

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ

Проректор по учебной работе
и образовательным инновациям

_____ О.И.Чуприс

«___» _____ 2019 г.

Регистрационный № УД-_____ /уч.

Анализ данных с использованием языка программирования R

**Учебная программа учреждения высшего образования
по учебной дисциплине для специальности:**

1-25 80 01 Экономика

Профилизация: Интеллектуальный анализ данных в цифровой экономике

2019 г.

Учебная программа составлена на основе ОСВО 1-25 80 01-2019, учебного плана № Е25-104/уч от 11.04.2019 г.

СОСТАВИТЕЛЬ:

Е.А. Минюкович, доцент кафедры цифровой экономики экономического факультета Белорусского государственного университета, к.э.н., доцент

Рецензент:

А.Б. Гедранович, технический директор ООО «ЭПИКА Девелопмент», к.э.н., доцент

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой цифровой экономики экономического факультета (протокол № 1 от 29.08.2019);

Научно-методическим Советом БГУ (протокол № 1 от 25.09.2019)

Зав.кафедрой _____ И.А.Карачун

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Цель учебной дисциплины – подготовка студентов к использованию методов и алгоритмов интеллектуального анализа и моделирования данных для решения практических задач и проведения теоретических исследований в предметных областях экономики.

Задачи учебной дисциплины:

1. Изучение методов и алгоритмов интеллектуального анализа и моделирования данных.
2. Рассмотрение реализации вышеназванных методов и алгоритмов в библиотеках языка программирования R.
3. Тренировка навыков интеллектуального анализа и моделирования реальных данных путем применения инструментария языка программирования R.

Место учебной дисциплины в системе подготовки специалиста с высшим образованием (магистра).

Учебная дисциплина относится к модулю «Анализ данных» компонента учреждения высшего образования.

Связи с другими учебными дисциплинами, включая учебные дисциплины компонента учреждения высшего образования, дисциплины специализации и др.

В рамках учебной дисциплины «Анализ данных с использованием языка программирования R» расширяются и углубляются знания и практические навыки, полученные при изучении учебной дисциплины «Анализ и прогнозирование временных рядов».

Требования к компетенциям

Освоение учебной дисциплины «Анализ данных с использованием языка программирования R» должно обеспечить формирование следующих специализированных компетенций:

СК-4. Уметь разрабатывать и анализировать математические модели, знать методы компьютерного анализа данных и современные информационные технологии на основе языка R, разрабатывать эффективные численные методы и алгоритмы и интегрировать их в компьютерные системы анализа данных.

В результате освоения учебной дисциплины студент должен:

знать: основные методы и алгоритмы интеллектуального анализа и моделирования данных;

уметь: проводить исследовательский анализ и моделирование данных с использованием инструментария языка программирования R;

владеть: навыками применения знаний и умений по анализу данных с использованием языка программирования R для решения актуальных экономических задач.

Структура учебной дисциплины

Дисциплина изучается во 2 семестре. Всего на изучение учебной дисциплины «Анализ данных с использованием языка программирования R» для очной формы получения высшего образования отведено 120 часов, в том числе 48 аудиторных часов, из них: лекции – 14 часов, практические занятия – 34 часа.

Трудоемкость учебной дисциплины составляет 3 зачетные единицы.

Форма текущей аттестации – экзамен.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Тема 1. Графический анализ данных

Импорт/экспорт данных. Библиотека ggplot2. Диаграммы рассеяния. Способы отображения множественных признаков. Гистограммы, диаграммы для отображения статистических характеристик и иные типы диаграмм.

Тема 2. Трансформация данных

Табличное представление данных и концепция «Tidy Data». Функции для фильтрации строк, выбора данных из столбцов, сортировки, добавления новых переменных, групповых операций.

Тема 3. Исследовательский анализ данных

Вопросы анализа. Дисперсионный и ковариационный анализ переменных. Пропуски в данных. Паттерны и модели. Визуализация.

Тема 4. Основы машинного обучения. Регрессионный анализ данных

4.1. Задачи «обучения с учителем»: регрессионный анализ и классификация данных. Задачи «обучения без учителя» - кластеризация данных. Проблемы «недообучения» и «переобучения». Подходы к оценке производительности моделей.

4.2. Задачи простой и множественной линейной регрессии. Модели регрессии: линейные и на основе деревьев решений, а также алгоритмы их построения. Метрики производительности моделей регрессии.

Тема 5. Подготовка данных для моделирования. Модели классификации данных

5.1: Типы переменных и возможные действия. Генерация новых переменных.

5.2 Задачи бинарной и мультикласс классификации. Модели классификации: линейные и на основе деревьев решений, а также алгоритмы их построения. Метрики производительности моделей классификации.

Тема 6. Модели кластеризации данных

Задача кластеризации. Модели кластеризации: «к-средних» и иерархической кластеризации. Метрики производительности моделей кластеризации.

Тема 7. Индивидуальные проекты по анализу данных на R

Выбор данных и постановка задачи машинного обучения. Исследовательский анализ данных и моделирование. Оценка результатов моделирования.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Дневная форма получения образования

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УСР	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное		
1	2	3	4	5	6	7	8	9
1	Графический анализ данных	2	4					практическая работа 1
2	Трансформация данных	2	4					практическая работа 2
3	Исследовательский анализ данных		4					практическая работа 3
4	Основы машинного обучения. Регрессионный анализ данных	4	6					дискуссия, практическая работа 4
5	Подготовка данных для моделирования. Модели классификации данных	4	6					дискуссия, практическая работа 5
6	Модели кластеризации данных	2	6					практическая работа 6
7	Индивидуальные и групповые проекты по анализу данных на R		4					индивидуальный проект
	Итого	14	34					

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Брюс, П. Практическая статистика для специалистов Data Science. 50 важнейших понятий / П. Брюс, Э. Брюс. — БХВ-Петербург, 2018. — 304 с.
2. Grolemund, G. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data / Garrett Grolemund, Hadley Wickham. — O'Reilly, 2017. — 520 p.
3. Gareth, J. An Introduction to Statistical Learning: with Applications in R / Gareth James, Daniela Witten, Trevor Hastie, Tibshirani. — Springer, 2017. — 426 p.
4. Мастицкий, С. Э. Статистический анализ и визуализация данных с помощью R / С. Э. Мастицкий, В. К. Шитиков. — ДМК-Пресс, 2015. — 496 с.
5. Hyndman, R Forecasting: Principles and Practice [Electronic resource] / Rob J Hyndman, George Athanasopoulos. — 2018. — Mode of access: <https://otexts.com/book/sfml>. – Date of access: 11.07.2019.

Перечень дополнительной литературы

1. Сигель, Э. Просчитать будущее: Кто кликнет, купит, соврет или умрет / Э. Сигель. — М.: Альпина Паблишер, 2018. — 374 с.

Перечень рекомендуемых средств диагностики и методика формирования итоговой оценки

Оценка практической работы формируется на основе следующих критериев: корректность полученных результатов и их интерпретацию, умение воспроизвести выполнение заданий, понимание практической применимости результатов работы, полнота ответов на вопросы. Итоговая оценка за практические работы рассчитывается путем усреднения оценок за запланированные к выполнению практические работы.

Оценка за проект формируется на основе следующих критериев: обоснованность используемых методов анализа данных, корректность трактовки полученных результатов, организация работы группы.

Формой текущей аттестации по дисциплине «Экономическая информатика» учебным планом предусмотрен экзамен.

При формировании итоговой оценки используется рейтинговая оценка знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая оценка предусматривает использование весовых коэффициентов для текущего контроля знаний и текущей аттестации студентов по дисциплине.

Примерные весовые коэффициенты, определяющие вклад текущего контроля знаний и текущей аттестации в рейтинговую оценку:

Формирование оценки за текущую успеваемость:

- практические работы – 50 %;
- дискуссия – 15 %;
- индивидуальный проект – 35 %.

Рейтинговая оценка по дисциплине рассчитывается на основе оценки текущей успеваемости и экзаменационной оценки с учетом их весовых коэффициентов. Вес оценка по текущей успеваемости составляет 50 %, экзаменационная оценка – 50 %.

Примерный перечень практических работ

- Практическая работа 1. Графический анализ данных
- Практическая работа 2. Трансформация данных
- Практическая работа 3. Исследовательский анализ данных
- Практическая работа 4. Регрессионный анализ данных
- Практическая работа 5. Модели классификации данных
- Практическая работа 6. Модели кластеризации данных

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используются следующие инновационные подходы и методы.

1. Практико-ориентированный подход, который предполагает:

- освоение содержания образования через решения практических задач;
- приобретение навыков эффективного выполнения разных видов профессиональной деятельности;
- ориентацию на генерирование идей, реализацию групповых студенческих проектов, развитие предпринимательской культуры;
- использование процедур, способов оценивания, фиксирующих сформированность профессиональных компетенций.

2. Метод проектного обучения, который предполагает:

- способ организации учебной деятельности студентов, развивающий актуальные для учебной и профессиональной деятельности навыки планирования, самоорганизации, сотрудничества и предполагающий создание собственного продукта;
- приобретение навыков для решения исследовательских, творческих, социальных, предпринимательских и коммуникационных задач.

3. Метод учебной дискуссии, который предполагает участие студентов в целенаправленном обмене мнениями, идеями для предъявления и/или согласования существующих позиций по определенной проблеме.

Использование метода обеспечивает появление нового уровня понимания изучаемой темы, применение знаний (теорий, концепций) при решении проблем, определение способов их решения.

4. Методы и приемы развития критического мышления, которые представляют собой систему, формирующую навыки работы с информацией в процессе чтения и письма; понимания информации как отправного, а не конечного пункта критического мышления.

Методические рекомендации по организации самостоятельной работы обучающихся

При изучении учебной дисциплины используются следующие формы самостоятельной работы:

- выполнение заданий, выдаваемых на практических занятиях;
- изучение материала, вынесенного на самостоятельную проработку;
- подготовка к практическим занятиям;
- подготовка индивидуальных и групповых проектов по анализу данных на R;
- подготовка к экзамену.

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)
1.			
2.			

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО
ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**

на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
_____ (протокол № ____ от _____ 201_ г.)

Заведующий кафедрой

УТВЕРЖДАЮ
Декан факультета
