

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Сибирский государственный индустриальный университет»
Кафедра прикладных информационных технологий и программирования

УТВЕРЖДАЮ
Проректор по учебной работе
_____ М.В. Темлянец
подпись
« ____ » _____ 20__ г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Python для анализа данных

09.04.01 «Информатика и вычислительная техника»
(направленность (профиль): «Информатика и вычислительная техника»)

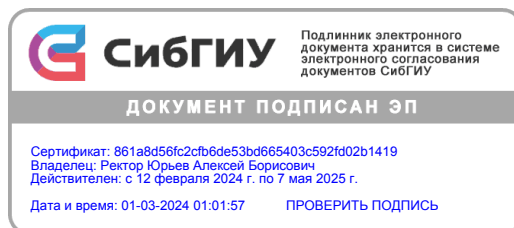
Квалификация выпускника
Магистр

Форма обучения
Очно-заочная форма

Срок обучения: 2 года 3 месяца

Год начала подготовки 2022

Новокузнецк
2022



1 Цели и задачи освоения учебной дисциплины

Целями учебной дисциплины являются:

- формирование у обучающихся знаний и навыков по подготовке данных к анализу, использованию высокоуровневых языков программирования для автоматизации операций анализа данных и их визуализации.

Задачами учебной дисциплины являются:

- изучение среды разработки Jupiter Notebook и пакетов NumPy и pandas;
- изучение методов подготовки данных к анализу средствами Jupiter Notebook;
- изучение методов визуализации данных средствами Jupiter Notebook;
- изучение методов агрегирования данных средствами Jupiter Notebook.

2 Место учебной дисциплины в структуре ООП по направлению подготовки (специальности)

Учебная дисциплина относится к учебным дисциплинам части, формируемой участниками образовательных отношений **Блока 1 «Дисциплины (модули)»** ООП по направлению подготовки (специальности) 09.04.01 «Информатика и вычислительная техника».

Учебная дисциплина опирается на базовые знания и компетенции, полученные в процессе получения предыдущего образования.

Учебная дисциплина дополняет знания, умения и навыки, получаемые по одновременно изучаемым и последующим дисциплинам:

- Машинное обучение для задач управления;
- Разработка и реализация проектов 2.

3 Планируемые результаты обучения по учебной дисциплине

Процесс изучения учебной дисциплины направлен на формирование следующих компетенций:

– Профессиональные компетенции

Наименование категории (группы) ПК	Код и наименование ПК	Код и наименование индикатора достижения ПК	Планируемые результаты обучения
	ПК-2: Способен разрабатывать и применять алгоритмы интеллектуального анализа больших объемов данных для управления	ПК-2.2 Собирает данные из различных источников и осуществляет их подготовку для анализа	– знать: типовые операции и их вариации для решения задач анализа данных; методы сбора и подготовки данных

	технологическими системами		<p>для анализа. – уметь: комбинировать и реализовывать типовые операции для сбора и анализа данных и их вариации в высокоуровневых языках программирования. – владеть: навыками реализации типовых операций для сбора и анализа данных в высокоуровневых языках программирования.</p>
		<p>ПК-2.3 Выбирает и использует методы и инструментальные средства для проведения анализа больших объемов данных, строит модели на основе данных</p>	<p>– знать: библиотеки и пакеты программ для анализа данных; методы реализации математических и статистических операции в высокоуровневых языках программирования. – уметь: устанавливать и использовать библиотеки и прикладные пакеты программ для анализа данных; применять логические условия в виде операций с массивами, математические и статистические операции, методы булевых массивов. – владеть: библиотеками и прикладными</p>

		пакетами программ для анализа данных; математическими и статистическими методами обработки данных в высокоуровневых языках программирования.
--	--	--

4 Объем и содержание учебной дисциплины

Учебные занятия по учебной дисциплине проводятся в форме контактной работы и в форме самостоятельной работы обучающихся.

Контактная работа обучающихся с педагогическим работником включает в себя занятия семинарского типа (семинары, практические занятия, практикумы), промежуточную аттестацию обучающихся и иную контактную работу, предусматривающую групповую или индивидуальную работу обучающихся с педагогическим работником. Контактная работа обучающихся с педагогическим работником может быть аудиторной, внеаудиторной, а также проводиться в электронной информационно-образовательной среде.

Объем учебной дисциплины

Семестр / курс		ИТОГО	1 семестр
Форма промежуточной аттестации			экзамен
Трудоёмкость	<i>академ. час.</i>	180	180
	<i>зачетных единиц</i>	5	5
Лекции, <i>академ. час.</i>		0	0
в форме практической подготовки		0	0
Лабораторные работы, <i>академ. час.</i>		0	0
в форме практической подготовки		0	0
Практические занятия, <i>академ. час.</i>		12	12
в форме практической подготовки		0	0
Курсовая работа / проект, <i>академ. час.</i>		0	0
в форме практической подготовки		0	0
Консультации, <i>академ. час.</i>		0	0
в форме практической подготовки		0	0
Самостоятельная работа, <i>академ. час.</i>		123	123
в форме практической подготовки		0	0
Контроль, <i>академ. час.</i>		45	45
в форме практической подготовки		0	0

Содержание учебной дисциплины

Раздел 1 Интерактивные вычисления и среда разработки Jupiter Notebook и основы NumPy;

Тема 1.1 Основы Jupiter Notebook (Основы синтаксиса и особенности среды разработки. Интроспекция. Исключения и обратная трассировка. Интеграция с matplotlib и режим pylab);

Тема 1.2 Средства разработки программ (Команды оболочки и псевдонимы. Интерактивный отладчик. Хронометраж программы: %time и %timeit.);

Тема 1.3 NumPy ndarray (Тип данных для ndarray. Операции между массивами и скалярами. Индексирование и вырезание. Булево индексирование. Транспонирование массивов и перестановка осей);

Тема 1.4 Обработка данных с применением массивов (Запись логических условий в виде операций с массивами. Математические и статистические операции. Методы булевых массивов. Сортировка. Устранение дубликатов и другие теоретико-множественные операции);

Тема 1.5 Линейная алгебра и генерация случайных чисел (Операции линейной алгебры. Работа с матрицами. Генерация случайных чисел: случайное блуждание. Моделирование сразу нескольких случайных блужданий);

Раздел 2 Основы pandas;

Тема 2.1 Введение в структуры данных pandas и базовая функциональность (Объект Series. Объект DataFrame. Индексные объекты. Переиндексация. Доступ по индексу, выборка и фильтрация. Арифметические операции и выравнивание данных. Применение функций и отображение. Сортировка и ранжирование);

Тема 2.2 Редукция и вычисление описательных статистик (Корреляция и ковариация. Уникальные значения, счетчики значений и членство. Фильтрация отсутствующих данных);

Раздел 3 Работа с данными: переформатирование, визуализация, агрегирование данных и групповые операции;

Тема 3.1 Переформатирование данных: очистка, преобразование, слияние, изменение формы (Комбинирование и слияние наборов данных. Изменение формы и поворот. Преобразование данных. Манипуляции со строками);

Тема 3.2 Построение графиков и визуализация (API библиотеки matplotlib. Функции построения графиков в pandas. Нанесение данных на карту. Инструментальная экосистема визуализации для Python);

Тема 3.3 Агрегирование данных и групповые операции (Механизм GroupBy. Агрегирование данных. Групповые операции и преобразования. Сводные таблицы и кросс-табуляция).

5 Перечень тем лекций

№ раздела / темы дисциплины	Темы лекций	Трудоемкость, <i>академ. час</i>	
		всего	в форме практической подготовки

	<i>Отсутствуют</i>		
Итого:		0	0

6 Перечень тем практических занятий (семинаров)

№ раздела / темы дисциплины	Темы практических занятий (семинаров)	Трудоемкость, <i>академ. час</i>	
		всего	в форме практической подготовки
Раздел 1.	Основы разработки в среде Jupiter Notebook	2	
Тема 2.1.	Работа с объектами Series и DataFrame	2	
Тема 2.2.	Редукция и вычисление описательных статистик в Jupiter Notebook	2	
Тема 3.1.	Переформатирование данных в Jupiter Notebook	2	
Тема 3.2.	Визуализация данных	2	
Тема 3.3.	Агрегирование данных в Jupiter Notebook	2	
Итого:		12	0

7 Перечень тем лабораторных работ

№ раздела / темы дисциплины	Темы лабораторных работ	Трудоемкость, <i>академ. час</i>	
		всего	в форме практической подготовки
	<i>Отсутствуют</i>		
Итого:		0	0

8 Перечень тем курсовых работ (проектов)

№ раздела / темы дисциплины	Темы курсовых работ (проектов)	Трудоемкость, <i>академ. час</i>	
		всего	в форме практической подготовки
	<i>Отсутствуют</i>		
Итого:		0	0

9 Виды самостоятельной работы

№ раздела / темы дисциплины	Виды самостоятельной работы	Трудоемкость, <i>академ. час</i>	
		всего	в форме практической подготовки
Раздел 1.	1. Изучение теоретического материала; 2. Подготовка к практическому занятию;	40	

	3. Прохождение тестирования.		
Раздел 2.	1. Изучение теоретического материала; 2. Подготовка к практическому занятию; 3. Прохождение тестирования.	41	
Раздел 3.	1. Изучение теоретического материала; 2. Подготовка к практическому занятию; 3. Прохождение тестирования.	42	
<i>Контроль</i>	<i>Подготовка к экзамену</i>	45	
Итого:		168	0

10 Учебно-методическое и информационное обеспечение учебной дисциплины

а) литература:

1 Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : учебное пособие / П. Флах. – Москва : ДМК-пресс, 2015. – 400 с. – ISBN 978-5-97060-273-7. – URL: <https://www.studentlibrary.ru/book/ISBN9785970602737.html> (дата обращения: 25.03.2022);

2 Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения : практическое пособие / С. Рашка. – Москва : ДМК-пресс, 2017. – 418 с. – ISBN 978-5-97060-409-0. – URL: <https://www.studentlibrary.ru/book/ISBN9785970604090.html> (дата обращения: 25.03.2022);

3 Хейдт, М. Изучаем pandas / М. Хейдт, А. В. Груздев. – Москва : ДМК-пресс, 2019. – ISBN 978-5-97060-670-4. – URL: <https://www.studentlibrary.ru/book/ISBN9785970606704.html> (дата обращения: 25.03.2022);

4 Маккинли, У. Python и анализ данных / У. Маккинли. – Москва : ДМК Пресс, 2015. – 482 с. – ISBN 978-5-97060-315-4. – URL: <https://www.studentlibrary.ru/ru/book/ISBN9785970603154.html> (дата обращения: 25.03.2022).

б) ресурсы информационно-телекоммуникационной сети «Интернет»:

1 Консультант студента. Электронная библиотека технического ВУЗа : электронно-библиотечная система / ООО «Политехресурс». –

Москва, [200 –]. – URL: <http://www.studentlibrary.ru>. – Режим доступа: для авторизир. пользователей;

2 ЛАНЬ : электронно-библиотечная система : [коллекция «Инженерно-технические науки»] / ООО «Издательство Лань». – Санкт-Петербург, [200 –]. – URL: <http://e.lanbook.com>. – Режим доступа: для авторизир. пользователей;

3 НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА eLIBRARY.RU : база данных / ООО «НЭБ». – Москва, [200 –]. – URL: <http://elibrary.ru>. – Режим доступа: по подписке;

4 Образовательная платформа ЮРАЙТ / ООО «Электронное издательство Юрайт». – Москва, [200 –]. – URL: <https://urait.ru>. – Режим доступа: для авторизир. пользователей;

5 Университетская библиотека онлайн : электронно-библиотечная система / ООО «Директ-Медиа». – Москва, [200 –]. – URL: <http://www.biblioclub.ru>. – Режим доступа: для авторизир. пользователей;

6 Электронная библиотека // Научно-техническая библиотека СибГИУ : сайт. – Новокузнецк, [200 –]. – URL: <https://library.sibsiu.ru/LibrELibraryFullText.asp>. – Режим доступа: для авторизир. пользователей;

7 Электронный каталог : сайт / Научно-техническая библиотека СибГИУ. – Новокузнецк, [199 –]. – URL: <http://libr.sibsiu.ru>.

в) лицензионное и свободно распространяемое программное обеспечение:

- 7-Zip;
- Adobe Acrobat Reader;
- Kaspersky Endpoint Security;
- Microsoft Office 2010;
- Microsoft Windows 7;
- Notepad++;
- PyCharm;
- PyPy.

г) базы данных и информационно-справочные системы:

1 КонсультантПлюс : справочно-правовая система / ООО «Информационный центр АНВИК». – Новокузнецк, [199 –]. – Режим доступа: компьютерная сеть библиотеки Сиб. гос. индустр. ун-та.;

2 Система ГАРАНТ : электронный периодический справочник / ООО «Правовой центр «Гарант». – Кемерово, [200 –]. – Режим доступа: компьютерная сеть Сиб. гос. индустр. ун-та.;

3 Техэксперт : информационно-справочная система / ООО «Группа компаний «Кодекс». – Кемерово, [200 –]. – Режим доступа: компьютерная сеть Сиб. гос. индустр. ун-та.;

4 Электронный реферативный журнал (ЭлРЖ) : база данных / ВИНТИ РАН. – Москва, [200 –]. – Режим доступа: компьютерная сеть библиотеки Сиб. гос. индустр. ун-та.

11 Материально-техническое обеспечение учебной дисциплины

Материально-техническое обеспечение учебной дисциплины включает учебные аудитории, оснащенные оборудованием, компьютерной техникой, и техническими средствами обучения, в том числе:

- учебную аудиторию для проведения занятий семинарского типа (практических занятий),оснащенную персональными компьютерами и выходом в сеть «Интернет»;
- учебную аудиторию (помещения) для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации;
- помещения для самостоятельной работы, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду, научно-техническую библиотеку СибГИУ.

Рабочая программа дисциплины составлена в соответствии с требованиями ФГОС ВО по направлению подготовки (специальности) 09.04.01 «Информатика и вычислительная техника».

Составитель(и):

доцент Бабичева Надежда Борисовна (кафедра прикладных информационных технологий и программирования).

Рабочая программа дисциплины рассмотрена и утверждена на заседании кафедры.

Приложение А

Аннотация

рабочей программы дисциплины «Python для анализа данных»

по направлению подготовки (специальности)

09.04.01 «Информатика и вычислительная техника»

(направленность (профиль): «Информатика и вычислительная техника»)

форма обучения – Очно-заочная форма

1 Цели и задачи освоения учебной дисциплины

Целями учебной дисциплины являются:

- формирование у обучающихся знаний и навыков по подготовке данных к анализу, использованию высокоуровневых языков программирования для автоматизации операций анализа данных и их визуализации.

Задачами учебной дисциплины являются:

- изучение среды разработки Jupiter Notebook и пакетов NumPy и pandas;
- изучение методов подготовки данных к анализу средствами Jupiter Notebook;
- изучение методов визуализации данных средствами Jupiter Notebook;
- изучение методов агрегирования данных средствами Jupiter Notebook.

2 Место учебной дисциплины в структуре ООП по направлению подготовки (специальности)

Учебная дисциплина относится к учебным дисциплинам части, формируемой участниками образовательных отношений **Блока 1 «Дисциплины (модули)»** ООП по направлению подготовки (специальности) 09.04.01 «Информатика и вычислительная техника».

Учебная дисциплина опирается на базовые знания и компетенции, полученные в процессе получения предыдущего образования.

Учебная дисциплина дополняет знания, умения и навыки, получаемые по одновременно изучаемым и последующим дисциплинам:

- Машинное обучение для задач управления;
- Разработка и реализация проектов 2.

3 Планируемые результаты обучения по учебной дисциплине

Процесс изучения учебной дисциплины направлен на формирование следующих компетенций:

- **Профессиональные компетенции**

Наименование	Код и наименование	Код и наименование	Планируемые
--------------	--------------------	--------------------	-------------

категории (группы) ПК	ПК	индикатора достижения ПК	результаты обучения
	ПК-2: Способен разрабатывать и применять алгоритмы интеллектуального анализа больших объемов данных для управления технологическими системами	ПК-2.2 Собирает данные из различных источников и осуществляет их подготовку для анализа	<p>– знать: типовые операции и их вариации для решения задач анализа данных; методы сбора и подготовки данных для анализа.</p> <p>– уметь: комбинировать и реализовывать типовые операции для сбора и анализа данных и их вариации в высокоуровневых языках программирования.</p> <p>– владеть: навыками реализации типовых операций для сбора и анализа данных в высокоуровневых языках программирования.</p>
		ПК-2.3 Выбирает и использует методы и инструментальные средства для проведения анализа больших объемов данных, строит модели на основе данных	<p>– знать: библиотеки и пакеты программ для анализа данных; методы реализации математических и статистических операции в высокоуровневых языках программирования.</p> <p>– уметь: устанавливать и использовать библиотеки и прикладные пакеты программ для анализа данных; применять логические условия в виде</p>

			операций с массивами, математические и статистические операции, методы булевых массивов. – владеть: библиотеками и прикладными пакетами программ для анализа данных; математическими и статистическими методами обработки данных в высокоуровневых языках программирования.
--	--	--	---

4 Объем учебной дисциплины

Семестр / курс		ИТОГО	1 семестр
Форма промежуточной аттестации			экзамен
Трудоёмкость	<i>академ. час.</i>	180	180
	<i>зачетных единиц</i>	5	5
Лекции, <i>академ. час.</i>		0	0
в форме практической подготовки		0	0
Лабораторные работы, <i>академ. час.</i>		0	0
в форме практической подготовки		0	0
Практические занятия, <i>академ. час.</i>		12	12
в форме практической подготовки		0	0
Курсовая работа / проект, <i>академ. час.</i>		0	0
в форме практической подготовки		0	0
Консультации, <i>академ. час.</i>		0	0
в форме практической подготовки		0	0
Самостоятельная работа, <i>академ. час.</i>		123	123
в форме практической подготовки		0	0
Контроль, <i>академ. час.</i>		45	45
в форме практической подготовки		0	0

5 Краткое содержание учебной дисциплины

В структуре учебной дисциплины выделяются следующие основные разделы (темы):

Раздел 1 Интерактивные вычисления и среда разработки Jupiter Notebook и основы NumPy;

Тема 1.1 Основы Jupiter Notebook (Основы синтаксиса и особенности среды разработки. Интроспекция. Исключения и обратная трассировка. Интеграция с matplotlib и режим pylab);

Тема 1.2 Средства разработки программ (Команды оболочки и псевдонимы. Интерактивный отладчик. Хронометраж программы: %time и %timeit.);

Тема 1.3 NumPy ndarray (Тип данных для ndarray. Операции между массивами и скалярами. Индексирование и вырезание. Булево индексирование. Транспонирование массивов и перестановка осей);

Тема 1.4 Обработка данных с применением массивов (Запись логических условий в виде операций с массивами. Математические и статистические операции. Методы булевых массивов. Сортировка. Устранение дубликатов и другие теоретико-множественные операции);

Тема 1.5 Линейная алгебра и генерация случайных чисел (Операции линейной алгебры. Работа с матрицами. Генерация случайных чисел: случайное блуждание. Моделирование сразу нескольких случайных блужданий);

Раздел 2 Основы pandas;

Тема 2.1 Введение в структуры данных pandas и базовая функциональность (Объект Series. Объект DataFrame. Индексные объекты. Переиндексация. Доступ по индексу, выборка и фильтрация. Арифметические операции и выравнивание данных. Применение функций и отображение. Сортировка и ранжирование);

Тема 2.2 Редукция и вычисление описательных статистик (Корреляция и ковариация. Уникальные значения, счетчики значений и членство. Фильтрация отсутствующих данных);

Раздел 3 Работа с данными: переформатирование, визуализация, агрегирование данных и групповые операции;

Тема 3.1 Переформатирование данных: очистка, преобразование, слияние, изменение формы (Комбинирование и слияние наборов данных. Изменение формы и поворот. Преобразование данных. Манипуляции со строками);

Тема 3.2 Построение графиков и визуализация (API библиотеки matplotlib. Функции построения графиков в pandas. Нанесение данных на карту. Инструментальная экосистема визуализации для Python);

Тема 3.3 Агрегирование данных и групповые операции (Механизм GroupBy. Агрегирование данных. Групповые операции и преобразования. Сводные таблицы и кросс-табуляция).

6 Составитель(и):

доцент Бабичева Надежда Борисовна (кафедра прикладных информационных технологий и программирования).