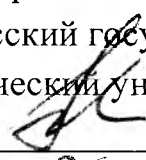


Учреждение образования
«Белорусский государственный экономический университет»

УТВЕРЖДАЮ

Ректор учреждения образования
«Белорусский государственный
экономический университет»

 В.Ю. Шутилин

«29» 06 2021 г.

Регистрационный № УД 4926-21/уч.

ИНСТРУМЕНТЫ РАБОТЫ С BIG DATA В СОЦИАЛЬНЫХ НАУКАХ

Учебная программа учреждения высшего образования
по учебной дисциплине для специальности 1-25 80 01 «Экономика»
(профилизация «Экономическая социология»)

Учебная программа составлена на основе ОСВО 1-25 01 12-2013 и учебного плана учреждения высшего образования по специальности 1-25 80 01 «Экономика» (профилизация «Экономическая социология»), дата утверждения 10.03.2020, регистрационный номер № 01Р-20.

СОСТАВИТЕЛЬ:

Акулич В.А., доцент кафедры экономической социологии и психологии предпринимательской деятельности учреждения образования «Белорусский государственный экономический университет», кандидат экономических наук, доцент.

РЕЦЕНЗЕНТЫ:

Смыкова Е.А., заведующий отделом социологии культуры Института социологии НАН Беларуси, кандидат социологических наук

Буховец Т.В., доцент кафедры национальной экономики и государственного управления, кандидат экономических наук, доцент

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой экономической социологии и психологии предпринимательской деятельности учреждения образования «Белорусский государственный экономический университет»
(протокол № 9 от «24» 04 2021);

Научно-методическим советом учреждения образования «Белорусский государственный экономический университет»
(протокол № 5 от «22» 06 2021).

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цель преподавания учебной дисциплины «Инструменты работы с Big data в социальных науках» является развитие навыков программирования и создание крепкой практической базы для обработки, визуализации и статистического анализа больших 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 больших данных – numpy, pandas, matplotlib, seaborn, scikit-learn.

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

Учебная дисциплина «Инструменты работы с Big data в социальных науках» 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 больших данных, строить предиктивные модели, создавать визуализации больших данных и работать с нейронными сетями. Программа рассчитана для магистрантов с начальными навыками программирования на Python. Тем магистрантам, которые имеют продвинутые навыки анализа Big data с помощью Python, будет предложено использование других инструментов обработки данных (R или SAS). В отдельных темах планируется использовать инструментальный стек из IBM SPSS Statistics и Python.

В результате изучения учебной дисциплины выпускник должен обладать следующими компетенциями:

- АК-1. Уметь применять базовые научно-теоретические знания для решения теоретических и практических задач;
- АК-2. Владеть системным и сравнительным анализом;
- АК-6. Владеть междисциплинарным подходом при решении проблем;
- АК-7. Иметь навыки, связанные с использованием технических устройств, управлением информацией и работой с компьютером;

В результате изучения учебной дисциплины студент 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 больших данных для их последующего анализа и их реализацию на Python;
- основы визуализации больших данных;
- основные понятия машинного обучения и их реализацию на языке программирования высокого уровня;

уметь:

- импортировать наборы больших данных;
- загружать большие 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 больших данных;
- решать простые задачи классификации, регрессии и кластеризации;
- выбирать корректные графики для визуализации больших 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 больших данных и построения предиктивных моделей;
- навыками построения визуализаций больших данных.

Изучение учебной дисциплины «Инструменты работы с Big data в социальных науках» предполагает наличие у студентов знаний курса «Компьютерные технологии анализа данных в социологии», знаний о системном и прикладном программном обеспечении, а также навыков работы с операционной системой MS Windows на уровне пользователя персонального компьютера.

В соответствии с учебным планом специальности 1-25 80 01 «Экономика» (профилизация «Экономическая социология») учебная программа рассчитана на 198 часов, из них аудиторных занятий 60 часов, в том числе: лекций – 30 часов, практических занятий – 10 часов, лабораторных занятий – 20 часов.

Форма текущей аттестации – экзамен.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Тема 1. ИМПОРТ НАБОРОВ БОЛЬШИХ ДАННЫХ.

Обзор основных аналитических инструментов работы с Big data в социальных науках (Python, R, SAS, и др.). 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.

Тема 2. ОЧИСТКА И ПОДГОТОВКА БОЛЬШИХ 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 БОЛЬШИМИ 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 БОЛЬШИХ 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
«ИНСТРУМЕНТЫ РАБОТЫ С BIG DATA В СОЦИАЛЬНЫХ НАУКАХ»
ДЛЯ ДНЕВНОЙ ФОРМЫ ПОЛУЧЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ II СТУПЕНИ
СПЕЦИАЛЬНОСТИ 1-25 80 01 «ЭКОНОМИКА» (ПРОФИЛИЗАЦИИ «ЭКОНОМИЧЕСКАЯ СОЦИОЛОГИЯ»)

Но мер разд ела, тем ы	Название раздела, темы	Количество аудиторных часов							Иное	Форма контроля знаний
		Лекции	Практические Занятия (ПЗ)	Семинарские занятия	Лабораторные Занятия (ЛЗ)	Количес т во часов УСР				
						Лек ции	ПЗ	ЛЗ		
1	Импорт наборов больших данных	2				2			[1, 2]	
2	Очистка и подготовка больших данных	2			2	2	2		[1, 2]	Контрольное задание на компьютере
3	Выполнение вычислений с помощью библиотеки NumPy	2			2	2	2		[1, 2]	Контрольное задание на компьютере
4	Манипуляции с большими данными с помощью пакета Pandas	4	2		2	2		2	[3, 4]	Контрольное задание на компьютере
5	Визуализация больших данных с помощью библиотеки Matplotlib	4	2		2	2		2	[3, 4]	Контрольное задание на компьютере
6	Машинное обучение с помощью библиотеки Scikit-learn	4	2		4	2		2	[1, 2]	Контрольное задание на компьютере
	Всего часов	18	6		12	12	4	8		Экзамен

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Методические рекомендации по организации самостоятельной работы студентов по учебной дисциплине «Инструменты работы с Big data в социальных науках»

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

Основными направлениями самостоятельной работы студента 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 (экзамену).

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)
4. Салин, В. Н. Статистический анализ данных цифровой экономики в системе «Statistica»: учебно-практическое пособие для направления бакалавриата «Экономика» / В. Н. Салин, Э. Ю. Чурилова. – М.: КНОРУС, 2019. – 237 с. – (Бакалавриат и магистратура). (Гриф РФ)

Дополнительная литература:

5. Бонцанини, М. Анализ социальных медиа на Python. – Birmingham: Packt Publishing Ltd., 2016. – 322 с.
6. Вандер, Д.П. Python для сложных задач. Наука о данных и машинное обучение / П.Д. Вандер. – СПб.: Питер, 2018. – 576 с.
7. Груздьев, А. Прогнозное моделирование в IBM SPSS Statistics, R и Python. Метод деревьев решений и случайный лес / А. Груздьев. – М.: ДМК Пресс, 2018. – 634 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. Дятлов, А. В. Анализ данных в социологии: учебник / А. В. Дятлов, Д. А. Гугуева. – Ростов-на-Дону ; Таганрог : Издательство Южного федерального университета, 2018. – 226 с. – ISBN 978-5-9275-2690-1. – Текст: электронный. – URL: <https://znanium.com/catalog/product/1039664> (дата обращения: 11.03.2021). – Режим доступа: по подписке.
11. Коэлью, Л.П. Построение систем машинного обучения на языке Python / Л.П. Коэлью, В. Ричарт. 2-е издание / Пер. с англ. Слинкин А.А. - М.: ДМК Пресс, 2018. - 302 с

12. Лесковец, Ю. Анализ больших наборов данных / Ю. Лесковец, А. Раджараман, Д. Ульман. [Пер. с англ. Слинкин А. А.]. – М.: ДМК Пресс, 2016. – 498 с.
13. Лонг, Д. Книга рецептов: Проверенные рецепты для статистики, анализа и визуализации данных / Д. 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, 2015. – 482 с.
15. Моосмюллер, Г. Маркетинговые исследования с SPSS: учебное пособие / Г. Моосмюллер, Н.Н. Ребик. – 2-е изд. – Москва: ИНФРА-М, 2019. – 200 с. (Высшее образование)
16. Орлова, И.В. Многомерный статистический анализ в экономических задачах: компьютерное моделирование в SPSS / В.И. Орлова. М.: ИНФРА-М. – 2021. – 310 с. (Вузовский учебник)
17. Рассел, М. Data Mining. Извлечение информации из Facebook, Twitter, LinkedIn, Instagram, GitHub / М. Рассел, М. Классен. – СПб.: Питер, 2020. – 464 с.
18. Силен, Д. Основы Data Science и Big Data. Python и наука о данных / Д. Силен, А. Мейсман, М. Али ; [перевел с англ. Е. Матвеев]. – СПб. [и др.] : Питер, 2020. – 334 с. : ил. – (Библиотека программиста).
19. Черткова, Е.А. Статистика. Автоматизация обработки информации: учебное пособие для вузов / Е.А. Черткова. – 2-е изд., испр. и доп. – Москва: Издательство Юрайт, 2021. – 195 с. – (Высшее образование).

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО

<p style="text-align: center;">Название учебной дисциплины, с которой требуется согласование</p>	<p style="text-align: center;">Название кафедры</p>	<p style="text-align: center;">Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине</p>	<p style="text-align: center;">Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)¹</p>
<p>Моделирование потребности в персонале на рынке труда и услуг</p>	<p>Экономическая и социальная информатика</p>	<p>Предложений нет</p> <p style="text-align: center;"><i>[Подпись]</i></p>	<p>Протокол № <u>9</u> от «<u>27</u>» <u>04</u> 20<u>21</u></p>

¹ При наличии предложений об изменениях в содержании учебной программы УВО

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

И.В. Лашук
(И.О. Фамилия)

УТВЕРЖДАЮ

Директор ИСГО

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

Д.Г. Доброродный
(И.О. Фамилия)

Зав. кафедрой

A handwritten signature in black ink, consisting of stylized Cyrillic letters, likely 'И.В. Лашук'.

И.В. Лашук