


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

Согласовано:  
Председатель УМК института

Зав. кафедрой  /Р.Х.Бахитова

 /Л.Р. Абзалилова

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

«Анализ больших данных в бизнесе»

Часть, формируемая участниками образовательных отношений

**Программа магистратуры**

Направление подготовки  
38.04.02 Менеджмент

Направленность (профиль) подготовки  
Управление бизнесом в цифровой экономике

Квалификация  
магистр

Разработчики (составители)

Доцент, к. т. н., доц.



Лакман И.А.

Доцент, к. т. н., доц.



Прудников В.Б.

Для приема 2021 г.

Уфа 2021 г.

Составитель / составители: Лакман И.А., 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 цифровой экономики и коммуникации протокол от «12» января 2022 г. №6.

Заведующий кафедрой



/ Р.Х. Бахитова

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. Цель и место дисциплины в структуре образовательной программы .....	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
5. Учебно-методическое и информационное обеспечение дисциплины .....	29
5.1. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.....	29
5.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и программного обеспечения, необходимых для освоения дисциплины .....	30
6. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине .....	31

# 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 (при наличии ОПК)	Формируемая компетенция (с указанием кода)	Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине
	<p><i>ПК-2 Способен использовать информационные технологии в организации производства и управлении бизнесом</i></p>	<p><i>ИПК 2.1 Разрабатывает, обосновывает и реализовывает управленческие решения, включая инновационного и новаторского направления деятельности компании</i></p>	<p><i>Знания:</i></p> <ol style="list-style-type: none"> <li>1. методы семплирования для получения сбалансированных выборок методы валидации и кросс-валидации при обучении алгоритмов машинного обучения</li> <li>2. методы и алгоритмы машинного обучения для решения задач финансовой аналитики – бинарные модели регрессии, байесовские классификаторы, алгоритмы бустинга, методы деревьев решений и случайного леса)</li> <li>3. метрики качества для оценки алгоритмов машинного обучения</li> <li>4. понятия, связанные с технологиями и особенностями обработки больших данных</li> </ol> <p><i>Умения:</i></p> <ol style="list-style-type: none"> <li>1. проводить качественную чистку данных, проводить восполнение данных; восполнять выборку до сбалансированного объема</li> <li>2. проводить классификацию и скоринг методами машинного обучения, в том числе ансамблирование алгоритмов машинного обучения для повышения точности решения задач классификации в скоринге</li> <li>3. проводить процедуры валидации и кросс-валидации для обучения алгоритмов машинного обучения</li> <li>4. использовать метрики качества для оценки алгоритмов машинного обучения</li> <li>5. адаптировать и реализовывать методы машинного обучения для использования в стриминговых системах в режиме реального времени</li> </ol> <p><i>Навыки:</i></p> <ol style="list-style-type: none"> <li>1. методами применения алгоритмов машинного обучения с использованием современных информационных инструментальных средств и сред</li> <li>2. построения стриминговой системы и витрин с использованием real-time данных</li> </ol>

## **2. Цель и место дисциплины в структуре образовательной программы**

Дисциплина «Анализ больших данных в бизнесе» относится к части, формируемой участниками образовательных отношений.

Дисциплина изучается при очно-заочной форме обучения на 2 курсе в 4 семестре и на 3 курсе в 5 семестре, при заочной форме обучения - на 2 и 3 курсах (3 и 4 семестры).

Целью изучения дисциплины является формирование знаний, навыков и умений, необходимых для понимания и практической реализации технологий машинного обучения, в том числе ансамблирования алгоритмов, проверке их алгоритмов с помощью процедур валидации и кросс-валидации, изучение техник сэмплирования в случае отсутствия сбалансированности классов; формирование системных знаний и навыков потоковой обработки больших данных, построения стриминговой системы и витрины с использованием real-time данных.

## **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 Анализ больших данных в бизнесе  
на 2 курс  
очно-заочной формы обучения

<b>Вид работы</b>	<b>Объем дисциплины</b>
Общая трудоемкость дисциплины (ЗЕТ / часов)	1/36
Учебных часов на контактную работу с преподавателем:	
лекций	6
практических/ семинарских	10
лабораторных	
других (групповая, индивидуальная консультация и иные виды учебной деятельности, предусматривающие работу обучающихся с преподавателем) (ФКР)	0,5
Учебных часов на самостоятельную работу обучающихся (СР)	19,5
Учебных часов на подготовку к экзамену/зачету/дифференцированному зачету (Контроль)	

Форма(ы) контроля:  
Контроль

№ 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 (коллоквиумы, контрольные работы, компьютерные тесты и т.п.)
		Всего	ЛК	ПР/СЕМ	ЛР	СРС			
1	2	3	4	5	6	7	8	9	10
1.	Преобразование и предварительная обработка данных. Оптимизация признакового пространства. Разделение выборок, кросс-валидация и метрики качества моделей. Селекция алгоритмов машинного обучения. ROC-анализ в финансах. Скор-карты. Технологии сэмплирования. ASMO, SMOTE. Особенности сэмплинга при работе с финансовыми данными	14	2	4		8	№№ 4, 5, из основного списка, №№ 1 из дополнительного списка	Работа с литературой и другими рекомендуемыми источниками. Выполнение кейсов 1, 2	Оценка кейсов 1, 2
2.	Алгоритмы дерева принятия решения. Решающее дерево. Жадная нисходящая стратегия «разделяй и властвуй». Алгоритм ID3. Недостатки жадной стратегии и способы их устранения. Проблема переобучения. Редукция	12	2	4		6	№№ 2, 4 из основного списка, №№ 1, 4 из дополнительного списка	Работа с литературой и другими рекомендуемыми источниками. Выполнение кейсов 3, 4	Проверка выполнения кейсов 3, 4

	<p>решающих деревьев:  предредукция и  постредукция, прунинг.  CART-деревья.  Алгоритмы случайного  леса (Random Forest).  Алгоритмы случайного  леса: косоугольный,  синтетический,  изолированный и  полностью  рандомизированный.  Определение важности  признаков по  алгоритмам, основанным  на деревьях решений).  Ансамблевые методы:  бустинги.  Экстремальный  градиентный бустинг.  Стохастический  градиентный бустинг.  EAD-, PD- и LGD-  моделирование на  основе алгоритмов  машинного обучения.</p>								
3.	<p>Задача коллекторского  скоринга на основе  моделей выживаемости.  Кривые Каплана-  Майера. Критерий  Вилкокона-Гехана.  Сравнение  выживаемости в  подгруппах.  Регрессионные модели  выживаемости.</p>	10	2	2		6	<p>№№ 2, 4 из основного  списка,  №№ 1, 4 из  дополнительного  списка</p>	<p>Работа с  литературой и  другими  рекомендуемыми  источниками.  Выполнение  кейса 5</p>	<p>Проверка выполнения  кейса 5</p>



	Выявление предикторов риска разорения на основе модели Кокса Задача прогноза оттока клиентов банка на основе модели выживаемости. Оценка качества моделей: CI- индекс Харела										
	<b>Всего часов:</b>	36	6	10		20					

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

**СОДЕРЖАНИЕ РАБОЧЕЙ ПРОГРАММЫ**

дисциплины Анализ больших данных в бизнесе  
на 3 курс  
очно-заочной формы обучения

<b>Вид работы</b>	<b>Объем дисциплины</b>
Общая трудоемкость дисциплины (ЗЕТ / часов)	2/72
Учебных часов на контактную работу с преподавателем:	
лекций	6
практических/ семинарских	8
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)	20,8
Учебных часов на подготовку к экзамену/зачету/дифференцированному зачету (Контроль)	36

Форма(ы) контроля:  
Экзамен 5 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 (коллоквиумы, контрольные работы, компьютерные тесты и т.п.)
		Всего	ЛК	ПР/СЕМ	ЛР	СРС			
1	2	3	4	5	6	7	8	9	10
1.	Концепция больших данных. Модели больших данных. Реляционные системы управления базами данных.	16	2	4		10	№№ 6, 7 и 8, из основного списка, №№ 5, 6 из дополнительного списка	Работа с литературой и другими рекомендуемыми источниками. Выполнение кейсов 6, 7	Оценка кейсов 6 и 7
2.	Аналитические базы данных. СУБД Vertica для хранения неструктурированных данных. Архитектура Data Lake хранения данных. Обработка данных в MPP системах. Поточковая обработка данных и стриминговые системы. Витрины данных MS SQL Server. Облачные технологии	20	4	4		12	№№ 6, 7 и 8, из основного списка, №№ 5, 6 из дополнительного списка	Работа с литературой и другими рекомендуемыми источниками. Выполнение кейса 8	Оценка кейса 8
	Экзамен	36				36			
	<b>Всего часов:</b>	72	6	8		58			

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

**СОДЕРЖАНИЕ РАБОЧЕЙ ПРОГРАММЫ**

дисциплины Анализ больших данных в бизнесе  
на 2 курс  
заочной формы обучения

<b>Вид работы</b>	<b>Объем дисциплины</b>
Общая трудоемкость дисциплины (ЗЕТ / часов)	2/72
Учебных часов на контактную работу с преподавателем:	
лекций	4
практических/ семинарских	4
лабораторных	
других (групповая, индивидуальная консультация и иные виды учебной деятельности, предусматривающие работу обучающихся с преподавателем) (ФКР)	
Учебных часов на самостоятельную работу обучающихся (СР)	64,0
Учебных часов на подготовку к экзамену/зачету/дифференцированному зачету (Контроль)	

Форма(ы) контроля:  
Контроль

№ 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 (коллоквиумы, контрольные работы, компьютерные тесты и т.п.)
		Всего	ЛК	ПР/СЕМ	ЛР	СРС			
1	2	3	4	5	6	7	8	9	10
1.	Преобразование и предварительная обработка данных. Оптимизация признакового пространства. Разделение выборок, кросс-валидация и метрики качества моделей. Селекция алгоритмов машинного обучения. ROC-анализ в финансах. Скор-карты. Технологии сэмплирования. ASMO, SMOTE. Особенности сэмплинга при работе с финансовыми данными	22	1	1		20	№№ 4, 5, из основного списка, №№ 1 из дополнительного списка	Работа с литературой и другими рекомендуемыми источниками. Выполнение кейсов 1, 2	Оценка кейсов 1, 2
2.	Алгоритмы дерева принятия решения. Решающее дерево. Жадная нисходящая стратегия «разделяй и властвуй». Алгоритм ID3. Недостатки жадной стратегии и способы их устранения. Проблема переобучения. Редукция	28	2	2		24	№№ 2, 4 из основного списка, №№ 1, 4 из дополнительного списка	Работа с литературой и другими рекомендуемыми источниками. Выполнение кейсов 3, 4	Проверка выполнения кейсов 3, 4

	<p>решающих деревьев:  предредукция и  постредукция, прунинг.  CART-деревья.  Алгоритмы случайного  леса (Random Forest).  Алгоритмы случайного  леса: косоугольный,  синтетический,  изолированный и  полностью  рандомизированный.  Определение важности  признаков по  алгоритмам, основанным  на деревьях решений).  Ансамблевые методы:  бустинги.  Экстремальный  градиентный бустинг.  Стохастический  градиентный бустинг.  EAD-, PD- и LGD-  моделирование на  основе алгоритмов  машинного обучения.</p>								
3.	<p>Задача коллекторского  скоринга на основе  моделей выживаемости.  Кривые Каплана-  Майера. Критерий  Вилкокона-Гехана.  Сравнение  выживаемости в  подгруппах.  Регрессионные модели  выживаемости.</p>	22	1	1		20	<p>№№ 2, 4 из основного  списка,  №№ 1, 4 из  дополнительного  списка</p>	<p>Работа с  литературой и  другими  рекомендуемыми  источниками.  Выполнение  кейса 5</p>	<p>Проверка выполнения  кейса 5</p>

	Выявление предикторов риска разорения на основе модели Кокса Задача прогноза оттока клиентов банка на основе модели выживаемости. Оценка качества моделей: CI- индекс Харела										
	<b>Всего часов:</b>	72	4	4		64					

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

**СОДЕРЖАНИЕ РАБОЧЕЙ ПРОГРАММЫ**

дисциплины Анализ больших данных в бизнесе  
на 3 курс  
заочной формы обучения

<b>Вид работы</b>	<b>Объем дисциплины</b>
Общая трудоемкость дисциплины (ЗЕТ / часов)	1/36
Учебных часов на контактную работу с преподавателем:	
лекций	4
практических/ семинарских	16
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)	5,8
Учебных часов на подготовку к экзамену/зачету/дифференцированному зачету (Контроль)	9

Форма(ы) контроля:  
Экзамен 5 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 (коллоквиумы, контрольные работы, компьютерные тесты и т.п.)
		Всего	ЛК	ПР/СЕМ	ЛР	СРС			
1	2	3	4	5	6	7	8	9	10
1.	Концепция больших данных. Модели больших данных. Реляционные системы управления базами данных.	16	2	4		10	№№ 6, 7 и 8, из основного списка, №№ 5, 6 из дополнительного списка	Работа с литературой и другими рекомендуемыми источниками. Выполнение кейсов 6, 7	Оценка кейсов 6 и 7
2.	Аналитические базы данных. СУБД Vertica для хранения неструктурированных данных. Архитектура Data Lake хранения данных. Обработка данных в MPP системах. Поточковая обработка данных и стриминговые системы. Витрины данных MS SQL Server. Облачные технологии	20	4	4		12	№№ 6, 7 и 8, из основного списка, №№ 5, 6 из дополнительного списка	Работа с литературой и другими рекомендуемыми источниками. Выполнение кейса 8	Оценка кейса 8
	Экзамен	36				36			
	<b>Всего часов:</b>	<b>72</b>	<b>6</b>	<b>8</b>		<b>58</b>			

#### 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 ПК-2 Способен использовать информационные технологии в организации производства и управлении бизнесом

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Критерии оценивания результатов обучения			
		2 («Не удовлетворительно»)	3 («Удовлетворительно»)	4 («Хорошо»)	5 («Отлично»)
ИПК 2.1 Разрабатывает, обосновывает и реализовывает управленческие решения, включая инновационного и новаторского направления деятельности компании	Знания: 1. методы семплирования для получения сбалансированных выборок методы валидации и кросс-валидации при обучении алгоритмов машинного обучения 2. методы и алгоритмы машинного обучения для решения задач финансовой аналитики – бинарные модели регрессии, байесовские классификаторы, алгоритмы бустинга, методы деревьев решений и случайного леса) 3. метрики качества для оценки алгоритмов машинного обучения 4. понятия, связанные с технологиями и особенностями обработки больших данных	Фрагментарные представления об основных методах семплирования для получения сбалансированных выборок, методах валидации и кросс-валидации при обучении алгоритмов машинного обучения, методах и алгоритмах машинного обучения для решения задач финансовой аналитики – бинарных моделях регрессии, байесовских классификаторах, алгоритмов бустинга, методы деревьев решений и случайного леса, метрик качества для оценки алгоритмов машинного обучения; основных понятиях, связанных с	Неполные представления об основных методах семплирования для получения сбалансированных выборок, методах валидации и кросс-валидации при обучении алгоритмов машинного обучения, методах и алгоритмах машинного обучения для решения задач финансовой аналитики – бинарных моделях регрессии, байесовских классификаторах, алгоритмов бустинга, методы деревьев решений и случайного леса, метрик качества для оценки алгоритмов машинного обучения; основных понятиях, связанных с	Сформированные, но содержащие отдельные пробелы представления об основных методах семплирования для получения сбалансированных выборок, методах валидации и кросс-валидации при обучении алгоритмов машинного обучения, методах и алгоритмах машинного обучения для решения задач финансовой аналитики – бинарных моделях регрессии, байесовских классификаторах, алгоритмов бустинга, методы деревьев решений и случайного леса, метрик качества для оценки алгоритмов машинного обучения; основных понятиях, связанных с	Сформированные систематические представления об основных методах семплирования для получения сбалансированных выборок, методах валидации и кросс-валидации при обучении алгоритмов машинного обучения, методах и алгоритмах машинного обучения для решения задач финансовой аналитики – бинарных моделях регрессии, байесовских классификаторах, алгоритмов бустинга, методы деревьев решений и случайного леса, метрик качества для оценки алгоритмов машинного обучения; основных понятиях,

		технологиями и особенностями обработки больших данных	особенностями обработки больших данных	технологиями и особенностями обработки больших данных	связанных с технологиями и особенностями обработки больших данных
	<p><i>Умения:</i></p> <p><i>1. проводить качественную чистку данных, проводить восполнение данных; восполнять выборку до сбалансированного объема</i></p> <p><i>2. проводить классификацию и скоринг методами машинного обучения, в том числе ансамблирование алгоритмов машинного обучения для повышения точности решения задач классификации в скоринге</i></p> <p><i>3. проводить процедуры валидации и кросс-валидации для обучения алгоритмов машинного обучения</i></p> <p><i>4. использовать метрики качества для оценки алгоритмов машинного обучения</i></p> <p><i>5. адаптировать и реализовывать методы машинного обучения для использования в стриминговых системах в режиме реального времени</i></p>	<p>Фрагментарные умения проводить качественную чистку данных, проводить восполнение данных; восполнять выборку до сбалансированного объема; проводить классификацию и скоринг методами машинного обучения, в том числе ансамблирование алгоритмов машинного обучения для повышения точности решения задач классификации в скоринге; проводить процедуры валидации и кросс-валидации для обучения алгоритмов машинного обучения; использовать метрики качества для оценки алгоритмов машинного обучения; адаптировать и реализовывать методы машинного обучения для использования в стриминговых системах в режиме реального времени</p>	<p>В целом успешное, но не систематическое использование умений проводить качественную чистку данных, проводить восполнение данных; восполнять выборку до сбалансированного объема; проводить классификацию и скоринг методами машинного обучения, в том числе ансамблирование алгоритмов машинного обучения для повышения точности решения задач классификации в скоринге; проводить процедуры валидации и кросс-валидации для обучения алгоритмов машинного обучения; использовать метрики качества для оценки алгоритмов машинного обучения; адаптировать и реализовывать методы машинного обучения для использования в стриминговых системах в режиме реального времени</p>	<p>В целом успешное, но содержащее отдельные пробелы использование умений проводить качественную чистку данных, проводить восполнение данных; восполнять выборку до сбалансированного объема; проводить классификацию и скоринг методами машинного обучения, в том числе ансамблирование алгоритмов машинного обучения для повышения точности решения задач классификации в скоринге; проводить процедуры валидации и кросс-валидации для обучения алгоритмов машинного обучения; использовать метрики качества для оценки алгоритмов машинного обучения; адаптировать и реализовывать методы машинного обучения для использования в стриминговых системах в режиме реального времени</p>	<p>Сформированное умение проводить качественную чистку данных, проводить восполнение данных; восполнять выборку до сбалансированного объема; проводить классификацию и скоринг методами машинного обучения, в том числе ансамблирование алгоритмов машинного обучения для повышения точности решения задач классификации в скоринге; проводить процедуры валидации и кросс-валидации для обучения алгоритмов машинного обучения; использовать метрики качества для оценки алгоритмов машинного обучения; адаптировать и реализовывать методы машинного обучения для использования в стриминговых системах в режиме реального времени</p>

	<p><i>Навыки:</i>  1. <i>методами применения алгоритмов машинного обучения с использованием современных информационных инструментальных средств и сред</i>  2. <i>построения стриминговой системы и витрин с использованием real-time данных</i></p>	<p>Фрагментарное владение навыками применения алгоритмов машинного обучения с использованием современных информационных инструментальных средств и сред; построения стриминговой системы и витрин с использованием real-time данных</p>	<p>В целом успешное, но не систематическое применение навыков владения алгоритмами машинного обучения с использованием современных информационных инструментальных средств и сред; построения стриминговой системы и витрин с использованием real-time данных</p>	<p>В целом успешное, но содержащее отдельные пробелы применение навыков владения алгоритмами машинного обучения с использованием современных информационных инструментальных средств и сред; построения стриминговой системы и витрин с использованием real-time данных</p>	<p>Успешное и систематическое применение навыков владения алгоритмами машинного обучения с использованием современных информационных инструментальных средств и сред; построения стриминговой системы и витрин с использованием real-time данных</p>
--	--	---	---	---	--

**4.2. 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
ПК 3.2 Владеет информационными технологиями и методами проведения и оценки эффективности работы по бизнес-анализу	<p><i>Знания:</i></p> <ol style="list-style-type: none"> <li>1. основные принципы использования машинного обучения для задач кредитного риск-менеджмента (PD, EAD, LGD)</li> <li>2. методы семплирования для получения сбалансированных выборок методы валидации и кросс-валидации при обучении алгоритмов машинного обучения</li> <li>3. методы и алгоритмы машинного обучения для решения задач финансовой аналитики – бинарные модели регрессии, байесовские классификаторы, алгоритмы бустинга, методы деревьев решений и случайного леса)</li> <li>4. метрики качества для оценки алгоритмов машинного обучения</li> </ol>	Проверка ответов на вопросы тестирования, решений кейсов, вопросы экзаменационного билета
	<p><i>Умения:</i></p> <ol style="list-style-type: none"> <li>1. применять инструменты машинного обучения для задач кредитного риск-менеджмента (PD, EAD, LGD), для задач кредитного скоринга, для задач прогноза оттока клиентов, для задач выявления мошеннических транзакций</li> <li>2. проводить качественную чистку данных, проводить восполнение данных;</li> <li>3. восполнять выборку до сбалансированного объема</li> <li>3. проводить классификацию и скоринг методами машинного обучения, в том числе ансамблирование алгоритмов машинного обучения для повышения точности решения задач классификации в скоринге</li> <li>4. проводить процедуры валидации и кросс-валидации для обучения алгоритмов машинного обучения</li> <li>5. использовать метрики качества для оценки алгоритмов машинного обучения</li> </ol>	Проверка ответов решений кейсов, вопросы экзаменационного билета
	<p><i>Навыки:</i></p> <ol style="list-style-type: none"> <li>1. применения алгоритмов машинного обучения в риск-менеджменте с использованием современных информационных инструментальных средств и сред</li> </ol>	Проверка ответов решений кейсов, вопросы экзаменационного билета

**Примерные вопросы для экзамена:**

1. Понятие методов машинного обучения
2. Инструментальные средства по машинному обучению
3. Предварительная подготовка данных
4. Переход к размеченным данным в выборке
5. Проведение Рос-анализа для бинарных моделей.
6. Расчет критериев специфичности и чувствительности модели.
7. Определение показателя AUC и коэффициента Джини для моделей.

8. Алгоритмы дерева принятия решения для решения задач классификации
9. Алгоритмы случайного леса для решения задач классификации
10. Ограничения для применения алгоритмов случайного леса
11. Признак и значения порога, по которому происходит оптимальное по заданному критерию разбиение в алгоритме randomforest.
12. Ансамбль алгоритмов randomforest в композиции методом простого голосования.
13. Ансамбль алгоритмов randomforest в композиции методом взвешенного голосования.
14. Метод *Out-Of-Bag* (ООВ) Бреймана для алгоритмов randomforest
15. Метод градиентного бустинга.
16. Алгоритм Adaboost.
17. Алгоритм Logitboost.
18. Метод экстремального бустинга.
19. Алгоритм, удаляющий определенное число образцов мажоритарного класса (undersampling).
20. Алгоритм, удаляющий определенное число образцов мажоритарного класса (oversampling).
21. Избавление от «зашумления» данных с помощью поиска связей Томека.
22. Алгоритм сэмплирования SMOTE.
23. Алгоритм сэмплирования ASMO.
24. Витрина данных.
25. SQL. Синтаксис языка.
26. Автоматизация отчетности в Metabase.
27. Модель данных и информационная архитектура предприятия.
28. ETL: автоматизация подготовки данных.
29. Инкрементальная и автоматическая выгрузка. Пайплайн.
30. Хранилище данных с несколькими источниками.
31. Документно-ориентированная СУБД MongoDB.
32. Использование неструктурированных данных в качестве источника.
33. Колоночно-аналитическая СУБД Vertica.
34. Apache Airflow как инструмент разработки, планирования и мониторинга рабочих процессов.
35. Озеро данных (Data Lake).
36. Hadoop для масштабируемой, распределенной обработки больших наборов данных.
37. Модель распределенных вычислений MapReduce.
38. Поточковая обработка данных. Kafka.
39. Облачные технологии для масштабирования систем. Яндекс. Облако

### **Экзаменационные билеты**

Экзамен является оценочным средством для всех этапов освоения компетенций.  
В экзаменационном билете – 3 вопроса.

### **Образец экзаменационного билета:**

Башкирский государственный университет

Направление подготовки 38.04.02  
«Менеджмент»

Институт экономики, финансов и бизнеса

Программа «Управление бизнесом в  
цифровой экономике»

Кафедра финансов и налогообложения

Дисциплина «Анализ больших данных в

Экзаменационный билет № 1

1. Алгоритм сэмплирования ASMO
2. Метод экстремального бустинга
3. Понятие витрины данных

Заведующий кафедрой

К.Е. Гришин

**Критерии 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.

**Кейс-задания (модуль 1)**

Для выполнения кейс-заданий рекомендуется использовать фреймворк RStudio (язык R) или Jupyter notebook/ Google Colab (язык Python).

**Кейс-задание 1. Предварительная обработка данных**

Цели практического задания:

- Узнать о возможностях загрузки данных в формате .csv и .xlsx
- Изучить выявление и устранение явных и неявных дубликатов
- Изучить выявление и устранение пропущенных значений, принципов заполнения
- Научиться преобразованию переменных
- Научиться проводить отбор переменных
- Научиться проводить фильтрацию наблюдений
- Научиться работать с командами группировки и суммирования

Порядок выполнения задания:

1. Ознакомиться с теоретическим материалом по теме
2. Скачать файл с исходным кодом скрипта
3. Открыть файл скрипта инструментальной среде и провести все действия по скрипту согласно описанию
4. В качестве задания создать новый скрипт, который загрузить в качестве ответа на кейс-задание
5. Используя встроенный набор данных 'cars' следует выполнить:
  - Провести перевод из американских единиц измерения в российские:
    - Перевести мили в час (mph) в километры в час (kph) (умножить скорость на 1.61)
    - Перевести футы в метры (умножить тормозной путь на 0.31)
    - (для пунктов 1 и 2 достаточно обновить существующие переменные и не создавать новые)
  - Создайте переменную ratio, которая будет равна тормозной путь (dist) / скорость (speed)
6. Используя встроенный набор 'swiss', создать новые таблицы данных, в которые отфильтруйте:
  - наблюдения, в которых доля католиков больше 50% и младенческая смертность меньше 20
  - наблюдения, в которых Examination или Education больше 20%
  - наблюдения, в которых фертильность больше 60 и младенческая смертность меньше или равна 18
  - наблюдения у которых Agriculture принимает значения 1.2 или 7.7

Контрольные вопросы

1. Какие стандарты представления данных позволяет обрабатывать R/Python?
2. Какие команды позволяют присоединить к датасету новые переменные ?
3. Какие команды позволяют отбрасывать из датасета переменные ?
4. Какие команды позволяют отфильтровать из датасета определенные наблюдения?
5. Зачем нужна группировка данных? Возможно ли проводить группировку одновременно по нескольким признакам?
6. Как можно поступать с пропущенными данными?

Результатом выполнения кейс-задания является отчет по кейсу 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.

## Кейс-задание 2. Технологии сэмплирования

Порядок выполнения задания

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 кейсу 2. К отчету предъявляются следующие требования:

1. Описать исходные данные.
2. Применить алгоритм сэмплирования SMOTE.
3. Применить алгоритм сэмплирования ASMO.

## Кейс-задание 3. Моделирование вероятности дефолта

Цель: Рассчитать вероятность дефолта (PD) заёмщиков на основе модели деревьев принятия решений CART. Провести рейтингование заёмщиков.

Провести классификацию заемщиков, для решения используя классификатор деревьев решения, предварительно проверив валидность исходных данных. Выбрать признак и значения порога, по которому происходит оптимальное по заданному критерию разбиение в алгоритме CART по вероятности дефолта (PD). Задать максимальное число объектов в вершине-листа дерева, для определения критерия останова алгоритма. Построить алгоритм на размеченных данных по кредитному скорингу. Провести классификацию заемщиков на основе расчета вероятности дефолта. Провести рейтингование заёмщиков.

Порядок выполнения задания:

1. Создайте новый скрипт и загрузите файл с данными, разделите выборки на train/test с соотношением 0.8, зависимая (целевая) переменная: Target
2. Для модели кредитного скоринга получите прогноз и матрицу сопряженности. Также необходимо определить, какие значения гиперпараметров оказались наилучшими.
3. Сформируйте параметры для кросс-валидации: число фолдов: 6
4. Провести моделирование, рейтингование заёмщиков

Оформление отчета

Отчет необходимо оформить в RMarkdown или тетрадке Jupyter notebook.

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

### Описание

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. 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 объектов, используя алгоритм бустинга `AdaBoost` с экспоненциальной функцией потерь

В практическом задании вы будете использовать классический набор данных `titanic`. Вам необходимо предсказать, какие пассажиры переживут крушение, а какие погибнут.

Порядок выполнения задания:

1. Создайте новый скрипт и загрузите исходные данные `titanic`
2. Разделите выборки на `train/test` с соотношением 0.8, зависимая (целевая) переменная: `Survived`
3. Проведите моделирование по нескольким типам моделей машинного обучения. Для каждой модели получите прогноз и матрицу сопряженности. Также для моделей построенных
4. через `caret` выпишите, какие значения гиперпараметров оказались наилучшими.
5. Сформируйте параметры для кросс-валидации (на выбор студента): число фолдов, число базовых алгоритмов и др.
6. Подведите итоги какая модель оказалась наилучшей

Оформление отчета

Отчет необходимо оформить в RMarkdown или тетрадке Jupyter notebook.

**Результатом выполнения задания** является отчет по кейсу 4.

К отчету предъявляются следующие требования:

1. Формулировка задач, 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. Применить алгоритм градиентного бустинга, используя правило жадного наращивания.
5. Найти пары наиболее оптимальных параметров, где под оптимизацией следует понимать принцип явной максимизации отступов, минимизировать функционал ошибки.
6. Провести классификацию объектов, для решения используя классификатор - метод экстремального бустинга с корректно подобранными весами.
7. Оценить качество классификаторов на тестовой выборке. Все графики и таблицы должны иметь сквозную нумерацию.
8. Четко сформулированные выводы по результатам выполнения кейса.

### **Кейс-задание 5. Построение модели коллекторского скоринга на основе моделей выживаемости (Кокса).**

1. Загрузите предоставленные преподавателем данные по коллекторскому скорингу.
2. Постройте графики функций выживаемости (времени до дефолта), оцененные методом Каплана–Майера.
3. Проанализируйте полученные результаты, сделав предварительный вывод о различии функций выживаемости для разных категорий заемщиков
4. Проведите логранговый тест Мантеля–Хензеля и тест Гехана–Вилкоксона на значимое различие в оценках выживаемости по подгруппам, сформированным по атрибутивному признаку (назначение кредита).
5. Постройте модель пропорциональных рисков Кокса по риску разорения, для оценки коэффициентов модели использовать метод частичного правдоподобия согласно методикам Эфрона или Брэслоу.
6. Оцените качество моделей на основе показателя  $R^2_{mer}$  (мера объясненной случайности,) и  $R^2_{mev}$  (мера объясненного отклонения), выберете лучшую методику оценки исходя из минимума информационных критериев Акайке и Шварца.
7. Проведите интерпретацию коэффициентов отобранной модели с точки зрения предикторов риска разорения.
8. Оформите отчёт в соответствии с шаблоном.

Оформление отчета

Отчет необходимо оформить в RMarkdown или тетрадке Jupyter notebook.

### **Вопросы для самоконтроля по модулю 2**

1. Витрина данных.
2. SQL. Синтаксис языка.
3. Автоматизация отчетности в Metabase.
4. Модель данных и информационная архитектура предприятия.
5. ETL: автоматизация подготовки данных.

6. Инкрементальная и автоматическая выгрузка. Пайплайн.
7. Хранилище данных с несколькими источниками.
8. Документно-ориентированная СУБД MongoDB.
9. Использование неструктурированных данных в качестве источника.
10. Колоночно-аналитическая СУБД Vertica.
11. Apache Airflow как инструмент разработки, планирования и мониторинга рабочих процессов.
12. Озеро данных (Data Lake).
13. Hadoop для масштабируемой, распределенной обработки больших наборов данных.
14. Модель распределенных вычислений MapReduce.
15. Поточковая обработка данных. Kafka.
16. Облачные технологии для масштабирования систем. Яндекс. Облако.

### Примерные вопросы для тестирования (модуль 2)

1. Реальные данные, которые попадают на вход модели после её разработки, называют:
  - тестовой выборкой;
  - обучающей выборкой;
  - валидационной выборкой.
2. Исходные данные, которые используются для проверки качества разработанной модели, называют:
  - тестовой выборкой;
  - обучающей выборкой;
  - валидационной выборкой.
3. Исходные данные, которые подаются на вход модели при обучении, называют:
  - тестовой выборкой;
  - обучающей выборкой;
  - валидационной выборкой.
4. Предобработка данных заключается в:
  - обработке пропущенных значений;
  - преобразования отдельных признаков (например, перевод категориальных признаков в количественные);
  - нормализация или стандартизация данных;
  - создание новых признаков на основе существующих;
  - все вышеперечисленное.
5. Вы решаете задачу прогнозирования индекса удовлетворённости клиентов (предполагается, что индекс – это целое число от 1 до 10). Данная задача обучения с учителем относится к:
  - задачам регрессии;
  - классификации.

6. Сопоставьте сущность и определение:

<b>Data Warehouse</b> (Хранилище данных)		Хранилище, в котором хранится огромное количество необработанных данных в собственном формате, включая структурированные, полуструктурированные и неструктурированные данные. Структура данных и требования не определены, пока данные не потребуются. Обеспечивают большую гибкость, чем более жесткое хранилище данных
<b>Data Mart</b> (Витрина)		База данных, оптимизированная для анализа

данных)	реляционных данных, поступающих из транзакционных систем и бизнес-приложений. Структура данных и схема определены заранее для оптимизации для быстрых запросов SQL, результаты которых обычно используются для оперативной отчетности и анализа. Данные очищаются, обогащаются и трансформируются, чтобы они могли выступать в качестве «единого источника истины», которому пользователи могут доверять
<b>Data Lake (Озеро данных)</b>	Подмножество хранилища данных, обычно для определенной бизнес-функции (маркетинг, продажи, складские запасы, финансы и т.п.)

7. Каково главное достоинство ELT-процесса в сравнении с ETL-процессом?
  - сокращение времени загрузки;
  - сокращение времени выгрузки.
8. Программирование сбора, хранения, обработки, поиска и визуализации данных:
  - Data enrichment;
  - Data Engineering;
  - Data Pipeline.
9. Описывает происхождение и изменения данных с течением времени (как данные были преобразованы, что изменилось и почему):
  - Data Lineage;
  - Data Management;
  - DataOps.
10. Перечислите модели данных:
  - модель Кимбалла;
  - модель Inmon;
  - подход Data Vault v.2.0;
  - подход Anchor Modelling.

## **Кейс-задание 6. Основы работы в MS SQL Server 2008**

### **Описание**

Кейс выполняется в СУБД MS SQL Server 2008, которую необходимо установить на основе информации, полученной от преподавателя (видео-инструкция). Для Microsoft SQL Server 2008 написание запросов производится в среде Microsoft SQL Server Management Studio.

### **Порядок выполнения**

- создайте новую БД, назовите ее Bikes
- создайте таблицы базы данных о продажах мотоциклов на основе предоставленного кода (БД Bikes)

- заполните таблицы данными вручную и с помощью оператора INSERT и предоставленного преподавателем кода. Проверьте, были ли добавлены записи.
- напишите три-четыре оператора обновления данных (UPDATE). Проверьте, были ли обновлены записи.
- напишите три оператора удаления данных (DELETE), чтобы удалить добавленные строки. Проверьте, были ли удалены записи.
- сохраните код на языке SQL в отдельном файле
- выполните операций вставки, обновления и удаления данных
- выполните операций выборки данных из одной таблицы (SELECT)
- напишите запрос на чтение всех столбцов из одной таблицы
- напишите запрос на чтение определенных столбцов из одной таблицы
- напишите запрос, который иллюстрирует целесообразность использования ключевого слова DISTINCT
- напишите запрос, который позволяет получить только определенные строки в таблице
- напишите запрос, который иллюстрирует применение ключевых слов IN/NOT IN
- напишите запрос, который иллюстрирует применение ключевого слова BETWEEN
- напишите два-три запроса, которые иллюстрирует применение ключевого слова LIKE. Покажите разницу между использованием символа нижнего подчеркивания «\_» и процента
- сохраните код на языке SQL в отдельном файл

**Результатом выполнения задания** является отчет по кейсу 6.

К отчету предъявляются следующие требования:

1. Отчет представляется в виде скрипта SQL с комментариями студента по результатам выполнения.
2. Формулировка задач, решение которых необходимо в ходе выполнения лабораторной работы.
3. Наличие четко сформулированных и корректных выводов по результатам выполнения кейса (запросов).

## **Кейс-задание 7. Выполнение запросов различной степени сложности в MS SQL Server 2008**

### **Описание**

Кейс выполняется в заранее установленной (кейс 6) СУБД MS SQL Server 2008, которую необходимо установить на основе информации, полученной от преподавателя (видео-инструкция). Для Microsoft SQL Server 2008 написание запросов производится в среде Microsoft SQL Server Management Studio.

### **Порядок выполнения**

- составить и выполнить следующие запросы к БД Vikes, согласно заданиям 1-15.

#### **Задание 1**

Вывести наименование товара, наименование бренда и цену, для товаров, выпущенных в 2018 году. Упорядочить по убыванию цены.

## **Задание 2**

Вывести имя, фамилию клиента, его телефон и дату заказа для всех клиентов. Упорядочить по фамилии клиента. Есть ли клиенты, не сделавшие ни одного заказа?

## **Задание 3**

Вывести наименование товара, название склада и количество товара на данном складе. Оставить все товары, но для товаров, которых никогда не было ни на одном складе должны указываться значения null, -- как для названия склада, так и для количества. Если товар на складе был, но остаток нулевой, то должны выводиться название склада и ноль. Упорядочить по двум столбцам 1) название склада и 2) наименование товара.

## **Задание 4**

Вывести наименование товара, название склада и количество товара на данном складе, в том числе вывести все товары и склады, для которых остаток нулевой. Упорядочить по двум столбцам 1) название склада и 2) наименование товара.

## **Задание 5**

Вывести наименование товара, цену товара, номер заказа для заказа 55. Упорядочить по убыванию цены товара.

## **Задание 6**

Вывести в формате "Имя Фамилия" для каждого сотрудника его начальника (manager) и его номер телефона. Упорядочить по "имя фамилия" начальника. Для самого главного начальника (над которым нет начальников) в соответствующих столбцах вывести null.

## **Задание 7**

Вывести наименование товара, его цену, год выпуска и остаток на складе для товаров, которых на складе осталось в 2 раза меньше, чем средний остаток по трем брендам, для которых остатки самые большие.

## **Задание 8**

Вывести информацию о номере заказа, дате заказа, id сотрудника, который принял заказ, но только для заказов, которые были оформлены сотрудниками, для которых менеджером является Fabiola Jackson. Вывести только первые 10 заказов, отсортировав по возрастанию даты заказа.

## **Задание 9**

Вывести наименование бренда, сумму остатков на всех складах по товарам данного бренда. Но только для трех топ-брендов, для которых сумма максимальна.

## **Задание 10**

Вывести средний остаток по трем брендам, для которых сумма остатков на всех складах максимальна.

## **Задание 11**

Выведите фамилию, имя и эл. почту менеджеров, которые за все время оформили более 50 заказов.

## **Задание 12**

Выведите наименование и цену товаров, которые оформлялись со скидкой не менее 15% и которых в заказе было не менее 2 шт.



### **Задание 13**

Выведите наименование, год выпуска и цену товаров, цена которых ниже, чем самая минимальная среди всех категорий товаров средняя цена.

### **Задание 14**

С использованием обобщенных табличных выражений выведите среднее и максимальное число заказов на одного клиента за весь период.

### **Задание 15**

С использованием обобщенных табличных выражений выведите суммарные продажи с учетом скидок за 2018 год, сгруппированных по наименованию бренда.

## **Кейс-задание 8. Построение витрины данных в MS SQL Server 2008 Microsoft Analysis Services**

### **Описание**

Кейс выполняется в заранее установленной (кейс 6) СУБД MS SQL Server 2008, которую необходимо установить на основе информации, полученной от преподавателя (видео-инструкция). Для Microsoft SQL Server 2008 написание запросов производится в среде Microsoft SQL Server Managment Studio.

Проектирование физической и логической структуры сущностей на основе базы данных осуществляется в MS Visual Studio.

### **Порядок выполнения**

- спроектировать витрину данных (Data Mart) с помощью MS Visual Studio для OLAP кубов;
- создать OLAP кубы на MS SQL Analysis Service в соответствии с разработанной логической структурой
- настроить показатели и измерения таким образом, чтобы было понятно для потребителей отчетов

**Примечание.** На практике при построении хранилищ данных и отчетности особое внимание необходимо уделить качеству исходных данных. Целесообразно проведение тестирования показателей и устранение расхождений.

**Результатом выполнения задания** является отчет по кейсу 8.

К отчету предъявляются следующие требования:

1. Отчет представляется в виде текстового файла с комментариями студента по результатам выполнения. В файле должны быть представлены скрин-шоты сводной таблицы, которая подключается к Analysis Service для вращения кубов.
2. Наличие четко сформулированных и корректных выводов по результатам выполнения кейса.

## 5. 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 (13 экз.)
2. Новиков, Федор Александрович. Символический искусственный интеллект: математические основы представления знаний : учеб. пособие для академического бакалавриата / Ф. А. Новиков .— Москва : Юрайт, 2019 .— 278 с. :
3. Искусственный интеллект и интеллектуальные системы управления / И. М. Макаров [и др.] ; РАН, Отделение информационных технологий и вычислительных систем .— М. : Наука, 2006 .— 333 с.
4. Кондрашов, Юрий Николаевич. Анализ данных и машинное обучение на платформе MS SQL Server : учебное пособие / Ю. Н. Кондрашов ; Финансовый университет при Правительстве РФ ; Департамент анализа данных, принятия решений и финансовых технологий .— Москва : РУСАЙНС, 2020 .— 303 с.
5. Криптоэкономика: мошенников выявит искусственный интеллект [[Текст]] // Эксперт. — 2018 .— № 43 .— С. 7
6. Келлехер, Д. Наука о данных: базовый курс : учебное пособие : [16+] / Д. Келлехер, Б. Тирни ; науч. ред. З. Мамедьяров ; пер. с англ. М. Белоголовского. — Москва : Альпина Паблицер, 2020. — 224 с. : схем., табл. — Режим доступа: по подписке. — URL: <https://biblioclub.ru/index.php?page=book&id=598235> (дата обращения: 29.09.2022). — ISBN 978-5-9614-3170-4. — Текст : электронный.
7. Сурова, Н. Ю. Искусственный интеллект / Н. Ю. Сурова, М. Е. Косов. — Москва : Юнити-Дана, 2021. — 408 с. : ил., табл., схем. — Режим доступа: по подписке. — URL: <https://biblioclub.ru/index.php?page=book&id=690578> (дата обращения: 29.09.2022). — Библиогр.: с. 383-385. — ISBN 978-5-238-03513-0. — Текст : электронный.
8. Туманов, В. Е. Проектирование хранилищ данных для систем бизнес-аналитики : учебное пособие / В. Е. Туманов. — Москва : Интернет-Университет Информационных Технологий (ИНТУИТ) : Бином. Лаборатория знаний, 2010. — 616 с. : ил., табл., схем. — (Основы информационных технологий). — Режим доступа: по подписке. — URL: <https://biblioclub.ru/index.php?page=book&id=233492> (дата обращения: 29.09.2022). — ISBN 978-5-9963-0353-3. — Текст : электронный.

#### Дополнительная литература

1. Чернов, Владимир Анатольевич . Интегрированное информационно-аналитическое обеспечение и искусственный интеллект в системе управления [[Текст]] = The integrated information and analytical providing and artificial intelligence in a control system / Чернов Владимир Анатольевич // Аудит. — 2017 .— № 1 .— С. 17-23
2. Бутенко, Екатерина Дмитриевна . Искусственный интеллект в банках сегодня: опыт и перспективы [[Текст]] / Е. Д. Бутенко // Финансы и кредит. — 2018 .— Т. 24, вып. 1 .— С. 143-153
3. Хорошилов, Евгений Евгеньевич. Новые технологии в финансовом секторе Канады: искусственный интеллект и "большие данные" [[Текст]] = New Technologies in Canadian Financial Sector: Artificial Intelligence and Big Data / Е. Е. Хорошилов // США. Канада. Экономика - политика - культура. — 2018 .— № 10 .— С. 50-65.
4. Винстон, Уэйн. Бизнес-моделирование и анализ данных. Решение актуальных задач с помощью Microsoft EXCEL : пер. с англ. яз. / У. Винстон ; перевод Ю. Бочиной .— 5-е изд. — Санкт-Петербург : Питер, 2018 .— 864 с. (8 экз.)

5. Манусов, В. З. Применение методов искусственного интеллекта в задачах управления режимами электрических сетей Smart Grid / В. З. Манусов, Н. Хазанзода, П. В. Матренин ; Новосибирский государственный технический университет. – Новосибирск : Новосибирский государственный технический университет, 2019. – 240 с. : ил., табл., схем., граф. – (Монографии НГТУ). – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=576716> (дата обращения: 29.09.2022). – Библиогр.: с. 216-228. – ISBN 978-5-7782-3911-1. – Текст : электронный.
6. Косарев, В. С. Нейронные сети в экономике и финансах / В. С. Косарев ; Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации. – Москва : Дело, 2021. – 118 с. : ил. – (Научные доклады: экономика). – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=694963> (дата обращения: 29.09.2022). – Библиогр. в кн. – ISBN 978-5-85006-382-5. – Текст : электронный.

## 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	<a href="http://www.gks.ru">www.gks.ru</a>
2.	Министерство финансов РФ	<a href="http://www.minfin.ru">www.minfin.ru</a>
3.	Международный валютный фонд	<a href="http://www.imf.org">www.imf.org</a>
4.	Центр макроэкономического анализа и краткосрочного прогнозирования	<a href="http://www.forecast.ru">www.forecast.ru</a>
5.	Территориальный орган Федеральной службы государственной статистики по РБ	<a href="http://www.bashstat.ru">www.bashstat.ru</a>
6.	Информационно-издательский центр «Статистика России»	<a href="http://www.infostat.ru">www.infostat.ru</a>
7.	Информационно-аналитический сайт в области информационных технологий	<a href="http://citforum.ru">citforum.ru</a>
8.	Издание о высоких технологиях	<a href="http://cnews.ru">cnews.ru</a>
9.	Библиотека Г. Верникова – все о менеджменте и ИТ - подборка аналитических материалов по вопросам экономики, менеджмента и информационных технологий.	<a href="http://vernikov.ru">vernikov.ru</a>
10.	Официальный портал ИТ-директоров (Реестр ИТ-поставщиков)	<a href="http://globalcio.ru">globalcio.ru</a>
11.	Журнал СЮ – руководитель информационной службы	<a href="http://cio-world.ru">cio-world.ru</a>
12.	Единый архив экономических и социологических данных ВШЭ	<a href="http://sophist.hse.ru/">http://sophist.hse.ru/</a>

В ходе подготовки к практическим занятиям, а также выполнении заданий для самостоятельной работы обучающиеся могут при необходимости использовать возможности информационно - справочных систем, электронных библиотек и архивов БашГУ, программного обеспечения:

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

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

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
<p><i>Учебная аудитория для проведения занятий лекционного типа:</i>                      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</p>	Лекции	<p><b>лаборатория социально-экономического моделирования № 107:</b>                      учебная мебель, доска, проекционный экран с светодиодом lumien master control, проектор casio, персональный компьютер пэвм кламас в комплекте – 18 шт.</p> <p><b>лаборатория анализа данных № 108:</b>                      учебная мебель, доска, персональный компьютер пэвм кламас в комплекте – 17 шт.</p> <p><b>аудитория № 110:</b>                      учебная мебель, доска, телевизор led.</p> <p><b>аудитория № 111:</b>                      учебная мебель, доска, телевизор led.</p> <p><b>аудитория № 114:</b></p>

<p>(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), лаборатория исследования процессов в экономике и управлении № 311а (гуманитарный корпус), лаборатория информационных технологий в экономике и управлении № 311в (гуманитарный корпус).</p>		<p>учебная мебель, доска.  <b>аудитория № 115:</b>  учебная мебель, колонки (2 шт.), динамики, dvd плеер toshiba, магнитола sony (4 шт.)  <b>аудитория №118:</b>  учебная мебель, проектор benq, колонки (2 шт.), музыкальный центр lg, флипчарт магнитно-маркерный на треноге  <b>аудитория № 122:</b>  учебная мебель, доска.  <b>аудитория № 204:</b>  учебная мебель, доска, проекционный экран с электродиодом lumien master control, проектор casio.  <b>аудитория № 207:</b>  учебная мебель, доска, телевизор led tcl.  <b>аудитория № 208:</b>  учебная мебель, доска, телевизор led tcl.  <b>аудитория № 209:</b>  учебная мебель, доска.  <b>аудитория № 210:</b>  учебная мебель, доска.  <b>аудитория № 212:</b>  учебная мебель, доска, проектор infocus.  <b>аудитория № 213:</b>  учебная мебель, доска, проекционный экран с электродиодом lumien master control, проектор casio.  <b>аудитория № 218:</b>  учебная мебель, доска, мультимедиа-проектор infocus.  <b>аудитория № 220:</b>  учебная мебель, доска.  <b>аудитория № 221</b>  учебная мебель, доска.  <b>аудитория № 222</b>  учебная мебель, доска.  <b>аудитория № 301</b>  учебная мебель, экран на штативе, проектор aser.  <b>аудитория № 302</b>  учебная мебель, персональный компьютер в комплекте hp, моноблок, персональный компьютер в комплекте моноблок igu.  <b>аудитория № 305</b>  учебная мебель, доска, проектор infocus.  <b>аудитория № 307</b>  учебная мебель, доска.  <b>аудитория № 308</b>  учебная мебель, доска.  <b>аудитория № 309</b>  учебная мебель, доска.</p>
<p><b>учебная аудитория для проведения занятий семинарского типа:</b> 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), лаборатория исследования процессов в экономике и управлении № 311а (гуманитарный корпус), лаборатория информационных технологий в экономике и управлении № 311в (гуманитарный корпус).</p>	<p>Практические/семинарские занятия</p>	
<p><b>учебная аудитория для проведения групповых и индивидуальных консультаций:</b> 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), лаборатория исследования процессов в экономике и управлении № 311а (гуманитарный корпус), лаборатория информационных технологий в экономике и управлении № 311в (гуманитарный корпус).</p>	<p>Групповые и индивидуальные консультации</p>	
<p><b>учебная аудитория для текущего контроля и промежуточной аттестации:</b> 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), лаборатория исследования процессов в экономике и управлении № 311а (гуманитарный корпус), лаборатория информационных технологий в экономике и управлении № 311в (гуманитарный корпус).</p>	<p>Текущий контроль и промежуточная аттестация</p>	
<p><b>помещения для хранения и профилактического обслуживания</b></p>	<p>Самостоятельная работа</p>	

<p><b>учебного оборудования:</b> аудитория № 115 (помещение, ул. Карла Маркса, д.3, корп.4), 118 (помещение, ул.Карла Маркса, д.3, корп.4)</p>		<p><b>лаборатория исследования процессов в экономике и управлении № 311а</b> учебная мебель, доска, персональный компьютер lenovo thinkcentre – 16 шт.</p> <p><b>лаборатория информационных технологий в экономике и управлении № 311в</b> учебная мебель, доска, персональный компьютер в комплекте № 1 i ru согр 510 – 14 шт.</p> <p><b>аудитория № 312</b> учебная мебель, доска.</p> <p>1. Windows 8 Russian. Windows Professional 8 Russian Upgrade. Договор № 104 от 17.06.2013 г. Лицензии – бессрочные.</p> <p>2. Microsoft Office Standard 2013 Russian. Договор № 114 от 12.11.2014 г. Лицензии – бессрочные.</p> <p>3. Система централизованного тестирования БашГУ (Moodle). GNU General Public License Version 3, 29 June 2007</p>
--	--	---