

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ИНСТИТУТ ЭКОНОМИКИ, ФИНАНСОВ И БИЗНЕСА

Утверждено:
на заседании кафедры
протокол от «17» февраля 2021 г. №5

Зав. кафедрой  /Р.Х.Бахитова

Согласовано:
Председатель УМК института

 /Л.Р. Абзалилова

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Интеллектуальный анализ данных»

Обязательная часть

Программа магистратуры

Направление подготовки
38.04.05 Бизнес-информатика

Направленность (профиль) подготовки
«Информационная бизнес-аналитика»

Квалификация
магистр

Разработчик (составитель)



Доцент, к. т. н., доц.

Прудников В.Б.

Для приема: 2021 г.

Уфа – 2021

Составитель / составители: Прудников В.Б.

Рабочая программа дисциплины утверждена на заседании кафедры цифровой экономики и коммуникации протокол от «17» февраля 2021 г. №5.

Дополнения и изменения, внесенные в рабочую программу дисциплины, утверждены на заседании кафедры _____,
протокол № ____ от « ____ » _____ 20 _ г.

Заведующий кафедрой _____ / _____ Ф.И.О/

Дополнения и изменения, внесенные в рабочую программу дисциплины, утверждены на заседании кафедры _____,
протокол № ____ от « ____ » _____ 20 _ г.

Заведующий кафедрой _____ / _____ Ф.И.О/

Список документов и материалов

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы	4
2. Место дисциплины в структуре ООП	5
3. Содержание рабочей программы (объем дисциплины, типы и виды учебных занятий, учебно-методическое обеспечение самостоятельной работы обучающихся)	5
4. Фонд оценочных средств по дисциплине	13
4.1. Перечень компетенций и индикаторов достижения компетенций с указанием соотнесенных с ними запланированных результатов обучения по дисциплине. Описание критериев и шкал оценивания результатов обучения по дисциплине.	13
4.2. Типовые контрольные задания или иные материалы, необходимые для оценивания результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине.	16
5. Учебно-методическое и информационное обеспечение дисциплины Ошибка! Закладка не определена.	
5.1. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины. Ошибка! Закладка не определена.	
5.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и программного обеспечения, необходимых для освоения дисциплины	Ошибка! Закладка не определена.
6. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине	13

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

По итогам освоения дисциплины обучающийся должен достичь следующих результатов обучения:

Категория (группа) компетенций ¹ (при наличии ОПК)	Формируемая компетенция (с указанием кода)	Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине
Проведение бизнес-анализа	ОПК-3 Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта	ОПК 3.1 Проводит анализ деятельности организации, факторов и условий, влияющих на деятельность организации, решает бизнес-задачи и задачи повышения эффективности функционирования организации на с использованием методов интеллектуального анализа данных. Оценивает результаты и последствия принятого решения	<p>Знать:</p> <ol style="list-style-type: none"> 1. классификацию типов наборов данных, методы сбора и подготовки исходных данных, технологии семплирования, методы машинного обучения решения задач в конкретных предметных областях 2. основные современные методы статистического анализа количественных и факторных данных: алгоритмы машинного обучения – байесовские классификаторы, методы опорных векторов, алгоритмы бустинга, методы деревьев решений и случайного леса 3. алгоритмы обработки входной информации, с применением современных инструментальных средств, методы предварительной обработки информации (метод главных компонент, кластерный анализ) для возможности применения продвинутых методов анализа данных <p>Уметь:</p> <ol style="list-style-type: none"> 1. формировать упорядоченные сводные массивы статистической информации в соответствии с целями исследования, готовить информационно-статистические материалы для выработки стратегических решений в области ИКТ 2. Уметь: подобрать в зависимости от качества и типа исходных данных соответствующие статистические методы и применить соответствующие методы визуализации данных, проверки гипотез, регрессионного и дисперсионного анализа для

¹ Указывается только для УК и ОПК (при наличии).

			<p><i>решения практических задач</i></p> <p><i>3. использовать современные средства статистического моделирования для проведения комплексного анализа данных и прогнозирования и интерпретировать получаемые результаты статистического анализа в виде аналитических текстов</i></p> <p><i>Владеть:</i></p> <p><i>1. навыками работы в прикладных программных средствах статистического моделирования, навыками сбора, анализа и обработки данных с помощью алгоритмов машинного обучения в средах статистического моделирования</i></p> <p><i>2. навыками работы с современными информационными средствами для получения и доставки информации, избавления от зашумления данных, опытом решения задач классификации с помощью алгоритмов машинного обучения</i></p> <p><i>3. методами устранения дисбаланса в формируемых выборках для обучения и тестирования модели, опытом формирования сводных массивов статистической информации, для анализа, архитектуры предприятий</i></p>
--	--	--	--

2. Место дисциплины в структуре ООП

Дисциплина «Интеллектуальный анализ данных» относится к обязательной части.

Дисциплина изучается на 1 курсе во 2 семестре и на 2 курсе в 3 семестре.

Цель: формирование у студентов теоретических знаний и практических навыков по сбору, статистической и интеллектуальной обработке и интерпретации информации.

3. Содержание рабочей программы (объем дисциплины, типы и виды учебных занятий, учебно-методическое обеспечение самостоятельной работы обучающихся)

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

СОДЕРЖАНИЕ РАБОЧЕЙ ПРОГРАММЫ

дисциплины ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ
на 2 семестр
очно-заочная форма обучения

Вид работы	Объем дисциплины
Общая трудоемкость дисциплины (ЗЕТ / часов)	4/144
Учебных часов на контактную работу с преподавателем:	
лекций	6
практических/ семинарских	12
лабораторных	
других (групповая, индивидуальная консультация и иные виды учебной деятельности, предусматривающие работу обучающихся с преподавателем) (ФКР)	1,2
Учебных часов на самостоятельную работу обучающихся (СР)	88,8
Учебных часов на подготовку к экзамену/зачету/дифференцированному зачету (Контроль)	36

Форма(ы) контроля:
Экзамен 2 семестр

№ п/п	Тема и содержание	Форма изучения материалов: лекции, практические занятия, семинарские занятия, лабораторные работы, самостоятельная работа и трудоемкость (в часах)					Основная и дополнительная литература, рекомендуемая студентам (номера из списка)	Задания по самостоятельной работе студентов	Форма текущего контроля успеваемости
		Всего	ЛК	ПР/СЕМ	ЛР	СРС			
1	2	3	4	5	6	7	8	9	10
Модуль 1. Методы и алгоритмы машинного обучения									
1.	Задачи, требующие решения на основе машинного обучения. Постановка задач обучения по прецедентам. Объекты и признаки. Типы шкал: бинарные, номинальные, порядковые, количественные. Типы задач: классификация, регрессия, прогнозирование. Основные понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль. Линейные модели регрессии и классификации. Метод наименьших квадратов. Полиномиальная регрессия. Примеры прикладных задач. Методика экспериментального исследования и сравнения алгоритмов на модельных и реальных данных. CRISP-DM — межотраслевой стандарт ведения проектов интеллектуального анализа данных	20	1		1	18	(1, 2)	Изучение теоретического материала, подготовка отчета по кейс-заданию 1	Проверка выполнения кейс-задания 1; курсовая работа
2.	Модели множественного выбора с неупорядоченными, бинарными и упорядоченными альтернативами. Условные логит-модель. Вложенные (nested) логит-модели. Оценивание логит-моделей. Мультиномиальная логит-модель. Примеры моделей с упорядоченными альтернативами. Упорядоченные пробит-модели.	20	1		1	18	(1, 2, 5)	Изучение теоретического материала, подготовка отчета по кейс-заданию 2	Проверка выполнения кейс-задания 2; курсовая работа
3.	Модели наивного байесовского классификатора.	21	1		2	18	(1, 2, 3)	Изучение	Проверка

	<p>Принцип максимума апостериорной вероятности. Теорема об оптимальности байесовского классификатора. Оценивание плотности распределения: три основных подхода. Непараметрическое оценивание плотности. Одномерный и многомерный случаи. Параметрическое оценивание плотности. Многомерное нормальное распределение, геометрическая интерпретация. Выборочные оценки параметров многомерного нормального распределения. Квадратичный дискриминант. Вид разделяющей поверхности. Подстановочный алгоритм, его недостатки и способы их устранения. Параметрический наивный байесовский классификатор. Смесь распределений.</p>							теоретического материала, подготовка отчета по кейс-заданию 3	выполнения кейс-задания 3; курсовая работа
4.	<p>Понятие логической закономерности. Параметрические семейства закономерностей: конъюнкции пороговых правил, синдромные правила, шары, гиперплоскости. Двухкритериальный отбор информативных закономерностей, парето-оптимальный фронт в (p,n)-пространстве. Алгоритмы дерева принятия решения. Решающее дерево. Жадная нисходящая стратегия «разделяй и властвуй». Алгоритм ID3. Недостатки жадной стратегии и способы их устранения. Проблема переобучения. Вывод критериев ветвления. Мера нечистоты (impurity) распределения. Энтропийный критерий, критерий Джини. Редукция решающих деревьев: <u>предредукция</u> и <u>постредукция</u>. Алгоритм C4.5. Алгоритмы случайного леса (Random Forest)</p>	20.5	0.5		2	18	(1, 2, 5)	Изучение теоретического материала, подготовка отчета по кейс-заданию 4	Проверка выполнения кейс-задания 4; курсовая работа
5.	<p>Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin). Случай линейной делимости и отсутствия линейной делимости. Связь с минимизацией регуляризованного эмпирического риска. Кусочно-линейная функция потерь. Задача квадратичного программирования и двойственная задача. Понятие опорных векторов. Рекомендации по</p>	21	1		2	18	(7, 8)	Изучение теоретического материала, подготовка отчета по кейс-заданию 5	Проверка выполнения кейс-задания 5; курсовая работа

	выбору константы C . Функция ядра (kernel functions), спрямляющее пространство, теорема Мерсера. Способы конструктивного построения ядер. Примеры ядер. SVM-регрессия.								
6.	Методы бустинга. Основные понятия: базовый алгоритм (алгоритмический оператор), корректирующая операция. Алгоритм AdaBoost. Экспоненциальная аппроксимация пороговой функции потерь. Процесс последовательного обучения базовых алгоритмов. Теорема о сходимости <u>бустинга</u> . Базовые алгоритмы в бустинге. Решающие пни. Варианты бустинга: <u>GentleBoost</u> , <u>LogitBoost</u> , <u>BrownBoost</u> , и другие. <u>Градиентный бустинг</u> . Стохастический градиентный бустинг. <u>Простое голосование</u> (комитет большинства). Алгоритм ComBoost. Идентификация нетипичных объектов (выбросов).	20.5	0.5		2	18	(5, 8)	Изучение теоретического материала, подготовка отчета по кейс-заданию 6	Проверка выполнения кейс-задания 6; курсовая работа
7.	Селекция алгоритмов машинного обучения. ROC-анализ. Чувствительность и специфичность. Ложноположительные и ложноотрицательные исходы. Площадь под кривой (Area under curve). Особенности применения ROC-кривых в медицинских исследованиях. Сравнение ROC-кривых между собой.	21	1		2	18	(6,8)	Изучение теоретического материала	Проверка выполнения кейс-заданий 1-6; курсовая работа
	Всего часов:	144	6		12	126			

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

СОДЕРЖАНИЕ РАБОЧЕЙ ПРОГРАММЫ

дисциплины ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ
на 3 семестр
очно-заочная форма обучения

Вид работы	Объем дисциплины
Общая трудоемкость дисциплины (ЗЕТ / часов)	5/180
Учебных часов на контактную работу с преподавателем:	
лекций	8
практических/ семинарских	24
лабораторных	
других (групповая, индивидуальная консультация и иные виды учебной деятельности, предусматривающие работу обучающихся с преподавателем) (ФКР)	3,2
Учебных часов на самостоятельную работу обучающихся (СР)	108,8
Учебных часов на подготовку к экзамену/зачету/дифференцированному зачету (Контроль)	36

Форма(ы) контроля:

 Экзамен 3 семестр

 курсовая работа 3 семестр

№ п/п	Тема и содержание	Форма изучения материалов: лекции, практические занятия, семинарские занятия, лабораторные работы, самостоятельная работа и трудоемкость (в часах)					Основная и дополнительная литература, рекомендуемая студентам (номера из списка)	Задания по самостоятельной работе студентов	Форма текущего контроля успеваемости
		Всего	ЛК	ПР/СЕМ	ЛР	СРС			
1	2	3	4	5	6	7	8	9	10
Модуль 2. Технологии сэмплирования									
1.	Процедуры подготовки данных для исследований. Упорядоченные и неупорядоченные данные. Транзакционные данные. Определение достаточного количества анализируемых объектов. «Грязные» данные. Избавление от «зашумления» данных. Проблемы очистки данных. Данные из единственного источника. Данные из различных источников. Определение способов трансформации потоков данных и правил отображения. Верификация. Трансформация. Обратная загрузка очищенных данных. Оптимизация признакового пространства. Трансформация признакового пространства.	23	2		9	49	(1, 2, 5)	Изучение теоретического материала, подготовка отчета по кейс-заданию 7	Проверка выполнения кейс-задания 7; курсовая работа
2.	Бустрап-анализ,	24	4		9	50	(3, 5)	Изучение	Проверка выполнения

	размножение выборок. Теоретические основы. Приложения бутстрапа. Аппроксимация стандартной ошибки выборочной оценки. Метод процентилей. Центрированный метод бутстрап-процентилей. Бутстрап t-критерий. Алгоритмы.							теоретического материала, подготовка отчета по кейс-заданию 8	кейс-задания 8; курсовая работа
3	Технологии сэмплирования. Проблема неполных данных. Восстановление пропуском. Метод ресамплинга. Цензурирование. Метод исключения некомплектных объектов. Методы с заполнением. Методы взвешивания. Методы, основанные на моделировании. Метод Бартлетта для заполнения пропусков.	21,8	2		6	49	(7, 8)	Изучение теоретического материала, подготовка отчета по кейс-заданию 9	Проверка выполнения кейс-задания 9; курсовая работа
	Экзамен	3.2							
	Контроль	36							
	Всего часов:	180	8		24	148			

4. Фонд оценочных средств по дисциплине

4.1. Перечень компетенций и индикаторов достижения компетенций с указанием соотнесенных с ними запланированных результатов обучения по дисциплине. Описание критериев и шкал оценивания результатов обучения по дисциплине.

Код и формулировка компетенции ОПК-3 Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Критерии оценивания результатов обучения			
		2 («Не удовлетворительно»)	3 («Удовлетворительно»)	4 («Хорошо»)	5 («Отлично»)
<p><i>ОПК 3.1</i> Проводит анализ деятельности организации, факторов и условий, влияющих на деятельность организации, решает бизнес-задачи и задачи повышения эффективности функционирования организации на с использованием методов интеллектуального анализа данных. Оценивает результаты и последствия</p>	<p><i>Знать:</i> 1. классификацию типов наборов данных, методы сбора и подготовки исходных данных, технологии семплирования, методы машинного обучения решения задач в конкретных предметных областях 2. основные современные методы статистического анализа количественных и факторных данных: алгоритмы машинного обучения – байесовские классификаторы, методы опорных векторов, алгоритмы бустинга, методы деревьев решений и случайного леса 3. алгоритмы обработки входной информации, с применением современных</p>	<p><i>Фрагментарные представления о типах данных, методах их сбора и технологий семплирования. Фрагментарные знания основных современных методов статистического анализа количественных и факторных данных. Фрагментарные знания алгоритмов обработки входной информации, и современных инструментальных средств их предварительной обработки</i></p>	<p><i>Неполные представления о типах данных, методах их сбора и технологий семплирования. Неполные знания основных современных методов статистического анализа количественных и факторных данных. Неполные знания алгоритмов обработки входной информации, и современных инструментальных средств их предварительной обработки</i></p>	<p><i>Сформированные, но содержащие отдельные пробелы представления о типах данных, методах их сбора и технологий семплирования. Сформированные, но содержащие отдельные пробелы знания основных современных методов статистического анализа количественных и факторных данных. Сформированные, но содержащие отдельные пробелы знания алгоритмов обработки входной информации, и современных инструментальных средств их предварительной обработки</i></p>	<p><i>Сформированные систематические представления о современных о типах данных, методах их сбора и технологий семплирования. Сформированные систематические знания основных современных методов статистического анализа количественных и факторных данных. Сформированные систематические знания основных современных методов статистического анализа количественных и факторных данных. Сформированные систематические знания алгоритмов обработки входной информации, и современных инструментальных средств их предварительной обработки</i></p>

<p>принятого решения</p>	<p>инструментальных средств, методы предварительной обработки информации (метод главных компонент, кластерный анализ) для возможности применения продвинутых методов анализа данных</p>				
	<p>Уметь:</p> <ol style="list-style-type: none"> 1. формировать упорядоченные сводные массивы статистической информации в соответствии с целями исследования, готовить информационно-статистические материалы для выработки стратегических решений в области ИКТ 2. подобрать в зависимости от качества и типа исходных данных соответствующие статистические методы и применить соответствующие методы визуализации данных, проверки гипотез, регрессионного и дисперсионного анализа для решения практических задач 3. использовать современные средства статистического моделирования для проведения комплексного анализа данных и 	<p>Фрагментарные умения формирования упорядоченных сводных массивов статистической информации. Фрагментарные умения подбора в зависимости от качества и типа исходных данных соответствующих статистических методов и умения применить для их анализа методы визуализации данных, проверки гипотез, проведения регрессионного и дисперсионного анализа. Фрагментарные умения использования современных средств статистического моделирования для проведения комплексного анализа данных, прогнозирования и интерпретации результатов.</p>	<p>В целом успешное, но не систематическое использование умений формирования упорядоченных сводных массивов статистической информации. В целом успешное, но не систематическое использование умений подбора в зависимости от качества и типа исходных данных соответствующих статистических методов и умения применить для их анализа методы визуализации данных, проверки гипотез, проведения регрессионного и дисперсионного анализа. В целом успешное, но не систематическое использование современных средств статистического моделирования для проведения комплексного анализа данных, прогнозирования и интерпретации результатов.</p>	<p>В целом успешное, но содержащее отдельные пробелы использование умений формирования упорядоченных сводных массивов статистической информации. В целом успешное, но содержащее отдельные пробелы использование умений подбора в зависимости от качества и типа исходных данных соответствующих статистических методов и умения применить для их анализа методы визуализации данных, проверки гипотез, проведения регрессионного и дисперсионного анализа. В целом успешное, но содержащее отдельные пробелы использование современных средств статистического моделирования для проведения комплексного анализа данных, прогнозирования и интерпретации результатов.</p>	<p>Сформированное системное умение формирования упорядоченных сводных массивов статистической информации. Сформированное системное умение подбора в зависимости от качества и типа исходных данных соответствующих статистических методов и умения применить для их анализа методы визуализации данных, проверки гипотез, проведения регрессионного и дисперсионного анализа. Сформированное системное умение использования современных средств статистического моделирования для проведения комплексного анализа данных, прогнозирования и интерпретации результатов.</p>

	<p><i>прогнозирования и интерпретировать получаемые результаты статистического анализа в виде аналитических текстов</i></p>				
	<p><i>Владеть:</i> 1. <i>навыками работы в прикладных программных средствах статистического моделирования, навыками сбора, анализа и обработки данных с помощью алгоритмов машинного обучения в средах статистического моделирования</i> 2. <i>навыками работы с современными информационными средствами для получения и доставки информации, избавления от зашумления данных, опытом решения задач классификации с помощью алгоритмов машинного обучения</i> 3. <i>методами устранения дисбаланса в формируемых выборках для обучения и тестирования модели, опытом формирования сводных массивов статистической информации, для анализа, архитектуры предприятий</i></p>	<p><i>Фрагментарное владение навыками работы в прикладных программных средствах статистического моделирования, навыками сбора, анализа и обработки данных. Фрагментарное владение навыками работы с современными информационными средствами для получения и доставки информации, избавления от зашумления данных, навыками применения алгоритмов машинного обучения. Фрагментарное владение методами устранения дисбаланса в формируемых выборках для обучения и тестирования моделей интеллектуального анализа данных.</i></p>	<p><i>В целом успешное, но не систематическое владение навыками работы в прикладных программных средствах статистического моделирования, навыками сбора, анализа и обработки данных. В целом успешное, но не систематическое владение навыками работы с современными информационными средствами для получения и доставки информации, избавления от зашумления данных, навыками применения алгоритмов машинного обучения. В целом успешное, но не систематическое владение методами устранения дисбаланса в формируемых выборках для обучения и тестирования моделей интеллектуального анализа данных.</i></p>	<p><i>В целом успешное, но содержащее отдельные пробелы владение навыками работы в прикладных программных средствах статистического моделирования, навыками сбора, анализа и обработки данных. В целом успешное, но содержащее отдельные пробелы владение навыками работы с современными информационными средствами для получения и доставки информации, избавления от зашумления данных, навыками применения алгоритмов машинного обучения. В целом успешное, но содержащее отдельные пробелы владение методами устранения дисбаланса в формируемых выборках для обучения и тестирования моделей интеллектуального анализа данных.</i></p>	<p><i>Сформированное системное владение навыками работы в прикладных программных средствах статистического моделирования, навыками сбора, анализа и обработки данных. Сформированное системное владение навыками работы с современными информационными средствами для получения и доставки информации, избавления от зашумления данных, навыками применения алгоритмов машинного обучения. Сформированное системное владение методами устранения дисбаланса в формируемых выборках для обучения и тестирования моделей интеллектуального анализа данных.</i></p>

4.2 2. Контрольные задания или иные материалы, необходимые для оценивания результатов обучения по дисциплине. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине.

Код и наименование индикатора достижения компетенции	Результаты обучения	Оценочные средства
<p><i>ОПК 3.1</i> Проводит анализ деятельности организации, факторов и условий, влияющих на деятельность организации, решает бизнес-задачи и задачи повышения эффективности функционирования организации на с использованием методов интеллектуального анализа данных. Оценивает результаты и последствия принятого решения..</p>	<p><i>Знать:</i></p> <ol style="list-style-type: none"> 1. классификацию типов наборов данных, методы сбора и подготовки исходных данных, технологии семплирования, методы машинного обучения решения задач в конкретных предметных областях 2. основные современные методы статистического анализа количественных и факторных данных: алгоритмы машинного обучения 3. алгоритмы обработки входной информации, с применением современных инструментальных средств, методы предварительной обработки информации 	<p>Выполнение индивидуальных кейс-заданий, курсовая работа</p>
	<p><i>Уметь:</i></p> <ol style="list-style-type: none"> 1. формировать упорядоченные сводные массивы статистической информации в соответствии с целями исследования, готовить информационно-статистические материалы для выработки стратегических решений в области ИКТ 2. подобрать в зависимости от качества и типа исходных данных соответствующие статистические методы и применить соответствующие методы визуализации данных, проверки гипотез, регрессионного и дисперсионного анализа 3. использовать современные средства статистического моделирования для проведения комплексного анализа данных и прогнозирования и интерпретировать получаемые результаты 	<p>Выполнение индивидуальных кейс-заданий; курсовая работа</p>
	<p><i>Владеть:</i></p> <ol style="list-style-type: none"> 1. навыками работы в прикладных программных средствах статистического моделирования, навыками сбора, анализа и обработки данных с помощью алгоритмов машинного обучения в средах статистического моделирования 2. навыками работы с современными информационными средствами для получения и доставки информации, избавления от зашумления данных, опытом решения задач классификации с помощью алгоритмов машинного обучения 3. методами устранения дисбаланса в формируемых выборках для обучения и тестирования модели, опытом формирования сводных массивов статистической информации 	<p>Выполнение индивидуальных кейс-заданий; курсовая работа</p>

Изучение дисциплины предполагает самостоятельное выполнение **девяти лабораторных работ (в виде решения кейс-заданий) по основным изученным темам.**

На оценку степени сформированности компетенции при выполнении заданий лабораторной работы влияет полнота и правильность выполнения всех необходимых заданий (пунктов) работы.

Формы и содержание текущего контроля:

- контроль посещаемости занятий;
- оценка степени подготовки к занятиям;
- выборочная проверка ответов на вопросы самоконтроля;
- оценка уровня развития компетенций в ходе анализа проблемных ситуаций и решения практических ситуаций.

Перед проведением итогового контроля преподаватель вычисляет процент полноты и правильности выполнения лабораторных работ (кейс-заданий).

Промежуточный и итоговый контроль по дисциплине «Интеллектуальный Анализ данных» проводится в виде экзамена. Оценка проводится по пятибалльной шкале (от 1 до 5).

Структура экзаменационного билета:

Экзаменационный билет содержит 3 вопроса.

Пример экзаменационного билета

Башкирский государственный университет	Направление подготовки 38.04.05 «Бизнес-информатика»
Институт экономики, финансов и бизнеса	Профиль «Информационная бизнес-аналитика»
Кафедра цифровой экономики и коммуникаций	Дисциплина «Интеллектуальный анализ данных»

Экзаменационный билет № 1

1. Общий вид и оценка гомпит-моделей
2. Алгоритмы дерева принятия решения для решения задач классификации
3. Метод экстремального бустинга

Зав. кафедрой

Р.Х. Бахитова

Критерии экзаменационной оценки (в баллах):

- **5 баллов** выставляется студенту, если студент дал полные, развернутые ответы на все теоретические вопросы билета, продемонстрировал знание функциональных возможностей, терминологии, основных элементов, умение применять теоретические знания при выполнении практических заданий. Студент без затруднений ответил на все дополнительные вопросы. Практическая часть работы выполнена полностью без неточностей и ошибок;

- **4 балла** выставляется студенту, если студент раскрыл в основном теоретические вопросы, однако допущены неточности в определении основных понятий. При ответе на дополнительные вопросы допущены небольшие неточности. При выполнении практической части работы допущены несущественные ошибки;

- 3 балла выставляется студенту, если при ответе на теоретические вопросы студентом допущено несколько существенных ошибок в толковании основных понятий. Логика и полнота ответа страдают заметными изъянами. Заметны пробелы в знании основных методов. Теоретические вопросы в целом изложены достаточно, но с пропусками материала. Имеются принципиальные ошибки в логике построения ответа на вопрос. Студент не решил задачу или при решении допущены грубые ошибки;

- 2 балла выставляется студенту, если ответ на теоретические вопросы свидетельствует о непонимании и крайне неполном знании основных понятий и методов. Обнаруживается отсутствие навыков применения теоретических знаний при выполнении практических заданий. Студент не смог ответить ни на один дополнительный вопрос.

Примерный перечень вопросов для подготовки к экзамену во 2 и 3 семестрах:

1. Понятие методов машинного обучения
2. Инструментальные средства по машинному обучению
3. Предварительная подготовка данных
4. Переход к размеченным данным в выборке
5. Применения метода главных компонент для снижения размерности признаков пространства.
6. Понятие моделей бинарного выбора.
7. Общий вид и оценка пробит-моделей.
8. Общий вид и оценка логит-моделей.
9. Общий вид и оценка гомпит-моделей
10. Оценивание моделей бинарного выбора методом максимального правдоподобия.
11. Проверка адекватности построенной модели бинарного выбора. Показатели качества оценки.
12. Селекция моделей на основе информационных критериев.
13. Критерий Хосмера-Лемешова.
14. Практический анализ модели бинарного выбора на основе маргинальных эффектов.
15. Модели множественного выбора с неупорядоченными альтернативами.
16. Практический анализ модели множественного выбора с неупорядоченными альтернативами на основе маргинальных эффектов.
17. Модели множественного выбора с упорядоченными альтернативами.
18. Практический анализ модели множественного выбора с упорядоченными альтернативами на основе маргинальных эффектов. Свойства маргинальных эффектов для модели множественного выбора с упорядоченными альтернативами
19. Проведение Рос-анализа для бинарных моделей.
20. Расчет критериев специфичности и чувствительности модели.
21. Определение показателя AUC и коэффициента Джини для моделей.
22. Модель наивного байесовского классификатора.
23. Оценка параметров для наивных байесовских моделей методом максимального правдоподобия.
24. Построение классификатора по вероятностной модели.
25. Апостериорное правило принятия решения (MAP).
26. Алгоритмы дерева принятия решения для решения задач классификации
27. Алгоритмы случайного леса для решения задач классификации
28. Ограничения для применения алгоритмов случайного леса
29. Признак и значения порога, по которому происходит оптимальное по заданному критерию разбиение в алгоритме randomforest.
30. Ансамбль алгоритмов randomforest в композиции методом простого голосования.
31. Ансамбль алгоритмов randomforest в композиции методом взвешенного голосования.
32. Метод *Out-Of-Bag* (OOB) Бреймана для алгоритмов randomforest
33. Классификации данных методом опорных векторов.
34. Разделение множества векторов на два класса с помощью линейной решающей функции.
35. Определение расстояния между граничными гиперплоскостями в методе опорных векторов.
36. Алгоритм линейной классификации Вапника.
37. Метод релевантных векторов.
38. Способ адаптации машины опорных векторов в случае линейно неразделимой выборки.
39. Понятие ядра и его свойства.
40. Ядра и спрямляющие пространства (виды).

41. Недостатки и ограничения применения метода опорных векторов.
42. Метод градиентного бустинга.
43. Алгоритм Adaboost.
44. Алгоритм Logitboost.
45. Метод экстремального бустинга.
46. Алгоритм, удаляющий определенное число образцов мажоритарного класса (undersampling).
47. Алгоритм, удаляющий определенное число образцов мажоритарного класса (oversampling).
48. Избавление от «зашумления» данных с помощью поиска связей Томека.
49. Избавление от «зашумления» данных с помощью правила сосредоточенного ближайшего соседа.
50. Генерирование псевдовыборок произвольного размера с помощью датчика псевдослучайных чисел.
51. Применение метода рецентрирования для нахождения бустрап-распределения.
52. Ресемплинг как метод бустрап-анализа.
53. Метод предискретизации при проведении бустрап-анализа.
54. Гладки и параметрический бустрапинг.
55. Скользящий блочный бустрапинг.
56. Алгоритм сэмплирования SMOTE.
57. Алгоритм сэмплирования ASMO.

Оценка, полученная при ответе на экзаменационные билеты, согласовывается с оценкой за выполнение лабораторных работ. Высчитывается средний балл (оценка), который округляется до ближайшего целого значения. В случае, если за ответ на экзамене и за текущий контроль (выполнение лабораторных работ) обучающийся получает разные оценки, а для округления среднего балла в большую сторону нет полных оснований (например, за лабораторную работу получено 4 балла, а за ответ на экзамене – 3 балла, средний балл составляет – 3,5), после ответа на вопросы билета, преподаватель может задать дополнительные вопросы, по результатам ответа на которые принимается окончательное решение об оценке.

Код и наименования индикатора достижения компетенции	Результаты	Оценка (по пятибалльной шкале)	Оценка
<i>ОПК 4.1 Осуществляет эффективную коммуникацию и получение обратной связи в ходе взаимодействия с клиентами и партнерами с помощью возможностей цифровой среды.</i>	Интегральная оценка полноты и правильности выполнения лабораторных работ	1-2	«неудовлетворительно»
		3	«удовлетворительно»
		4	«хорошо»
		5	«отлично»
	Оценка полноты и правильности ответа на вопросы экзаменационного билета	1-2	«неудовлетворительно»
		3	«удовлетворительно»
		4	«хорошо»
		5	«отлично»

Окончательная оценка вклада дисциплины «Интеллектуальный анализ данных» в формирование каждой компетенции проводится на основании суммы балла (оценки) полноты и правильности:

- 1) выполнения лабораторных работ,
- 2) ответов на соответствующий вопрос экзаменационного билета.

Ниже приводится последовательность (этапы) решения кейс-заданий лабораторных работ, а также критерии оценивания.

Оценочные материалы для кейс-заданий (лабораторных работ)

Кейс-задание 1.

Раздел (тема) дисциплины: **Модели множественного выбора**

Задание выполняется на лабораторной работе №1.

Построение классификатора на основе моделей бинарной регрессии.

1. Провести предварительный анализ исходных данных. Исключить аномальные наблюдения (если такие есть), заполнить пропуски (если они имеются). Провести корреляционный анализ независимых переменных, исключив переменные, значительно коррелирующие с другими переменными ($>0,9$).

2. Построить статистически значимую модель бинарной регрессии, оценив параметры методом максимального правдоподобия, применяя метод пошагового исключения, в которой все переменные будут статистически значимы. Подобрать функцию распределения, описывающую вероятность положительной альтернативы (например, выживет пациент или умрет) между нормальным распределением (пробит), логистическим (логит) и экстремальным (гомпит) на основе минимума информационных критериев.

3. Проверить качество отобранной модели, подтвердив его значениями коэффициентов R^2 МакФаддена, тестом отношения правдоподобия (LR-тестом), результатами теста Хосмера-Лемешоу и любым тестом на нормальность распределения остатков (например, Колмогорова-Смирнова или Бера-Жарка).

4. Рассчитать маргинальные эффекты и провести интерпретацию коэффициентов модели.

5. Оформить отчет о выполнении задания с приведением условия задачи, результатов решения и выводов.

Результатом выполнения кейс-задания является отчет по лабораторной работе № 1. К отчету предъявляются следующие требования:

1. Четкое формулирование поставленной цели исследования
2. Формулирование задач, решение которых необходимо для достижения поставленной цели.
3. Описание в виде пунктов, тех действий, которые требуются для решения поставленных задач. Все рисунки и таблицы последовательно нумеруются и описываются. Каждый пункт решения поставленных задач сопровождается анализом принятого решения. При проведении статистических тестов, обязательно выписывается нулевая и альтернативная гипотеза, формулируется принятие решения на обосновано выбранном уровне значимости, указывается критическая область отказа от нулевой гипотезы в пользу альтернативной.
4. В заключении выписывается отобранная адекватная модель с оцененными коэффициентами с указанием под оценками коэффициентов значений t -статистик в скобках или стандартных ошибок коэффициентов. Также приводятся значения маргинальных эффектов и дается их интерпретация.

Кейс-задание 2

Раздел (тема) дисциплины: **Модели множественного выбора**

Задание выполняется на лабораторной работе № 2.

Построение классификатора на основе моделей с упорядоченными альтернативами.

1. Провести предварительный анализ исходных данных. Исключить аномальные наблюдения (если такие есть), заполнить пропуски (если они имеются). Зависимая переменная должна иметь несколько упорядоченных альтернатив.
2. Построить статистически значимую модель упорядоченной регрессии, оценив параметры модели методом максимального правдоподобия, применяя метод пошагового исключения, в которой все переменные будут статистически значимы. Проверить статистическую значимость границ латентой переменной. Подобрать функцию распределения, описывающую вероятность каждой альтернативы между

нормальным распределением (пробит), логистическим (логит) и экстремальным (гомпит) на основе минимума информационных критериев.

3. Проверить качество отобранной модели, подтвердив его значениями коэффициентов R^2 МакФаддена, тестом отношения правдоподобия (LR-тестом).
4. Рассчитать маргинальные эффекты для каждой упорядоченной альтернативы и провести интерпретацию коэффициентов модели. Проверить равенство нулю суммы маргинальных эффектов для каждой альтернативы по каждой переменной.
5. Оформить отчет о выполнении задания с приведением условия задачи, результатов решения и выводов.

Результатом выполнения кейс-задания является отчет по лабораторной работе № 2.

К отчету предъявляются следующие требования:

1. Четкое формулирование поставленной цели исследования.
2. Формулирование задач, решение которых необходимо для достижения поставленной цели.
3. Описание в виде пунктов, тех действий, которые требуются для решения поставленных задач. Все рисунки и таблицы последовательно нумеруются и описываются. Каждый пункт решения поставленных задач сопровождается анализом принятого решения. При проведении статистических тестов, обязательно выписывается нулевая и альтернативная гипотеза, формулируется принятие решения на обоснованном выбранном уровне значимости, указывается критическая область отказа от нулевой гипотезы в пользу альтернативной.
4. В заключении выписывается отобранная адекватная модель с оцененными коэффициентами в виде системы с указанием под оценками коэффициентов и границ латентной переменной значений t -статистик в скобках или стандартных ошибок коэффициентов, Также приводятся значения маргинальных эффектов для каждой упорядоченной альтернативы и дается их интерпретация.

Кейс-задание 3. Фильтрация на основе наивного байесовского классификатора

Раздел (тема) дисциплины: **Байесовская классификация**

Задание выполняется на лабораторной работе № 3.

Провести классификацию IT-приложений для решения подобных задач согласно наивному байесовскому классификатору, предварительно проверив валидность исходных данных. Сформировать правила классификации, сочетаемые с классом априорных вероятностей по данным тестовой выборки. Пересчитать на основе оцененной модели классификации апостериорные вероятности для полученных ранее правил на основе данных обучающей выборки. Вывести матрицу неточности для каждой зависимой переменной. Проверить гипотезу о корректности сформированной модели байесовской классификации. Построить график проведения байесовской классификации на обучающей выборке. Построить график предсказания, апостериорной вероятности. Построить сценарный прогноз отнесения IT-приложений к определенным классам, используя найденную модель.

Результатом выполнения задания является отчет по лабораторной работе № 2. К отчету предъявляются следующие требования:

1. Формулирование задачи, решение которых необходимо в ходе выполнения лабораторной работы (например, провести классификацию программного обеспечения автоматизирования технологических процессов от уровня проектирования до внедрения).
2. Описание данных для тестовой и обучающей выборок, заключение о валидности данных.
3. Четко сформулированные выводы по результатам выполнения лабораторной работы.
4. Оценить качество классификатора на тестовой выборке. Все графики и таблицы должны иметь сквозную нумерацию.

Кейс-задание 4. Алгоритмы дерева принятия решения, алгоритмы случайного леса

Раздел (тема) дисциплины: **алгоритмы случайного леса**

Задание выполняется на лабораторной работе № 4

Провести классификацию объектов, для решения используя классификатор деревьев решения и алгоритм randomforest, предварительно проверив валидность исходных данных. Выбрать признак и значения порога, по которому происходит оптимальное по заданному критерию разбиение в алгоритме randomforest. Задать максимальное число объектов в вершине-листа дерева, для определения критерия останова алгоритма. Построить алгоритм на размеченных данных.

Результатом выполнения задания является отчет по лабораторной работе № 4. К отчету предъявляются следующие требования:

1. Формулирование задачи, решение которых необходимо в ходе выполнения лабораторной работы (например, провести классификацию программного обеспечения автоматизирования технологических процессов от уровня проектирования до внедрения).
2. Описание данных для тестовой и обучающей выборок, заключение о валидности данных.
3. Обосновать выбор признака и значения порога, по которому происходит оптимальное по заданному критерию разбиение в алгоритме randomforest
4. Четко сформулированные выводы по результатам выполнения лабораторной работы.
5. Оценить качество классификатора на тестовой выборке. Все графики и таблицы должны иметь сквозную нумерацию.

Кейс-задание 5. Метод опорных векторов

Раздел (тема) дисциплины:**Метод опорных векторов.**

Задание выполняется на лабораторной работе № 5.

Провести классификацию объектов, для решения используя классификатор метод опорных векторов, предварительно проверив валидность исходных данных. Проверить являются ли данные линейно разделимыми, и если да, то следует по обучающей выборке построить линейную решающую функцию, предварительно разметив данные. Применить алгоритм Вапника. В случае отсутствия линейного разделения, выбрать ядро классификатора: полиномиальное, сигмоидное, радиальное и радиальное Гауссово.

Результатом выполнения задания является отчет по лабораторной работе № 5. К отчету предъявляются следующие требования:

1. Формулирование задачи, решение которых необходимо в ходе выполнения лабораторной работы (например, провести классификацию программного обеспечения автоматизирования технологических процессов от уровня проектирования до внедрения).
2. Описание данных для тестовой и обучающей выборок, заключение о валидности данных.
3. Проверить являются ли данные линейно разделимыми, и если да, то следует по обучающей выборке построить линейную решающую функцию, предварительно разметив данные
4. В случае отсутствия линейного разделения, выбрать ядро классификатора: полиномиальное, сигмоидное, радиальное и радиальное Гауссово.
5. Оценить качество классификатора на тестовой выборке. Все графики и таблицы должны иметь сквозную нумерацию.

Кейс-задание 6. Метод бустинга

Раздел (тема) дисциплины:**Бустинг.**

Задание выполняется на лабораторной работе № 6.

Провести классификацию объектов, для решения используя классификатор метод бустинга, предварительно проверив валидность исходных данных. Применить алгоритм градиентного бустинга, используя правило жадного наращивания. Найти пары наиболее оптимальных параметров, где под оптимизацией следует понимать принцип явной максимизации отступов, минимизировать функционал ошибки. Провести классификацию объекта, используя алгоритм бустинга AdaBoost с экспоненциальной функцией потерь.

Результатом выполнения задания является отчет по лабораторной работе № 6. К отчету предъявляются следующие требования:

1. Формулирование задачи, решение которых необходимо в ходе выполнения лабораторной работы (например, провести классификацию программного обеспечения автоматизирования технологических процессов от уровня проектирования до внедрения).
2. Описание данных для тестовой и обучающей выборок, заключение о валидности данных.
3. Применить алгоритм градиентного бустинга, используя правило жадного наращивания.
4. Найти пары наиболее оптимальных параметров, где под оптимизацией следует понимать принцип явной максимизации отступов, минимизировать функционал ошибки.
5. Применить в качестве альтернативного алгоритма – алгоритм AdaBoost или LogitBoost.
6. Провести классификацию объектов, для решения используя классификатор - метод экстремального бустинга с корректно подобранными весами.
7. Оценить качество классификатора на тестовой выборке. Все графики и таблицы должны иметь сквозную нумерацию.

Кейс-задание 7. Методы подготовки данных для исследования

Раздел (тема) дисциплины **подготовка данных, избавление от «зашумления».**

Задание выполняется на лабораторной работе № 7

Определить в исходной информации количество случаев мажоритарного и миноритарного класса. Применить алгоритмы сэмплирования *undersampling* и *oversampling*. В соответствии подходом связи Томека, все без исключения мажоритарные наблюдения удалить с комплекта сведений, то есть убрать «зашумление». Провести такую же процедуру, но используя правило сосредоточенного ближайшего соседа. Выбрать лучший алгоритм избавления от «зашумления» данных.

Результатом выполнения задания является отчет по лабораторной работе № 7.

К отчету предъявляются следующие требования:

1. Описать исходные данные.
2. Избавиться от «зашумления» в данных с помощью алгоритма связей Томека.
3. Избавиться от «зашумления» в данных с помощью правило сосредоточенного ближайшего соседа.
4. Выбрать лучший алгоритм избавления от «зашумления» данных.

Кейс-задание 8.

Раздел (тема) дисциплины: **Использование метода рецентрирования для нахождения бутстрап-распределения**

Задание выполняется на лабораторной работе № 8.

1. Присвоить значение индекса, соответствующее попаданию в отрезки $[0; 1/n]; [1/n; 2/n]; \dots; [1 - 1/n; 1]$.
2. Рассчитать среднюю выборочную на выборках, используя бутстрапинг, в частности, передискретизацию, чтобы получить распределение \bar{X} .
 - 2.1 Сначала провести повторную выборку данных для получения передискретизованной выборки.
 - 2.2 Вычислить среднюю этой бутстрап выборки.
 - 2.3 Повторить этот процесс 100 раз, чтобы получить следующие бутстрап выборки вычислить 100 точек бутстрап оценок.
3. Для полученного эмпирического бутстрап-распределения получить доверительный интервал оценки средней.

Результатом выполнения кейс-задания является отчет по лабораторной работе № 8.

К отчету предъявляются следующие требования:

Четкое формулирование поставленной цели исследования. Формулирование задач, решение которых необходимо для достижения поставленной цели.

Кейс-задание 9. Технологии сэмплирования

Раздел (тема) дисциплины **технологии сэмплинга**

Задание выполняется на лабораторной работе № 9

Определить в исходной информации количество случаев мажоритарного и миноритарного класса. Применить алгоритмы одностороннего сэмплирования. Применить способ повышения количества образцов миноритарного класса – метод SMOTE (Synthetic Minority Oversampling Technique). Выбрать лучший алгоритм сэмплирования. Применить метод адаптивного искусственного увеличения числа примеров миноритарного класса ASMO (Adaptive Synthetic Minority Oversampling). Выбрать лучший алгоритм сэмплирования.

Результатом выполнения задания является отчет по лабораторной работе № 9. К отчету предъявляются следующие требования:

1. Описать исходные данные.
2. Применить алгоритм сэмплирования SMOTE.
3. Применить алгоритм сэмплирования ASMO.
4. Выбрать лучший алгоритм избавления от «зашумления» данных

Критерии оценки выполнения кейс-заданий:

Критерии оценивания	Количество баллов
Лабораторная работа подготовлена к выполнению, обучаемый знает цель лабораторной работы; задания решены без ошибок с первого раза, правильно выбраны решения заданий; правильно выполнены расчёты, обучающийся понимает, что они значат; полно даны ответы на контрольные вопросы; отчёт оформлен аккуратно, сделаны выводы.	5
Лабораторная работа подготовлена к выполнению, обучаемый знает цель лабораторной работы; задания решены с ошибками, потребовалась дополнительная помощь преподавателя, правильно выбраны методики решения заданий; расчёты выполнены с консультацией преподавателя; полно даны ответы на контрольные вопросы; отчёт оформлен аккуратно, сделаны выводы	3-4
Лабораторная работа подготовлена к выполнению, обучаемый знает цель лабораторной работы; задания выполнены с ошибками, потребовалась дополнительная помощь преподавателя, правильно выбраны методики решения заданий; с ошибками выполнены расчёты, даже с консультацией преподавателя или обучающийся не может объяснить, как выполнялись расчеты; даны ответы на контрольные вопросы	2
Лабораторная работа подготовлена к выполнению, обучаемый не знает цель лабораторной работы; задачи решены с ошибками, потребовалась дополнительная помощь преподавателя, неверно выбраны методы решения задач; не выполнены расчёты; не даны ответы на устные контрольные вопросы; отчёт оформлен небрежно, выводы не сделаны	менее 2

Курсовая работа являются одной из форм отчётности по итогам курса, позволяет структурировать знания обучаемых.

Требования к содержанию и оформлению курсовой работы

Выбор темы. Очень важно правильно выбрать тему. Выбор темы не должен носить формальный характер, а иметь практическое и теоретическое обоснование, иметь интерес для студента.

Примерные темы курсовых работ по дисциплине «Интеллектуальный анализ данных»

1. Статистический анализ качества и уровня жизни семей в современной России
2. Статистический анализ жилищных условий различных типов домохозяйств и оценка эффективности жилищной политики
3. Статистический анализ факторов рождаемости
4. Исследование особенностей жизнедеятельности семей в современной России
5. Структурные особенности потребительских расходов домохозяйств разных типов
6. Анализ демографических факторов уровня жизни
7. Социально-демографический профиль бедности
8. Гендерные особенности экономической активности
9. Сравнительный анализ формальной и неформальной занятости
10. Статистический анализ уровня образования россиян: тенденции и дифференциация
11. Тенденции преступности несовершеннолетних в РФ
12. Статистическое исследование состояние здоровья населения
13. Статистический анализ социальных девиаций
14. Исследование факторов суицида различных социально-демографических групп населения
15. Применение статистических методов в анализе социальной и демографической политики
16. Структура и тенденции в сфере статистики судимости
17. Статистический анализ эффективности мер борьбы со злоупотреблением алкоголя
18. Статистический анализ демографической ситуации в регионе (стране).
19. Экономико-статистический анализ развития рынка страхования в России.
20. Оценка результатов проведения пенсионной реформы в России.
21. Экономико-статистический анализ российского рынка страховых продуктов.
22. Гендерный аспект в проведении тарифной политики негосударственного пенсионного страхования.
23. Экономико-статистический анализ надежности страховых компаний России.
24. Экономико-статистический анализ российских инвестиционных институтов.
25. Статистический анализ условий жизни домохозяйств в Российской Федерации.
26. Статистический анализ кредитного поведения домохозяйств в Российской Федерации.
27. Статистический анализ брачности в Российской Федерации.
28. Статистический анализ уровня счастья людей в Российской Федерации.
29. Статистический анализ выездного туризма в Российской Федерации.
30. Здоровый образ жизни россиян
31. Безопасность и удовлетворенность работой правоохранительных органов в России.
32. Медицинское обслуживание в России
33. Информационные технологии в жизни россиян
34. Статистический анализ досуговой активности россиян
35. Статистический анализ рейтинга научных журналов в базе данных Scopus.
36. Возрастные аспекты оценки удовлетворенности жизнью в России

Выбранная тема не должна быть слишком общей, поскольку небольшой объем работы (до 20 страниц) не позволит раскрыть ее в полной мере.

При выборе темы необходимо учитывать полноту ее освещения в имеющейся научной литературе. Для этого можно воспользоваться тематическими каталогами библиотек и библиографическими указателями литературы, периодическими изданиями и ежемесячными указателями экономической литературы, либо справочно-библиографическими ссылками изданий посвященных данной теме.

После выбора темы составляется список изданной по теме (проблеме) литературы, опубликованных статей, необходимых справочных источников.

Общие требования к содержанию курсовой работы

Во введении на 2–3 страницах должна быть показана цель и актуальность курсовой работы, указаны задачи, которые ставит перед собой студент при ее написании. Кратко следует коснуться содержания отдельных параграфов работы, отметить особенность использования в работе методики, охарактеризовать в общих чертах основные источники, которые нашли в ней свое отражение. Желательно также дать краткую характеристику объекта исследования материалов, по которым пишется курсовая работа. Первый параграф, как правило, носит теоретический характер. Все сказанное желательно иллюстрировать цифровыми данными и статистическими данными из статистических справочников, монографий, журнальных статей и других источников, не забывая при этом давать соответствующие ссылки на источники.

В следующих параграфах рассматриваются дальнейшие вопросы курсовой работы. Изложение должно быть последовательно и логично. Оно должно быть также конкретным, целиком оправданным и опираться на действующую практику. Важно не простое переписывание, а критический разбор излагаемых вопросов.

Содержание курсовой работы должно быть конкретным и целиком опираться на расчеты. Следует обратить внимание на стилистику, язык работы, её оформление. Цифровой материал приводится в работе в виде аналитических таблиц, для наглядности рекомендуется строить схемы и графики. Все таблицы, схемы и графики следует пронумеровать. Ссылки на приложения целесообразно давать с указанием номера. Все разделы работы должны быть связаны между собой. Поэтому особое внимание нужно обращать на логические переходы от одного параграфа к другому и внутри параграфа от вопроса к вопросу. Все части курсовой работы должны быть связаны между собой, должны дополнять и углублять одна другую.

В заключении следует сделать общие выводы и кратко изложить предложения.

После заключения студент должен поставить дату окончания работы и свою подпись, а также принести список использованной литературы и приложения. Очень тщательно должно быть выполнено все остальное оформление работы.

Оформление курсовой работы

Материал в работе располагается в следующей последовательности:

1. Титульный лист (заполняется по утвержденной кафедрой форме).
2. Задание на курсовую работу.
3. План-оглавление.
4. Текстовое изложение курсовой работы (по параграфам).
5. Список литературы.
6. Приложения, которые содержат практический материал, использованный в работе (если он не помещен по ходу изложения).

Курсовая работа выполняется в текстовом процессоре Word через 1,5 интервала шрифтом Times New Roman (кегель 14). Цвет шрифта должен быть черным. По всем сторонам листа оставляются поля. Размер левого поля - 30 мм, правого - 10 мм, верхнего - 20 мм, нижнего - 20 мм.

Все листы курсовой работы должны быть пронумерованы. Каждый раздел в тексте должен иметь заголовок в точном соответствии с наименованием в плане-оглавлении.

Новый вопрос можно начинать на той же странице, на которой кончился предыдущий, если на этой странице кроме заголовка поместится несколько строк текста.

Цифровые данные в сгруппированном и систематизированном виде представляются в таблицах и графиках. Таблицы обычно помещаются по ходу изложения, после ссылки на них, однако не рекомендуется переносить таблицы с одной страницы на другую; тем более недопустимо разрывать заголовок с таблицей, помещая их на разных страницах. Таблицы

должны иметь порядковый номер, заголовок – отражать их содержание, а примечание – ссылку на источник.

Примеры представления таблиц, рисунков и формул приведены в соответствующих методических рекомендациях.

Количество цифрового материала должно соответствовать содержанию курсовой работы. Не следует приводить данных, не имеющих прямого отношения к излагаемому вопросу.

В таблицах и в тексте следует избегать полного написания больших чисел. Для этого целесообразно укрупнять единицы измерения в соответствии с необходимой точностью.

В работе можно использовать только общепринятые сокращения и условные обозначения.

Использованные в работе цифровые данные, выводы, мысли других авторов в пересказе и цитаты в обязательном порядке должны сопровождаться ссылками на использованные работы. Эти ссылки могут быть сделаны в виде сносок в нижней части страницы с указанием автора, названия работы, издательства, года издания и номера страницы, где находится данное высказывание, или с указанием в скобках сразу же после высказывания номера источника в списке литературы, если речь идет о содержании всего источника, например [1]. Если же дается цитата, то приводится в скобках как номер источника, так и номер страницы или страниц, например [1, с. 2].

Пересказ мыслей и выводов других авторов следует делать без искажения этих мыслей. Цитаты должны быть тщательно выверены и заключены в кавычки. Студент несет ответственность за точность приносимых данных, а также за объективность изложения мыслей других авторов.

Общий объем работы 20-25 страниц машинописного текста (без учета приложений).

Список использованной литературы и других источников составляется в следующей последовательности:

1. Законы, постановления правительства и Государственной думы.
2. Нормативные акты, инструктивные материалы, официальные справочники.
3. Специальная экономическая литература в алфавитном порядке по фамилиям авторов или названиям, если на титульном листе книги автора нет (монографии, статьи).
4. Периодические издания с указанием года и месяца журналов и газет (если статьи из них не приведены в предыдущем разделе списка литературы).

Последним этапом выполнения курсовой работы является ее внешнее оформление. Курсовая работа должна быть подписана студентом до представления преподавателю.

4. Учебно-методическое и информационное обеспечение дисциплины

5.1. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Основная литература

1. Анализ данных : учебник для академического бакалавриата / ГУ - Высшая школа экономики; под ред. В. С. Мхитаряна .— Москва : Юрайт, 2016 .— 490 с. (13 экз в библиотеке)

2. Ананьев, В. А. Анализ экспериментальных данных [Электронный ресурс] : учеб. пособие / В. А. Ананьев .— Кемерово : Кемеровский государственный университет, 2009 .— 102 с. [Электронный ресурс] URL=https://biblioclub.ru/index.php?page=book_red&id=232208

3. Чашкин, Ю.Р. Математическая статистика. Анализ и обработка данных : учеб. пособие для студ. высших учеб. заведений .— 2-е изд., перераб. и доп. — Ростов н/Д : Феникс, 2010 .— 236с. . (3 экз в библиотеке)

4. Барский, А.Б. Логические нейронные сети : учебное пособие.— Москва : Интернет-Университет Информационных Технологий, 2007.— 352 с. [Электронный ресурс] URL=https://biblioclub.ru/index.php?page=book_red&id=232983

Дополнительная литература

5. Макшанов, А. В., Журавлев А.Е. Технологии интеллектуального анализа данных: Учебное пособие. — СПб. : Издательство «Лань», 2018 .— 212 с. [Электронный ресурс] URL=<https://e.lanbook.com/reader/book/109617/#2>

6. Сидняев, Н.И, Вилисова Н.Т. Введение в теорию планирования эксперимента: учеб. пособие.— М. : МГТУ им. Баумана. Золотая коллекция, 2011. – 463 с. [Электронный ресурс] URL=<https://e.lanbook.com/book/106359#authors>

7. Бонцанини М. Анализ социальных медиа на Python / пер. с англ. А.В. Логунова.— М : ДМК Пресс, 2018.— 288 с.: ил. . [Электронный ресурс] URL=<https://e.lanbook.com/reader/book/108129/#4>

8. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / пер. с англ. А.А. Слинкина. – М.: ДМК Пресс, 2015. – 400 с.: ил. [Электронный ресурс] URL=<https://e.lanbook.com/reader/book/69955/#4>

5.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и программного обеспечения, необходимых для освоения дисциплины

Пользователям библиотеки БашГУ предоставляется возможность использования следующих электронных информационных ресурсов:

№	Наименование Интернет-ресурса	Ссылка (URL) на Интернет ресурс
1.	Федеральная служба государственной статистики	www.gks.ru
2.	Министерство финансов РФ	www.minfin.ru
3.	Международный валютный фонд	www.imf.org
4.	Центр макроэкономического анализа и краткосрочного прогнозирования	www.forecast.ru
5.	Территориальный орган Федеральной службы государственной статистики по РБ	www.bashstat.ru
6.	Информационно-издательский центр «Статистика России»	www.infostat.ru
7.	Информационно-аналитический сайт в области информационных технологий	citforum.ru
8.	Издание о высоких технологиях	cnews.ru
9.	Библиотека Г. Верникова – все о менеджменте и IT - подборка аналитических материалов по вопросам	vernikov.ru

	экономики, менеджмента и информационных технологий.	
10.	Официальный портал ИТ-директоров (Реестр ИТ-поставщиков)	globalcio.ru
11.	Журнал СIO – руководитель информационной службы	cio-world.ru
12.	Электронно-библиотечная система	ZNANIUM.COM
13.	Единый архив экономических и социологических данных ВШЭ	http://sophist.hse.ru/

1. База данных периодических изданий на платформе EastView: «Вестники Московского университета», «Издания по общественным и гуманитарным наукам» - <https://dlib.eastview.com/>
2. Информационная система «Единое окно доступа к образовательным ресурсам» - <http://window.edu.ru>
3. Научная электронная библиотека eLibrary.ru - <http://elibrary.ru/defaultx.asp>
4. Справочно-правовая система Консультант Плюс - <http://www.consultant.ru/>
5. Электронная библиотечная система «Университетская библиотека онлайн» – <https://biblioclub.ru/>
6. Электронная библиотечная система «ЭБ БашГУ» – <https://elib.bashedu.ru/>
7. Электронная библиотечная система издательства «Лань» – <https://e.lanbook.com/>
8. Электронный каталог Библиотеки БашГУ – <http://www.bashlib.ru/catalogi>.
9. Архивы научных журналов на платформе НЭИКОН (Cambridge University Press, SAGE Publications, Oxford University Press) - <https://archive.neicon.ru/xmlui/>
10. Издательство «Annual Reviews» - <https://www.annualreviews.org/>
11. Издательство «Taylor&Francis» - <https://www.tandfonline.com/>
12. Windows 8 Russian. Windows Professional 8 Russian Upgrade Договор №104 от 17.06.2013 г. Лицензии бессрочные.
13. Microsoft Office Standard 2013 Russian. Договор №114 от 12.11.2014 г. Лицензии бессрочные.
14. Windows 8 Russian. Windows Professional 8 Russian Upgrade. Договор № 104 от 17.06.2013 г. Лицензии бессрочные.
15. R-studio GNU General Public License Version 3, 19 November 2007.
16. R GNU General Public License Version 2, June 1999.

6. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
<p>учебная аудитория для проведения занятий лекционного типа: лаборатория социально-экономического моделирования № 107 (помещение, ул.Карла Маркса, д.3, корп.4), лаборатория анализа данных № 108 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 110 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 111 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 114 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 122 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 204 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 207 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 208 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 209 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 210 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 212 (гуманитарный корпус), аудитория № 213 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 218 (гуманитарный корпус), аудитория № 220 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 221 (гуманитарный корпус), аудитория № 222 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 301 (гуманитарный корпус), аудитория № 305 (гуманитарный корпус), аудитория № 307 (гуманитарный корпус), аудитория № 308 (гуманитарный корпус), аудитория № 309 (гуманитарный корпус), аудитория № 110 (гуманитарный корпус), лаборатория исследования процессов в экономике и управлении № 311а (гуманитарный корпус), лаборатория информационных технологий в экономике и управлении № 311в (гуманитарный корпус).</p>	Лекции	<p>лаборатория социально-экономического моделирования № 107: учебная мебель, доска, проекционный экран с светодиодом lumien master control, проектор casio, персональный компьютер пэвм кламас в комплекте – 18 шт.</p> <p>лаборатория анализа данных № 108: учебная мебель, доска, персональный компьютер пэвм кламас в комплекте – 17 шт.</p> <p>аудитория № 110: учебная мебель, доска, телевизор led.</p> <p>аудитория № 111: учебная мебель, доска, телевизор led.</p> <p>аудитория № 114: учебная мебель, доска.</p> <p>аудитория № 115: учебная мебель, колонки (2 шт.), динамики, dvd плеер toshiba, магнитола sony (4 шт.)</p> <p>аудитория №118: учебная мебель, проектор benq, колонки (2 шт.), музыкальный центр lg, флипчарт магнитно-маркерный на треноге</p> <p>аудитория № 122: учебная мебель, доска.</p> <p>аудитория № 204: учебная мебель, доска, проекционный экран с светодиодом lumien master control, проектор casio.</p> <p>аудитория № 207: учебная мебель, доска, телевизор led tcl.</p> <p>аудитория № 208: учебная мебель, доска, телевизор led tcl.</p> <p>аудитория № 209: учебная мебель, доска.</p> <p>аудитория № 210: учебная мебель, доска.</p> <p>аудитория № 212: учебная мебель, доска, проектор infocus.</p> <p>аудитория № 213: учебная мебель, доска, проекционный экран с светодиодом lumien master control, проектор casio.</p> <p>аудитория № 218: учебная мебель, доска, мультимедиа-проектор infocus.</p> <p>аудитория № 220: учебная мебель, доска.</p> <p>аудитория № 221 учебная мебель, доска.</p>
<p>учебная аудитория для проведения занятий семинарского типа: лаборатория социально-экономического моделирования № 107 (помещение, ул.Карла Маркса, д.3, корп.4), лаборатория анализа данных № 108 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 110 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 111 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 114 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 122 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 204 (помещение, ул.Карла</p>	Практические/семинарские занятия	

<p>Маркса, д.3, корп.4), аудитория № 207 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 208 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 209 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 210 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 212 (гуманитарный корпус), аудитория № 213 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 218 (гуманитарный корпус), аудитория № 220 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 221 (гуманитарный корпус), аудитория № 222 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 301 (гуманитарный корпус), аудитория № 305 (гуманитарный корпус), аудитория № 307 (гуманитарный корпус), аудитория № 308 (гуманитарный корпус), аудитория № 309 (гуманитарный корпус), аудитория № 110 (гуманитарный корпус), лаборатория исследования процессов в экономике и управлении № 311а (гуманитарный корпус), лаборатория информационных технологий в экономике и управлении № 311в (гуманитарный корпус).</p>		<p>аудитория № 222 учебная мебель, доска.</p> <p>аудитория № 301 учебная мебель, экран на штативе, проектор aser.</p> <p>аудитория № 302 учебная мебель, персональный компьютер в комплекте hp, моноблок, персональный компьютер в комплекте моноблок itu.</p> <p>аудитория № 305 учебная мебель, доска, проектор infocus.</p> <p>аудитория № 307 учебная мебель, доска.</p> <p>аудитория № 308 учебная мебель, доска.</p> <p>аудитория № 309 учебная мебель, доска.</p> <p>лаборатория исследования процессов в экономике и управлении № 311а учебная мебель, доска, персональный компьютер lenovo thinkcentre – 16 шт.</p> <p>лаборатория информационных технологий в экономике и управлении № 311в учебная мебель, доска, персональный компьютер в комплекте № 1 itu corp 510 – 14 шт.</p> <p>аудитория № 312 учебная мебель, доска.</p> <p>1. Windows 8 Russian. Windows Professional 8 Russian Upgrade. Договор № 104 от 17.06.2013 г. Лицензии – бессрочные. Microsoft Office Standard 2013 Russian. Договор № 114 от 12.11.2014 г. Лицензии – бессрочные.</p>
<p>учебная аудитория для проведения групповых и индивидуальных консультаций: лаборатория социально-экономического моделирования № 107 (помещение, ул.Карла Маркса, д.3, корп.4), лаборатория анализа данных № 108 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 110 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 111 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 114 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 122 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 204 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 207 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 208 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 209 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 210 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 212 (гуманитарный корпус), аудитория № 213 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 218 (гуманитарный корпус), аудитория № 220 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 221 (гуманитарный корпус), аудитория № 222 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 301 (гуманитарный корпус), аудитория № 305 (гуманитарный корпус), аудитория № 307 (гуманитарный корпус), аудитория № 308 (гуманитарный корпус), аудитория № 309 (гуманитарный корпус), аудитория № 110 (гуманитарный корпус), лаборатория исследования процессов в экономике и управлении № 311а (гуманитарный корпус), лаборатория</p>	<p>Групповые и индивидуальные консультации</p>	

информационных технологий в экономике и управлении № 311в (гуманитарный корпус).		
<p>учебная аудитория для текущего контроля и промежуточной аттестации: лаборатория социально-экономического моделирования № 107 (помещение, ул.Карла Маркса, д.3, корп.4), лаборатория анализа данных № 108 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 110 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 111 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 114 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 122 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 204 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 207 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 208 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 209 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 210 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 212 (гуманитарный корпус), аудитория № 213 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 218 (гуманитарный корпус), аудитория № 220 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 221 (гуманитарный корпус), аудитория № 222 (помещение, ул.Карла Маркса, д.3, корп.4), аудитория № 301 (гуманитарный корпус), аудитория № 305 (гуманитарный корпус), аудитория № 307 (гуманитарный корпус), аудитория № 308 (гуманитарный корпус), аудитория № 309 (гуманитарный корпус), аудитория № 110 (гуманитарный корпус), лаборатория исследования процессов в экономике и управлении № 311а (гуманитарный корпус), лаборатория информационных технологий в экономике и управлении № 311в (гуманитарный корпус).</p>	Текущий контроль и промежуточная аттестация	
<p>помещения для самостоятельной работы: аудитория № 302 читальный зал (гуманитарный корпус).</p>	Самостоятельная работа	