Минобрнауки России

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Оренбургский государственный университет имени В.А. Бондаренко»

Кафедра прикладной математики

РАБОЧАЯ ПРОГРАММА

ДИСЦИПЛИНЫ

«Б1.Д.Б.22 Методы интеллектуального анализа данных»

Уровень высшего образования

БАКАЛАВРИАТ

Направление подготовки
38.03.05 Бизнес-информатика
(код и наименование направления подготовки)

Информационные системы и бизнес-аналитика
(наименование направленности (профиля) образовательной программы)

Квалификация
Бакалавр

Форма обучения
Очная

Год набора 2026

Рабочая программа дисциплины «*Б1.Д.Б.22 Методы интеллектуального анализа данных*» рассмотрена и утверждена на заседании кафедры

Кафедра прикладной математики
наименование кафедры

протокол № 11 от "12" марта 2026 г.

Заведующий кафедрой
Кафедра прикладной математики _____ И.П. Болодурина
наименование кафедры *подпись* *расшифровка подписи*

Исполнители:
Доцент каф ПМ, к.т.н. _____ Л.С. Гришина
должность *подпись* *расшифровка подписи*

_____
должность *подпись* *расшифровка подписи*

СОГЛАСОВАНО:
Председатель методической комиссии по направлению подготовки
38.03.05 Бизнес-информатика _____ И.П. Болодурина
код наименование *личная подпись* *расшифровка подписи*

Заведующий отделом формирования фонда и научной обработки документов
главный библиограф _____ Г.Е. Гасайтис
личная подпись *расшифровка подписи*

Уполномоченный по качеству института
_____ С.Н. Морозова
личная подпись *расшифровка подписи*

№ регистрации _____

© Гришина Л.С., 2026
© ОГУ, 2026

1 Цели и задачи освоения дисциплины

Цель (цели) освоения дисциплины:

формирование практических навыков по применению методов и алгоритмов интеллектуального анализа данных (Data Mining) для извлечения скрытых знаний, закономерностей и аналитических инсайтов из больших массивов бизнес-данных.

Задачи:

1. ознакомление с основными понятиями методологии CRISP-DM, задачами классификации, регрессии, кластеризации и поиска ассоциативных правил, математическими основами алгоритмов машинного обучения;
2. формирование и развитие на этой основе навыков проведения разведочного анализа данных (EDA), предобработки и визуализации данных, оценки качества моделей;
3. систематизация приемов и методов работы с программными средствами Data Mining, инструментами интерпретации результатов анализа для принятия управленческих решений;
4. ознакомление с современными технологиями интеллектуального анализа данных, основными понятиями, методами и принципами построения предиктивных моделей, анализом временных рядов, перспективными направлениями развития Data Mining в задачах бизнес-аналитики.

2 Место дисциплины в структуре образовательной программы

Дисциплина относится к базовой части блока Д «Дисциплины (модули)»

Пререквизиты дисциплины: *Б1.Д.Б.14 Математика, Б1.Д.Б.15 Теория вероятностей и математическая статистика, Б1.Д.Б.23 Моделирование бизнес-процессов*

Постреквизиты дисциплины: *Б1.Д.В.10 Имитационное моделирование экономических систем, Б1.Д.В.Э.3.1 Технология Big Data, Б1.Д.В.Э.3.2 Инструментальные средства искусственного интеллекта, Б2.П.Б.У.1 Научно-исследовательская работа (получение первичных навыков научно-исследовательской работы)*

3 Требования к результатам обучения по дисциплине

Процесс изучения дисциплины направлен на формирование следующих результатов обучения

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
ОПК-4 Способен понимать принципы работы информационных технологий; использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений	ОПК-4-В-2 Знать основные этапы и принципы принятия управленческого решения	**Знать:** - основные этапы интеллектуального анализа данных; - математические основы алгоритмов классификации, кластеризации и регрессии; - методы предобработки и визуализации данных; - метрики качества моделей машинного обучения. **Уметь:** - проводить разведочный анализ данных (EDA); - выбирать и обучать модели для

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
		решения бизнес-задач; - оценивать качество моделей с использованием соответствующих метрик; - интерпретировать результаты анализа для принятия управленческих решений. **Владеть:** - инструментарием Data Mining: bi- платформы - методами поиска ассоциативных правил и анализа временных рядов; - навыками представления результатов анализа бизнес-пользователям.
ОПК-6 Способен выполнять отдельные задачи в рамках коллективной научно-исследовательской, проектной и учебно-профессиональной деятельности для поиска, выработки и применения новых решений в области информационно-коммуникационных технологий	ОПК-6-В-3 Знает основы системного анализа для определения области исследования и моделирования для генерации и применения новых научных и проектных решений	**Знать:** - основы системного анализа в задачах Data Mining; - методологию CRISP-DM. **Уметь:** - формулировать задачи интеллектуального анализа данных на основе бизнес-требований; - выбирать методический аппарат исследования. **Владеть:** - методами научного исследования в области анализа данных; - навыками работы в команде над проектами Data Mining.

4 Структура и содержание дисциплины

4.1 Структура дисциплины

Общая трудоемкость дисциплины составляет 4 зачетные единицы (144 академических часа).

Вид работы	Трудоемкость, академических часов	
	5 семестр	всего
Общая трудоёмкость	**144**	**144**
Контактная работа:	**51,25**	**51,25**
Лекции (Л)	34	34
Практические занятия (ПЗ)	16	16
Консультации	1	1
Промежуточная аттестация (зачет, экзамен)	0,25	0,25
Самостоятельная работа: - выполнение индивидуального творческого задания (ИТЗ); - самоподготовка (проработка и повторение лекционного материала и	**92,75**	**92,75**

Вид работы	Трудоемкость, академических часов	
	5 семестр	всего
материала учебников и учебных пособий; - изучение разделов курса в системе электронного обучения; - подготовка к практическим занятиям; - подготовка к рубежному контролю и т.п.)		
Вид итогового контроля (зачет, экзамен, дифференцированный зачет)	**экзамен**	

Разделы дисциплины, изучаемые в 5 семестре

№ раздела	Наименование разделов	Количество часов				
		всего	аудиторная работа			внеауд. работа
			Л	ПЗ	ЛР	
1	Введение в Data Mining. Жизненный цикл проекта (CRISP-DM).	36	8	4		24
2	Задачи классификации и регрессии.	36	10	6		20
3	Кластерный анализ и поиск аномалий.	36	8	4		24
4	Поиск ассоциативных правил и анализ временных рядов.	36	8	2		26
	Итого:	144	34	16		94
	Всего:	144	34	16		94

4.2 Содержание разделов дисциплины

Раздел 1. Введение в Data Mining. Жизненный цикл проекта (CRISP-DM).

Теоретические основы интеллектуального анализа данных. Понятие, предмет и задачи интеллектуального анализа данных (Data Mining). Место Data Mining в системе наук о данных. Отличие методов интеллектуального анализа данных от классической математической статистики и эконометрики. Основные классы задач Data Mining: классификация, регрессия, кластеризация, поиск ассоциативных правил, анализ временных рядов, обнаружение аномалий.

Методология CRISP-DM. Жизненный цикл проекта интеллектуального анализа данных в соответствии с методологией Cross-Industry Standard Process for Data Mining (CRISP-DM): понимание бизнес-задачи, понимание данных, подготовка данных, моделирование, оценка модели, внедрение. Формирование требований к данным на основе бизнес-целей. Построение карты данных.

Разведочный анализ данных (Exploratory Data Analysis, EDA). Принципы и задачи разведочного анализа данных. Одномерный и многомерный статистический анализ. Визуализация распределений, корреляционных зависимостей и структурных особенностей данных. Методы графического представления: гистограммы, диаграммы размаха, диаграммы рассеяния, матрицы корреляций.

Предобработка данных. Обработка пропущенных значений: методы удаления, импутации. Обнаружение и обработка выбросов: статистические методы (правило трёх сигм, межквартильный размах), метод изолирующего леса. Нормализация и стандартизация количественных признаков. Кодирование категориальных признаков. Понижение размерности: метод главных компонент (PCA).

BI-платформы как среда реализации задач Data Mining. Обзор современных BI-платформ и аналитических систем: отечественные решения (Visiology, BaseGroup Analytics, Deductor Studio, Yandex DataLens, Directum Analytics, LogiRU) и зарубежные аналоги (Microsoft Power BI, Tableau, Qlik Sense, IBM SPSS Modeler). Архитектура BI-платформ: источники данных, хранилища данных, витрины данных, аналитические модели, презентационный уровень. Возможности BI-платформ по реализации методов Data Mining без написания программного кода (no-code / low-code подход).

Раздел 2. Задачи классификации и регрессии.

Постановка задачи классификации. Классификация как задача обучения с учителем. Формализация задачи: пространство признаков, пространство решений, обучающая выборка. Понятие разделяющей поверхности. Бинарная и многоклассовая классификация.

Методы классификации. Линейные методы. Методы, основанные на деревьях: деревья решений (CART, C4.5, C5.0), критерии разбиения (прирост информации, индекс Джини). Метод опорных векторов (SVM). Ансамблевые методы. Метод k-ближайших соседей (k-NN).

Метрики качества моделей классификации. Матрица ошибок. Accuracy, Precision, Recall, F1-score, Fβ-score. ROC-кривая и площадь под ней (ROC-AUC). PR-кривая (Precision-Recall Curve). Коэффициент Каппа (Cohen's Kappa). Логарифмические потери (Log Loss).

Постановка задачи регрессии. Регрессия как задача прогнозирования непрерывной числовой целевой переменной. Формализация задачи. Методы регрессии. Линейные методы: метод наименьших квадратов (МНК), гребневая регрессия (Ridge), лассо-регрессия (Lasso), Elastic Net. Регрессионные деревья: CART для регрессии, случайный лес регрессии. Градиентный бустинг для регрессии. Метод опорных векторов для регрессии (SVR). Метод k-ближайших соседей для регрессии (k-NN Regressor).

Метрики качества моделей регрессии. Средняя абсолютная ошибка (MAE), среднеквадратичная ошибка (MSE, RMSE), средняя абсолютная процентная ошибка (MAPE), коэффициент детерминации (R^2), скорректированный R^2. Анализ остатков. Переобучение и регуляризация.

Реализация задач классификации и регрессии в BI-платформах. Построение предиктивных моделей в среде Microsoft Power BI (интеграция с Azure ML), Visiology, Deductor Studio, BaseGroup Analytics. Визуализация результатов классификации и регрессии: матрицы ошибок, ROC-кривые, графики предсказанных и фактических значений, анализ остатков. Интерпретация моделей для бизнес-пользователей: важность признаков, SHAP-значения.

Раздел 3. Кластерный анализ и поиск аномалий.

Постановка задачи кластеризации. Кластеризация как задача обучения без учителя. Формализация задачи: разбиение множества объектов на непустые непересекающиеся подмножества (кластеры). Критерии качества кластеризации: компактность, разделимость. Меры близости. Методы кластеризации. Центроидные методы. Иерархическая кластеризация. Методы, основанные на плотности. Метрики качества кластеризации.

Понятие аномалии, выброса и новизны. Классификация аномалий. Применение кластерного анализа и поиска аномалий в бизнес-аналитике. Сегментация клиентской базы (RFM-анализ, поведенческая сегментация). Кластерный анализ товарных категорий. Обнаружение мошеннических операций в финансовой сфере. Выявление аномалий в производственных процессах и IT-инфраструктуре.

Реализация в BI-платформах. Кластерный анализ в среде Deductor Studio, Visiology, BaseGroup Analytics. Визуализация кластеров: диаграммы рассеяния с цветовой кодировкой, радарные диаграммы, тепловые карты. Интерпретация профилей кластеров для бизнес-пользователей. Встроенные средства обнаружения аномалий в Microsoft Power BI и Yandex DataLens.

Раздел 4. Поиск ассоциативных правил и анализ временных рядов.

Задача поиска ассоциативных правил. Алгоритмы Apriori и FP-Growth. Метрики качества ассоциативных правил. Поддержка (support), достоверность (confidence), лифт (lift), убеждение (conviction), леверидж (leverage). Фильтрация и ранжирование правил. Визуализация ассоциативных правил: матрицы, графы, тепловые карты.

Применение ассоциативных правил в бизнес-аналитике. Анализ покупательского спроса в розничной торговле. Оптимизация выкладки товаров в торговых залах. Анализ паттернов поведения пользователей в цифровых сервисах.

Анализ временных рядов. Понятие временного ряда, его компоненты: тренд, сезонность, циклическая составляющая, случайная компонента. Стационарность временных рядов: строгая и слабая. Тесты на стационарность (Дики-Фуллера, КПСС). Методы анализа и прогнозирования временных рядов.

4.3 Практические занятия (семинары)

№ занятия	№ раздела	Тема	Кол-во часов
1-2	1	Разведочный анализ данных (EDA) и предобработка.	4
3	2	Построение моделей регрессии	2
4-5	2	Построение моделей классификации (Логистическая регрессия, Случайный лес).	4
6-7	3	Кластеризация клиентской базы (K-Means).	4
8	4	Поиск ассоциативных правил (Алгоритм Apriori) в розничных продажах.	2
		Итого:	16

5 Учебно-методическое обеспечение дисциплины

5.1 Основная литература

1. Павлов, С. И. Системы искусственного интеллекта: учебное пособие: [16+] / С. И. Павлов. – Томск: Томский государственный университет систем управления и радиоэлектроники, 2011. – Часть 1. – 175 с. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=208933 (дата обращения: 02.06.2026).

2. Павлов, С. И. Системы искусственного интеллекта: учебное пособие / С. И. Павлов. – Томск: Томский государственный университет систем управления и радиоэлектроники, 2011. – Часть 2. – 194 с. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=208939 (дата обращения: 02.06.2026).

3. Кревецкий, А. В. Основы технологий искусственного интеллекта: учебное пособие: [16+] / А. В. Кревецкий, Ю. А. Ипатов, Н. И. Роженцова; под общ. ред. А. В. Кревецкого ; Поволжский государственный технологический университет. – Йошкар-Ола: Поволжский государственный технологический университет, 2023. – 272 с.: ил., табл., схем. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=714624 (дата обращения: 02.06.2026).

5.2 Дополнительная литература

1. Ясницкий, Л. Н. Интеллектуальные системы: учебник / Л. Н. Ясницкий. – 2-е изд. – Москва : Лаборатория знаний, 2020. – 224 с.: ил., табл., схем. – (Учебник для высшей школы). – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=712949 (дата обращения: 02.04.2026).

2. Машинное обучение: учебник: [16+] / Е. Ю. Бутырский, В. В. Цехановский, Н. А. Жукова [и др.]. – Москва : Директ-Медиа, 2023. – 368 с.: ил., табл., схем., граф. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=701807 (дата обращения: 02.04.2026).

3. Ручкин, В. Н. Универсальный искусственный интеллект и экспертные системы / В. Н. Ручкин, В. А. Фулин . - СПб.: БХВ-Петербург, 2009. - 238 с.

4. Джарратано, Д. Экспертные системы: принципы разработки и программирование / Д. Джарратано, Г. Райли.: пер. с англ.-М.: ООО «И.Д.Вильямс», 2007.-1152 с.

5. Чулюков, В. А. Системы искусственного интеллекта. Практический курс: Учеб. пособие для вузов / В.А. Чулюков. - М.: Бином, 2008. - 293 с. : ил.

6. Мещерина, Е. В. Системы искусственного интеллекта: методические указания для обучающихся по образовательным программам высшего образования по направлениям подготовки 02.03.01 Математика и компьютерные науки, 02.03.02 Фундаментальная информатика и 7 информационные технологии / Е. В. Мещерина; М-во науки и высш. образования Рос. Федерации, Федер. гос. бюджет. образоват. учреждение высш. образования "Оренбург. гос. ун-т", Каф. алгебры и дискрет. математики. - Оренбург: ОГУ. - 2019. - 33 с.- [Электронный ресурс]. – URL: http://artlib.osu.ru/web/books/metod_all/107694_20190802.pdf

7. Мещерина, Е. В. Системы искусственного интеллекта: учебно-методическое пособие для обучающихся по образовательным программам высшего образования по направлениям подготовки 02.03.01 Математика и компьютерные науки, 02.03.02 Фундаментальная информатика и информационные технологии, специальности 10.05.01 Компьютерная безопасность / Е. В. Мещерина; М-во науки и высш. образования Рос. Федерации, Федер. гос. бюджет. образоват. учреждение высш. образования "Оренбург. гос. ун-т". - Оренбург : ОГУ. - 2019. - 96 с.- То же [Электронный ресурс] – URL: http://artlib.osu.ru/web/books/metod_all/109690_20191002.pdf

8. Сидоркина И. Г. Системы искусственного интеллекта: учеб.пособие для вузов. - М.:КноРус, 7 2011. - 245 с..

5.3 Периодические издания

1. Вестник компьютерных и информационных технологий: журнал. - М.: Изд. "Спектр"
2. Информационные технологии: журнал. - М.: Изд. "Новые технологии".
3. Автоматизация в промышленности: журнал. - М. :Изд. дом "Инфоавтоматизация"

5.4 Интернет-ресурсы

1. www.aiportal.ru – портал искусственного интеллекта. На сайте представлены статьи и файлы по основным направлениям исследований в области искусственного интеллекта.
2. www.intuit.ru/department/ds/fuzzysets - сайт Национального Открытого Университета «Интуит», курс «Нечеткие множества»;
3. http://www.intuit.ru/studies/courses/1122/167/info - сайт Национального Открытого Университета «Интуит», курс «Проектирование систем искусственного интеллекта;
4. https://www.coursera.org/learn/vvedenie-v-iskusstvennyi-intellekt - «Coursera», МООК: «Введение в искусственный интеллект»;
5. https://openedu.ru/course/hse/INTRAI/ - «Открытое образование», Каталог курсов, МООК: «Введение в искусственный интеллект».

5.5 Программное обеспечение, профессиональные базы данных и информационные справочные системы

1. Операционная система РЕД ОС
2. Пакет офисных приложений LibreOffice
3. Университетская платформа электронного обучения «Электронные курсы ОГУ в системе обучения Moodle» (http://moodle.osu.ru);
4. Автоматизированная интерактивная система сетевого тестирования - АИССТ (зарегистрирована в РОСПАТЕНТ, Свидетельство о государственной регистрации программы для ЭВМ №2011610456, правообладатель – Оренбургский государственный университет), режим доступа http://aist.osu.ru.

6 Материально-техническое обеспечение дисциплины

Учебные аудитории для проведения занятий лекционного типа, для проведения групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации.

Для проведения занятий семинарского типа используется компьютерный класс, оснащенный компьютерами с минимальными характеристиками: оперативная память: не менее 2 Gb; процессор не 8 менее чем на 2 ядра и частотой не менее 1,6 Ghz; объём памяти видеокарты не менее 512 Mb; жесткий диск не менее чем на 200Gb; наличие Usb – разъема на лицевой стороне системного блока (вверху); диагональ ЖК монитора не менее 17. 8.

Аудитории оснащены комплектами ученической мебели, техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Помещение для самостоятельной работы обучающихся оснащены компьютерной техникой, подключенной к сети "Интернет", и обеспечением доступа в электронную информационно-образовательную среду ОГУ.