыков программирования и создание крепкой практической базы для обработки, визуализации и статистического анализа данных. Магистранты познакомятся с аналитическими задачами и разными алгоритмами машинного обучения. В этом курсе будут углублены навыки программирования на языке Python, навыки работы со специализированными библиотеками для анализа и визуализации данных – numpy, pandas, matplotlib, seaborn, scikit-learn.

Задачи учебной дисциплины «Анализ данных в Python» – развить и закрепить навыки программирования на языке Python; сформировать и развить навыки работы со специализированными библиотеками для обработки, визуализации и анализа данных; развить навыки работы с данными (сбор, обработка, визуализация, разведывательный анализ); освоить терминологию области машинного обучения и познакомиться с базовыми алгоритмами.

Учебная дисциплина «Анализ данных в Python» является одной из ряда специальных дисциплин в области компьютерных технологий, изучаемых магистрантами экономических специальностей на протяжении всего курса обучения.

Магистранты научатся основным конструкциям языка и другим особенностям программирования на Python, углубятся в объектно-ориентированное программирование, научатся проводить анализ данных, строить предиктивные модели, создавать визуализации данных и работать с нейронными сетями. Программа рассчитана для магистрантов с начальными навыками программирования на Python.

В результате изучения учебной дисциплины магистрант должен обладать следующей специализированной компетенцией:

– СК-9. Уметь пользоваться языком Python для решения аналитических задач.

В результате изучения учебной дисциплины магистрант должен:

знать:

– терминологию области машинного обучения и базовые алгоритмы машинного обучения;

– основные понятия программирования включая стандартные алгоритмы, и их реализацию на Python;

– современные средства для разработки программ на Python;

– основы сбора, подготовки и очистки данных для их последующего анализа и их реализацию на Python;

– основы визуализации данных;

– основные понятия машинного обучения и их реализацию на языке программирования высокого уровня;

уметь:

– импортировать наборы данных;

– загружать данные в Pandas и работать с ними (фильтрация, агрегация, заполнение пропущенных значений);

- пользоваться языком Python для решения аналитических задач;
- подсчитывать описательные статистики, оценивать распределения, интерпретировать корреляции;
- проводить разведывательный анализ данных;
- решать простые задачи классификации, регрессии и кластеризации;
- выбирать корректные графики для визуализации данных, уметь кастомизировать их внешний вид, интерпретировать графики;
- создавать интерактивные визуализации с помощью Matplotlib, seaborn;
- определять тип задачи машинного обучения, выбирать корректные модели для ее решения, осуществлять подбор параметров и выбирать лучшую модель, создавать модели машинного обучения с помощью Scikit-learn;
- решать задачи машинного обучения от постановки исследовательского вопроса до интерпретации результатов.

владеть:

- основными конструкциями языка программирования Python;
- навыками объектно-ориентированного программирования;
- навыками анализа данных и построения предиктивных моделей;
- навыками построения визуализаций данных.

Изучение учебной дисциплины «Анализ данных в Python» предполагает наличие у магистрантов знаний курса «Объектно-ориентированное программирование и структуры данных», знаний о системном и прикладном программном обеспечении, а также навыков работы с операционной системой MS Windows на уровне пользователя персонального компьютера.

В соответствии с учебными планами для специальности 1-25 80 01 «Экономика» профилизация «Экономическая информатика» и 1-25 80 10 «Статистика и анализ» учебная программа рассчитана на 108 часов, из них аудиторных занятий 36 часов, в том числе: лекций – 18 часов, в т.ч. УСРС – 8 часов, лабораторных занятий – 18 часов, в т.ч. УСРС – 6 часов.

Форма текущей аттестации – зачет.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Тема 1. ИМПОРТ НАБОРОВ ДАННЫХ

Чтение и запись данных, форматы файлов. Загрузка данных из разных источников. Взаимодействие с базами данных. Чтение данных из Excel. Работа с CSV файлами и данными в формате JSON. Парсинг простых XML-данных. Чтение данных из таблиц HTML. Чтение данных из файла SAS. Взаимодействие с HTML и Web API. Аспекты национальной безопасности, защиты информации в сфере ИКТ.

Тема 2. ОЧИСТКА И ПОДГОТОВКА ДАННЫХ

Переформатирование данных: очистка, преобразование, слияние, изменение формы. Решение проблемы пропущенных данных. Устранение дубликатов. Замена значений. Обнаружение и фильтрация выбросов.

Тема 3. ВЫЧИСЛЕНИЯ С ПОМОЩЬЮ БИБЛИОТЕКИ NUMPY

Обзор популярных библиотек для математических расчетов, анализа и визуализации данных. Функционал библиотеки Numpy. Введение в массивы библиотеки NumPy. Выполнение вычислений над массивами. Агрегирование данных. Сортировка массивов. Структурированные данные. NumPy как средство эффективной и SciPy как средство интеллектуальной обработки данных. Дополнительные функции библиотеки SciPy.

Тема 4. МАНИПУЛЯЦИИ С ДАННЫМИ С ПОМОЩЬЮ ПАКЕТА PANDAS

Возможности и функционал библиотеки Pandas. Ключевые структуры данных (Series, DataFrames). Операции над данными в библиотеке Pandas. Обработка отсутствующих значений. Индексация и выборка данных. Иерархическая индексация. Объединение наборов данных. Агрегация и группировка. Сводные таблицы. Работа с временными рядами.

Тема 5. ВИЗУАЛИЗАЦИЯ ДАННЫХ С ПОМОЩЬЮ БИБЛИОТЕКИ MATPLOTLIB

Введение в API библиотеки Matplotlib. Построение графиков в Matplotlib. Линейные графики. Диаграммы рассеяния. Гистограммы. Настройки легенд и цветов на графиках. Построение трехмерных графиков. Отображение географических данных. Визуализация с помощью библиотеки Seaborn. Другие географические библиотеки языка Python.

Тема 6. МАШИННОЕ ОБУЧЕНИЕ С ПОМОЩЬЮ БИБЛИОТЕКИ SCIKIT-LEARN

Виды машинного обучения. Основные библиотеки машинного обучения Python (Scikit-learn, Keras, TensorFlow). Создание тренировочных наборов – предобработка данных. Точность и достоверность модели. Выбор лучшей модели.

Шаги типичного практического сценария машинного обучения. Загрузка набора данных. Исследование данных с помощью Pandas. Визуализация признаков с помощью Matplotlib. Разбиение данных для обучения и тестирования. Создание модели. Обучение модели. Тестирование модели. Настройка параметров модели и оценка ее точности. Формирование прогнозов на основании «живых» данных, которые еще неизвестны модели.

Функционал библиотеки Scikit-learn. Классификация с помощью k-соседей. Линейные модели для регрессии и классификации (модель линейной регрессии, логистическая регрессия, и др.). Наивные байесовские классификаторы. Деревья решений и случайный лес. Метод опорных векторов. Основы нейронных сетей. Метод главных компонент. Алгоритмы кластеризации (кластеризация методом k-средних, иерархическая кластеризация, и др.).

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ
«АНАЛИЗ ДАННЫХ В PYTHON»
ДЛЯ ДНЕВНОЙ ФОРМЫ ПОЛУЧЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
Специальности 1-25 80 01 «Экономика»
Специальности 1-25 80 10 «Статистика и анализ»

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов						Иное	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Количество часов УСП			
						Лекции	ЛЗ		
1	Импорт наборов данных					2		[1, 2, 20-21]	
2	Очистка и подготовка данных	2			2			[1, 2]	Контрольное задание на компьютере
3	Выполнение вычислений с помощью библиотеки NumPy	2			2			[1, 2]	Контрольное задание на компьютере
4	Манипуляции с данными с помощью пакета Pandas	2			2	2	2	[3, 4]	Контрольное задание на компьютере
5	Визуализация данных с помощью библиотеки Matplotlib	2			2	2	2	[3, 4]	Контрольное задание на компьютере
6	Машинное обучение с помощью библиотеки Scikit-learn	2			4	2	2	[1, 2]	Контрольное задание на компьютере
	Всего часов	10			12	8	6		Зачет

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ
«АНАЛИЗ ДАННЫХ В PYTHON»
ДЛЯ ЗАОЧНОЙ ФОРМЫ ПОЛУЧЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
Специальности 1-25 80 01 «Экономика»

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов						Иное	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Количество часов УСП			
						Лекции	ЛЗ		
1.	Импорт наборов данных							[1, 2, 20-21]	
2.	Очистка и подготовка данных							[1, 2]	Контрольное задание на компьютере
3.	Выполнение вычислений с помощью библиотеки NumPy	1						[1, 2]	Контрольное задание на компьютере
4.	Манипуляции с данными с помощью пакета Pandas	1			2			[3, 4]	Контрольное задание на компьютере
5.	Визуализация данных с помощью библиотеки Matplotlib	1			2			[3, 4]	Контрольное задание на компьютере
6.	Машинное обучение с помощью библиотеки Scikit-learn	1			2			[1, 2]	Контрольное задание на компьютере
	Всего часов	4			6				Зачет

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Методические рекомендации по организации самостоятельной работы магистрантов по учебной дисциплине «Анализ данных в Python»

В овладении знаниями учебной дисциплины важным этапом является самостоятельная работа магистрантов. Рекомендуется бюджет времени для самостоятельной работы в среднем 1 час на 2-х часовое аудиторное занятие.

Основными направлениями самостоятельной работы магистранта являются:

- первоначально подробное ознакомление с программой учебной дисциплины;
- ознакомление со списком рекомендуемой литературы по учебной дисциплине в целом и ее разделам, изучение необходимой литературы по темам учебной программы, подбор необходимой информации в дополнительной литературе;
- подготовка к лабораторным занятиям;
- подготовка к формам контроля знаний (контрольные задания, тесты);
- подготовка к форме текущей аттестации (зачету).

ЛИТЕРАТУРА

Основная:


1. Федоров, Д.Ю. Программирование на языке высокого уровня Python: учебное пособие для вузов / Д. Ю. Федоров. – 2-е изд., перераб. и доп. – Москва: Издательство Юрайт, 2021. – 161 с. – (Высшее образование, Гриф УМО ВО)
2. Гниденко, И.Г. Технологии и методы программирования: учебное пособие для вузов / И.Г. Гниденко, Ф.Ф. Павлов, Д.Ю. Федоров. – Москва: Издательство Юрайт, 2021. – 235 с. – (Высшее образование, Гриф УМО ВО)
3. Жуков, Р.А. Язык программирования Python. Практикум: учебное пособие / Р.А. Жуков. – Москва: ИНФРА-М, 2021. – 216 с. (Высшее образование)

Дополнительная литература:

4. Бхаргава, А. Грокаем алгоритмы. Иллюстрированное пособие для программистов и любопытствующих. – СПб.: Питер, 2017. – 288 с.
5. Брюс, П. Практическая статистика для специалистов Data Science: Пер. с англ. / П. Брюс, Э. Брюс. – СПб.: БХВ-Петербург, 2018. – 304 с.
6. Вандер, Д.П. Python для сложных задач. Наука о данных и машинное обучение / П.Д. Вандер. – СПб.: Питер, 2018. – 576 с.
7. Васильев, А.Н. Python на примерах: практический курс по программированию. – 3-е изд. – СПб.: Наука и техника, 2019. – 428 с.
8. Гуриков, С.Р. Основы алгоритмизации и программирования на Python: учебное пособие / С.Р. Гуриков. – Москва: ИНФРА-М, 2021. – 261 с. (Высшее образование)
9. Дейтел, П. Python: Искусственный интеллект, большие данные и облачные вычисления / П. Дейтел, Х. Дейтел. – СПб.: Питер, 2020. – 864 с.
10. Коэлю, Л.П. Построение систем машинного обучения на языке Python / Л.П. Коэлю, В. Ричарт. 2-е издание / Пер. с англ. Слинкин А.А. - М.: ДМК Пресс, 2018. - 302 с
11. Криволапов, С.Я. Математика на Python. (Бакалавриат, Магистратура). Учебник. / С.Я. Криволапов, М.Б. Хрипунова. – Москва: КноРус, 2021. – 456 с. – (Высшее образование, Гриф УМО ВО)
12. Кун, М. Предиктивное моделирование на практике / М. Кун, К. Джонсон. – СПб.: Питер, 2019. – 640 с.
13. Лонг, Д. Книга рецептов: Проверенные рецепты для статистики, анализа и визуализации данных / Дж. Лонг, П. Титор. [пер. с англ. Д. А. Беликова]. – М.: ДМК Пресс, 2020. – 510 с.

14. Маккинли, У. Python и анализ данных / [Пер. с англ. Слинкин А. А.]. – М.: ДМК Пресс, 2018. – 482 с.
15. Мюллер, А. Введение в машинное обучение с помощью Python: Руководство для специалистов по работе с данными / А. Мюллер, С. Гвидо. – М., 2017. – 393 с.
16. О’Нил, К. Data Science. Инсайдерская информация для новичков. Включая язык R / К. О’Нил, Р. Шатт. – СПб.: Питер, 2019. – 368 с.
17. Поляков, К.Ю. Программирование. Python. C++. В 4-х частях: учебное пособие для общеобразовательных организаций / К.Ю. Поляков. – Москва: БИНОМ. Лаборатория знаний, 2019. – 720 с.
18. Рашка, С. Python и машинное обучение: машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow 2 / С. Рашка, В. Мирджалили. 3-е изд.: Пер. с англ. – СПб. : ООО «Диалектика», 2020. – 848 с.
19. Шарден Б., Массарон Л., Боскетти А. Крупномасштабное машинное обучение вместе с Python / Б. Шарден, Л. Массарон, А. Боскетти / пер. с англ. А.В. Логунова. – М.: ДМК Пресс, 2018. – 358 с.
20. Закон Республики Беларусь «Об информации, информатизации и защите информации» от 10.11.2008 № 455-3.
21. Постановление Совета безопасности Республики Беларусь «О концепции информационной безопасности Республики Беларусь» от 18.03.2019 № 1

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Количественные методы анализа в бизнесе	Математических методов в экономике	Предложений нет  Г.О. Читая	Протокол № 12 от 29.04.2021 г.

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ УВО

на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры экономической информатики (протокол № ____ от _____ 20__ г.)

Заведующий кафедрой

к.э.н., доцент _____
(ученая степень, ученое звание) (подпись)

А.М. Зеневич
(И.О. Фамилия)

УТВЕРЖДАЮ

Директор ИМП

к.и.н., доцент _____
(ученая степень, ученое звание) (подпись)

В.А. Беляев
(И.О. Фамилия)