

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «БАЙКАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

Проректор по учебной работе
д.юр.н., доц. Васильева Н.В.



21.06.2024г.

Рабочая программа дисциплины
Б1.Э.2. Data Mining и бизнес-аналитика в экономике

Направление подготовки: 09.03.03 Прикладная информатика
Направленность (профиль): Системы искусственного интеллекта
Квалификация выпускника: бакалавр
Форма обучения: очная, заочная

	Очная ФО	Заочная ФО
Курс	3	3
Семестр	32	32
Лекции (час)	18	4
Практические (сем, лаб.) занятия (час)	36	10
Самостоятельная работа, включая подготовку к экзаменам и зачетам (час)	54	94
Курсовая работа (час)		
Всего часов	108	108
Зачет (семестр)		
Экзамен (семестр)	32	32

Иркутск 2024

Программа составлена в соответствии с ФГОС ВО по направлению 09.03.03
Прикладная информатика.

Автор Е.В. Аксенюшкина

Рабочая программа обсуждена и утверждена на заседании кафедры
математических методов и цифровых технологий

Заведующий кафедрой А.В. Родионов

1. Цели изучения дисциплины

Цель дисциплины сформировать у студентов компетенции в области применения технологий обработки данных (в том числе big data) и машинного обучения к решению прикладных задач, современных проблем прикладной математики и информатики, проблем обработки и анализа информации.

Задачи дисциплины:

- формирование у студентов представления о типах задач, возникающих в области интеллектуального анализа данных (Data Mining);
- изучение основных подходов и алгоритмов решения задач анализа данных и особенностей их применения к решению реальных задач;
- получение студентами навыка по выявлению, формализации и успешному решению практических задач анализа данных, возникающие в процессе их профессиональной деятельности;
- получение практического навыка в работе с существующими программными пакетами по анализу данных.

2. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Компетенции обучающегося, формируемые в результате освоения дисциплины

Код компетенции по ФГОС ВО	Компетенция
ПК-4	Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС
ПК-8	Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях

Структура компетенции

Компетенция	Формируемые ЗУНы
ПК-4 Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У. Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС Н. Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС
ПК-8 Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях	З. Знать, как руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях У. Уметь разрабатывать проекты по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях Н. Владеть навыками руководства проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий

3. Место дисциплины (модуля) в структуре образовательной программы

Принадлежность дисциплины - БЛОК 1 ДИСЦИПЛИНЫ (МОДУЛИ): Элективная дисциплина.

4. Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 3 зач. ед., 108 часов.

Вид учебной работы	Количество часов (очная ФО)	Количество часов (заочная ФО)
Контактная(аудиторная) работа		
Лекции	18	4
Практические (сем, лаб.) занятия	36	10
Самостоятельная работа, включая подготовку к экзаменам и зачетам	54	94
Всего часов	108	108

5. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

5.1. Содержание разделов дисциплины

Заочная форма обучения

№ п/п	Раздел и тема дисциплины	Семестр	Лекции	Семинар Лаборат. Практич.	Самостоят. раб.	В интерактивной форме	Формы текущего контроля успеваемости
1	Методы и стадии Data Mining	32	0	2	20		Лабораторная работа №1 по теме 1
2	Методы классификации и прогнозирования	32	2	4	27		Лабораторная работа №2 по теме 2
3	Метод кластерного анализа	32	2	2	27		Лабораторная работа №3 по теме 3
4	Методы поиска ассоциативных правил	32	0	2	20		Лабораторная работа №4 по теме 4. Теоретический тест по темам 1-4
	ИТОГО		4	10	94		

Очная форма обучения

№ п/п	Раздел и тема дисциплины	Семестр	Лекции	Семинар Лаборат. Практич.	Самостоят. раб.	В интерактивной форме	Формы текущего контроля успеваемости
-------	--------------------------	---------	--------	---------------------------------	--------------------	--------------------------	--------------------------------------

№ п/п	Раздел и тема дисциплины	Семестр	Лекции	Семинар Лаборат. Практич.	Самостоят. раб.	В интерактивной форме	Формы текущего контроля успеваемости
1	Методы и стадии Data Mining	32	4	10	12		Лабораторная работа №1 по теме 1
2	Методы классификации и прогнозирования	32	6	14	14		Лабораторная работа №2 по теме 2
3	Метод кластерного анализа	32	4	10	14		Лабораторная работа №3 по теме 3
4	Методы поиска ассоциативных правил	32	4	2	14		Лабораторная работа №4 по теме 4. Теоретический тест по темам 1-4
	ИТОГО		18	36	54		

5.2. Лекционные занятия, их содержание

№ п/п	Наименование разделов и тем	Содержание
1	Технологии анализа данных	Введение в анализ данных. Принципы анализа данных. Структурированные данные. Подготовка данных к анализу. Технологии Data Mining. Аналитические платформы. Введение в алгоритмы Data Mining.
2	Очистка и предобработка данных	Оценка качества данных. Технологии и методы оценки качества данных. Очистка и предобработка. Фильтрация данных. Введение в сокращение размерности. Сокращение числа признаков. Сэмплинг.
3	Классификация и регрессия. Статистические методы	Введение в классификацию и регрессию. Линейная регрессия. Оценка соответствия линейной регрессии реальным данным. Множественная линейная регрессия. Методы отбора переменных в регрессионные модели. Основы логической регрессии. Интерпретация модели логической регрессии.
4	Классификация и регрессии. Машинное обучение	Введение в деревья решений. Алгоритмы построения деревьев решений. Ансамбли моделей. Бэггинг. Бустинг. Альтернативные методы построения ансамблей.
5	Анализ и прогнозирование временных рядов	Введение в прогнозирование. Временные ряды и его компоненты. Модели прогнозирования. Прогнозирование в торговле и логистике.
6	Методы кластерного анализа. Иерархические методы	Методы кластерного анализа. Меры сходства. Методы объединения и связи. Иерархический кластерный анализ.
7	Методы кластерного анализа. Итеративные методы	Алгоритм k-средних (k-means). Алгоритм PAM. Факторный анализ. Новые алгоритмы и некоторые модификации кластерного анализа
8	Задача ассоциация	Ассоциативные правила. Алгоритм Apriori. Иерархические ассоциативные правила. Последовательные шаблоны.
9	Методы поиска ассоциативных	Характеристика ассоциативных правил. Границы поддержки и достоверности ассоциативных правил. Методы поиска

№ п/п	Наименование разделов и тем	Содержание
	правил	ассоциативных правил.

5.3. Семинарские, практические, лабораторные занятия, их содержание

№ раздела и темы	Содержание и формы проведения
1	Введение в Data Mining. Базовые типы данных. Структуры данных (списки, словари), функции, модули. Работа с Python.
1	Загрузка и первичный анализ данных. Чтение данных из CSV, Excel, SQL. Типы данных. Пропущенные значения. Описательная статистика. Визуализация данных (гистограммы, диаграммы рассеяния).
1	Очистка данных. Обработка пропущенных значений (замена на среднее, медиану). Поиск и обработка выбросов. Работа с дубликатами.
1	Преобразование признаков. Работа с категориальными признаками. Нормализация числовых признаков. Создание новых признаков.
1	Визуализация данных для Data Mining. Глубокое погружение в matplotlib и seaborn. Создание информативных графиков для анализа данных (ящик с усами, тепловые карты, парные диаграммы).
2	Введение в классификацию. Типы задач классификации. Метрики оценки качества классификации. переобучение и недообучение.
2	Логистическая регрессия. Математические основы. Реализация в Python. Интерпретация результатов. Применение в бинарной классификации.
2	Метод опорных векторов. Принцип работы алгоритма. Настройка параметров. Ядра SVM. Применение для линейно неразделимых данных.
2	Деревья решений. Построение дерева решений. Энтропия и информационный выигрыш. Преимущества и недостатки. Настройка параметров (глубина дерева, минимальное число объектов в листе).
2	Ансамбли алгоритмов. Bagging и Random Forest. Boosting. Сравнение качества ансамблей с одиночными моделями.
2	Применение классификации на практике. Распознавание спама. Кредитный скоринг. Медицинская диагностика.
3	Введение в кластеризацию. Типы задач кластеризации. Алгоритм k-means (реализация, выбор оптимального числа кластеров). Метрики оценки качества кластеризации.
3	Иерархическая кластеризация. Алгоритмы агломеративной кластеризации. Дендрограммы. Выбор оптимального числа кластеров. Сравнение с k-means.
3	Кластеризация на основе плотности. Принципы работы алгоритма. Настройка параметров. Применение для поиска кластеров сложной формы.
3	Применение кластеризации на практике. Сегментация клиентов. Кластеризация документов. Анализ изображений.
4	Ассоциативные правила. Алгоритм Apriori. Метрики оценки правил (поддержка, уверенность, лифт). Визуализация правил. Применение для анализа рыночных корзин.

6. Фонд оценочных средств для проведения промежуточной аттестации по дисциплине (полный текст приведен в приложении к рабочей программе)

6.1. Текущий контроль

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1...З.п, У.1...У.п, Н.1...Н.п))	Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы (Наименование оценочного средства)	Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания (по 100-балльной шкале)
1	1. Методы и стадии Data Mining	ПК-8	У. Уметь разрабатывать проекты по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях Н. Владеть навыками руководства проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях	Лабораторная работа №1 по теме 1	Лабораторная работа №1 оценивается в 20 баллов (20)
2	2. Методы классификации и прогнозирования	ПК-8	У. Уметь разрабатывать проекты по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях Н. Владеть навыками руководства проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях	Лабораторная работа №2 по теме 2	Лабораторная работа №2 оценивается в 20 баллов (20)
3	3. Метод кластерного анализа	ПК-4	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У. Уметь разрабатывать	Лабораторная работа №3 по теме 3	Лабораторная работа №3 оценивается в 20 баллов. (20)

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1...З.п, У.1...У.п, Н.1...Н.п)	Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы (Наименование оценочного средства)	Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания (по 100-балльной шкале)
			архитектуру, дизайна, информационное обеспечение и прототип ИС Н. Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС		
4	4. Методы поиска ассоциативных правил	ПК-4	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У. Уметь разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС Н. Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС	Лабораторная работа №4 по теме 4	Лабораторная работа №4 оценивается в 20 баллов. (20)
5		ПК-8	З. Знать, как руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях	Теоретический тест по темам 1-4	Теоретический тест состоит из 10 вопросов и оценивается в 20 баллов (20)
				Итого	100

6.2. Промежуточный контроль (зачет, экзамен)

Рабочим учебным планом предусмотрен Экзамен в семестре 32.

ВОПРОСЫ ДЛЯ ПРОВЕРКИ ЗНАНИЙ:

1-й вопрос билета (30 баллов), вид вопроса: Тест/проверка знаний. Критерий: Правильный ответ на вопрос теста оценивается в 3 балла.

Компетенция: ПК-4 Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

Знание: Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

1. Достоинства и недостатки регрессионных моделей.
2. Достоинства и недостатки тематического моделирования.
3. Особенности тематического моделирования.
4. Суть метода наименьших квадратов (МНК).
5. Суть модели скользящего среднего (МСС).
6. Цель применения метода МНК.
7. Цель применения тематического моделирования.

Компетенция: ПК-8 Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях

Знание: Знать, как руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях

8. Алгоритм K-means.
9. Алгоритм иерархической кластеризации.
10. Виды регрессионных моделей.
11. Достоинство и недостатки кластерного подхода.
12. Меры близости в кластерном анализе.
13. Направления в кластерном анализе.
14. Принцип выбора адекватной модели в прогнозировании.
15. Суть регрессионной модели (смысл коэффициентов регрессии).
16. Цели и задачи кластерного анализа.
17. Цели применения СС.
18. Цель применения регрессионной модели.
19. Что такое data mining, компоненты data mining.

ТИПОВЫЕ ЗАДАНИЯ ДЛЯ ПРОВЕРКИ УМЕНИЙ:

2-й вопрос билета (30 баллов), вид вопроса: Задание на умение. Критерий: Правильно выполненное задание оценивается в 30 баллов.

Компетенция: ПК-4 Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

Умение: Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС

Задача № 1. Выполните кластеризацию данных

Компетенция: ПК-8 Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях

Умение: Уметь разрабатывать проекты по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях

Задача № 2. Выполните кластеризацию данных

ТИПОВЫЕ ЗАДАНИЯ ДЛЯ ПРОВЕРКИ НАВЫКОВ:

3-й вопрос билета (40 баллов), вид вопроса: Задание на навыки. Критерий: Правильно выполненное задание оценивается в 40 баллов.

Компетенция: ПК-4 Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

Навык: Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС

Задание № 1. Выполните классификацию данных

Компетенция: ПК-8 Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях

Навык: Владеть навыками руководства проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях

Задание № 2. Выполните классификацию данных

ОБРАЗЕЦ БИЛЕТА

Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования «БАЙКАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ» (ФГБОУ ВО «БГУ»)	Направление - 09.03.03 Прикладная информатика Профиль - Системы искусственного интеллекта Кафедра математических методов и цифровых технологий Дисциплина - Data Mining и бизнес- аналитика в экономике
---	--

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 1

1. Тест (30 баллов).
2. Выполните кластеризацию данных (30 баллов).
3. Выполните классификацию данных (40 баллов).

Составитель _____ Е.В. Аксеньюшкина

Заведующий кафедрой _____ А.В. Родионов

7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

а) основная литература:

1. Шпер В. Л. Data Mining/ В. Л. Шпер// Номер журнала, N 7, С. 56-57, 2011, ч.з 2-202
2. Чубукова И.А. Data Mining. учебное пособие. Электронный ресурс/ И.А. Чубукова.- Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), 2016.-470 с.
3. [Федин, Ф. О. Анализ данных. Часть 1. Подготовка данных к анализу : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — Москва : Московский городской педагогический университет, 2012. — 204 с. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : \[сайт\]. — URL: <https://www.iprbookshop.ru/26444.html> \(дата обращения: 06.06.2021\). — Режим доступа: для авторизир. пользователей](https://www.iprbookshop.ru/26444.html)
4. [Федин, Ф. О. Анализ данных. Часть 2. Инструменты Data Mining : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — Москва : Московский городской педагогический университет, 2012. — 308 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART :](https://www.iprbookshop.ru/26444.html)

[сайт]. — URL: <https://www.iprbookshop.ru/26445.html> (дата обращения: 29.05.2023). — Режим доступа: для авторизир. пользователей

б) дополнительная литература:

1. Чубукова И. А. Data Mining. учеб. пособие/ И. А. Чубукова.- М.: БИНОМ. Лаборатория знаний, 2006.-382 с.
2. Чубукова И. А. Data Mining. 2-е изд., испр./ И.А. Чубукова.- Москва: Интернет-Университет Информационных Технологий, 2008.-383 с.
3. Братищенко В.В. Информационные технологии в бизнес-аналитике с использованием Orange3 и Python.- Иркутск: Изд-во БГУ, 2021.- 120 с.
4. [Воронов, В. И. Data Mining - технологии обработки больших данных : учебное пособие / В. И. Воронов, Л. И. Воронова, В. А. Усачев. — Москва : Московский технический университет связи и информатики, 2018. — 47 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : \[сайт\]. — URL: \[http://www.iprbookshop.ru/81324.html\]\(https://www.iprbookshop.ru/81324.html\)](https://www.iprbookshop.ru/81324.html)

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля), включая профессиональные базы данных и информационно-справочные системы

Для освоения дисциплины обучающемуся необходимы следующие ресурсы информационно-телекоммуникационной сети «Интернет»:

- Сайт Байкальского государственного университета, адрес доступа: <http://bgu.ru/>, доступ круглосуточный неограниченный из любой точки Интернет
- Электронно-библиотечная система IPRbooks, адрес доступа: <https://www.iprbookshop.ru>. доступ неограниченный

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Изучать дисциплину рекомендуется в соответствии с той последовательностью, которая обозначена в ее содержании. Для успешного освоения курса обучающиеся должны иметь первоначальные знания в области информатики и математики.

На лекциях преподаватель озвучивает тему, знакомит с перечнем литературы по теме, обосновывает место и роль этой темы в данной дисциплине, раскрывает ее практическое значение. В ходе лекций студенту необходимо вести конспект, фиксируя основные понятия и проблемные вопросы.

Практические (семинарские) занятия по своему содержанию связаны с тематикой лекционных занятий. Начинать подготовку к занятию целесообразно с конспекта лекций. Задание на практическое (семинарское) занятие сообщается обучающимся до его проведения. На семинаре преподаватель организует обсуждение этой темы, выступая в качестве организатора, консультанта и эксперта учебно-познавательной деятельности обучающегося.

Изучение дисциплины (модуля) включает самостоятельную работу обучающегося.

Основными видами самостоятельной работы студентов с участием преподавателей являются:

- текущие консультации;
- прием и разбор домашних заданий (в часы практических занятий);
- прием и защита лабораторных работ (во время проведения занятий).

Основными видами самостоятельной работы студентов без участия преподавателей являются:

- формирование и усвоение содержания конспекта лекций на базе рекомендованной лектором учебной литературы, включая информационные образовательные ресурсы (электронные учебники, электронные библиотеки и др.);

- самостоятельное изучение отдельных тем или вопросов по учебникам или учебным пособиям;
- подготовка к семинарам и лабораторным работам;
- выполнение домашних заданий в виде решения отдельных задач, проведения типовых расчетов, расчетно-компьютерных и индивидуальных работ по отдельным разделам содержания дисциплин и др.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения

В учебном процессе используется следующее программное обеспечение:

- MS Office,
- Python,

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю):

В учебном процессе используется следующее оборудование:

- Помещения для самостоятельной работы, оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду вуза,
- Учебные аудитории для проведения: занятий лекционного типа, занятий семинарского типа, практических занятий, выполнения курсовых работ, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные специализированной мебелью и техническими средствами обучения,
- Компьютерный класс