

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.О.26. Технологии машинного обучения и искусственного интеллекта в бизнесе

Направление подготовки: 38.03.05 Бизнес-информатика

Направленность (профиль): Автоматизация и цифровая трансформация бизнеса

Квалификация выпускника: бакалавр

Форма обучения: очно-заочная

Курс	4
Семестр	41-42
Лекции (час)	64
Практические (сем, лаб.) занятия (час)	36
Самостоятельная работа, включая подготовку к экзаменам и зачетам (час)	116
Курсовая работа (час)	
Всего часов	216
Зачет (семестр)	41
Экзамен (семестр)	42

Иркутск 2024

Программа составлена в соответствии с ФГОС ВО по направлению 38.03.05
Бизнес-информатика.

Автор В.В. Братищенко

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

Заведующий кафедрой А.В. Родионов

1. Цели изучения дисциплины

Целями освоения дисциплины «Информационные технологии бизнес-аналитики» являются следующие:

- получение теоретических знаний в области информационных технологий применяемых для аналитической обработки данных;
- овладение аналитическими возможностями языка SQL;
- приобретение навыков использования хранилищ данных;
- получение знаний и навыков использования моделей Data Mining.

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 проектировать, разрабатывать и внедрять компоненты ИТ-инфраструктуры предприятия, обеспечивающие совершенствование и поддержку бизнес-процессов, в том числе с применением инновационных цифровых технологий

Структура компетенции

Компетенция	Формируемые ЗУНы
ПК-4 Способен проектировать, разрабатывать и внедрять компоненты ИТ-инфраструктуры предприятия, обеспечивающие совершенствование и поддержку бизнес-процессов, в том числе с применением инновационных цифровых технологий	З. Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия У. Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н. Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия

3. Место дисциплины (модуля) в структуре образовательной программы

Принадлежность дисциплины - БЛОК 1 ДИСЦИПЛИНЫ (МОДУЛИ): Обязательная часть.

Дисциплины, использующие знания, умения, навыки, полученные при изучении данной: "Использование систем искусственного интеллекта для анализа бизнеса"

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 6 зач. ед., 216 часов.

Вид учебной работы	Количество часов
Контактная(аудиторная) работа	
Лекции	64

Практические (сем, лаб.) занятия	36
Самостоятельная работа, включая подготовку к экзаменам и зачетам	116
Всего часов	216

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

№ п/п	Раздел и тема дисциплины	Семестр	Лекции	Семинар Лаборат. Практич.	Самостоят. раб.	В интерактивной форме	Формы текущего контроля успеваемости
1	Общие сведения об аналитических информационных технологиях	41	4	0	10		Тест 1
2	Исследование зависимостей (Data Mining)	41	24	0	34		Задание 1. Исследование и преобразование данных. Задание 2. Решение задачи регрессии. Задание 3. Классификация. Задание 4. Решение задачи кластеризации, изучение кластеров, определение кластера объекта по атрибутам. Задание 5. Решение задачи ассоциации. Прогнозирование следующей ассоциации. Задание 6. Решение задачи прогнозирования, определение точности прогнозирования. Тест 2
3	Обработка текстов на естественном языке	42	36	36	72		Задание 7. Применение Regexp для выделения лексем.. Задание 8.

№ п/п	Раздел и тема дисциплины	Семестр	Лекции	Семинар Лаборат. Практич.	Самостоят. раб.	В интерактивной форме	Формы текущего контроля успеваемости
							Синтаксический разбор с применение стеммеров.. Задание 9. Выполнить параметризацию корпуса текстов в виде «мешка слов».. Задание 10. Настроить эмбединговую модель по корпусу текстов.. Задание 11. Настроить эмбединговую модель предложений по корпусу текстов.. Задание 12. Классификация корпуса текстов.. Задание 13. Кластеризация корпуса текстов.. Тест 3.. Тест 4.. Тест 5.. Тест 6.
	ИТОГО		64	36	116		

5.2. Лекционные занятия, их содержание

№ п/п	Наименование разделов и тем	Содержание
1	Общие сведения об аналитических информационных технологиях	Описываются традиционные аналитические технологии, применяемые в экономических информационных системах. Приводится обзор новых технологий: многомерный анализ данных, модели Data Mining. Даются основные этапы проектирования аналитических технологий по стандарту CRIPS.
2	Исследование зависимостей. Обзор	Описываются поиск зависимостей с помощью моделей Data Mining. Рассматривается технология определения моделей. Разбираются наиболее распространенные модели и области их применения
3	Преобразование и очистка данных	Преобразование структуры данных (множество таблиц – таблица фактов, таблицы измерений). Агрегирование (функции агрегирования). Перевод значений (замена кодов). Создание новых данных (вычисление показателей на основе имеющихся). Очистка данных . Снижение размерности.
4	Задачи	Задача классификации. Tree – дерево решений. Random Forest

№ п/п	Наименование разделов и тем	Содержание
	классификации и кластеризации	– случайный лес. kNN – метод k ближайших соседей. SVM – метод опорных векторов. Логистическая регрессия. Naive Bayes – наивная байесовская модель. AdaBoost – композиция алгоритмов обучения. Neural Network – нейронная сеть. Stochastic Gradient Descent – метод стохастического градиентного спуска. Матрица ошибок (Confusion Matrix) классификации. Показатели качества классификации. ROC-функция (ROC Analysis). Лифт-функция (Lift Curve). График калибровки вероятности (Calibration Plot). Сравнение моделей. Предсказание класса. Регрессионные модели. Кластеризация. Hierarchical Clustering – иерархическая кластеризация. k-Means – кластеризация методом k-средних. DBSCAN - Основанная на плотности пространственная кластеризация с выделением шума. Выбросы.
5	Задачи и модели поиска ассоциаций	Алгоритм взаимосвязей или ассоциативных правил (AssociationRules). Транзакции. Предметный набор. Ассоциативный правила. Характеристики модели.
6	Анализ временных рядов	Временные ряды . Преобразования рядов. Характеристики рядов. Модели рядов.
7	Методы разбора текстов.	Язык Regexp описания синтаксических единиц. Выделение лексем. Синтаксический разбор текстов. Формирование термов на основе лемм, грамем, стемм. Неоднозначность разбора. Программные средства синтаксического разбора. Меры близости термов.
8	Параметрическое описание текстов.	Описание текста в виде «мешка слов». Числовые характеристик термов. Эмбединговые модели термов и текстов. Онтологические модели и их применение для параметризации тестов.
9	Классификация текстов.	Задачи, решаемые на основе классификации текстов. Модели классификации текстов. Обучение моделей. Показатели качества классификаторов.
10	Кластеризация текстов.	Задачи, решаемые методами кластеризации текстов. Модели кластеризации. Показатели качества кластеризации.

5.3. Семинарские, практические, лабораторные занятия, их содержание

№ раздела и темы	Содержание и формы проведения
1	Изучение вариантов применения информационных технологий аналитики.. Обсуждение вариантов применения информационных аналитических технологий. Для выбранной предметной области рассматриваются варианты использования аналитических технологий для решения задач управления. Определение исходных данных для различных аналитических моделей. Выявление источников исходных данных. Обсуждение проблем преобразования исходных данных к виду удобному для анализа.
2	Знакомство с моделями и задачами Data Mining. Определение целесообразности решения задач исследования зависимостей для выбранных предметных областей: виды зависимостей и моделей определения зависимостей, проверка адекватности модели, применение моделей для

№ раздела и темы	Содержание и формы проведения
	решения аналитических задач.
2	Настройка параметров и интерпретация результатов моделей классификации. Описание моделей для решения задач классификации. Для каждой модели изучаются исходные данные для выявления зависимостей, настройка параметров модели, проверка адекватности, интерпретация результатов моделирования, применение модели для решения практических задач.
2	Настройка параметров и интерпретация результатов моделей регрессии. Описание моделей для решения задач регрессии. Для каждой модели изучаются исходные данные для выявления зависимостей, настройка параметров модели, проверка адекватности, интерпретация результатов моделирования, применение модели для решения практических задач.
2	Настройка параметров и интерпретация результатов моделей кластеризации. Описание моделей для решения задач кластеризации. Для каждой модели изучаются исходные данные для выявления зависимостей, настройка параметров модели, проверка адекватности, интерпретация результатов моделирования, применение модели для решения практических задач.
2	Настройка параметров и интерпретация результатов моделей поиска ассоциаций. Описание моделей для решения задач поиска ассоциаций. Для каждой модели изучаются исходные данные для выявления зависимостей, настройка параметров модели, проверка адекватности, интерпретация результатов моделирования, применение модели для решения практических задач.
2	Настройка параметров и интерпретация результатов моделей случайных рядов. Описание моделей для решения задач моделирования случайных рядов. Выделение трендов и сезонных колебаний, проверка адекватности, интерпретация результатов моделирования, применение модели для решения практических задач.
3	Синтаксический разбор текстов.. Применение регулярных выражений для выделения лексем. Использование программных средств синтаксического разбора. Вычисление близости термов по написанию.
3	Применение параметрического описания текстов.. Применение программ параметризации текстов в виде «мешка слов». Эмбединговые модели слов и текстов.
3	Применение классификаторов текстов.. Обучение классификаторов. Оценка качества классификации. Сравнение классификаторов.
3	Применение кластеризации текстов.. Настройка моделей кластеризации. Определение показателей качества кластеризации. Сравнение моделей по качеству кластеризации. Выделение стереотипов в кластерах.

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. Общие сведения об аналитических информационных технологиях	ПК-4	З.Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия	Тест 1	Доля правильных ответов (10)
2	2. Исследование зависимостей (Data Mining)	ПК-4	З.Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия У.Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н.Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия	Задание 1. Исследование и преобразование данных	Полнота и оптимальность преобразования (10)
3		ПК-4	З.Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия У.Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н.Владеть навыками проектирования, разработки и внедрения инновационных	Задание 2. Решение задачи регрессии	Правильность постановки задачи (15)

№ 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)
			цифровых компонент ИТ-инфраструктуры предприятия		
4		ПК-4	З.Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия У.Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н.Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия	Задание 3. Классификация	Обоснованность выбора (15)
5		ПК-4	З.Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия У.Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н.Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия	Задание 4. Решение задачи кластеризации, изучение кластеров, определение кластера объекта по атрибутам	Правильность постановки задачи (15)
6		ПК-4	З.Знать современные	Задание 5. Решение	Правильность

№ 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)
			тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия У. Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н. Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия	задачи ассоциации. Прогнозирование следующей ассоциации	постановки задачи (10)
7		ПК-4	3. Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия У. Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н. Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия	Задание 6. Решение задачи прогнозирования, определение точности прогнозирования	Правильность постановки задачи (15)
8		ПК-4	3. Знать современные тенденции развития и применения инновационных цифровых компонент	Тест 2	Доля правильных ответов (10)

№ 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)
			ИТ-инфраструктуры предприятия		
				Итого	100
9	3. Обработка текстов на естественном языке	ПК-4	З.Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия У.Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н.Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия	Задание 7. Применение Regexp для выделения лексем.	Корректность распознавания (10)
10		ПК-4	З.Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия У.Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н.Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия	Задание 8. Синтаксический разбор с применением стеммеров.	Корректность разбора (10)
11		ПК-4	З.Знать современные	Задание 9. Выполнить	Корректность

№ 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)
			тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия У. Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н. Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия	параметризацию корпуса текстов в виде «мешка слов».	программы (10)
12		ПК-4	З. Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия У. Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н. Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия	Задание 10. Настроить эмбединговую модель по корпусу текстов.	Корректность модели (10)
13		ПК-4	З. Знать современные тенденции развития и применения инновационных цифровых компонент	Задание 11. Настроить эмбединговую модель предложений по корпусу текстов.	Корректность программы (10)

№ 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)
			ИТ-инфраструктуры предприятия У. Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н. Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия		
14		ПК-4	З. Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия У. Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н. Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия	Задание 12. Классификация корпуса текстов.	Корректность моделей (15)
15		ПК-4	З. Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия У. Уметь проектировать,	Задание 13. Кластеризация корпуса текстов.	Корректность моделей (15)

№ 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)
			разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия Н. Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия		
16		ПК-4	З. Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия	Тест 3.	Доля правильных ответов (5)
17		ПК-4	З. Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия	Тест 4.	Доля правильных ответов (5)
18		ПК-4	З. Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия	Тест 5.	Доля правильных ответов (5)
19		ПК-4	З. Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия	Тест 6.	Доля правильных ответов (5)
				Итого	100

6.2. Промежуточный контроль (зачет, экзамен)

Рабочим учебным планом предусмотрен Зачет в семестре 41.

ВОПРОСЫ ДЛЯ ПРОВЕРКИ ЗНАНИЙ:

1-й вопрос билета (40 баллов), вид вопроса: Тест/проверка знаний. Критерий: Количество баллов, умноженное долю правильных ответов.

Компетенция: ПК-4 Способен проектировать, разрабатывать и внедрять компоненты ИТ-инфраструктуры предприятия, обеспечивающие совершенствование и поддержку бизнес-процессов, в том числе с применением инновационных цифровых технологий

Знание: Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия

1. Алгоритм взаимосвязей или ассоциативных правил: исходные данные, характеристики ассоциативных правил.
2. Задача кластеризации и ее применение.
3. Модель авторегрессии и интегрированного скользящего среднего нестационарного случайного ряда.
4. Модель кластеризации
5. Назовите этапы разработки аналитических технологий по стандарту CRISP. Определите работы, выполняемые на каждом этапе.
6. Общие сведения по анализу временных рядов. Детерминированная и случайная компоненты временного ряда.
7. Перечислите задачи исследования зависимостей, укажите исходные данные, результаты решения каждой задачи.
8. Перечислите классы программных продуктов, решающих аналитические задачи. Укажите особенности их применения.
9. Постановка задачи классификации.
10. Приведите общий вид модели аналитической обработки.
11. Применение дерева решений для задачи классификации.
12. Применение логистической регрессии для решения задачи классификации.
13. Применение нейронной сети для решения задачи классификации.
14. Регрессионная модель и ее решение.
15. Решение задачи классификации упрощенным алгоритмом Байеса.
16. Стационарные временные ряда, модель авторегрессии и скользящего среднего стационарного ряда.
17. Точность и эффективность классификации.

ТИПОВЫЕ ЗАДАНИЯ ДЛЯ ПРОВЕРКИ УМЕНИЙ:

2-й вопрос билета (30 баллов), вид вопроса: Задание на умение. Критерий: Корректность решения.

Компетенция: ПК-4 Способен проектировать, разрабатывать и внедрять компоненты ИТ-инфраструктуры предприятия, обеспечивающие совершенствование и поддержку бизнес-процессов, в том числе с применением инновационных цифровых технологий

Умение: Уметь проектировать, разрабатывать и внедрять инновационные цифровые компоненты ИТ-инфраструктуры предприятия

Задача № 1. Для предложенной предметной обрасти определить набор кубов. Описать структуру каждого куба: перечислить показатели с определением для каждого функцию агрегирования, перечислить измерения описать каждое измерение: вид применяемых шкал, иерархическую структуру. Определить источники данных и запросы для извлечения данных измерений и значений показателей. Задать измерения и определить куб.

Задача № 2. Для предложенных данных построить модели классификации четырьмя разными алгоритмами (дерево решений, байесовская модель, логистическая регрессия, нейронная сеть), сравнить модели на основе числовых характеристик эффективности классификации. Выполнить прогнозирование класса для предложенных исходных данных.

Задача № 3. Для предложенных данных построить модель ассоциаций изучить характеристики разных предметных наборов. Предложить способы применения результатов решения задачи ассоциации

Задача № 4. Для предложенных данных построить модель кластеризации. Исследовать и определить каждый полученный кластер. Применить кластеризацию для выявления исключений. Выполнить прогнозирование кластера по исходным данным.

Задача № 5. Для предложенных данных построить модель прогнозирования. Изучить дерево решений прогноза. Оценить степень разброса прогноза.

ТИПОВЫЕ ЗАДАНИЯ ДЛЯ ПРОВЕРКИ НАВЫКОВ:

3-й вопрос билета (30 баллов), вид вопроса: Задание на навыки. Критерий: Эффективность решений.

Компетенция: ПК-4 Способен проектировать, разрабатывать и внедрять компоненты ИТ-инфраструктуры предприятия, обеспечивающие совершенствование и поддержку бизнес-процессов, в том числе с применением инновационных цифровых технологий

Навык: Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия

Задание № 1. Для выбранной предметной области определить систему показателей, признаков (факторов) и зависимости показателей. Определить источники (структуры) данных для анализа. Разработать модели поиска закономерностей. Объединить разработанные компоненты в единую технологию.

ОБРАЗЕЦ БИЛЕТА

Министерство науки и высшего образования
Российской Федерации
Федеральное государственное бюджетное
образовательное учреждение
высшего образования
**«БАЙКАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ»
(ФГБОУ ВО «БГУ»)**

Направление - 38.03.05 Бизнес-
информатика
Профиль - Автоматизация и цифровая
трансформация бизнеса
Кафедра математических методов и
цифровых технологий
Дисциплина - Технологии машинного
обучения и искусственного интеллекта в
бизнесе

БИЛЕТ № 1

1. Тест (40 баллов).
2. Для предложенных данных построить модели классификации четырьмя разными алгоритмами (дерево решений, байесовская модель, логистическая регрессия, нейронная сеть), сравнить модели на основе числовых характеристик эффективности классификации. Выполнить прогнозирование класса для предложенных исходных данных. (30 баллов).
3. Для выбранной предметной области определить систему показателей, признаков (факторов) и зависимости показателей. Определить источники (структуры) данных

для анализа. Разработать модели поиска закономерностей. Объединить разработанные компоненты в единую технологию. (30 баллов).

Составитель _____ В.В. Братищенко

Заведующий кафедрой _____ А.В. Родионов

Рабочим учебным планом предусмотрен Экзамен в семестре 42.

ВОПРОСЫ ДЛЯ ПРОВЕРКИ ЗНАНИЙ:

1-й вопрос билета (30 баллов), вид вопроса: Тест/проверка знаний. Критерий: Количество баллов, умноженное долю правильных ответов.

Компетенция: ПК-4 Способен проектировать, разрабатывать и внедрять компоненты ИТ-инфраструктуры предприятия, обеспечивающие совершенствование и поддержку бизнес-процессов, в том числе с применением инновационных цифровых технологий

Знание: Знать современные тенденции развития и применения инновационных цифровых компонент ИТ-инфраструктуры предприятия

1. Выделение лексем. Синтаксический разбор текстов. Формирование термов на основе лемм, грамем, стемм. Неоднозначность разбора.
2. Задачи, решаемые методами кластеризации текстов.
3. Задачи, решаемые на основе классификации текстов.
4. Меры близости термов.
5. Модели классификации текстов.
6. Модели кластеризации. Показатели качества кластеризации.
7. Обучение моделей. Показатели качества классификаторов.
8. Онтологические модели и их применение для параметризации тестов.
9. Описание текста в виде «мешка слов». Числовые характеристик термов.
10. Программные средства синтаксического разбора.
11. Эмбединговые модели термов и текстов.
12. Язык Regexp описания синтаксических единиц.

ТИПОВЫЕ ЗАДАНИЯ ДЛЯ ПРОВЕРКИ НАВЫКОВ:

2-й вопрос билета (30 баллов), вид вопроса: Задание на навыки. Критерий: Корректность решения.

Компетенция: ПК-4 Способен проектировать, разрабатывать и внедрять компоненты ИТ-инфраструктуры предприятия, обеспечивающие совершенствование и поддержку бизнес-процессов, в том числе с применением инновационных цифровых технологий

Навык: Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия

Задание № 1. Выполнить кластеризацию корпуса текстов. Вычислить показатели качества кластеризации.

Задание № 2. Выполнить обучение классификатора. Определить показатели качества классификации.

Задание № 3. Выполнить параметризацию корпуса текстов в виде «мешка слов». Выбрать 5 наиболее близких текстов.

Задание № 4. Выполнить параметризацию текста с использованием онтологической модели. Определить 5 наиболее близких текстов.

Задание № 5. Настроить эмбединговую модель по корпусу текстов. Определить 5 наиболее близких термов.

Задание № 6. Настроить эмбединговую модель предложений по корпусу текстов. Определить 5 наиболее близких предложений.

Задание № 7. Применение Regexp для выделения лексем текста.

Задание № 8. Синтаксический разбор с применение стеммеров.

3-й вопрос билета (40 баллов), вид вопроса: Задание на навыки. Критерий: Эффективность решений.

Компетенция: ПК-4 Способен проектировать, разрабатывать и внедрять компоненты ИТ-инфраструктуры предприятия, обеспечивающие совершенствование и поддержку бизнес-процессов, в том числе с применением инновационных цифровых технологий

Навык: Владеть навыками проектирования, разработки и внедрения инновационных цифровых компонент ИТ-инфраструктуры предприятия

Задание № 1. Предложить использование обработки текстов на естественном языке для определенной предметной области.

Задание № 2. Предложить использование синтаксического разбора и параметризации текстов на естественном языке для определенной предметной области.

ОБРАЗЕЦ БИЛЕТА

Министерство науки и высшего образования
Российской Федерации
Федеральное государственное бюджетное
образовательное учреждение
высшего образования
**«БАЙКАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ»
(ФГБОУ ВО «БГУ»)**

Направление - 38.03.05 Бизнес-
информатика
Профиль - Автоматизация и цифровая
трансформация бизнеса
Кафедра математических методов и
цифровых технологий
Дисциплина - Технологии машинного
обучения и искусственного интеллекта в
бизнесе

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 1

1. Тест (30 баллов).

2. Выполнить параметризацию текста с использование онтологической модели.

Определить 5 наиболее близких текстов. (30 баллов).

3. Предложить использование синтаксического разбора и параметризации текстов на естественном языке для определенной предметной области. (40 баллов).

Составитель _____ В.В. Братищенко

Заведующий кафедрой _____ А.В. Родионов

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

а) основная литература:

1. Чубукова И. А. Data Mining. учеб. пособие/ И. А. Чубукова.- М.: БИНОМ. Лаборатория знаний, 2006.-382 с.

2. Паклин Н. Б. Николай Борисович, Орешков В. И. Вячеслав Игоревич Бизнес-аналитика: от данных к знаниям. учеб. пособие. 2-е изд., испр./ Н. Паклин, В. Орешков.- СПб.: Питер, 2013.-701 с.
3. Братищенко В.В. Информационные технологии в бизнес-аналитике.- Иркутск: Изд-во БГУ, 2019.- 127 с.
4. [Батура, Т. В. Математическая лингвистика и автоматическая обработка текстов на естественном языке : учебное пособие / Т. В. Батура. — Новосибирск : Новосибирский государственный университет, 2016. — 166 с. — ISBN 978-5-4437-0548-4. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : \[сайт\]. — URL: <https://www.iprbookshop.ru/93489.html> \(дата обращения: 27.05.2024\). — Режим доступа: для авторизир. пользователей](#)
5. [Ганегедара, Т. Обработка естественного языка с TensorFlow / Т. Ганегедара ; перевод В. С. Яценков. — Москва : ДМК Пресс, 2020. — 382 с. — ISBN 978-5-97060-756-5. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : \[сайт\]. — URL: <https://www.iprbookshop.ru/130336.html>](#)
6. [Йылдырым, С. Осваиваем архитектуру Transformer. Разработка современных моделей с помощью передовых методов обработки естественного языка / С. Йылдырым, М. Асгари-Ченаглу. — Москва : ДМК Пресс, 2022. — 319 с. — ISBN 978-5-93700-106-1. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : \[сайт\]. — URL: <https://www.iprbookshop.ru/125183.html> \(дата обращения: 27.05.2024\). — Режим доступа: для авторизир. пользователей](#)
7. [Нестеров С.А. Интеллектуальный анализ данных средствами MS SQL Server 2008 \[Электронный ресурс\] / С.А. Нестеров. — Электрон. текстовые данные. — М. : Интернет-Университет Информационных Технологий \(ИНТУИТ\), 2016. — 303 с. — 2227-8397. — Режим доступа: <http://www.iprbookshop.ru/62813.html>](#)

б) дополнительная литература:

1. [Маккинли, Уэс Python и анализ данных / Уэс Маккинли ; перевод А. Слинкина. — 2-е изд. — Саратов : Профобразование, 2019. — 482 с. — ISBN 978-5-4488-0046-7. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : \[сайт\]. — URL: <http://www.iprbookshop.ru/88752.html> \(дата обращения: 29.01.2020\)](#)

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
- Gensim. Documentation, адрес доступа: https://radimrehurek.com/gensim/auto_examples/index.html#core-tutorials-new-users-start-here.
- 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 программирования и баз данных.

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;
- коллоквиум как форма контроля освоения теоретического содержания дисциплин: (в часы консультаций, предусмотренные учебным планом);
- прием и разбор домашних заданий (в часы практических занятий);
- прием и защита лабораторных работ (во время проведения занятий);
- выполнение курсовых работ в рамках дисциплин (руководство, консультирование и защита курсовых работ в часы, предусмотренные учебным планом) и др.

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:

- Deductor-academic,
- Python,
- Scilab,

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,
- Мультимедийный класс,
- Компьютерный класс,
- Наборы демонстрационного оборудования и учебно-наглядных пособий