

Белорусский государственный университет

УТВЕРЖДАЮ

Проректор по учебной работе
и образовательным инновациям

О.Н. Здрок



30 июня 2021 г.

Регистрационный № УД- 9963 /уч.

РАЗВЕДОЧНЫЙ АНАЛИЗ ДАННЫХ

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

1-25 01 12 Экономическая информатика

2021 г.

Учебная программа составлена на основе ОСВО 1-25 01 12-2013, учебного плана рег. № E25-289/уч. от 16.03.2020.

СОСТАВИТЕЛИ:

И.А. Карачун, заведующий кафедрой цифровой экономики экономического факультета БГУ, кандидат экономических наук, доцент

РЕЦЕНЗЕНТЫ:

В.О. Сувалов, ведущий специалист ОАО «Банк БелВЭБ»

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой цифровой экономики
(протокол № 10 от 29.06.2021);
Научно-методическим Советом БГУ
(протокол № 7 от 30.06.2021)

Заведующий кафедрой цифровой экономики

И.А. Карачун

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Цель учебной дисциплины – формирование у студентов комплекса теоретических знаний и методологических основ в области проведения разведочного анализа данных с использованием интеллектуальных систем обработки информации и визуализации данных.

Задачи учебной дисциплины:

- Получить представление о современных подходах и процедурах разведочного анализа данных.
- Познакомиться с основными методами интеллектуального анализа данных.
- Изучить основные технологии визуализации и представления данных.
- Выработать навыки использования современных информационных технологий и программных средств, реализующих алгоритмы интеллектуального анализа данных.

Место учебной дисциплины в системе подготовки специалиста с высшим образованием.

Учебная дисциплина относится к циклу специальных дисциплин (дисциплин по выбору студента) компонента учреждения высшего образования.

Связи с другими учебными дисциплинами, включая учебные дисциплины компонента учреждения высшего образования, дисциплины специализации и др. Базовыми дисциплинами для изучения курса «Разведочный анализ данных» являются: «Высшая математика», «Теория вероятностей и математическая статистика», «Статистика», «Компьютерные информационные технологии».

Требования к компетенциям

Освоение учебной дисциплины «Разведочный анализ данных» должно обеспечить формирование следующих академических, социально-личностных и профессиональных компетенций:

академические компетенции:

АК-1. Уметь применять базовые научно-теоретические знания для решения теоретических и практических задач.

АК-2. Владеть системным и сравнительным анализом.

АК-3. Владеть исследовательскими навыками.

АК-4. Уметь работать самостоятельно.

АК-5. Быть способным порождать новые идеи (обладать креативностью).

АК-6. Владеть междисциплинарным подходом при решении проблем.

АК-7. Иметь навыки, связанные с использованием технических устройств, управлением информацией и работой с компьютером.

АК-9. Уметь учиться, повышать свою квалификацию в течение всей жизни.

социально-личностные компетенции:

СЛК-2. Быть способным к социальному взаимодействию.

СЛК-3. Обладать способностью к межличностным коммуникациям.

СЛК-4. Владеть навыками здоровьесбережения.

СЛК-5. Быть способным к критике и самокритике.

СЛК-6. Уметь работать в команде.

профессиональные компетенции:

ПК-13. Оценивать эффективность решений в сфере информатизации.

ПК-14. Использовать информационные технологии для повышения эффективности обработки исходных данных, проведения математических и статистических расчётов, ведения документооборота и маркетинговых исследований.

ПК-26. Осуществлять проектирование, тестирование, сопровождение и эксплуатацию информационных систем, разрабатывать техническую документацию к программному обеспечению и требования к внедрению тиражируемых информационных систем.

ПК-27. Проводить научные исследования в области использования информационных технологий в экономике.

ПК-28. Проводить научные исследования с целью совершенствования методов проектирования, тестирования, оценки качества, внедрения и сопровождения прикладного программного обеспечения.

В результате освоения учебной дисциплины студент должен:

знать: основные понятия математического анализа, теории вероятностей и математической статистики, теоретические основы информатики, программирования; место и роль интеллектуальных систем обработки данных в архитектуре предприятия; основные модели и технологии, основанные на обработке знаний в системах прикладного искусственного интеллекта, области и границы их применения; основные исторические и современные подходы анализа данных; основные методы, на основе которых проводится анализ данных; основные задачи разведочного анализа данных; основные модели описания и представления данных; основные программные средства разведочного анализа; основные визуальные средства представления данных;

уметь: использовать математические и инструментальные программные средства для решения задач анализа информации, создания электронных таблиц, отчётов и презентаций; формулировать требования к характеристикам интеллектуальной системы; выбирать шкалу представления данных; формулировать задачу разведочного анализа; проводить сравнение функциональных возможностей программных средств разведочного анализа данных; использовать методы формальной логики для представления предметной области; формировать визуальные представления статических данных; формировать визуальные представления временных рядов; формировать визуальные представления финансовых данных; интерпретировать и синтезировать комплексные модели, построенные на основе технологий прикладного искусственного интеллекта;

владеть: основными классами современных и перспективных интеллектуальных систем, входящих в ИТ-инфраструктуру предприятия; основным содержанием дисциплин микроэкономики, менеджмента, архитектуры предприятия, теоретическими основами информатики; программными средствами (минимум одним) разведочного анализа данных; методами обнаружения ошибок в визуальных представлениях.

Структура учебной дисциплины

Дисциплина изучается в 4 семестре. Всего на изучение учебной дисциплины «Разведочный анализ данных» отведено:

для очной формы получения высшего образования – 146 часов, в том числе 72 аудиторных часа, из них: лекции – 28 часов, практические занятия – 36 часов, управляемая самостоятельная работа (дистанционно) – 8 часов.

Трудоемкость учебной дисциплины составляет 4 зачетные единицы.

Форма текущей аттестации – *экзамен*.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Тема 1. Основные понятия, задачи, модели и методы разведочного анализа

Краткая историческая справка. Основные понятия и определения. Сравнительная классификация задач, решаемых человеком и искусственным интеллектом. Эволюция информационных систем. Предмет и содержание курса. Связь курса с другими дисциплинами. Обзор моделей представления знания и предметной области. Интеллектуальные методы обработки информации. Проблемы управления знаниями предприятия.

Тема 2. Программные средства разведочного анализа

Программные средства, реализующие интеллектуальные методы обработки. Сравнительный анализ программных продуктов, на примере продуктов: Microsoft Excel, Wolfram Mathematica, ЯП R и Python. Визуализация и подготовка графических материалов для отчётов.

Тема 3. Типы и качество данных

Понятие «добыча данных» (Data Mining). Хранилища данных. Предварительное исследование данных. Правила оформления датасета.

Тема 4. Визуальный анализ данных

Визуализация результатов статистической обработки. Виды графиков. Визуализация временных рядов. Интерполяционные техники. Визуализация финансовых рядов. Графики типа «японские свечи» и OHLC. Визуализация многомерных данных. Основные ошибки визуализации. Погружение в тему и контекст.

Тема 5. Характеристики данных

Метрические и оценочные показатели. Оценка центрального положения. Оценки вариабельности. Корреляция. Обследование распределения данных. Обследование двоичных и категориальных данных.

Тема 6. Основы Python

Создание типа данных «класс». Работа с базами данных. Связанные структуры данных. Работа с графикой. Сортировки. Визуализация данных на языке Python с помощью библиотеки matplotlib.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Дневная форма получения образования с применением дистанционных образовательных технологий

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	Основные понятия, задачи, модели и методы разведочного анализа	4			2			Коллоквиум, реферат, тест, дискуссия
2	Программные средства разведочного анализа	4			4			Контрольная работа, тест, дискуссия
3	Типы и качество данных	4			6			проект, дискуссия
4	Визуальный анализ данных	6			8		4 (ДО)	Открытое эвристическое задание, лабораторная работа, устный опрос, презентация
5	Характеристики данных	4			8			Проект, тест, лабораторная работа
6	Основы Python	6			8		4 (ДО)	Открытое эвристическое задание, доклад, дискуссия
		28			36		8	

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Миркин Б. Г. Введение в анализ данных : учебник и практикум. – М.: Издательство Юрайт, 2019. – 174 с.
2. Козлов А.Ю., Мхитарян В.С., Шишов В.Ф. Статистический анализ данных в MS Excel. Учебное пособие. – М.: ИНФРА-М, 2019. – 320 с.
3. Казанский А.А. Прикладное программирование на Excel 2019. Учебное пособие для вузов. – М.: Издательство Юрайт, 2019. – 171 с.
4. Груздев А.В. Прогнозное моделирование в IBM SPSS Statistics, R и Python: метод деревьев решений и случайный лес. – М.: ДМК Пресс, 2018. – 642с.
5. Федоров Д.Ю. Программирование на языке высокого уровня Python. – М.: Юрайт, 2019. – 161 с.
6. Грас Дж. Data Science. Наука о данных с нуля: Пер. с англ. – СПб.: БХВ-Петербург, 2019. – 336 с.

Перечень дополнительной литературы

1. Калинина В.Н., Соловьев В.И. Анализ данных. Компьютерный практикум (для бакалавров). Учебное пособие. – М.: КноРус, 2017. – 166 с.
2. Соловьев В.И. Анализ данных в экономике: Теория вероятностей, прикладная статистика, обработка и анализ данных в Microsoft Excel. – М.: КноРус, 2019. – 497с.
3. Фрэнкс Б. Укрощение больших данных. Как извлекать знания из массивов информации с помощью глубокой аналитики. – М.: Манн, Иванов и Фербер, 2014. – 349 с.
4. Grolemund G., Wickham H. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. – O'Reilly, 2017. – 520 p.

Перечень рекомендуемых средств диагностики и методика формирования итоговой оценки

Формой текущей аттестации по дисциплине «Разведочный анализ данных» 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.

Оценка за письменные и устные ответы на лекциях (опрос) включает в себя корректность и полноту ответа, обоснованность аргументов, наличие примеров из практики. Оценка за выполнение лабораторных работ формируется на основе следующих критериев: корректность полученных результатов и их интерпретацию, умение воспроизвести выполнение заданий, защиты выполненных индивидуальных заданий, полнота ответов на вопросы. Итоговая оценка за лабораторные работы рассчитывается путем усреднения оценок за запланированные к выполнению

лабораторные работы. Контрольные работы оцениваются исходя из полноты выполнения заданий, корректности полученных результатов, качества исполнения, проявления креативности. Итоговая оценка за контрольные работы (тесты) рассчитывается путем усреднения оценок по всем контрольным работам (тестам).

Формирование оценки за текущую успеваемость:

- выполнение проектов, открытых эвристических заданий – 20 %;
- контрольные работы (тесты), коллоквиумы – 20%;
- доклады, участие в дискуссиях на лекциях – 20 %;
- выполнение лабораторных работ – 40 %.

Итоговая оценка по дисциплине рассчитывается на основе оценки текущей успеваемости (рейтинговой оценки) и экзаменационной оценки с учетом их весовых коэффициентов и использования правил математического округления.

Примерный перечень заданий для управляемой самостоятельной работы студентов

Управляемая самостоятельная работа (консультационно-методическая поддержка и контроль) осуществляется в дистанционной форме и обеспечивается средствами образовательного портала БГУ LMS Moodle. В отдельных случаях управляемая самостоятельная работа проводится в форме аудиторных занятий, согласно утвержденному графику. Объем часов на составление и размещение заданий, консультации и контроль, осуществляемые с использованием технологий дистанционного обучения, планируется в пределах учебных часов, отведенных на УСР. Приоритетным направлением для разработки УСР в дистанционной форме являются открытые задания как основной содержательный элемент эвристического обучения.

Тема 4. Визуальный анализ данных.

Презентация проекта (4 часа)

Задание, формирующее *достаточные знания* по изученному учебному материалу на уровне узнавания.

Эссе «Визуализация курсового исследования».

Задание: Проанализируйте свою курсовую работу на предмет соответствия визуализации своего исследования и его результатов (графиков и таблиц, схем и пр.) основным принципам качественной презентации. Возможно, часть сведений, представленных в текстовом виде, можно было организовать в виде графических объектов. Напишите подробный «разбор полетов». В задание прикладывайте исходный файл курсовой работы и файл с текстом эссе.

Задание, формирующее *компетенции на уровне воспроизведения и применения полученных знаний*.

Форма контроля: устный опрос, презентация.

Тема 6. Основы Python.
Презентация проекта (4 часа)

Задание, формирующее *достаточные знания* по изученному учебному материалу на уровне узнавания.

Устный опрос по разделу Основы Python:

1. Очистка и форматирование данных.
2. Разведочный анализ данных.
3. Конструирование и выбор признаков.
4. Сравнение метрик нескольких моделей машинного обучения.
5. Гиперпараметрическая настройка лучшей модели.
6. Оценка лучшей модели на тестовом наборе данных.
7. Интерпретирование результатов работы модели.
8. Выводы и работа с документами.

Задание, формирующее *компетенции на уровне воспроизведения и применения полученных знаний*.

Форма контроля: устный опрос.

Примерная тематика лабораторных работ и открытых эвристических заданий

Форма контроля: устный опрос и презентация проектов.

1. Самостоятельная оценка параметров макроэкономических показателей.

Задание: Каждому студенту (индивидуально) нужно рассчитать статистики, рассмотренные на лекциях, для основных макроэкономических показателей базы данных проекта Мэддисон для нескольких стран/территорий, сравнить и проанализировать во времени, выявить закономерности, предположить, почему они имеют место и т.п.

2. Статья для обсуждения.

Задание: Читаем статью, обсуждаем содержание. С какими тезисами вы готовы согласиться или не согласиться, что замечали за собой и своими знакомыми, ваше мнение относительно рекомендаций. Нажмите на ссылку <https://www.sb.by/articles/zybkoe-tsifrovoe-schaste.html>, чтобы открыть ресурс.

3. Оценка центрального положения и отклонений.

Задание: Выбираем собственный датасет и делаем для него оценки центрального положения и отклонений. Анализируем результаты, делаем выводы и представляем на занятии.

4. Корреляции, сравнения, аналитика.

Задание: Исследуем индивидуальные датасеты на особенности эмпирического распределения (гистограммы), наличие связей, ищем закономерности в данных, анализируем и представляем результаты.

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используются:

эвристический подход, который предполагает: осуществление студентами лично-значимых открытий окружающего мира; демонстрацию многообразия решений большинства профессиональных задач и жизненных проблем; творческую самореализацию обучающихся в процессе создания образовательных продуктов; индивидуализацию обучения через возможность самостоятельно ставить цели, осуществлять рефлексию собственной образовательной деятельности;

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;

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;

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;

метод группового обучения, который представляет собой форму организации учебно-познавательной деятельности обучающихся, предполагающую функционирование разных типов малых групп, работающих как над общими, так и специфическими учебными заданиями.

Методические рекомендации по организации самостоятельной работы обучающихся

Для организации самостоятельной работы студентов по учебной дисциплине следует использовать современные информационные ресурсы: разместить на образовательном портале комплекс учебных и учебно-методических материалов (учебно-программные материалы, учебное издание для теоретического изучения дисциплины, методические указания к лабораторным занятиям, материалы текущего контроля и текущей аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательных стандартов высшего образования и учебно-программной документации, в т.ч. вопросы для подготовки к зачету, задания, тесты, вопросы для самоконтроля, список рекомендуемой литературы, информационных ресурсов).

Самостоятельная (практическая) работа студентов по изучению дисциплины «Разведочный анализ данных» выполняется в форме аудиторной и внеаудиторной работы. Студентам предлагается самостоятельное изучение ряда вопросов, что предполагает углубленное изучение основной и дополнительной литературы. Эффективность самостоятельной работы студентов проверяется в ходе текущего и итогового контроля знаний. При изучении учебной дисциплины рекомендуется использовать следующие формы самостоятельной (практической) работы, предоставленной в системе дистанционного обучения: поиск (подбор) и обзор литературы и электронных источников по индивидуально-заданной проблеме курса; работы, предусматривающие решение задач и выполнение упражнений, выдаваемых на практических занятиях; подготовка к контрольным работам; подготовка к экзамену.

Примерное экзаменационное задание – представление индивидуального проекта

Типовой проект должен включать следующие разделы:

- Изучение постановки задачи (business analysis).
- Выявление (аудит) источников данных.
- Изучение затрагиваемых процессов и связанных с ними данных (system analysis).
- Разведочный анализ данных и выдвижение гипотез (explorative data analysis & data mining).
- Визуализация решения.
- Рекомендации по реализации, внедрению, поддержке решения.

На 10 баллов необходима качественная проработка всех указанных задач, приветствуется наличие публикаций по теме исследования. Защита проекта в форме дискуссии.

В соответствующий раздел курса LMS Moodle прикрепляется презентация проекта, файлы с решениями и исходники данных.

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)
Корпоративные информационные системы	Цифровой экономики	Изменений в учебной программе не требуется	29.06.2021, протокол № 10

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО
ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**
на 2021/2022 учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры цифровой экономики (протокол № _____ от _____ 202_ г.)

Заведующий кафедрой
к.э.н., доцент

И.А. Карачун

УТВЕРЖДАЮ
Декан факультета
к.ф.-м.н., доцент

А.А. Королева